A Validation Study of Vaisala's Wind Energy Assessment Methods

DECEMBER 2015

*Authors: Mark Stoelinga, Senior Scientist, and
Matthew Hendrickson, Manager of Energy Assessment*

EXECUTIVE SUMMARY

Vaisala has conducted a significant validation study of its due diligence wind energy assessment methodology. The study was based on 127 years of energy from commercial wind farm operations at 30 different wind farms in the United States, Europe, and Asia. Pre-construction assessments were performed with Vaisala's current methodology in a blind retrospective forecast framework. The mean bias error for all wind farm years (where "error" is defined as actual energy produced minus the pre-construction long-term estimate, expressed as a percent) was +0.1%, with a 95% confidence interval of ±4.3%. The standard deviation of the 1-year errors was 8.8%, somewhat lower than Vaisala's mean estimated 1-year uncertainty on energy of 10.6%, indicating that Vaisala's estimated uncertainties are somewhat conservative.

In this study, Vaisala has also examined whether there is a relationship between prediction error and pre-construction estimate uncertainty. To do this, a Monte Carlo model was developed that enabled simulation of thousands of validation studies similar to the one actually carried out. A simple metric for quantifying the uncertainty/error relationship was defined, which is the slope of the best-fit line on a scatter plot of absolute wind farm year errors versus 1-year uncertainty estimates. The slope in Vaisala's study was 0.57, and this slope implied a median "uncertainty model skill factor" (i.e. the fraction of variance in the uncertainty explained by the uncertainty model, with 0.0 showing no skill and 1.00 showing perfect skill) of 0.51.

Put in other words, Vaisala's result was the median result expected by a "half-perfect" uncertainty model. However, given the sample size of the current study, the 95% confidence lower bound was a skill factor of only 0.09 (i.e. Vaisala is 95% confident that its uncertainty model is at least 9% perfect). These results illustrate how difficult it is to confirm a strong positive uncertainty/error relationship. Even with 1000 additional wind farm years and the current results still standing, Vaisala would still only be able to make a statement with 95% confidence that we have an uncertainty model that is at least 25% perfect.

This study, is good evidence that Vaisala's wind energy assessment process is calibrated to actual production while showing that there is at least some positive relationship between uncertainty and prediction error.

1 MOTIVATION

Wind resource assessment is a critical component of the project development process, with a number of its important steps depending on the results of the resource assessment, including project layout and micrositing, turbine suitability analysis, financing, and turbine warrantees. It is incumbent upon experts within the wind energy industry not only to provide

the most accurate resource assessments possible, but to demonstrate that accuracy with validation studies that compare pre-construction wind energy assessments to operational wind farms. Answering the simple question, "What did the wind farm generate compared to what was predicted?" is very important.

Furthermore, as the industry has matured, experts have increasingly been relied upon to estimate not only the expected mean energy production of a wind farm, but to accurately estimate the uncertainty surrounding that mean estimate. This shift has come as the industry has matured and investors are gaining confidence that the mean estimate is calibrated to actual generation. There is general industry acceptance that taking steps to reduce the uncertainty of a project's energy estimate, such as installing more meteorological masts at higher heights or using more sophisticated modeling techniques for long-term energy estimation at turbine locations, is advantageous. However, we are not aware of any studies that have shown, either empirically or theoretically, that reduced uncertainty estimates translates into lower error of estimation.

Several important validation studies have emerged from the wind resource assessment industry in recent years [1-3], which have shown three key results, the first two of which are encouraging and the third of which is not:

1. There has been a positive trend, from mean overestimation of production by as much as 10% during the 2000's, to mean correct estimation of energy production at the present time. In other words, the industry is now perceived to be calibrated with respect to P50, whereas it was not before.

2. There has been a confirmation that, averaged over many wind farms, 1-year errors in pre-construction energy assessment are close to what is expected from the average estimates (across all wind farms) of uncertainty, although possibly slightly conservative.

3. There has been a weak, or in some cases lack, of apparent relationship between pre-construction estimates of 1-year uncertainty for individual wind farms and observed 1-year errors of energy production. In other words, the industry is demonstrating a lack of ability to distinguish low uncertainty projects from high uncertainty projects.

In response to the motivations described above, Vaisala has undertaken a major validation study of our wind energy assessment approach, including both the P50 energy estimates and the associated 1-year uncertainty estimates. Additionally, Vaisala has taken the step of analyzing the statistical significance of our results, by developing a Monte Carlo simulator that produces thousands of instances of the same validation study configuration, using the known statistical behavior of wind farms in the actual study. The Monte Carlo simulator provides an important determination of the confidence interval on the mean of our validation histogram, on the standard deviation of that histogram, and on the relationship between our pre-construction uncertainties and yearly errors of our long-term estimates. It also provides a view into how many wind farm years (WFYs) would be needed in the validation study to reduce these confidence intervals to any desired level.

The study described herein provides several benefits, both internal and external to Vaisala.

• Internally, this study validates and informs the science that underpins Vaisala's methodology, and provides a test bed for ongoing improvements to our methodology. Vaisala is continually innovating all components of its renewable energy assessment techniques, and it is important to track the effect of these improvements on a

diverse portfolio of projects, to ensure the stability and reliability of our system. The infrastructure Vaisala has built to conduct this study has been retained and streamlined for continual testing in the future.

- Externally, this validation study provides the industry with confirmatory support for the accuracy of Vaisala's methodology.

- This validation study also adds to the body of knowledge in the wind energy assessment industry as a whole, providing a new "data point" on the degree of calibration of wind resource assessment techniques within the industry, as well as new information on the confidence of validation study results.

2 VAISALA'S WIND RESOURCE ASSESSMENT METHODOLOGY

Vaisala's approach to wind resource assessment follows many of the standard best practices familiar to the industry, but is also founded on some key innovative technologies:

- Numerical Weather Prediction (NWP, also known as mesoscale modeling), for both spatial modeling and multi-decadal weather and climate simulation. Recent studies independent of Vaisala [4,5] have demonstrated the reduced error of NWP-based methods for wind resource assessment.

- Treatment of wind modeling, power modeling, wake modeling, and other losses in a multi-decadal hourly time series domain rather than a frequency domain; and

- Full propagation of uncertainties throughout the entire modeling process, including error dependencies.

These newer techniques advance the scientific foundation of wind resource assessment, but at the same time, require diligence in validation to ensure the industry that our wind resource assessment techniques are indeed providing accuracy, reliability, and value to the process.

3 DATASET DETAILS

This validation study was enabled through partnerships developed between Vaisala and a number of wind farm owner/operators, who provided pre- and post-construction data from a diverse set of wind farms specifically for the purposes of this study. The dataset gathered for this study is summarized below, and in Figs. 1-2:

- The dataset included 30 wind farms, approximately located as shown in Fig. 1. A diverse set of geographies and climates are represented, primarily over the United States, but also including a few wind farms in Europe and one in Southeast Asia. Nine of the sites are in complex terrain, six are in rolling terrain, and 15 are in flat terrain.

- Installed capacity ranges from 20 to 300 MW (Fig. 2a), with an average of 126 MW.

- Hub heights are mostly 80 m, but with a few higher and lower hub heights as well (Fig 2b). Nine different turbine manufacturers were represented.

- The commercial operation dates (CODs) of the wind farms ranged from 2007 to 2014, but a majority of the wind farms began commercial operations in 2010 or 2011 (Fig. 2c). This time period was by intent, as we sought wind farms that were new enough to have benefitted from a relatively state-of-the-art pre-construction meteorological campaign,

but in existence long enough to provide several years of energy validation data. This distribution of CODs yielded a total of 127 WFYs of commercial energy production, which were distributed among the wind farms as shown in Fig. 2d, for an average of 4.23 WFYs per wind farm.

Our wind farm partners provided the following data for the study:

- Pre-construction data, including met mast locations, raw measurement data, commissioning and calibration documents, turbine layout, turbine specifications (power curve, thrust curve, etc.), and information on nearby turbines.

- Post-construction data, including monthly totals of metered energy, as well as monthly estimates of percent system availability and of externally curtailed energy.

Figure 1. Map showing approximate regional distribution of wind farms included in the study. The number inside each marker specifies the number of wind farms in the region indicated.
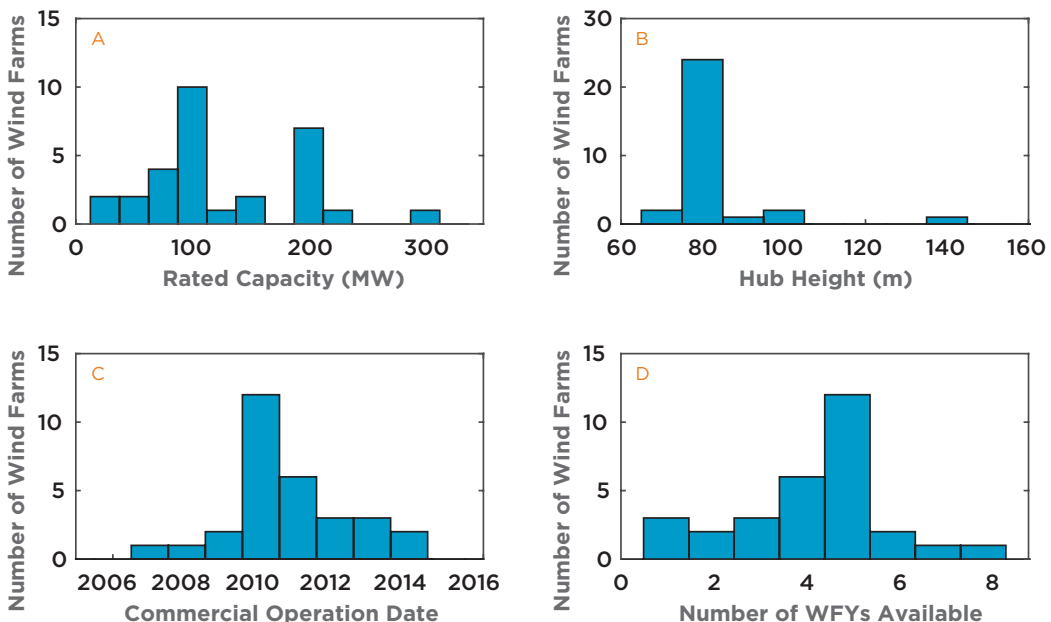

Figure 2. Histograms of the characteristics of wind farms included in the study: (a) by rated capacity (MW); (b) by turbine hub height (m); (c) by commercial operation date; and (d) by number of WFYs of energy production data available to the study.

4 PROCEDURE

The study was conducted as a blind retrospective forecast. We generated all "pre-construction estimates" as part of the study, using Vaisala's current methods, but with a view only into the data available prior to wind farm commissioning. While this approach does not address the question of how a methodology has improved or regressed over time, it does give the community a clearer picture into how Vaisala's current methodology will perform in present applications, because all 30 wind energy assessments in the study were performed in the exact same way. It also allowed Vaisala to build the study quickly by obtaining data and performing the assessments and analysis all at once, rather than waiting for the dataset to populate organically from commissioned wind energy assessments over time.

The only exception to the blind restriction is that we generated a different pre-construction estimate for each known configuration of external turbines that has occurred since commercial operation of a wind farm began, and then to validate each WFY, we used the estimate appropriate to that time period. The reason for this approach is that we are not attempting to validate our ability to predict future wind energy build-out surrounding a wind farm, so we chose to remove the uncertainty of unknown new external wakes from the validation. Vaisala's assessment reports will typically call out the likelihood of new wind development in the vicinity to the extent it can be predicted, but it is not included quantitatively in the energy assessment, unless the impending build-out is clear and imminent.

With the data in hand, Vaisala produced a wind energy assessment for each wind farm. We performed our standard due diligence wind energy assessment process, including quality control of all data provided, long-term modeling, spatial modeling, wake modeling, loss calculations, and uncertainty modeling. The key final numbers for each wind farm assessment are the long-term mean net annual energy production (the P50 AEP estimate) in gigawatt-hours; and the uncertainty in percent for any one year of production. The 1-year uncertainty arises both from uncertainties in the energy estimation technique, and from interannual variability in wind resource and losses. Within this current work, it should be noted that we use the term error to describe differences between observed generation and the P50 AEP estimate. While we do this for convenience, the annual differences that arise from natural climate variability, as well as from losses such as system availability that are known to fluctuate annually, are expected and are not truly "errors" of estimation; whereas the bias that materializes over time after variability is averaged out, can be considered errors in the assessment methodology, the pre-construction measurements, or possibly observed generation.

With regard to the production data, at each wind farm, we gathered the monthly data into groups of 12 consecutive months (not necessarily January through December), and aggregated the monthly production data into these "wind farm year" (WFY) values. Our due diligence assessment reports can include qualitative guidance on future external curtailments, but they do not often quantitatively include external curtailments or its variability in the future energy estimate, except where regionally consistent instability exists. Therefore, to provide a consistent validation dataset, we added the wind farm-reported estimates of external curtailment back into the metered production numbers, to create a compatible comparison with the pre-construction assessment estimates, assuming no curtailment. This step does add additional noise to the validation data that is dependent on the quality of the external curtailment estimates. However, our wind farm partners are all keenly interested in accurate curtailment estimates, and are presumed to have developed reliable estimation techniques for their own purposes, so this approach is not seen as significantly detrimental to the study.

5 RESULTS

5.1 Wind Energy Error Histogram

First, we present what has become a standard depiction of wind energy assessment validation studies in the industry, which is the error histogram (Fig. 3). This histogram is populated by calculating the difference of each WFY energy value from the pre-construction P50 estimate for that wind farm and expressing it as a percent of the P50. These values are referred to as the 1-year errors. Each histogram bin covers 2% in width, and the counts (*y*-axis) add up to the full set of 127 WFYs. The statistical properties of the mean and standard deviation of the 1-year energy errors are also calculated, and are annotated in the figure. The mean of all the 1-year errors, referred to hereafter as the mean bias error, is very close to 0%, meaning that on average, the assessment method provides a well calibrated P50 estimate of energy. A unique aspect of this study is that we have used Monte Carlo simulations (described in section 6) to calculate the 95% confidence interval on this mean, shown at the top of the figure. While this particular validation study showed near-perfect calibration, we can only say with 95% confidence that the mean error is within ±4.3% of zero.
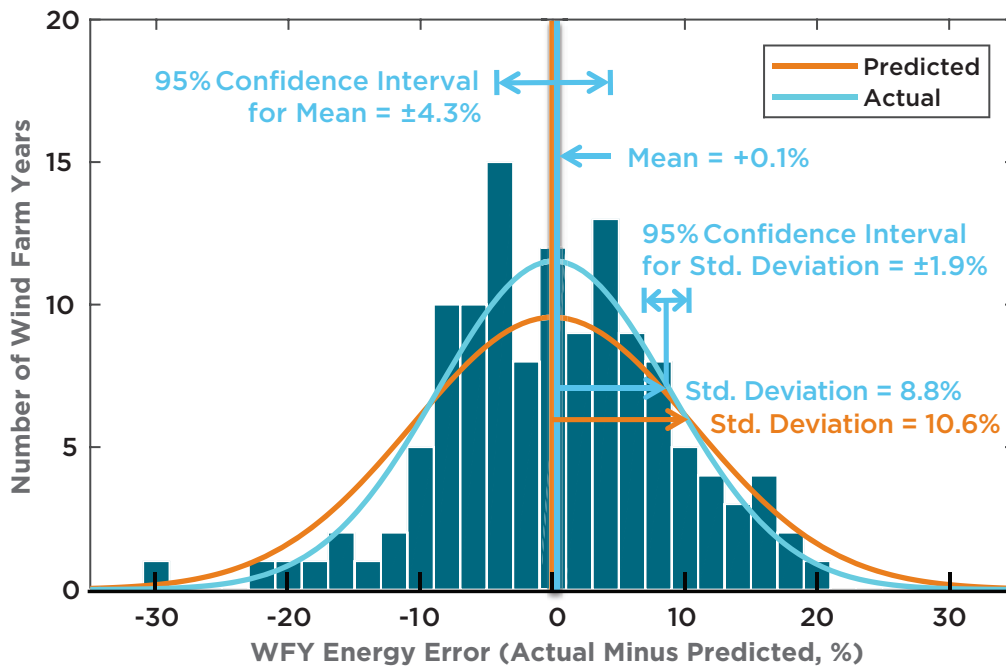

Figure 3. Histogram of WFY energy errors (1-year actual minus long-term predicted, %). Vertical lines show means of actual (blue, *x* = +0.1%) and predicted (orange, *x* = 0.0%) distributions. Bell curves depict actual and expected normal distributions, centered on actual and predicted means, and using standard deviations from actual data (blue) and from mean predicted uncertainty (orange), respectively. The 95% confidence intervals on the mean and standard deviation of the actual distribution, which are ±4.3% and ±1.9%, respectively, are indicated by the labeled intervals.

The histogram shape follows that of a normal distribution, though not perfectly so. Deviations from a perfect normal distribution are expected, due to both the limited sample size, and to other effects such as nonlinearity of the power curve and upper limits of certain losses like availability. However, the distribution passes simple tests of normality and is assumed to be so for the purpose of this study. Two normal distributions are depicted as bell curves in the figure. The orange curve represents what one would expect for the histogram shape from Vaisala's predictions, assuming perfect calibration of the mean bias error (i.e. centered on zero) and a standard deviation equal to the mean 1-year uncertainty

on energy for all 30 wind farms (10.6%). The blue curve represents what was obtained from actual data, centered on the mean 1-year bias error (+0.1%) and with a standard deviation of the actual 1-year errors (8.8%). It can be seen that the two curves are similar, though the actual distribution (8.8%) is narrower than the expected uncertainty (10.6%), meaning that on average, Vaisala's predictions are slightly more accurate than the uncertainty predictions give credit for, at least for this sample of wind farms. The Monte Carlo simulations (section 6) yield a 95% confidence interval on the standard deviation of 1-year errors of ±1.9%, so the difference between the standard deviation obtained in this study and the average uncertainty estimate of 1.8% (10.6% minus 8.8%) is not quite large enough to be significantly different at the 95% confidence level. Therefore, while this result motivates continued testing of Vaisala's uncertainty model, additional data should be collected before an across-the-board reduction in uncertainty to bring the mean uncertainty into "calibration" is effected.

5.2 Adjustment for Annually Varying Windiness and System Availability

As mentioned above, the 1-year uncertainty of the pre-construction energy estimates includes both the uncertainty in the long-term energy estimate for a particular wind farm, as well as the predicted interannual fluctuations in windiness, system availability, and other losses. The most straightforward analysis of the data is to simply calculate 1-year errors and show the results, with an understanding that both of these sources of uncertainty are included. However, it is also of interest to factor out annually varying contributions to the error, to the extent possible, to get a clearer view into the error of the long-term estimate itself, and we have done so in this study.

For the system availability adjustment, we used wind farm-reported monthly system availability percentages, and adjusted our WFY energy predictions accordingly. For windiness, Vaisala's time series approach to wind energy assessment is well suited to provide the necessary information, because we were able to simply extend our NWP-based energy assessment time series past the COD up to the present time, and use that time series to provide a wind-adjusted energy estimate for each operational WFY. We also recalculated our uncertainty estimates excluding interannual variability of windiness, system availability, and both effects simultaneously.

The results of this analysis are shown in Table 1. Each effect accounted for separately moves the mean bias error downward (toward underperformance) by about 1%, and the combined effect by about 2%. This suggests that the particular WFYs used in the study may have been slightly windier than the long-term mean, and system availability may have been slightly higher than assumed in our reports on average. However, it should be noted that both the estimate of windiness of a period and the estimate of availability, as presented by the sponsors, come with their own uncertainty. In any event, the decreases in mean bias error for these effects, either separately or combined, fall well within the 95% confidence interval on the mean bias error of ±4.3%.

Figure 4 shows a scatter plot of the root mean-squared 1-year error versus the adjusted mean 1-year uncertainty for each of the experiments. Although the points lie slightly below the one-to-one line due to the previously noted lower than estimated standard deviation of 1-year errors, the points line up nearly parallel to the one-to-one line, meaning that error spread and uncertainty decrease roughly proportionally to each other as availability and/or windiness are removed from the analysis. This result indicates that the uncertainty model properly captures the annually varying portion of the overall 1-year uncertainty.

Experiment	Mean Bias Error (Actual Minus Predicted)
No Adjustment	+0.1%
Adjustment for System Availability	-1.2%
Adjustment for Windiness	-0.9%
Adjustment for Both Availability and Windiness	-2.2%

Table 1. Mean bias error of energy estimates (actual minus predicted), for four different experiments
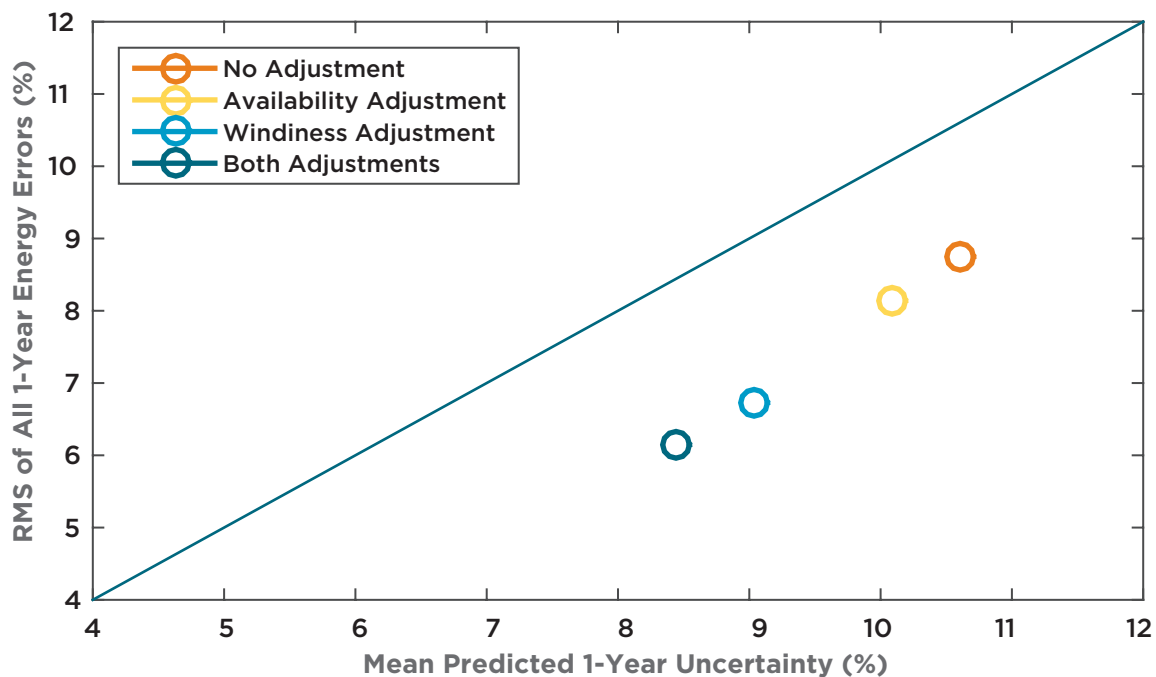

Figure 4. Plot of root mean-squared WFY energy errors from actual data (*y*-axis) versus mean predicted 1-year uncertainty (*x*-axis), for the four experiments shown in the legend in the upper left corner.

5.3 Relationship of Pre-Construction Uncertainty Estimates to 1-Year Errors

Recently, there has been concern within the industry about the apparent lack of a relationship between errors in generation at operating wind farms and the pre-construction uncertainty estimates. Since estimates of uncertainty are the primary way by which risk is quantified by the investment committee for a planned project, this lack of an uncertainty/error relationship calls into question the industry's skill at characterizing risk. In response to this concern, several new or updated presentations and reports on wind energy validation studies have also examined this relationship to test whether uncertainty estimates exhibit a relationship to observed error, and the results have been mixed.

The typical way by which this relationship is assessed is to construct a scatter plot of 1-year errors in production on the *y*-axis, against the pre-construction 1-year energy uncertainty on the *x*-axis. This has been done for the present study (Fig. 5). If the correct relationship exists, then the 1-year errors should spread out more in the vertical direction on the plot the farther to the right one looks. One way the correctness of the spread can be assessed is to shade an envelope bounded by the lines for which production error is equal to plus or minus the uncertainty, and at any point along the *x*-axis, approximately 68% of the data points should lie within that envelope. There are several difficulties with this approach. First, for typical validation studies, the sample size is small enough such that it is difficult to visually determine the relative density of points inside and outside the envelope as one moves laterally along the *x*-axis. Second, there are usually more WFYs with mid-range values of uncertainty than with very small or very large values, so visually comparing the behavior of low, medium, and high uncertainty projects is difficult. One can compare errors to uncertainty within several bins of uncertainty, but this yields multiple numbers rather than a single metric, and the expected statistical behavior of the result is not readily apparent.

One approach that has been used for similar purposes in the realm of probabilistic weather forecasting is to depict a similar scatter plot, except showing absolute error (always positive, rather than positive or negative), as is shown here in Fig. 6; and then fitting a line to the data. This plot will subsequently be referred to as an uncertainty validation plot. Although the scatter plot has a very low correlation (r = 0.22), the increased spread of the errors with increased uncertainty should yield a positive slope to the fitted line. In fact, it can be shown either through derivation or with a large Monte Carlo simulation (as will be seen later) that the slope of this fitted line is expected to be $\sqrt{2/\pi}$, or 0.80, for a "perfect uncertainty model" in a very large validation study.

In contrast, for an uncertainty model with zero skill, the spread of the errors will not change as one goes from low to high uncertainty, and the expected slope of the fitted line will be zero. Thus, the slope of the fitted line is a useful metric for establishing the strength of the uncertainty/error relationship, because it is a single metric, and has known values for both a "perfect uncertainty model" and a "zero skill uncertainty model." In the present study, the fitted line in the validation plot yielded a slope of 0.57, which is positive, but less than perfect.

A key question is this: Even if we knew hypothetically that we had a perfect uncertainty model, what might we expect the possible values of this slope to be in any single validation study of a limited sample size, such as the present one with 127 WFYs? To address this question, we turn to Monte Carlo simulations of validations studies.

Figure 5. Scatter plot of WFY energy errors (*y*-axis) versus predicted 1-year uncertainty of energy estimates (*x*-axis). Light-blue shaded region is bounded by the lines along which energy errors equal uncertainty, or in other words, the one-standard-deviation bounds of the expected distribution of energy errors. 68% of the data points are expected to lie within the shaded region.
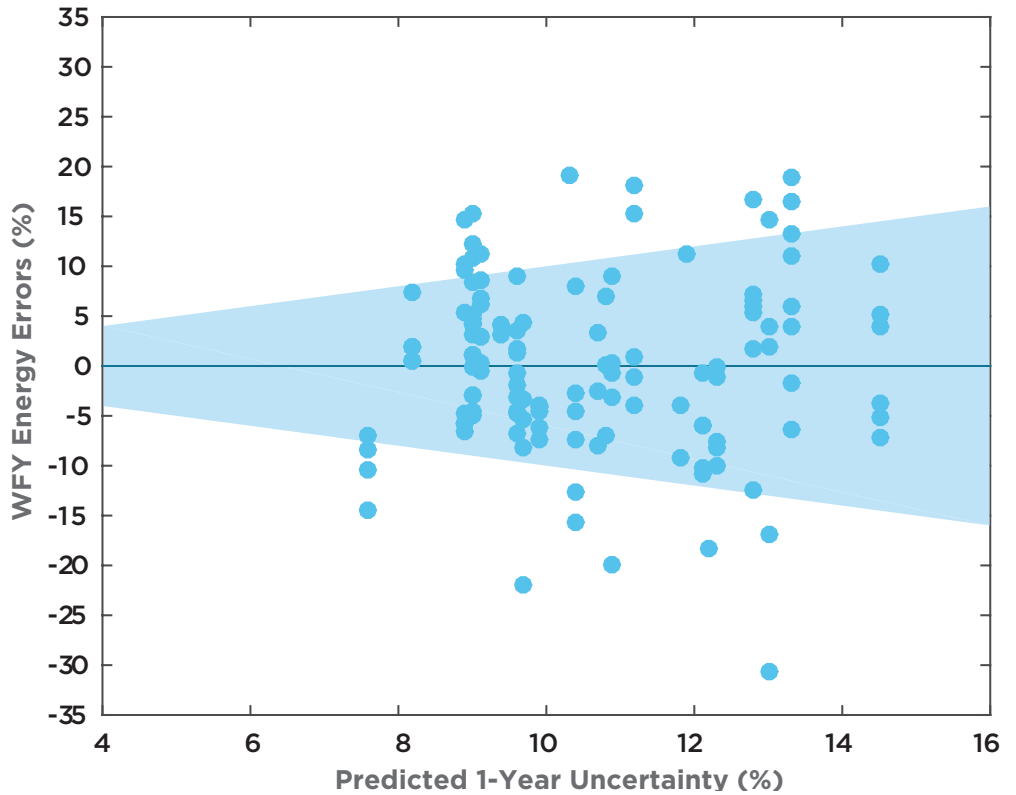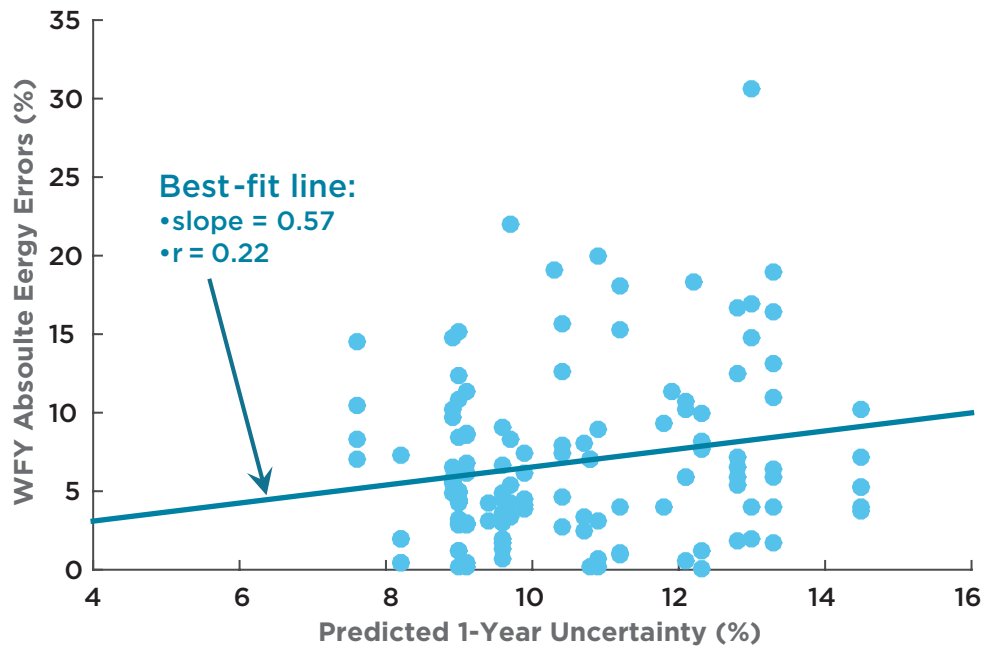

Figure 6. Scatter plot of WFY absolute energy errors (*y*-axis) versus predicted 1-year uncertainty (*x*-axis). Dark blue line is the best-fit line to the data points, with slope and correlation indicated in inset.

Best-fit line:
•slope = 0.57
•r = 0.22

6 MONTE CARLO SIMULATIONS

6.1 Motivation and Model Description

Validation studies such as Vaisala's have been produced with varying sizes of datasets. This fact begs the question, how do the results depend on the size of the dataset? We have presented three key results in the foregoing section, namely, that our mean bias error is close to zero, that the standard deviation of our 1-year errors is close to the mean estimated uncertainty, and that we show a positive relationship between our uncertainty estimates and the 1-year errors at individual wind farms. But considering the sample size, how robust are these results? What is the likelihood that we obtained a "lucky draw," or "unlucky draw"? To address this question, we examine statistical significance and confidence intervals on our results using a Monte Carlo simulator that can quickly generate thousands of wind farm energy predictions and errors that match the known statistical behavior of the wind farms in our actual study. Here we describe the model assumptions and present results of the simulations.

The following parameters are set for the model. The values that pertain to simulating Vaisala's actual validation study are shown in parentheses, although these can be varied to simulate the assumed size or behavior of any validation study.

1. Number of wind farms (30) and WFYs (127).

2. Mean and standard deviation of WFYs per wind farm (4.23 ± 2.0)

3. True mean bias error in the validation study (assumed 0.0%)

4. True average uncertainty of WFY energy errors (9.68%)

5. Standard deviation of true uncertainty for WFYs (1.82%)

6. Uncertainty model skill factor (varying levels tested, from 0.00 to 1.00)

7. Fraction of contribution to total uncertainty contributed by "fixed" as opposed to "annually varying" errors for a wind farm (0.33)

8. Assumed maximum correlation of errors among nearby wind farms (0.48)

9. Inter-farm distance of drop-off to half-amplitude correlation (1000 km).

The meaning and function of these parameters become clear as the procedure is described for one iteration of the Monte Carlo model. We first randomly select from a normal distribution the number of production years for each wind farm, using the specified average and standard deviation of number of WFYs per wind farm. The numbers are tweaked to always force the desired number of wind farms and WFYs. For each wind farm, we then randomly assign an uncertainty for the 1-year prediction, using the assumed true mean (0.0%) and true standard deviation. We separately assign two different random values of uncertainty, the "predicted" version and the "unknown" version, completely uncorrelated with each other, in order to give the ability to adjust uncertainty model skill. If the skill factor is set to 1.0 ("assumed perfect"), all of the actual 1-year errors for a wind farm are randomly drawn from a distribution defined by the "predicted" uncertainty, ensuring a perfect uncertainty/error relationship. Conversely, if skill factor is set to 0.0, all of the actual 1-year errors are randomly drawn from a distribution defined by the "unknown" uncertainty, ensuring no uncertainty/error relationship. Skill values between 0.0 and 1.0 draw from a distribution whose spread ranges proportionally between the "predicted" and "unknown"

distributions, ensuring a "fractional skill." Mathematically, the skill factor is defined as the fraction of variance among wind farms of 1-year uncertainty that is correctly explained by the uncertainty model, but it can be understood more simply as the ability of the uncertainty model to correctly distinguish low-uncertainty projects from high-uncertainty projects.

Once the true uncertainty is defined for each wind farm, the WFY errors are randomly drawn. Two sources of error dependence are accounted for: fixed versus annually varying contributions to WFY errors, and spatial correlation of WFY errors. Regarding the first, it is assumed (as discussed previously) that 1-year errors arise from both a "fixed error" in the wind energy assessment that is characteristic of a wind farm and does not change from year to year; and an "annually varying" error that results both from interannual variability of windiness, losses, etc., as well as annually varying error. The apportioning of uncertainty and errors among these two sources affects the relative independence of WFY errors and the confidence intervals obtained, and must be accounted for.

In examining the actual 1-year errors in our study, we determined that approximately one-third of the variance in annual error arises from the "fixed" contribution, and two-thirds from the "varying" contribution. When errors are randomly selected in the model, they are selected from two different distributions following these proportions, with the fixed error being the same for all years within the same wind farm, and the varying error changing from year to year. With regard to spatial correlation, we examined the correlations of both fixed and varying errors (in the same years) from wind farms based on geographic distance, and found, not surprisingly, that those errors correlate for nearby wind farms. The varying errors correlated more strongly than the fixed. This relationship must also be accounted for in randomly selecting WFY errors as it also reduces the relative independence of WFYs. We defined a correlation matrix based on inter-farm distance, and using a technique known as Cholesky decomposition, imposed that correlation matrix on all the WFY energy errors that were randomly chosen, adding dependence to the errors.

6.2 Results for the Confidence Interval on the Mean and Standard Deviation of 1-Year Energy Errors

To determine the 95% confidence interval on the mean bias error (i.e., the ±4.3% interval shown at the top of Fig. 3), we ran the Monte Carlo simulation as configured above for 5000 iterations, and calculated the 95% confidence interval as half the distance between the 2.5 and 97.5 percentile values of mean bias error. We performed three different experiments, the results of which are shown in Table 2. In the first experiment, we set the "fixed" proportion of wind farm energy errors to zero, and turned off inter-farm correlation of energy errors, effectively rendering WFYs completely independent of one another. This yields a 95% confidence interval on the mean bias error of ±1.7%, although with perfect independence, one can easily calculate this result as $1.96 \times \sigma / \sqrt{n}$, where σ is the assumed standard deviation of WFY energy errors (9.68%), n is the number of WFYs (127), and 1.96 is the number of standard deviations from the mean that encompass 95% of the data in a normal distribution.

However, if we sequentially "turn on" each of the effects that limits the independence of the WFY errors, it is seen that the 95% confidence interval increases to ±2.6% when a realistic proportion of the energy errors is assumed to be "fixed" for a particular wind farm, and increases an even greater amount when realistic inter-farm correlations of errors are accounted for, up to our final estimate of ±4.3%. This result illustrates the importance of

including these complex dependency effects as we have done here via the Monte Carlo model, because the simple formula assuming complete independence significantly underestimates the confidence intervals.

Another way to view these results is to use the simple formula above to back out the effective number of independent WFYs based on the larger 95% confidence interval values. Accounting for the "fixed" versus "varying" wind farm errors drops the effective number of independent data points from 127 to 56, and accounting for inter-farm correlation drops it all the way down to 22. This latter drop occurs as a result of very modest inter-farm correlations, which average only 0.1 among all the different wind farm pairings. This highlights the need for continual validation and to make these studies as large as possible to narrow the confidence interval on the results.

The Monte Carlo model can also be used to answer the question, if we want to achieve a target 95% confidence interval on our mean bias error, how large a study do we need to have? This can be answered by running the same Monte Carlo simulation as described above, except with increased numbers of WFYs (and proportionally increased number of wind farms). The results of this experiment are shown in Fig. 7. On the *x*-axis is shown both the number of WFYs and the effective number of independent WFYs based on the dependency effects discussed above. The circle shows the result for the actual size of the Vaisala study previously stated: 127 WFYs and a 95% confidence interval of ±4.3%. Our study is on a relatively steep part of the curve, meaning there is the potential to significantly reduce the 95% confidence interval down to 2% (less than half its current value) by increasing the study by a few hundred WFYs. Beyond that, the confidence interval starts to decrease much less rapidly, not reaching 1% until >2000 WFYs are included.

We used the same set of Monte Carlo simulations to determine the 95% confidence interval on the standard deviation of 1-year errors. The value of the standard deviation for our actual study was found to be 8.8% (Fig. 3). The Monte Carlo simulations indicated that the 95% confidence interval for a validation study with all error dependency effects included was ±1.9%.

Assumption	95% Confidence Interval on Mean Bias Error	Effective Number of Independent WFYs
All WFYs are completely independent of each other.	±1.7%	127 (full data set)
WFYs within a wind farm share a "fixed" error for that wind farm, but wind farms are completely independent from each other.	±2.6%	56
WFYs within a wind farm share a "fixed" error for that wind farm, and wind farms that are within ~1000 km of each other have correlated errors.	±4.3%	22

Table 2. 95% confidence interval on mean bias error, and effective number of independent WFYs, under three different assumptions of WFY dependence.

Figure 7. Plot of the 95% confidence interval for mean bias error, as a function of the number of WFYs used in the Monte Carlo simulations. The orange circle indicates the number of WFYs (127) and estimated confidence interval (±4.3%) for the actual validation study performed by Vaisala.
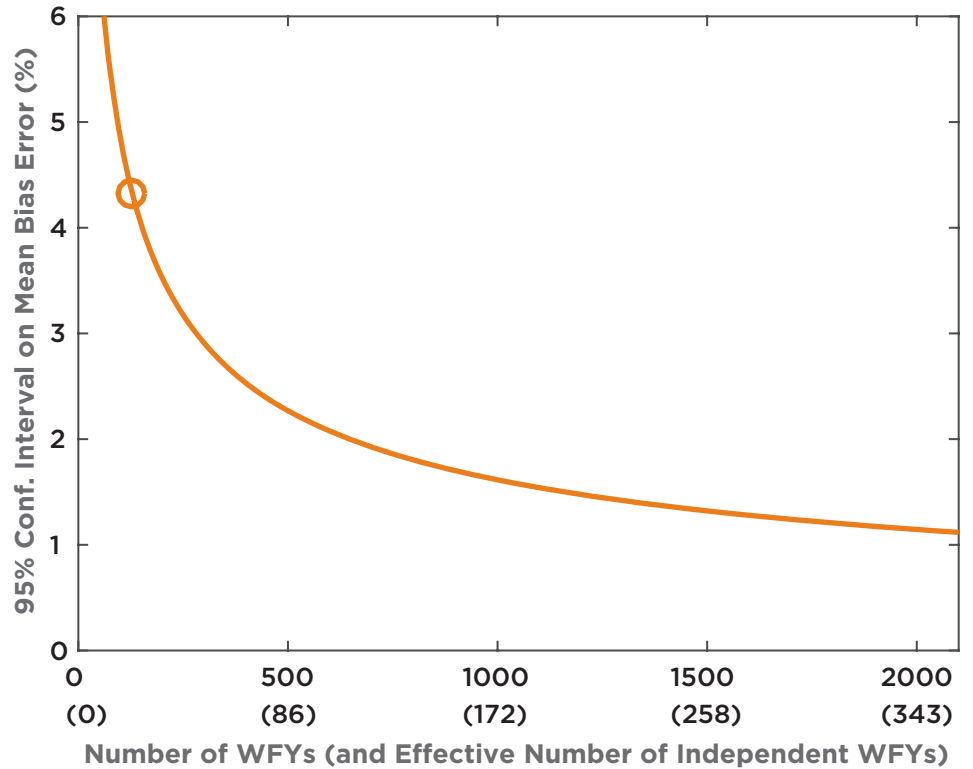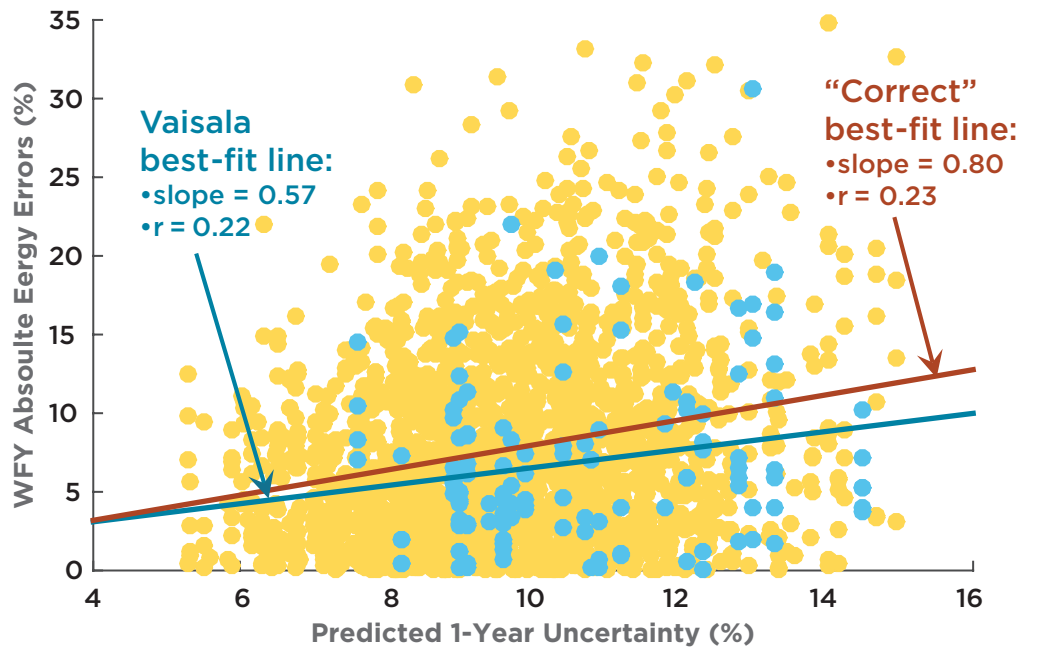

Figure 8. As in Fig. 6, but with the results superimposed from the simulated "super validation study" (2000 WFYs, "perfect uncertainty model"). Slope and correlation of the best-fit lines for both the actual Vaisala validation study (blue dots and line) and the simulated "super validation study" (light orange dots and tan line) are indicated in figure.

6.3 Results For Error Versus Uncertainty

In section 5.3, we presented the "uncertainty validation plot," which was a scatter plot of WFY absolute energy errors on the y-axis versus the pre-construction 1-year uncertainty estimate on the x-axis. Using the Monte Carlo simulator, we simulated a "super validation study" with 2000 WFYs, over 15 times larger than our actual study. We also assumed a perfect uncertainty model. The uncertainty validation plot for this simulated single large study is shown in Fig. 8, and is overlaid with the validation study plot from Vaisala's actual study (the exact same data seen in Fig. 6). The large study illustrates that, even though it reflects the behavior of a "perfect uncertainty model," and has 2000 WFYs, the expected asymmetry of the scatter plot (i.e. larger spread on the right hand side) is only slightly apparent. If one focuses attention on a narrower range of uncertainty, such as between 8 and 12% along the x-axis, the asymmetry is almost not apparent at all, lost in the noise of the random WFY errors. Therefore, it should not be surprising that such plots for studies with much more limited sample size do not readily exhibit a strong visual relationship, even if the uncertainty model has significant skill.

Also discussed in section 5.3 were the benefits of using the slope of the best-fit line to the scatter plot as a metric for how well the uncertainty model captures the spread of the WFY energy errors. The best-fit lines (along with slopes and correlations) for both the Vaisala study and the large study simulation are shown in the plot. Several iterations of the large study were performed, and with this large a study, the slope of the line converges to a consistent value of 0.8 for a perfect uncertainty model. This can also be derived analytically for normally distributed errors, where the expected slope of the best-fit line can be shown to be $\sqrt{2/\pi}$, or 0.8. An important question, then, is what is the likely value of the slope of this line for a more limited sample size such as in Vaisala's or other validation studies performed across the industry?

Returning to the Monte Carlo simulations of a validation study the same size as the one Vaisala performed (127 WFYs), Fig. 9 shows uncertainty validation scatter plots for nine random iterations of the model, which were generated assuming a perfect uncertainty model. Even with a perfect uncertainty model, the slopes of the fitted lines vary considerably. Most have an assortment of different positive slopes, but one is flat (giving the impression of zero skill) and one is even negative (suggesting anti-skill). If 5000 iterations are run, the obtained slopes can be displayed in a histogram (Fig. 10a). It is evident that although the distribution is centered on the expected slope for a perfect uncertainty model (0.8), the tails of the distribution are significant, with a non-negligible portion of iterations producing very small or even negative slopes. The experiment can be repeated by setting the skill factor to lower values, such as 0.5 (where the uncertainty model explains half the variance of the true uncertainty, Fig. 10b), or 0.0 (representing an uncertainty model with no skill, Fig. 10c). As the prescribed skill is lowered, the histogram shifts to the left. But even with half or zero prescribed skill, many iterations still produce slopes near the correct value for a perfect uncertainty model.

These plots highlight the relationship between prescribed uncertainty model skill and the slopes obtained for individual validation studies of a finite sample size. The expectation value of the resulting slope relates directly to the prescribed model skill, but with considerable spread in the distribution. This relationship works in the reverse direction as well: if we perform a single validation study of a given sample size and obtain a value of the slope of the best-fit line in the uncertainty validation plot (as we have done with the actual Vaisala study), this slope should relate to an underlying uncertainty model skill, but in a probabilistic way. A large positive slope does not necessarily translate to definite high skill, nor does a small or negative slope translate to definite low skill.
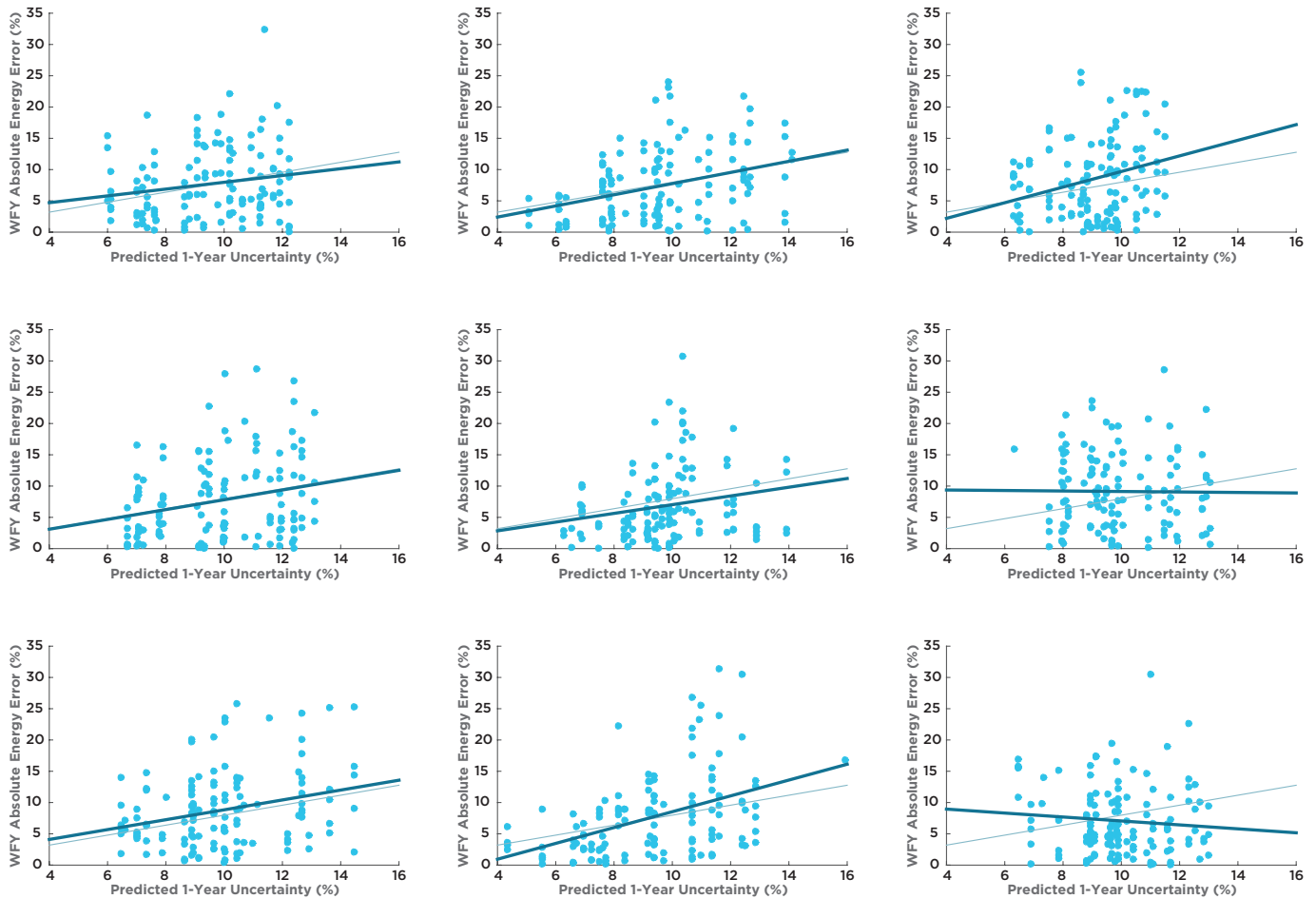
Figure 9. Uncertainty validation plots from nine random realizations from the Monte Carlo validation study simulator, using the study size and distribution characteristics of Vaisala's actual validation study. Each panel follows what is shown in Fig. 6, but for a simulated random "validation study," rather than the actual study result shown in Fig. 6. The Monte Carlo simulation is configured to simulate a "perfect uncertainty model." In each plot, the heavy line is the best-fit line for the data in that plot, whereas the thin line is the "correct" line for a very large validation study and a perfect uncertainty model (slope = 0.80).

To quantify what this probabilistic relationship is, we performed a set of Monte Carlo simulations with 50,000 iterations, in which each iteration used a different value of prescribed uncertainty model skill factor, ranging uniformly between 0.0 and 1.0. For each iteration, we saved the prescribed value of the model skill factor and the obtained slope of the line in the uncertainty validation plot. We then binned the results by the line slope, and analyzed the distribution of underlying model skills in the iterations that produced each binned value of slope. Finally, we repeated this entire process for ten different sizes of validation studies, ranging from 50 to 3200 WFYs.

The results of this experiment are shown in Fig. 11. The left plot indicates the median value of uncertainty model skill factor (*y*-axis) that produced the different values of fitted line slope (*x*-axis). The different curves correspond to the experiments with increasing numbers of WFYs. Increasing the number of WFYs leads to steeper curves, showing that the fitted line slope is a stronger indicator of model skill as the study size increases. The curve labeled "127" WFYs corresponds to the size of Vaisala's study, and the slope obtained in the Vaisala study of 0.57 (shown by the vertical blue line) crosses the 127 WFY curve at median skill factor = 0.51.
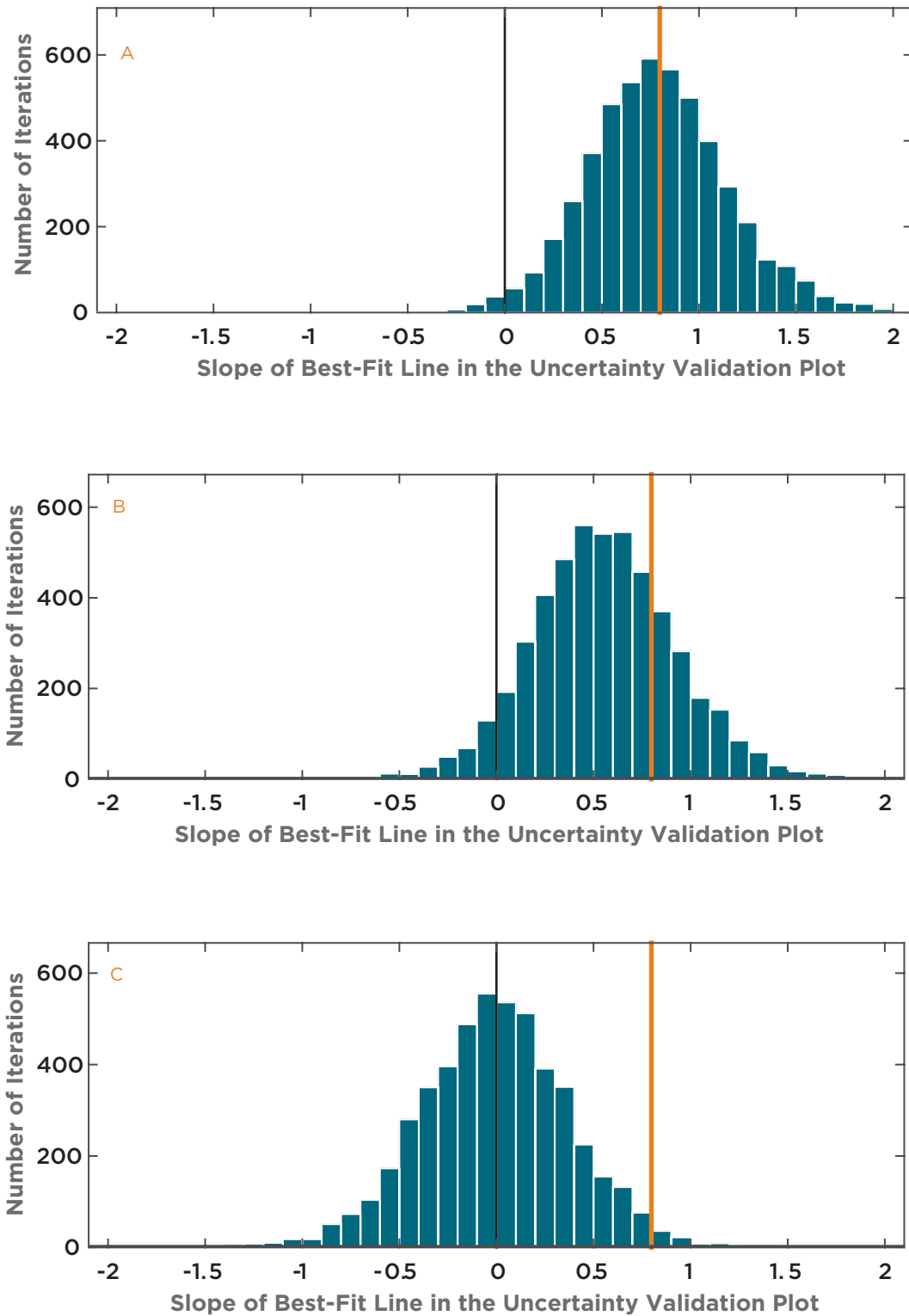
Figure 10. Monte Carlo simulation results showing histograms of the slope of the best-fit line in the uncertainty validation plot (e.g., those in Fig. 9) for three different values of prescribed uncertainty model skill factor: (a) skill factor = 1.0 ("perfect uncertainty model"); (b) skill factor = 0.5 ("half-perfect uncertainty model"); and (c) skill factor = 0.0 ("no-skill uncertainty model"). Orange vertical line shows the slope expected for a "perfect uncertainty model" validation study with a very large number of WFYs ($\sqrt{2/\pi}$, or 0.80).
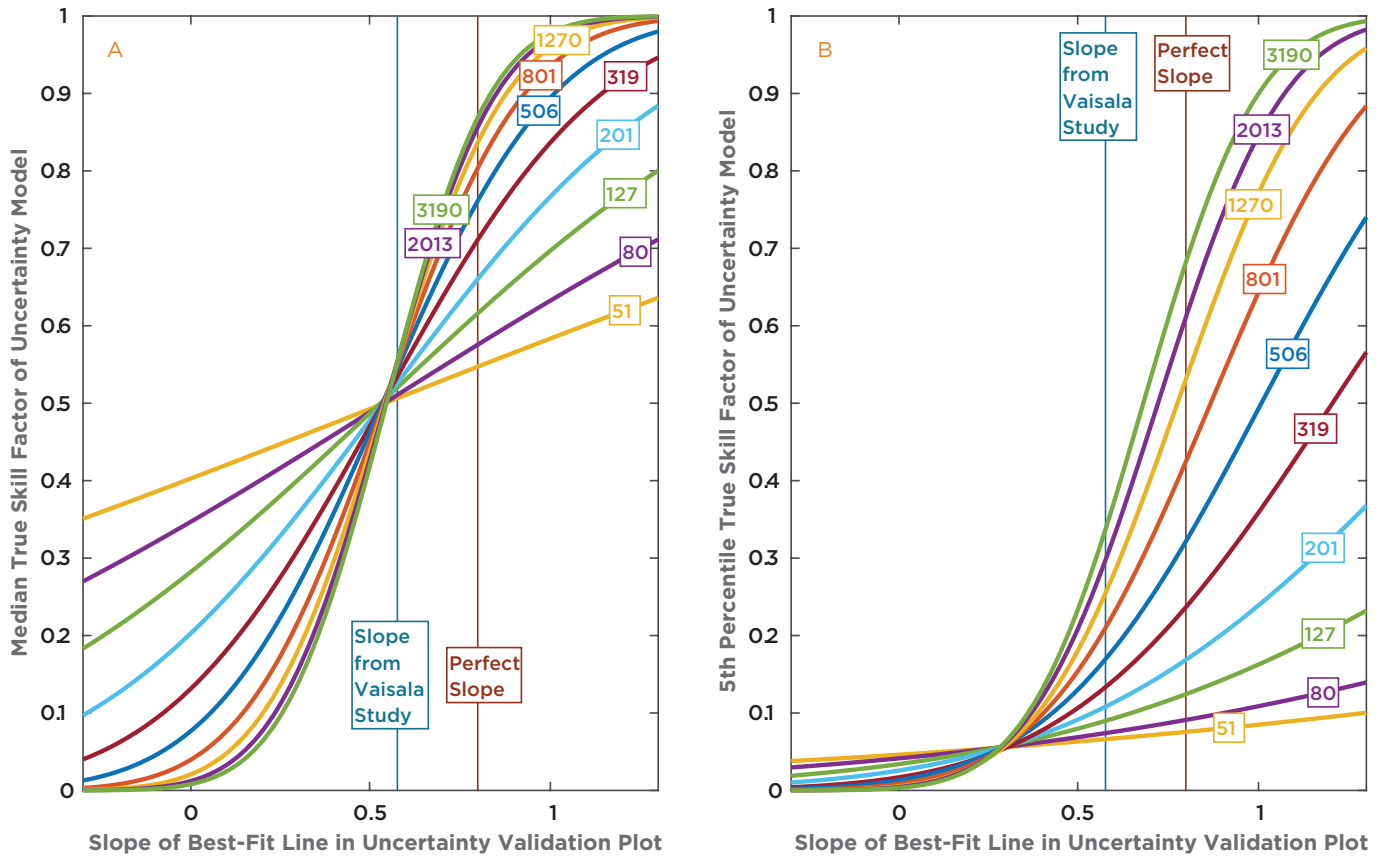
Figure 11. Monte Carlo results showing the statistical relationship of the true skill factor of an uncertainty model (*y*-axis) and the slope of the best-fit line in the uncertainty validation plot from a single validation study (*x*-axis), as a function of the number of WFYs included in the study (different colored curves, labeled with number of WFYs). (a) Median true skill. (b) 5th percentile (or 95%-confidence lower bound) of true skill. Vertical lines indicate the actual slope obtained in this Vaisala validation study (0.57), and the slope expected for a very large study using a "perfect uncertainty model" (0.80).

The right plot is the same, except instead of showing the median value of model skill factor, it shows the 5th percentile value, which can also be described as the 95% confidence lower bound on model skill. This lower bound on model skill for the Vaisala study sits at the intersection of the 127-WFY curve and the vertical line indicating the Vaisala study's fitted line slope, where the skill factor is 0.09. That is a relatively low value within the range of possible values of 0.0–1.0, indicating that there is considerable uncertainty in assessing the skill of the uncertainty model for a validation study of 127 WFYs.

As with the discussion of the 95% confidence interval on the mean bias error (Fig. 7), we can ask the question, how will our confidence in uncertainty model skill improve as the size of the study increases? One unknown here is that if we enlarge the study size, we would get a different random draw, and would likely obtain a different slope from the uncertainty validation scatter plot. But assuming the slope stays the same, if we move up the vertical "Vaisala Study" line in Fig. 11b, from the 127-WFY curve up to the 506-WFY curve, a modest gain of about 0.1 in the 95% confidence lower bound on model skill factor can be obtained. Beyond that, similar increases are possible but only at the expense of exponentially increasing study size (note the number of WFYs increases nonlinearly from one curve to the next). So we could conclude, as with the 95% confidence interval on the mean bias error, that increasing the study size by perhaps a few hundred WFYs would be valuable, but would reach a point of diminishing returns beyond that size.

Another way to interpret this study is to look at it in terms of what it would take to have 95% confidence that an uncertainty model is "half perfect" (an uncertainty model skill factor of 0.5) given a certain study size. In Fig. 11b, looking at where the uncertainty model skill factor of 0.5 intersects the various curves, one can see what slope would be required for various study sizes to be able to make the "half perfect" statement. For a study of 3190 WFYs, one would need to "draw" a slope of 0.68. For a study of 1270 WFYs, one would need to "draw" a near-perfect slope of 0.77 (recall that a "perfect slope" is 0.80). For a study of 506 WFYs, one would need to "draw" a beyond-perfect slope of 1.01 to have 95% confidence that the study is at least "half perfect."

7 SUMMARY

Using a dataset of 127 years of commercial wind farm operation from 30 different wind farms, mostly within the United States, Vaisala has conducted a significant validation study of its due diligence wind energy assessment methodology. Pre-construction assessments were performed with Vaisala's current methodology in a blind retrospective forecast framework. These assessments of long-term net annual energy production (the P50 AEP) were compared with yearly energy production values from the participating wind farms, yielding a set of 127 wind farm year (WFY) percent energy errors (actual energy generated minus long-term prediction). The mean bias error across the entire study was +0.1%, with a 95% confidence interval of ±4.3%. Plants slightly overgenerated, compared to our estimate. The standard deviation of the 1-year errors was 8.8%, somewhat lower than Vaisala's mean estimated 1-year uncertainty on energy of 10.6%, indicating that Vaisala is slightly more accurate than is given credit for in our current uncertainty model. However, the mean estimated 1-year uncertainty lies just slightly within the 95% confidence interval for the standard deviation of 1-year errors of 8.8% ±1.9%. Adjusting the results for annual windiness and annual system availability narrowed the width of the error histograms commensurate with the implied reduction in mean uncertainty. These two adjustments also led to slight (1% each) shifts toward negative mean bias error, but the overall result was still well within the 95% confidence interval on our estimate of the mean bias error.

A Monte Carlo model was developed that enabled simulation of thousands of validation studies with realistic uncertainty properties. The Monte Carlo model was used to establish the 95% confidence interval stated above. It was also used to explore the nature of the relationship between pre-construction estimates of uncertainty and the yearly wind energy errors from those estimates. A simple metric for quantifying this uncertainty/error relationship was defined, which is the slope of the best-fit line on a scatter plot of absolute WFY errors versus 1-year uncertainty estimates. It was found that the slope of such a line, while having an expected value of 0.80 for a very large study with a perfect uncertainty model, can vary considerably due to random "luck of the draw" with a finite sized validation study. The slope in Vaisala's study was 0.57, and this slope implied a median "uncertainty model skill factor" of 0.51. Here, the uncertainty model skill factor is defined as the fraction of variance among wind farms of 1-year uncertainty explained by the uncertainty model, with 1.0 indicating perfect skill, and 0.0 indicating no skill. However, the range of possible "skill factors" of the uncertainty model implied by this slope is very broad, given the sample size of 127 WFYs. The 95% confidence lower bound on skill factor for the Vaisala study was only 0.09, reflecting the difficulty in pinning down uncertainty model skill for validation studies of the size Vaisala conducted.

Finally, in terms of both the 95% confidence interval on mean bias error and the 95% confidence lower bound on the inferred skill of the uncertainty model, the Monte Carlo results indicated that larger validation studies will only improve confidence intervals. Continual validation of our methods and collecting as much data as possible is required to have confidence that our models maintain calibration with actual production and are required to be able to stay abreast of changing technologies and newer methodologies that improve accuracy.

8 REFERENCES

1. DNV/GL, 2014: Wind project performance white paper, actual versus predicted: 2014 update. White paper available from DNV/GL at http://www.gl-garradhassan.com/assets/img/content/DNV_GL_Wind_Power_Project_Performance_White_Paper.pdf.

2. AWS TruePower, 2009 (and 2012 update): Closing the gap on underperformance: A review and calibration of AWS TruePower's Energy Estimation Methods. White paper available from AWS TruePower at https://www.awstruepower.com/knowledge-center/technical-papers.

3. Istchenko, R. (WSP), 2014: Wind resource assessment uncertainty validation. Presentation given at AWEA Wind Resource and Project Energy Assessment Seminar, Orlando, FL, December 3, 2014.

4. Poulos, G., 2013: Investigating the wind flow modeling experiment. Presentation at the AWEA Wind Resource and Project Energy Assessment Seminar, Las Vegas, NV, December 10, 2013.

5. Bernadett, D., 2014: Which to choose? Comparing the performance of wind flow models in varied conditions. Presentation given at AWEA WindPower 2014 Conference and Exhibition, Las Vegas, NV, May 7, 2014