

# Understanding Biases in Pre-Construction Estimates

Monte Lunacek, M. Jason Fields, Anna Craig, Joseph C. Y. Lee,  
John Meissner, Caleb Philips, Shuangwen Sheng, and Ryan King

National Renewable Energy Lab, Golden, CO. USA

E-mail: monte.lunacek@nrel.gov

## Abstract.

The pre-construction energy generation of a wind farm (P50) is difficult to estimate and evaluate. This paper presents a methodology to measure the accuracy of the p50 prediction, which we call the Historical Validation Survey (HVS), for several wind farms in the continental United States. Our results indicate that there is a bias between predicted and measured energy, even when controlling for factors like grid curtailment and resource variability. We also find that our results depend on the assumptions we make during analysis, which we quantify with a sensitivity analysis. This method allows the estimation of uncertainty we have in our findings. When we account for reasonable ranges of model assumptions, we find that, in the most optimistic case, there is still a bulk  $-5.5\%$  bias when estimating pre-construction energy generation. When controlling for grid curtailment this number reduces to a range of  $-3.5$  to  $-4.5\%$ .

## 1. Introduction

Exceedance probabilities are often used in the wind industry to describe the expected Annual Energy Production (AEP) for a potential site [1]. The uncertainty associated with this estimate forms the foundation of the investment business risk model for wind plants [2]. Over estimating the pre-construction AEP can create financial loss for wind farm owners and investors. At the same time, the wind energy industry as a whole lacks standardization for verifying pre-construction estimates [2]. We started the Historical Validation Survey (HVS), as part of the U.S. Department of Energys Performance, Risk, Uncertainty and Finance (PRUF) project, to provide an open and transparent methodology that quantifies the average biases in the current wind resource assessment process.

In this paper, we compare monthly production data, which we obtained from the Energy Information Administration (EIA), with the pre-construction p50 estimate that we collected from 56 projects in the United States. Our results indicate that there is a significant bias between predicted and measured energy, even when controlling for factors like commission date and grid curtailment. We also find that the results are highly dependent on the exogenous variables we use during our analysis, which we account for with a sensitivity analysis over a range of reasonable values. This allows us to estimate the uncertainty we have in our findings. Even over a broad range of assumptions, our results show that the pre-construction production estimates are higher than the actual production values by  $6.7\%$ .

Around 2010 the industry became increasingly aware of an overprediction bias [3] and has since taken steps to improve resource assessment practices. These efforts have culminated in efforts to improve prediction accuracy through a bias correction and accounting for additional



sources of potential energy loss. Projects in this analysis that started after 2011 show progress in reducing this bias, but are still overestimating production by 5.5%.

This paper is organized in the following way. In the next section, we describe the methodology we use to measure bias while controlling for the uncertainty in our model. This includes how we address outliers in the EIA data and control for the experienced, but unpredictable extremes of the wind resource. Then in Section 3, we discuss our results, under a variety of assumptions and filters, before concluding the paper in Section 4.

## 2. Method

The P50 estimate is the average annual energy production value that a wind farm should be expected to produce during its lifetime [2]. Our primary goal is to compare the p50 estimates with the corresponding wind farm annual production data, in order to calculate bias. The raw EIA data, however, is not fully representative of the values of interest. Specifically the EIA includes cases of extreme curtailment and availability driven downtime, which are typically outside the scope of the p50 prediction. We also need to account for extreme periods of below- or above-average windiness at the wind farms.

In this section, we describe three ways that we compare the EIA data with the p50 estimate: 1) using the raw data, 2) correcting outliers in the raw data, and 3) estimating 20-year production based on a full range of external wind data. We refer to each of these as raw, outlier corrected, and long-term corrected respectively. We also discuss the assumptions we make when using these methods and how we measure the uncertainty in our results.

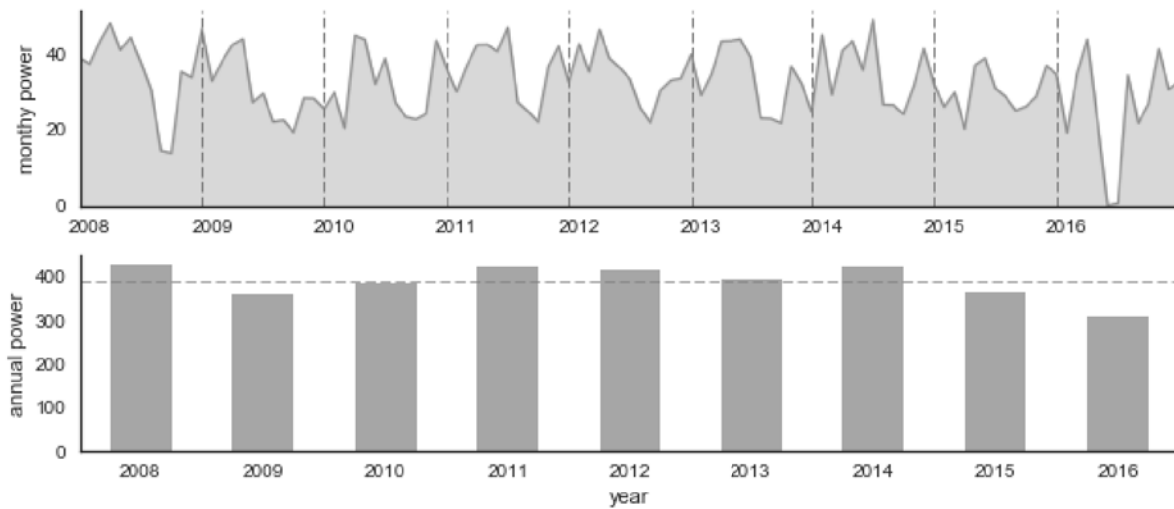
### 2.1. Comparison methods

Comparing the raw EIA data with the P50 estimate is the most straightforward and precise analysis, as it requires minimal assumptions. Starting with the most recent data, we form groups of 12 months, and the sum of these months becomes the set of annual values for which we compute the percent difference between each year in the set and the P50 value. Figure 1 shows this process graphically. The percent difference makes it possible to compare wind farms of different magnitudes and different number of years.

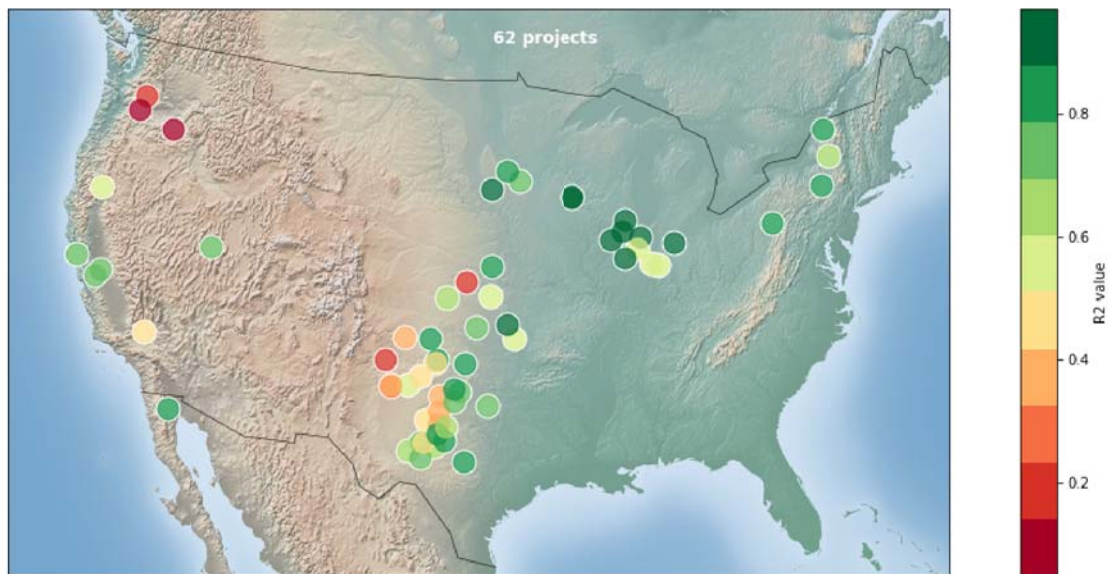
One way to make the comparison with EIA data more fair is to correct months that are likely outliers in the data. We do this by comparing each plants monthly production data against monthly wind data, and then correct points that are statistical outliers. Our second analysis method, *outlier corrected*, uses a simple linear model that estimates monthly power production based on average monthly wind speed. Points that are outside a portion of the standard deviations from the estimate are considered outliers and are corrected by replacing them with the models estimate, which may include some variation based on the original data.

We estimated the wind using several different wind datasets, which we discuss later in this section, and assign each plant to the closest wind measurement location. This typically provides a reasonable estimate for locations where the EIA and wind data are close, or where the wind plant is in a region with consistent wind (such as the middle of the country), but may be a poor estimate otherwise. Figure 2 shows the projects in this analysis colored by the correlation between wind and EIA data.

Some wind plants only have a few years of data. Our third method, *long-term corrected*, estimates the wind plants production for the last 20 years by filling in missing values with an estimate from a linear model. This method uses a combination of outlier-corrected values for the years we have data, and combines this with imputed data from the linear model for the remaining years.



**Figure 1.** Converting EIA data into annual values for a single project. The monthly values (top graph) are grouped into a set of annual values (lower graph). The percent difference between the set of annual values and the P50 estimate creates a set of percent difference values for a specific project.

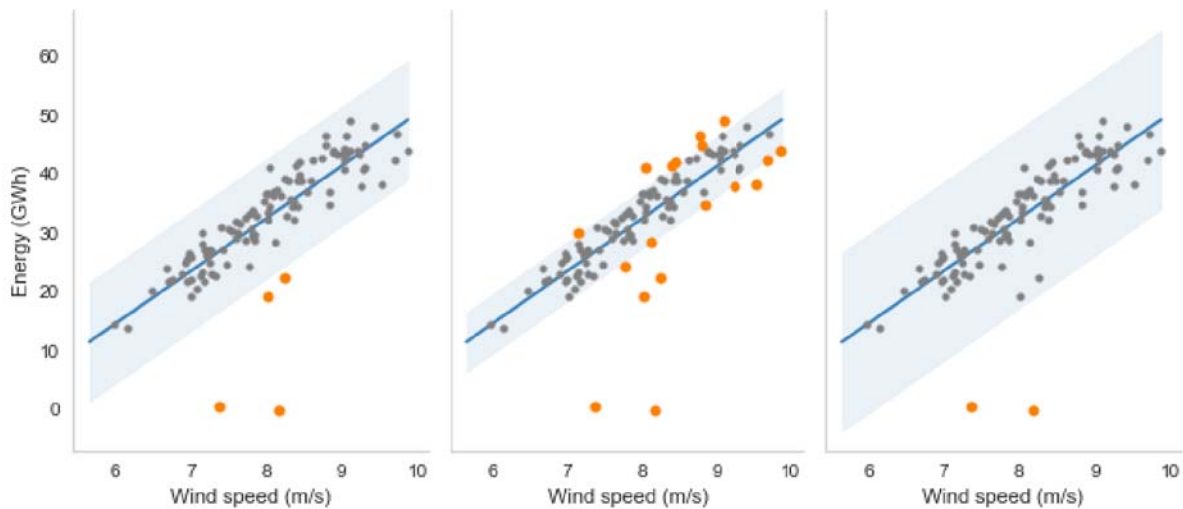


**Figure 2. Project locations and wind correlation:** The 62 projects we have P50 estimates for are shown above colored by how well the EIA data correlates with the wind data. Some of the projects have a much stronger correlation.

### 2.2. Model Uncertainty

Each type of comparison described above has a set of exogenous variables that impact the final bias. We account for the uncertainty in these parameters by sampling from a distribution of reasonable values and then assessing whether each parameter impacts our final bias or not.

We apply two filters that remove certain wind farms from the set of projects we analyze. The



**Figure 3.** Outlier detection based on the *upper* and *lower* bounds of the linear model. The left-most graph uses two standard deviations from the model to detect outliers ( $\sigma = 2$ ). This method can be made less flexible (middle figure,  $\sigma = 1$ ) or more forgiving (right-most figure,  $\sigma = 3$ ).

first year of data may not be representative of a wind farm’s actual production, so we add this as a condition in our reporting. We also impose a minimum number of months that a project must have in order to be included. For this paper we consider 12 or 24 months to be the minimum.

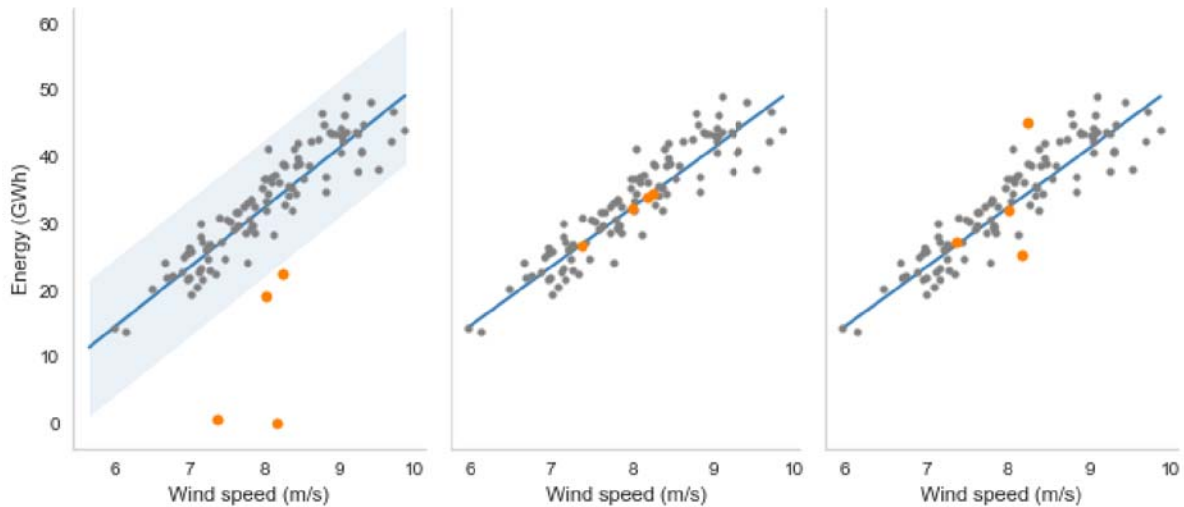
While the filters apply to all the comparison methods, there are a set of variables that only impact the *corrected* and *long-term* comparisons. In these methods, the linear model may have a large impact on the final result and we need to understand the variability in our choices. If a project has  $n$  months, then we bootstrap the linear model by selecting  $n$  months from this set with replacement [4]. This will cause some variation in the linear model and the overall results will be less impacted by any single month. We only use this technique to fit the linear model; once the model is fit, all the points are still corrected and extrapolated as described below.

Figure 3 shows how our choice of an upper and lower bound of the standard deviation impact the data. Larger bounds correct fewer points. Of the parameters we can assign values to, the *upper-* and *lower-bound* parameters have the most impact on our final results. We sample these parameters in the range of  $[2, 4]$ , which represents a standard deviation from the linear model.

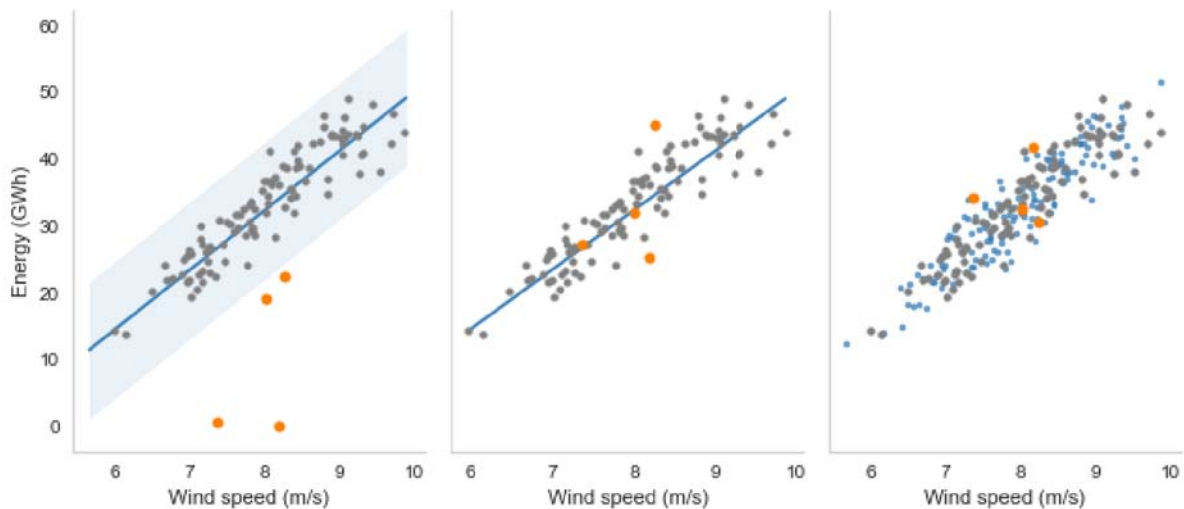
The outliers can be corrected with some degree of noise. We call this parameter *noise*. When the value is zero, the corrected point is placed directly on the linear model. The larger the positive value for *noise* is, the larger the variation is from the linear model. Our simulations show that this parameter does not have a significant impact on our results. Figure 4 graphically explains the impact of the *noise* parameter. We sample this parameter in the range of  $[0, 2]$ .

When we estimate the long-term corrected values we can add noise using a method similar to the corrected noise, *noise*. We refer to this parameter as *wnoise*. Figure 5 shows the impact of this parameter. Like the *noise* parameter, we show that this attribute does not have a significant impact on our overall results when sampled in the range of  $[0, 2]$ .

The final variable we sample is the wind data for each project. There are four different wind datasets we sample from: NASA’s MERRA2 reanalysis dataset [5], MERRA2 with a derived the monthly average wind speeds at 80 meters above the surface, the ERA 80 meter data [6], and a final data set that is a mixture of the previous three, which we refer to as optimistic. The optimistic dataset assigns each project the wind dataset that has the highest correlation with



**Figure 4.** Outlier correction based on the *noise* parameter. The left-most graph shows the original data with detected outliers. The middle graph uses a noise value of zero to place, or correct, the location of the outliers to the linear model. The right-most graph places the outliers near the linear model at a distance based on a positive value for *noise* ( $noise = 0.5$ ).



**Figure 5.** Windiness correction projects the model back in time to estimate the output of a wind farm over the last 20 years. We can do this without noise (middle graph,  $noise = 0.5$  and  $wnoise = 0.0$ ) or with noise (right-most graph,  $noise = 0.5$  and  $wnoise = 1.0$ ).

it's production data. We refer to these datasets as [ merra, merra80, era, opt]. The table 1 summarizes the exogenous variables that we control during our analysis.

### 2.3. Assumptions

In this analysis, we assume that the EIA data is a good estimate of the wind plant's production. In some instances, we may have a limited amount of data and are therefore assessing a wind plants performance based on limited information. The outlier and long-term corrected analysis

Parameter	Values	Description
Remove first year	Boolean	Remove first 12 months of data.
Ensure	12 or 24	Minimum number of months needed.
Bootstrap	Boolean	Select $n$ months with replacement.
Upper bound	[2, 4]	Upper standard deviation for outlier correction.
Lower bound	[2, 4]	Lower standard deviation for outlier correction.
Corrected noise	[0, 2]	Noise standard deviation for correcting outlier points.
Long-term noise	[0, 2]	Noise standard deviation for long-term projection points.
Wind	1 of 4	Specific wind dataset to use.

**Table 1.** The exogenous variables in our model. The values represent reasonable conditions for each parameter. The wind variable is one of the four data sets we curated.

make two additional assumptions: that the wind data is a close proxy for the actual monthly wind at the site and that the linear model is a reasonable estimate of how wind and energy are related.

In some of our results we only include wind farms that have a stronger than 0.7 coefficient of determination ( $r^2$ ) between energy production and wind speed. In other words, if there is a weak correlation between power and wind then we discard the wind farm from our results. We do this to ensure that we are not correcting or projecting values in our model that are based on wind speeds that may not be indicative of the wind located at the plant.

#### 2.4. Metrics

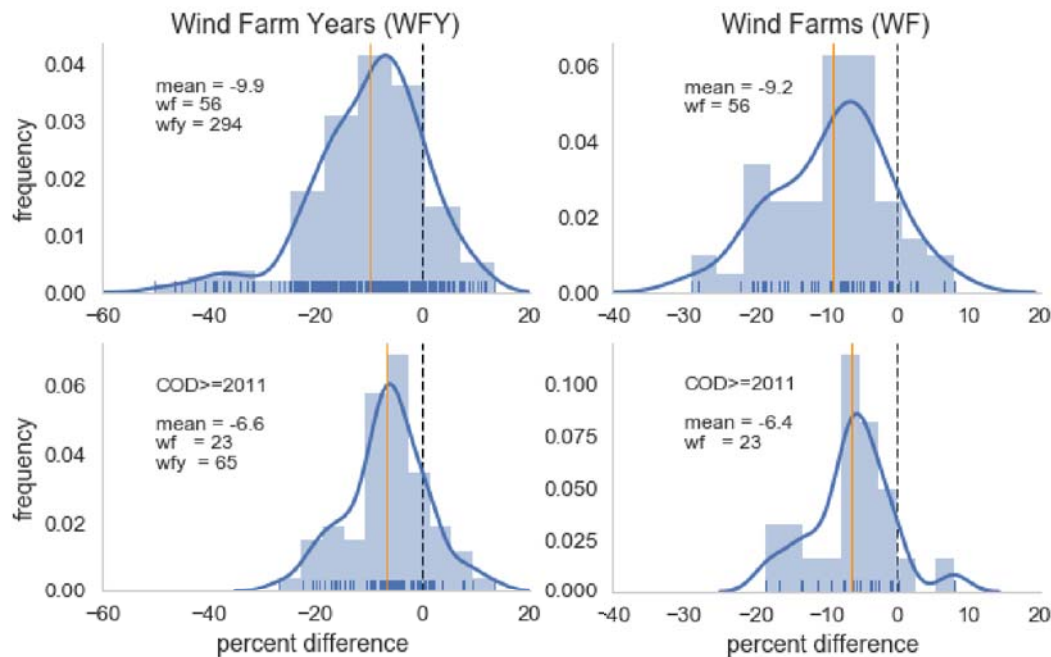
We assess bias in two ways. We define a Wind Farm Year (WFY) as any given annual data point treated as an independent sample. A Wind Farm (WF), on the other hand, is the aggregate of WFYs for any individual wind farm. For each WFY, we calculate the percent difference between the EIA energy production data and the P50 estimate, which illuminates bias across all WFYs. We also investigate the overall bias for any single project in our sample (e.g. each WF).

### 3. Results

Our results indicate that there is a significant bias when looking at all projects across the range of analysis methods we developed for the HVS project. The raw analysis is only impacted by two of the exogenous variables in Table 1: remove first year and minimum duration. This is because these inputs act as filters on the data.

To quantify the uncertainty in this approach, we sampled 13,000 iterations of the method described in the previous section by randomly selecting values from Table 1. Some of these variables have an observable impact on our final results and others were not significantly different. Our primary goal, however, is to show that over a large range of appropriate values, we still have a significant bias in P50 estimates.

We performed an Analysis of Variance (ANOVA) on each variable group in our sample. Removing the first year improved all results and as did enforcing a minimum duration of 12 months. If we excluded the Merra 50m data set, there was not a significant difference in the wind dataset used for the analysis. This makes sense because the other datasets are estimated at the 80 meter level. The noise parameters were significantly different at the wind farm level (WF), but were not significantly different at the WFY metric. The values at the WF level, while different, were close in mean. Finally, the choice of upper and lower bound for removing outliers has a significant impact on our bias result. The higher values for the upper bound created a smaller bias. Lower values for the lower bound created a smaller bias. This makes sense: we



**Figure 6. Raw results:** The mean bias for WFYs and WFs is  $-9.9$  and  $-9.2$  respectively. The bias drops to  $-6.6$  and  $-6.4$  for projects that started after 2011.

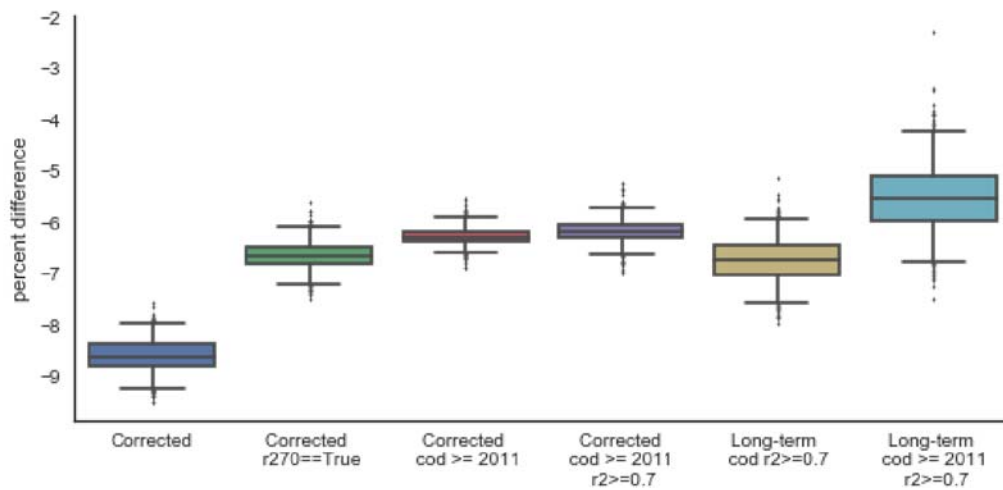
leave large outliers alone, but correct, or improve, marginal outliers on the lower section (e.g. increase the power).

The results we present here are more optimistic than the general case in that we removed the first year and enforced a minimum of only 12 months. In some instances, we look at only projects that started after 2011, which we label  $COD \geq 2011$ , and projects that have an  $r^2$  value greater than 0.7, which we label  $r^2 \geq 0.7$ . This is because we are using the linear model to correct or project points and this ensures that our model is a reasonable estimate.

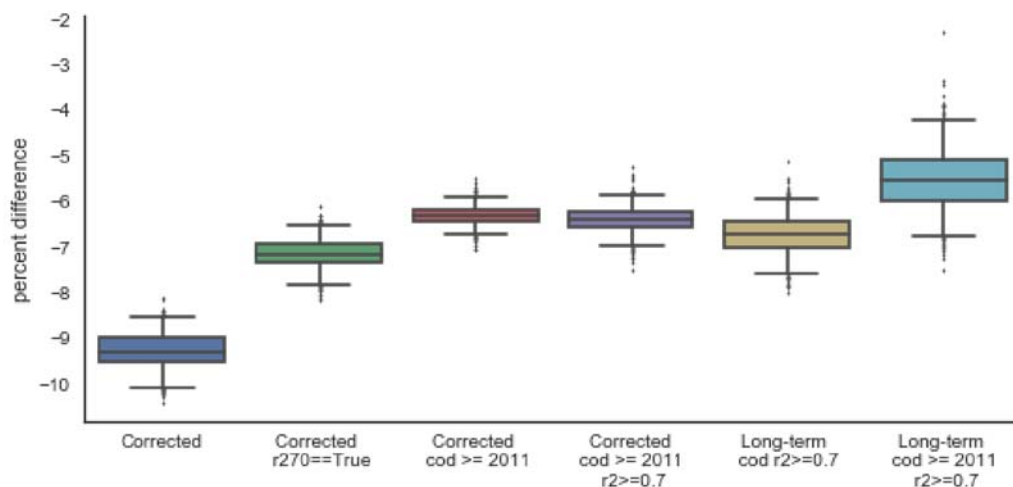
Figure 6 shows a histogram of the raw results based on WFY (left) and WF (right). This represents a single interaction of our model because we have removed the first year and enforced a minimum of 12 months. No other variables impact the raw bias. The mean bias for WFY and WY is  $-9.9$  and  $-9.2$  respectively. For projects that have started after 2011, the bias drops to  $-6.6$  and  $-6.4$  for WFY and WF respectively. Notice that the distribution of WF has less variance because its value is the average of all the WFY in that project.

The corrected and long-term analysis depend on the variables of our iterative sampling. Figures 7 and 8 show the distribution of final bias results for a variety of filters. In both instances, we improve by filtering projects that have an  $r^2 \geq 0.7$  and started after 2011. The long-term correction is an improvement over the corrected analysis when filtering both the  $r^2$  values and the  $COD \geq 2011$ . This improvement in the post-2011 results show progress in the wind industry towards more accurate pre-construction estimates.

The main observation in this paper is this: Figures 7 and 8 show that even over a variety of reasonable exogenous variable settings, correcting outliers, and accounting for long-term performance using four different wind datasets, we still have a significant bias. The WFY and WF long-term corrected bias, with the  $r^2$  values filtered is  $-6.7\% \pm 0.8\%$ . If we only look at wind farms since 2011 we improve, but are still over estimating by  $-5.5\% \pm 1.28\%$ .



**Figure 7. Wind Farms:** Average bias aggregated by Wind Farm. Even limiting to recent, post 2011, wind farms and ensuring a reasonably high correlation with the wind model used, we still find a 5.5% negative bias in the p50 prediction.



**Figure 8. Wind Farm Years:** Average bias aggregated by all the years combined, from each wind farm, as a single set. We observe a similar trend as with Figure 7 which is aggregated at the wind farm level.

#### 4. Conclusions and Future Work

In this paper we provide a methodology for evaluating the bias in pre-construction estimates that accounts outliers, long-term wind impacts, and an uncertainty in our model parameters. When we account for wind resource variability, we calculate a bias of  $-6.7\% \pm 0.8\%$ . Projects that have started after 2011 have improved since the industry became aware of the over-prediction bias in 2010. Here the calculated bias reduces to  $-5.5\% \pm 1.28\%$ . If we assume that grid curtailment is around 1% – 2%, then we still have a negative bias problem. We have applied this method to projects in the Continental United States, but we believe this method is applicable elsewhere.



We conclude that systematic negative bias of pre-construction energy production is evident in the EIA data regardless of the date of commission, the locations of wind farm, or the type of analysis used to reach the conclusion. These benchmarking results lay the foundation for the overall PRUF project and provide a methodology for conducting an evaluation of pre-construction wind farms, which includes uncertainty quantification of the overall results.

It is clear from this analysis that data quality and assumptions play an important role in the conclusions we draw from our analysis. Future work will utilize a more robust data set and address a larger set of assumptions to understand the sensitivity of this greater set of inputs.

## 5. References

- [1] A. Dobos, P. Gilman, and M. Kasberg, "P50/p90 analysis for solar energy systems using the system advisor model," in *2012 World Renewable Energy Forum*, 2012.
- [2] A. Clifton, A. Smith, and M. Fields, "Wind plant preconstruction energy estimates. current practice and opportunities," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2016.
- [3] K. Dickinson, L. Delle Monache, K. McCormack, and P. Magontier, "Wind energy resource assessment: Information production, uses, and value-survey report," 2014.
- [4] B. Efron and R. J. Tibshirani, "An introduction to the bootstrap," No. 57 in *Monographs on Statistics and Applied Probability*, Chapman & Hall, 1993.
- [5] M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G.-K. Kim, *et al.*, "Merra: Nasas modern-era retrospective analysis for research and applications," *Journal of climate*, vol. 24, no. 14, pp. 3624–3648, 2011.
- [6] D. P. Dee, S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, *et al.*, "The era-interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the royal meteorological society*, vol. 137, no. 656, pp. 553–597, 2011.