


RESEARCH ARTICLE

Data assimilation impact of in situ and remote sensing meteorological observations on wind power forecasts during the first Wind Forecast Improvement Project (WFIP)

James M. Wilczak¹  | Joseph B. Olson² | Irina Djalalova² | Laura Bianco² | Larry K. Berg³ | William J. Shaw³ | Richard L. Coulter⁴ | Richard M. Eckman⁵ | Jeff Freedman⁶ | Catherine Finley⁷ | Joel Cline⁸

¹Physical Science Division, NOAA Earth System Research Laboratory, Boulder, Colorado, USA

²Cooperative Institute for Research in Environmental Sciences at the NOAA Earth System Research Laboratory, University of Colorado, Boulder, Colorado, USA

³Pacific Northwest National Laboratory, Department of Energy, Richland, Washington, USA

⁴Argonne National Laboratory, Department of Energy, Lemont, Illinois, USA

⁵Air Resources Laboratory, National Oceanic and Atmospheric Administration, Idaho Falls, Idaho, USA

⁶AWS Truepower, Albany, New York, USA

⁷WindLogics Inc, St. Paul, Minnesota, USA

⁸Energy Efficiency and Renewable Energy, Department of Energy, Washington, D.C., USA

Correspondence

James M. Wilczak, NOAA/Earth System Research Laboratory, 325 Broadway, Boulder, Colorado 80305, USA.

Email: james.m.wilczak@noaa.gov

Present Address

Jeff Freedman, Atmospheric Sciences Research Center/University of Albany, State University of New York, Albany, New York, USA.

Catherine Finley, Department of Earth and Atmospheric Sciences, St. Louis University, St. Louis, Missouri, USA.

Joel Cline, NOAA/National Weather Service, Washington D.C., USA.

Abstract

During the first Wind Forecast Improvement Project (WFIP), new meteorological observations were collected from a large suite of instruments, including wind velocities measured on networks of tall towers provided by wind industry partners, wind speeds measured by cup anemometers mounted on the nacelles of wind turbines, and wind profiles by networks of Doppler sodars and radar wind profilers. Previous data denial studies found a significant improvement of up to 6% root mean squared error (RMSE) reduction for short-term wind power forecasts due to the assimilation of all of these observations into the National Oceanic and Atmospheric Administration (NOAA) Rapid Refresh (RAP) forecast model using a 3D variational data assimilation scheme. As a follow-on study, we now investigate the impacts of assimilating into the RAP model either the additional remote sensing observations (sodars and radar wind profilers) alone or assimilating the industry-provided in situ observations (tall towers and nacelle anemometers) alone, in addition to routinely available standard meteorological data sets. The more numerous tall tower/nacelle observations provide a relatively large improvement through the first 3 to 4 hours of the forecasts, which diminishes to a negligible impact by forecast hour 6. In comparison, the sparser vertical profiling sodars/radars provide an initially smaller impact that decays at a much slower rate, with a positive impact present through the first 12 hours of the forecast. Large positive assimilation impacts for both sets of instruments are found during daytime hours, while small or even negative impacts are found during nighttime hours.

KEYWORDS

data assimilation, forecasting, observations, wind power

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2019. This article is a U.S. Government work and is in the public domain in the USA. Wind Energy Published by John Wiley & Sons, Ltd.

Funding information

National Oceanic and Atmospheric Administration; Office of Energy Efficiency and Renewable Energy, Grant/Award Numbers: DE-AC02-06CH11357 and DE-AC05-76RL01830; US Department of Energy (Office of Science, Office of Basic Energy Sciences and Energy Efficiency and Renewable Energy, Solar Energy Technology Program)

1 | INTRODUCTION

Because of the variability of wind and solar energy and uncertainties in our ability to forecast how much energy they will provide at any given moment, forecasts for these renewable energy sources can alter the cost of energy, with improvements to these forecasts having the potential to reduce costs. For example, Jonsson et al¹ demonstrated that wind power forecasts have a considerable impact on day-ahead electricity spot market prices, while Gonzalez-Aparicio and Zucker² showed that the variability of wind power forecast errors highly influences fluctuations of the intraday (11-20 h) price of electricity. McGarrigle and Leahy³ showed that for the Irish grid system with an anticipated 33% wind penetration in 2020, for the day-ahead market, a reduction of wind energy forecast mean absolute error from 8% to 4% would reduce the total system cost of energy between 0.5% and 1.6%, with a 9% reduction in wind curtailment and a corresponding reduction in gas turbine electricity generation. Similarly, Brancucci Martinez-Anido and Hodge⁴ showed that a 25% reduction in solar forecast errors in the day-ahead market with an 18% solar penetration level would reduce electricity generation costs by \$0.50 per megawatt hour of solar power generation. These studies focused on the day-ahead or intraday markets, where cost savings can be analyzed in the unit commitment process. Evaluating the economic value of shorter-term forecasting is more challenging due to the complexity of applying an accurate monetary value to system reliability. In one of the few studies to investigate the economic value of shorter-term wind forecasts, Hodge et al⁵ evaluated the cost savings of improving ultra-short-term (less than 40 min) forecasts on the CAISO system and found that for a presumed future scenario with 25% wind energy penetration, a 50% forecast skill improvement would result in annual savings of \$146 million.

With the goal of improving wind energy forecasts in the 0- to 6-hour time frame, the US Department of Energy (DOE) sponsored the first Wind Forecast Improvement Project (WFIP). This was a public-private research program, with partners that included the National Oceanic and Atmospheric Administration (NOAA), private forecasting companies (WindLogics and AWS Truepower), DOE national laboratories, grid operators, and universities.⁶ One path towards improving the skill of wind power forecasts based on numerical weather prediction models is to improve the initial model states in those forecasts. WFIP evaluated this approach through the collection of special observations that were assimilated into weather forecast models. The new observations were collected during concurrent yearlong field campaigns in two high wind energy resource areas of the United States and included 12 radar wind profilers, 12 sodars, 184 instrumented tall towers (nominally 60 m high), and over 400 nacelle anemometers.^{6,7} WFIP focused independently on two geographic regions: (a) the Northern Study Area (NSA), comprising the Dakotas, Minnesota, and Iowa (Figure 1, top panel), and (b) the Southern Study Area (SSA), located in west-central Texas (Figure 1, bottom panel). Because the primary interest was on short-term forecasts up to 6 hours, the principal NOAA models used during WFIP were the hourly updated 13-km resolution Rapid Refresh (RAP), the 3-km resolution High-Resolution Rapid Refresh (HRRR), and the North American Mesoscale (NAM) 4-km nest, although only the RAP model is used in the present analysis. Although the primary goal was to improve forecasts out to 6 hours, the models have been evaluated to 15 hours to show potential benefits at longer forecast horizons.

In our previous study,⁶ the combined impact of assimilating the instrumented tall tower, turbine nacelle anemometer, and sodar and radar wind profiler observations into the 13-km resolution NOAA RAP model was determined through a set of data denial (DD) experiments. The DD experiments included six episodes covering all four seasons, each approximately 9 days in length, for a total of 55 days of simulations with 24 separate simulations per day. The results demonstrated a positive impact due to the assimilation of the WFIP observations of up to 6% improvement of relative root mean squared error (RMSE) wind power for forecast hour 1 and approximately 3% improvement when averaged over forecast hours 1 to 6. Even larger improvements (6% averaged over forecast hours 1-6) were found when evaluating the spatially aggregate power averaged over domains as large as 500 × 600 km, which may be of more relevance for grid balancing. Since the characteristics of the in situ and remote sensing instruments are very different and because the in situ data could be routinely provided by industry groups for assimilation in operational weather forecast models at relatively little cost, we now investigate the impacts separately for each of the two data types. Questions that this study attempts to address are as follows: What are the relative benefits of the in situ versus remote sensing observations? Are they redundant or complementary? Do they provide similar improvements to model forecast skill? Is it potentially worth investing in the more expensive remote sensing observations?

2 | METHODOLOGY

Of the six episodes used in the previously reported WFIP all-or-nothing DD study, two were selected for this current study. The first was in the winter, spanning 9 days between January 7 and January 15, 2012. This episode was selected because of the occurrence of two strong cold fronts

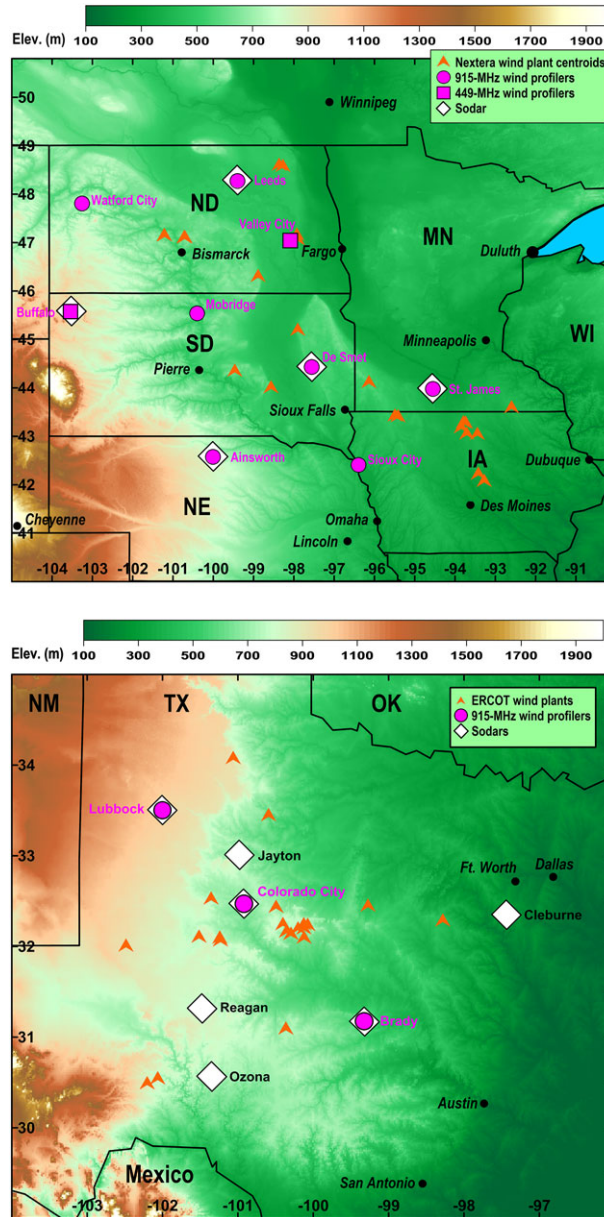


FIGURE 1 Base maps of the Northern Study Area (NSA) (top panel) and the Southern Study Area (bottom panel), showing the topography and the locations of the wind plants, radar wind profilers, and sodars [Colour figure can be viewed at wileyonlinelibrary.com]

traversing both the NSA and SSA domains on January 7 and 11 and their associated up and down ramp events. The second DD episode was in the autumn, spanning 9 days between October 13 and October 21, 2011. This episode was selected because of the occurrence of strong low-level jets in the SSA on 6 of the 9 days. We note that with 24 hourly forecasts per day and 18 days total in the two episodes, there were 432 separate 15-hour forecasts available for a statistical analysis.

2.1 | Observational data sets

The special observational data set was divided into two categories: remote sensing and industry-provided in situ instrumentation. The remote sensing observations in the NSA consisted of nine radar wind profilers (seven were 915-MHz boundary layer profilers and two were 449-MHz profilers) and five Doppler sodars. The radar wind profilers also measured temperature profiles using the Radio Acoustic Sounding System (RASS) technique. The in situ industry-provided instrumentation in the NSA consisted of 133 tall towers providing vector winds (the vast majority with instruments at the 60 m level) and 405 turbine nacelle-mounted anemometers (at approximately 85 m) for which only wind speeds but not directions were available. As was done in the previous study, a simple “blade-wash” correction was applied to the turbine nacelle anemometers to account for the effect of the turbine blades immediately upwind of each nacelle. Height and range gate spacing for the various instruments are listed in Table 1.

TABLE 1 Observation heights and range gate spacing for the in situ observations (tall towers and turbine nacelle anemometers and for the remote sensing observations (915- and 449-MHz radar wind profilers measuring wind velocity and temperature profiles) and sodar wind velocity profiles^a [Colour table can be viewed at wileyonlinelibrary.com]

	Height Range	Range Gate Spacing
Tall tower winds	30-80 m	
Nacelle winds	85 m	
RWP 915-MHz winds	140 m to 4 km	60 m, 100 m
RWP 915-MHz temperatures	130 m to 1.6 km	60 m
RWP 449-MHz winds	150 m to 9 km	100 m, 200 m
RWP 449-MHz temperatures	180 m to 2.2 km	100 m
Sodar winds	30 m to 200 m	20 m

^aRadar wind profiler winds were measured in two modes with different vertical resolution (range gate spacing).

In the SSA, three radar wind profilers (all 915 MHz), seven Doppler sodars, and vector winds from 51 tall towers were available for assimilation. There were no turbine nacelle measurements in the SSA. Finally, although a single lidar system was deployed intermittently in the NSA and several were sited in the SSA for a portion of the field campaign, they were not used in the current DD study because of the limited time periods for which their data were available.

All of the observations were carefully reviewed before assimilation, and spurious measurements (approximately 5% of the total) were removed. Such removals could be due to, for example, radio frequency interference, ground clutter, and bird contamination of the radar wind profiler data, rain or excessive winds for the sodars, or direction biases or icing events for the tall towers. We note that Ancell et al⁸ also investigated the impact of assimilating WFIP observations, using both ensemble Kalman filter and 3D variational methods within the Weather Research and Forecasting (WRF) model, and found mixed results, which may be attributable to the fact that unlike the current study, they did not use fully quality controlled observational data.

2.2 | Model simulations

Because the focus of WFIP was on short-term forecasts, the principal operational NOAA models used were those that are reinitialized and run every hour, with the 13-km resolution NOAA RAP model used for the current data assimilation study. The model is a particular implementation of the Advanced Research Weather Research and Forecasting (WRF-ARW) model, which uses the 3D variational Gridpoint Statistical Interpolation (GSI) data assimilation system.⁹ The RAP model was chosen because the higher-resolution HRRR did not yet have its own independent assimilation system during WFIP as it was initialized through interpolation of the 13-km RAP initial grids to the HRRR 3-km grid. Also, since the RAP is run at coarser resolution, its DD simulations could be run faster than with the HRRR.

To reduce deleterious high-frequency noise, the GSI data assimilation system uses a coarsened grid (approximately three times lower resolution than the native RAP model) with large decorrelation length scales when it creates new initial fields. If there are multiple observations of the same variable within a single cell of that coarser grid, it will thin the observations down to a single value, as ultimately only one value can be assimilated per grid cell, and to guarantee convergence to a representative solution. To better utilize the spatially dense nacelle wind speeds, they were combined into a set of nine “super-obs” by averaging observations that fell within boxes three times the RAP grid cell dimension on a side (39 × 39 km) and using a mean speed value after eliminating speed outliers greater than two standard deviations from the mean. Tower and nacelle observations were then mapped into the model level corresponding to the observation height. All range gates from the remote sensor instruments were interpolated to the model vertical grid and then assimilated at those heights.

Four sets of simulations were run with the RAP model, all of which assimilated standard National Weather Service observations including rawinsondes, radar wind profilers, scanning radar (velocity azimuth display winds and reflectivity), aircraft, surface meteorological stations, and Geostationary Operational Environmental Satellite (GOES) and Advanced Microwave Sounding Unit (AMSU) satellite data.^{10,11} The first set, the remote sensing simulations, additionally assimilated the WFIP radar wind profilers and sodars. The second set, the in situ data simulations, additionally assimilated only the industry-provided tall tower and nacelle anemometer observations. The third set, the full WFIP observations simulations, additionally assimilated both the WFIP radar wind profilers and sodars and the tall tower and nacelle anemometer observations. The fourth set of simulations was a control, in which only the standard observations were assimilated, but none of the special WFIP observations.

Assimilation of local observations can improve forecasts through a reduction of bias, or through improving the model's spatiotemporal correlation with the observations. Since forecasts for wind energy applications typically have some type of postprocessing bias correction technique applied, we would like to assess the impact of the local data assimilation on improving model skill beyond a simple bias reduction. Therefore,

we apply a rudimentary bias correction to the forecasts, in which the bias is estimated as the average wind speed calculated independently for each of the 15 forecast hours over an entire DD simulation at each tower or sodar, minus the verifying tall tower or sodar wind speeds at the same times. More complicated methods were also investigated, but since the same bias correction method is applied to both the control and experimental simulations, the impact on model skill improvement was found to be similar for all methods. We note that by removing the bias, we make it more challenging to show an improvement due to the data assimilation.

2.3 | Wind power conversion

Since wind power rather than wind speed is of primary interest for the wind energy enterprise, wind speeds from the model, tall towers, and sodars were converted to power using a generic power curve. This generic power curve was created by AWS Truepower (now UL Renewables), and represents a composite of several different manufactures' International Electrotechnical Commission (IEC) Class II turbines, with values given at 1-ms^{-1} intervals. A seventh-order polynomial was then fit to these data and used to convert speed to power, with polynomial coefficient values provided in Appendix A. Comparisons with actual wind plant power output were not made because we did not have access to information on turbines that were not generating power due to malfunctions, maintenance, or curtailment and because it would have required use of uncertain turbine wake loss estimates. The use of a single mean power curve applied to the tall tower and sodar data allows us to only focus on meteorological effects.

3 | RESULTS

The impact of the data assimilation is assessed for the in situ data alone, remote sensing data alone, and the two combined, using bulk RMSE, probability density functions (PDFs), scatter plots showing the time synchronization of the improvements, and the diurnal variation of the improvement as a function of forecast lead time. All results are for the combined northern and southern study areas.

3.1 | DD episode representativeness

The first step of the analysis was to demonstrate that the two selected DD episodes were representative of the larger data set of six DD episodes used in Wilczak et al.⁶ Figure 2 compares the average RMSE percent improvement of wind power from the assimilation of all of the WFIP data for the two selected DD episodes (black bars) with that found for all six DD episodes (red bars), using the tall tower data for verification. The pattern

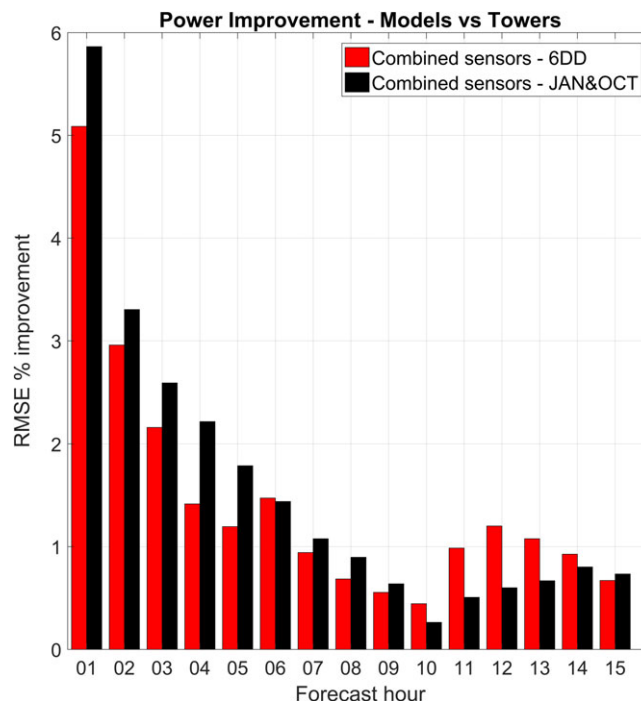


FIGURE 2 Root mean squared error (RMSE) percent improvement for Rapid Refresh (RAP) simulations due to the assimilation the Wind Forecast Improvement Project (WFIP) radar wind profilers, sodars, tall towers vector winds, and nacelle anemometer scalar wind speeds. The red bars are the average for all six data denial episodes, while the black bars are for the average of the subset of two data denial episodes (in January and October) [Colour figure can be viewed at wileyonlinelibrary.com]

of improvement found in the two DD episodes is quite similar to the larger data set, with improvements starting at 5% to 6% at forecast hour 1, trailing off to 1% at forecast hour 15. We note that these percentage improvements are meaningful when compared with the historic rates of improvement in 12-hour forecasts of 850-hPa winds of 0.4%/year for the NOAA Global Forecast System (GFS) and 0.7%/year for the NOAA North American Mesoscale (NAM) over a 10-year period from 2004 to 2013.⁶ Of course the improvements from the assimilation of the special WFIP observations likely will be confined to a small geographic region surrounding the two study areas and also have been evaluated only in terms of wind speed, not other meteorological variables.

3.2 | Bulk RMSE improvements

Next, the DD simulations were repeated for the two selected episodes, first assimilating the WFIP remote sensing data and second assimilating the WFIP industry-provided in situ data. Both sets of simulations also assimilated all other conventional observations and again used the tall tower data for verification. Figure 3 shows the percent improvement of wind power for the remote sensing data (green bars), in situ data (magenta bars), and compares the sum of these to the simulations when both sets were assimilated simultaneously (black bars), where the percent improvement is defined as $100 \times (\text{RMSE}_{\text{exp}} - \text{RMSE}_{\text{cntl}}) / \text{RMSE}_{\text{cntl}}$. The in situ data have a large impact in the first forecast hour of about 6%, which decreases rapidly with time, with no discernable impact after forecast hour 7. In contrast, the remote sensing data have a smaller initial impact of 1% to 1.5% in the first several hours, but that improvement remains through the length of the simulations, slightly decreasing to approximately 0.75% in forecast hours 10 to 15. Also, we note that the sum of the improvements of the in situ and remote sensing data assimilated separately (magenta plus green bars) is generally close to, but somewhat larger than, that when the data are assimilated together, indicating that there is only a slight redundancy in the information that they provide. The initially larger impact of the in situ observations in the first few forecast hours is likely due to the fact that there are many more in situ observing sites than remote sensing sites, while the more lasting benefit of the remote sensing observations is likely due to the fact that those instruments measure a deeper layer of the atmosphere, thus impacting the surface over a longer period of time.

We investigated whether the larger improvement found for the in situ data assimilation (towers and nacelle anemometers) for the first several forecast hours in Figure 3 is due to the fact that the tall tower observations are used both for assimilation as well as for verification. To test this hypothesis, we repeated the analysis, but now using the 80-m wind speed data from the 12 sodars for verification instead of the 184 tall towers (Figure 4). In comparison with Figure 3, the in situ data (magenta bars) still have a larger impact than the remote sensing data especially for the earlier forecast hours, indicating that this larger impact is a robust result, likely due to the much greater number of tall tower and nacelle sites

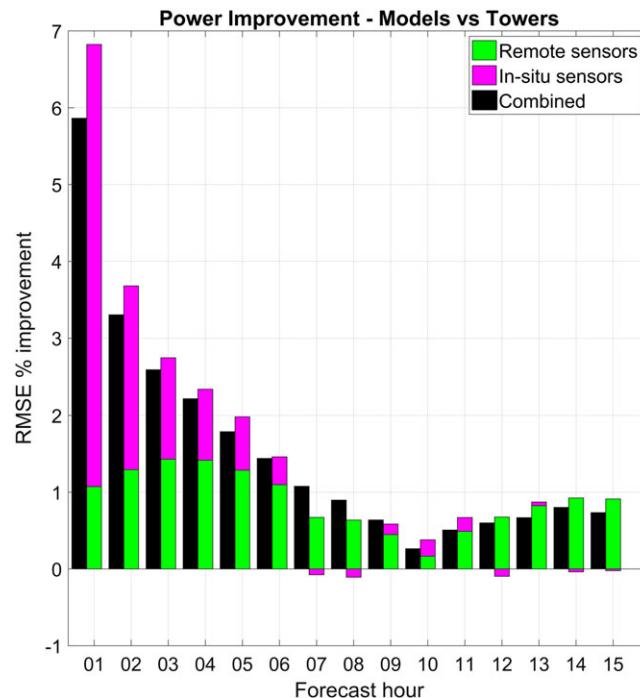


FIGURE 3 Root mean squared error (RMSE) percent improvement for Rapid Refresh (RAP) simulations that assimilate various combinations of the Wind Forecast Improvement Project (WFIP) data, averaged for the October and January data denial (DD) episodes. The black bars are the same as those shown in Figure 2 in which the radar wind profilers, sodars, tall towers vector winds, and nacelle anemometer scalar wind speeds are assimilated; the magenta bars show the impact of assimilating only the tower and nacelle observations, and the green bars show the impact of assimilating only the wind profiling radar and sodar observations [Colour figure can be viewed at wileyonlinelibrary.com]

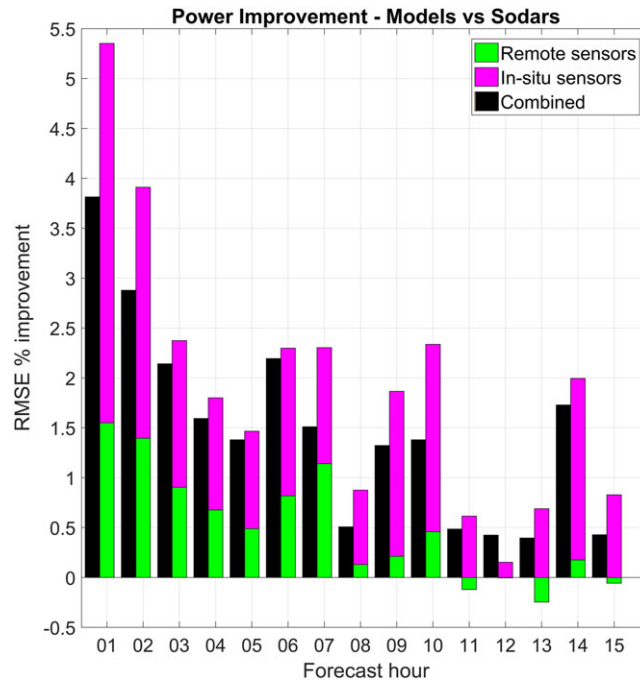


FIGURE 4 As in Figure 3 except using 80-m winds at the 12 sodar sites, converted to power, for verification [Colour figure can be viewed at wileyonlinelibrary.com]

assimilated than remote sensing sites. We also note that the overall magnitude of the improvement when assimilating all of the new WFIP observations is somewhat smaller in Figure 4 (using sodar verification) than in Figure 3 (using tall tower verification). This can be explained by the fact that the remote sensing observations spanned a larger area than the tall towers and nacelle observations and were often on the periphery of the study areas, whereas the towers and nacelles were more generally located towards the center of the study areas (Figure 1). Thus, model skill at the tall tower and nacelle verification sites always benefitted from the remote sensors, but model skill at the remote sensor verification sites did not always benefit from the tall tower and nacelle observations. Finally, we note that the assimilation impact in Figure 4 becomes noisier for forecast hours 08 to 15 and that this is likely due to the much smaller number of sodar verification sites available than tall tower sites. For the remainder of our analysis, we use the tall tower data for verification.

3.3 | PDFs of the improvement

PDFs for the in situ and remote sensing assimilation results, using hourly values of RMSE percent improvement of wind power, are shown in Figure 5, for forecast hours 0 to 5. Key points are that first the width of each distribution is large compared with its mean value, indicating that although the overall impact is moderately positive, there are times when the assimilation of the new WFIP data can lead to either significantly improved or degraded forecasts. Second, PDFs for the in situ data indicate these observations are in most instances likely to improve the forecast through forecast hour 2. Beyond hour 3, the narrowness of the in situ PDFs relative to the remote sensing PDFs indicates that the in situ observations have a smaller impact, good or bad. This demonstrates that the remote sensing data have the ability to alter the model forecasts to a greater extent at longer forecast hours, which is reasonable given that they cover a deeper layer of the atmosphere.

3.4 | Time synchronization of improvements

To better understand whether the impact of the in situ data and the remote sensing data are synchronized in time, we next consider scatter plots of the hourly RMSE percent improvement of wind power in Figure 6, again shown for forecast hours 00 to 05. The numbers in each quadrant show the percentage of hourly data points that fall into that quadrant, while the sum of the two values in either the right or left halves, or the top or bottom halves, of each panel indicate the number of in situ or remote sensing observations that have a positive or negative impact. The largest quadrant percentage for each forecast hour is when the in situ and remote sensing observations both have a positive impact, due, for example, to there being a meteorological feature missing in the control simulation that is consistently observed by both the in situ and remote sensing instrumentation, and that improves the simulations when it more accurately appears in the initialization. Other than the initialization time, the minimum number of occurrences for all forecast hours is when the in situ data have a negative impact but the remote sensing data a positive

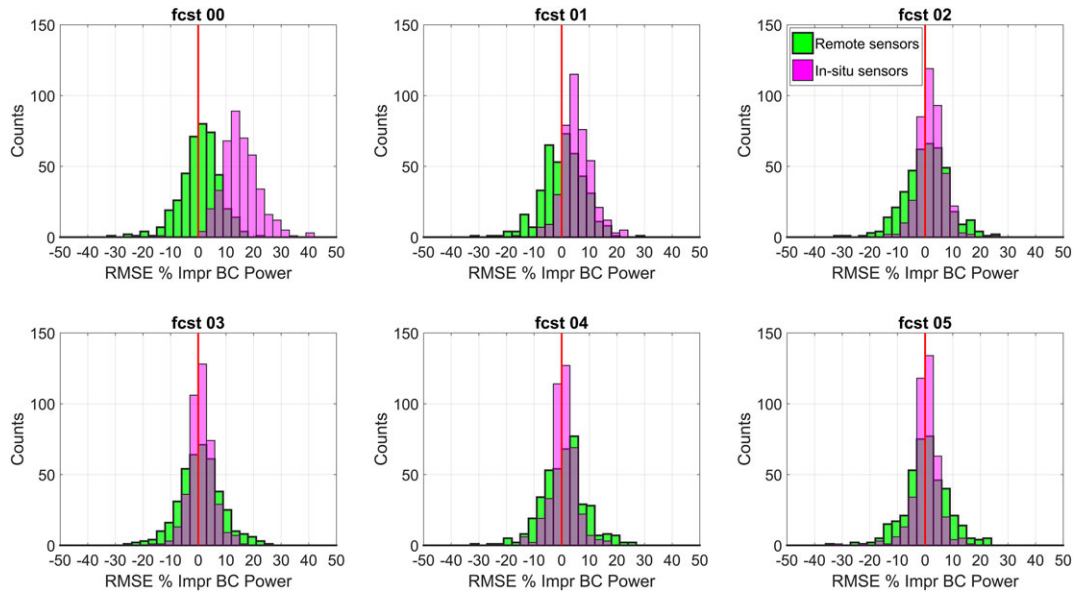


FIGURE 5 Probability density functions (PDFs) of the root mean squared error (RMSE) percent improvement for Rapid Refresh (RAP) simulations that assimilate either the tower and nacelle observations (magenta) or that assimilate the radar wind profilers and sodars (green), averaged for the October and January data denial (DD) episodes, and for forecast hours 00 through 05 [Colour figure can be viewed at wileyonlinelibrary.com]

impact. This indicates that if the in situ instruments observe a meteorological structure or phenomenon that worsens model skill, it is more likely that the remote sensing observations will agree with the in situ analysis (also providing a negative impact) than they will disagree with the in situ data. One possible explanation would be that they both observe some atmospheric structure that is incompatible with the model physics/dynamics, and assimilation of those data degrade model skill. We also note that the sum of the positive-positive and negative-negative

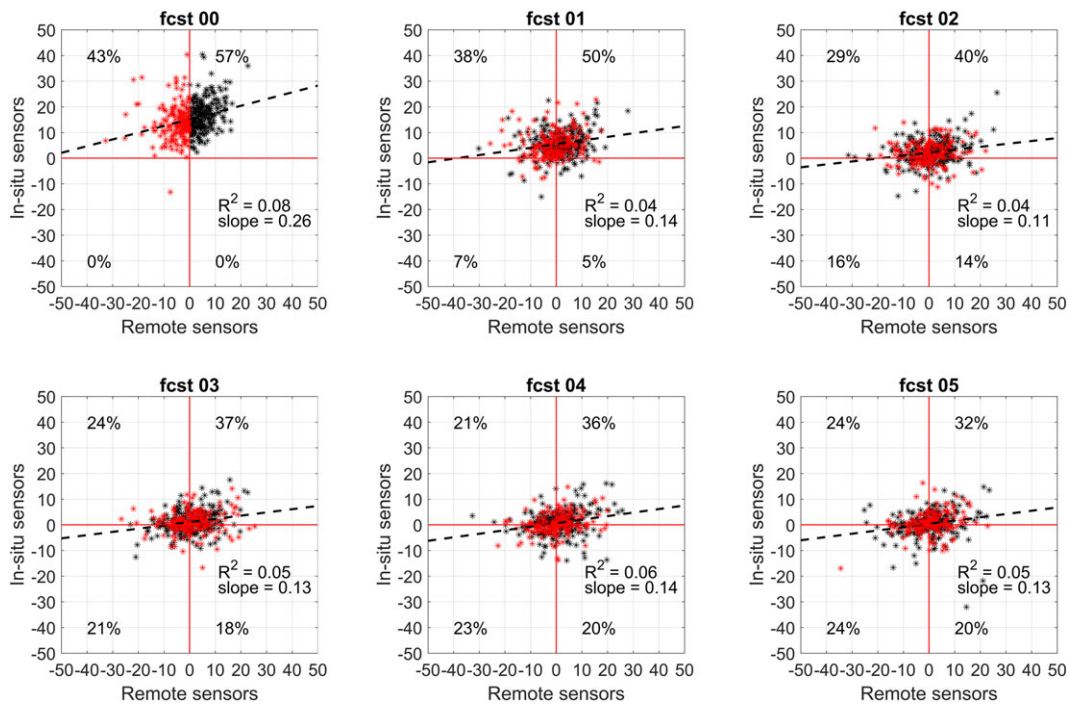


FIGURE 6 Scatter plots of the root mean squared error (RMSE) percent improvement for Rapid Refresh (RAP) simulations that assimilate either the in situ tower and nacelle observations (ordinate) or the remote sensing radar wind profilers and sodars (abscissa), for the October and January data denial (DD) episodes, and for forecast hours 00 through 05. The percentage of points that fall within each quadrant is shown, as is a best fit line, slope, and coefficient of determination. Data points for forecasts for which the remote sensing observations degraded the model initialization, when using the in situ data for verification, are shown as red, while those forecasts for which the remote sensing observations improved the initialization are shown as black [Colour figure can be viewed at wileyonlinelibrary.com]

percentages is nearly constant in all panels, ranging between 56% and 59%, while best fit lines have a consistently positive slope, indicating that the signs of the impacts of the in situ and remote sensing data have a tendency to be synchronized in time.

For the initialization time (hour 00) in Figure 6, we note that 43% of the time the remote sensing observations degrade the initialization when verified using the tall tower observations. An interesting question is whether eliminating data that degrade the initialization compared with all of the remaining observations would then improve the forecast at a later time. Such a procedure might be possible by, for example, using adjoint methods such as the Forecast Sensitivity to Observation Impact (FSOI) approach¹²⁻¹⁴ or the Ensemble Forecast Sensitivity for Observation (EFSO) formulation.^{15,16} Although running a full test of this concept would require many more simulations and is beyond the scope of this paper, we can determine whether eliminating those forecasts for which the remote sensing data degraded the initialization, when using the in situ observations as verification, would then improve on average the skill of the remaining forecasts. To do this, we simply eliminated from the statistics all forecasts that had a negative impact from the remote sensing observations at the initialization time in Figure 6, indicated by the points in red. This did improve the skill of the remaining forecasts for forecast hours 01 and 02, where the sum of the percentages with positive remote sensing impact increased from 55% to 64% and from 54% to 59%, but not for later hours. Also, we note that it is possible that the remote sensing data that degraded the 80-m power forecasts may have improved other aspects of the forecasts, such as temperature or winds at other levels.

3.5 | Statistical significance of improvements

The statistical significance of the impacts of assimilating the additional in situ data is shown in Figure 7. Here, the top panels show the RMSE for power (left) and vector wind (right), for the control (red line) and when assimilating the additional in situ data (blue line). The lower panels show the RMSE percent improvement, with error bars representing the 95% confidence intervals defined as $(\pm 1.96\sigma/\sqrt{n})$ where n has been reduced by the autocorrelation of the time series to account for the fact that the simulations are not statistically independent. The positive impact of the in situ data assimilation is seen to be statistically significant for power until forecast hour 05 and for vector wind until forecast hour 04. Figure 8 shows the same statistics when assimilating the remote sensing observations. For power (left panels), the positive impact is borderline statistically significant at the 95% confidence level except for forecast hours 06 to 10, while for vector wind, the impact is statistically significant for almost all 15 forecast hours. The reduced statistical significance when assimilating the remote sensing observations compared with the in situ observations is due to the smaller number of remote sensing sites. For future work, it would be useful to repeat the analysis, especially for the assimilation of the remote sensing observations, using longer or more numerous DD episodes covering more meteorological conditions.

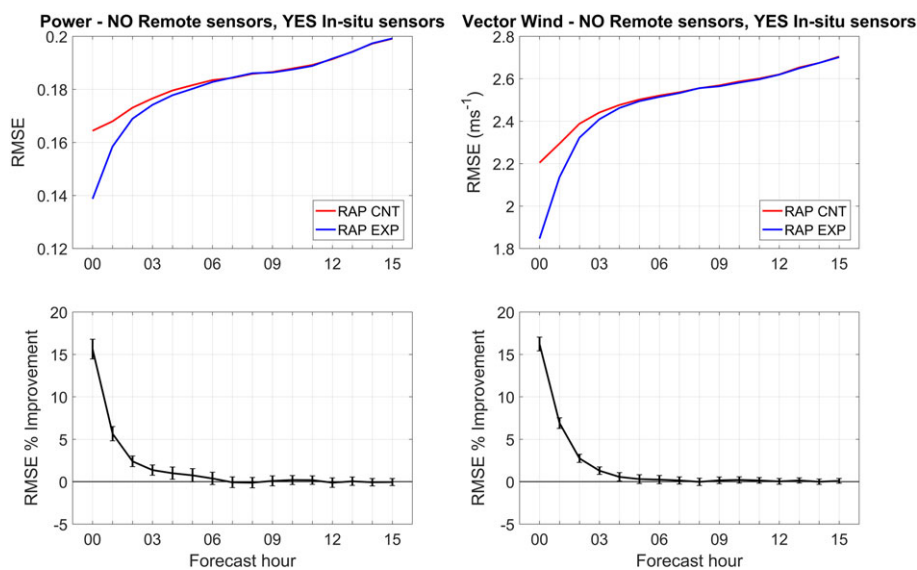


FIGURE 7 Top: root mean squared error (RMSE) for hub-height power, expressed as a decimal fraction of maximum power (rated) capacity (left), and for vector wind (right), for the January and October data-denial episode days. The red curve is for the control simulations that assimilated the standard observations, and the blue curve is for the experimental simulations that also assimilated the Wind Forecast Improvement Project (WFIP) in situ observations. Bottom: the difference in RMSEs between the control and experimental simulations expressed as a percentage improvement. Error bars represent the 95% confidence intervals defined as $(\pm 1.96\sigma/\sqrt{n})$ where n is reduced by the autocorrelation of the time series [Colour figure can be viewed at wileyonlinelibrary.com]

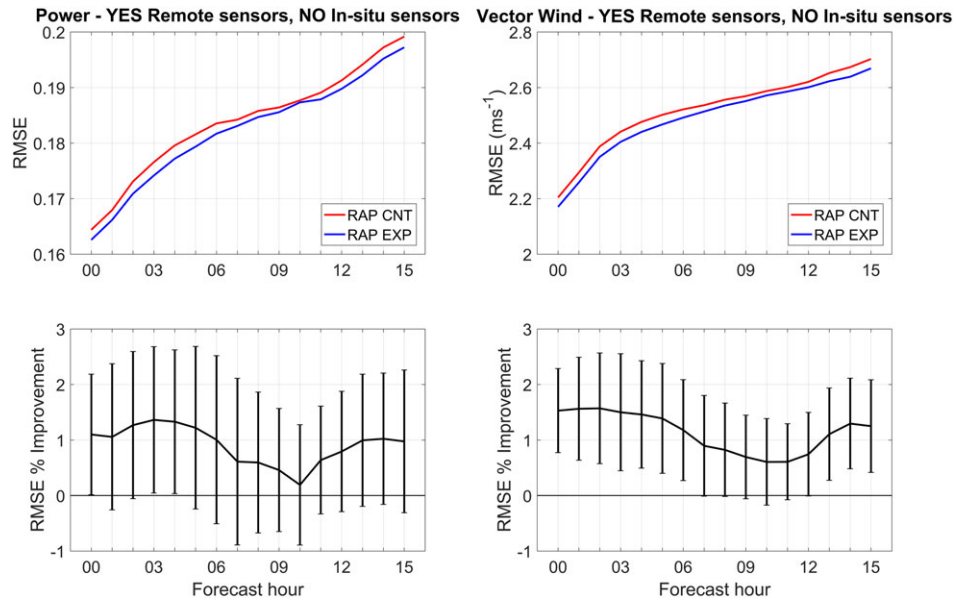


FIGURE 8 As in Figure 7 except the red curve is for the control simulations, and the blue curve is for the experimental simulations that also assimilated the Wind Forecast Improvement Project (WFIP) radar wind profilers and sodar observations [Colour figure can be viewed at wileyonlinelibrary.com]

3.6 | Diurnal variation of assimilation impact

To understand the diurnal variation of the impact of the WFIP data assimilation, we show in Figure 9 the RMSE percent improvement as a function of both forecast horizon and forecast validation hour of the day. Sunrise for the centers of the northern and southern study areas for the January DD episode were approximately 14:08 and 13:48 UTC, while for the October episode, they were 12:51 and 12:50 UTC, while sunset times were 23:12 and 23:43 UTC for the northern area and 23:57 and 00:08 UTC for the southern area. The average of the two area's sunrise and sunset times is shown by the horizontal black lines in Figure 9. The in situ observations (left panel) show the greatest improvements during the daytime hours of 14 to 20 UTC (08-14 local standard time [LST]) and for shorter lead times, with a weak minimum for the later forecast horizon hours with validation times prior to sunrise 09 to 12 UTC (03-06 LST). In comparison, the percent RMSE

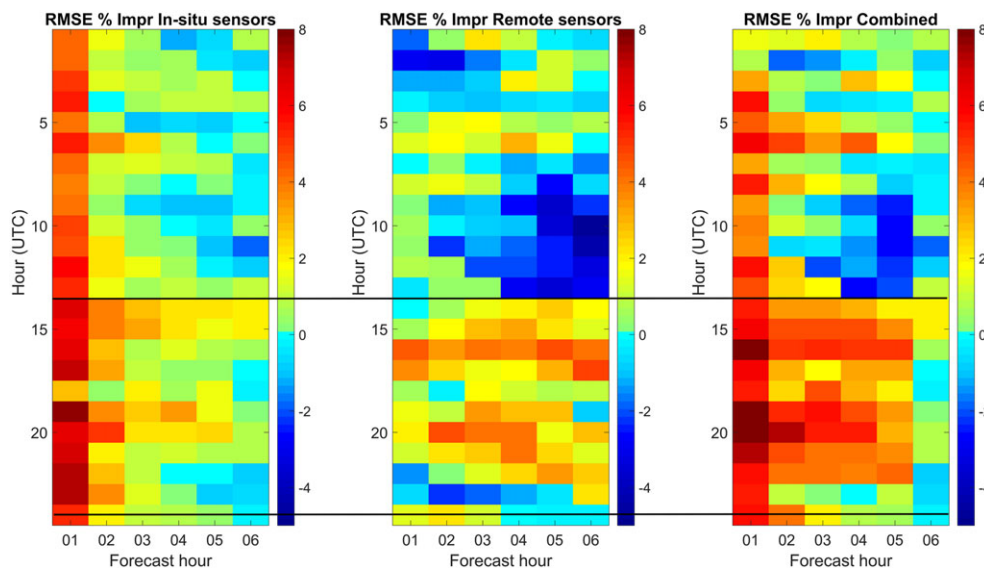


FIGURE 9 Root mean squared error (RMSE) percent improvement as a function of forecast length (abscissa) and forecast hour validation time (ordinate), when assimilating the additional in situ observations (left panel), the remote sensing observations (middle panel), and when assimilating both (right panel). Horizontal black lines indicate the times of sunrise and sunset averaged between the centers of the two study areas and the dates of the two data denial episodes [Colour figure can be viewed at wileyonlinelibrary.com]

improvement due to assimilation of the remote sensing observations (center panel) shows a similar large positive improvement during the daytime hours between 14 and 22 UTC (08–16 LST) but over a longer range of forecast hours. It also has a more pronounced minimum and even degradation of forecast skill for the later forecast horizon hours with nighttime and early morning validation times between 09 and 13 UTC (03–07 LST). The right panel shows the percent RMSE improvement when assimilating both the in situ and remote sensing observations, which again shows a similar diurnal variation but with larger overall positive impacts.

The daytime greater positive impact of the in situ and remote observations is likely due to the fact that they are more representative of winds through the entire well-mixed boundary layer when they are assimilated and also that they are more representative at the verification hour. In contrast, nighttime measurements in the stable boundary layer can be more influenced by even relatively gentle variations in the topography, leading to observations, at least in the lower portion of the boundary layer, that are not representative of the larger scale flow. Although assimilation of those observations may produce an initial improvement in model skill, at later forecast hours, the nonrepresentative nature of those observations can lead to decreased forecast skill. It is also possible that forcing the model's stable boundary layer with observations that are inconsistent with the model's boundary layer parameterizations, in that they depict vertical gradients of temperature and wind that cannot be resolved spatiotemporally by the model, will eventually lead to greater forecast errors than if the observations were never assimilated at all. Here, we note that numerical weather prediction models have difficulty accurately simulating the low-level jet,^{17,18} as shown in particular for the WFIP data set by Wilczak et al.⁶ and Mirocha et al.¹⁹ The potential presence of gravity waves in the observations might also have a similar effect.

4 | SUMMARY

Observations from regional networks of in situ (tall tower and turbine nacelle-mounted anemometers) and remote sensors (radar wind profilers and sodars) collected as part of WFIP were assimilated independently and in combination into the hourly updated NOAA RAP model for two DD episodes of 9 days each in January 2012 and October 2011. The assimilation impacts are evaluated for idealized wind power forecasts, using the tall tower and sodar data sets for verification.

For forecast hours 01 to 03, the numerous in situ observing sites are found to have a greater impact than the remote sensing data, of which there are far fewer sites. However, for longer forecast lead times, the impact of the in situ observations rapidly decreases, becoming indiscernible after forecast hour 07, while the impact of the remote sensing observations, which provide measurements through a deeper layer of the atmosphere, remains positive and statistically significant for most hours through forecast hour 15. The larger positive impact of the in situ observations in the early forecast hours is independent of whether the validation is done using the tall tower or sodar observations. Histograms of RMSE percent improvement for wind power show that on an hourly basis the impact of the assimilation can be strongly positive or negative, with the net positive improvement evident after significant temporal averaging (see also Djalalova et al.²⁰ for an example of forecast degradation through data assimilation). Scatter plots of RMSE percent improvement from the two observational categories show that improvements from the in situ and remote sensing observations are weakly temporally correlated. Finally, an analysis of the diurnal variation of the data assimilation impact demonstrates that both the in situ and remote sensing observations have the largest positive impact for daytime validation hours, with a minimum or even negative impact for validation hours in the late nighttime hours and early morning hours.

The significant short-term increases in forecast skill found from assimilating industry-provided tall tower and turbine nacelle wind speed measurements offer a relatively easy and inexpensive pathway towards improvement of wind energy forecasts. Since most wind plant owners/operators/developers maintain networks of tall towers measuring wind speed (for wind turbine output assessment or for wind prospecting purposes) and since nearly every wind turbine provides nacelle wind speed measurements, these represent a great untapped potential for improving wind power forecasts. In most cases, the only new requirements would be to invest in data transmission infrastructure and maintenance, so that the observations can be made available in real time to either national weather forecasting centers or to industry forecasters that run their own numerical weather prediction models. Further skill, especially at longer lead times, could be added with networks of remote sensors such as sodars, radar wind profilers, or potentially lidars, but at higher cost.

ACKNOWLEDGMENTS

The Wind Forecast Improvement Project was supported by the US Department of Energy, Office of Energy Efficiency and Renewable Energy, and by the National Oceanic and Atmospheric Administration. The authors wish to thank the many engineers, technicians, and support scientists that contributed to the collection of the observational and model data sets. The Pacific Northwest National Laboratory is operated for the US Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830. Argonne National Laboratory is operated for the US Department of Energy by UChicago Argonne, LLC under contract DE-AC02-06CH11357.

ORCID

James M. Wilczak  <https://orcid.org/0000-0002-9912-6396>

REFERENCES

1. Jonsson T, Pinson P, Madsen H. On the market impact of wind energy forecasts. *Energy Econ.* 2010;32(2):313-320. <https://doi.org/10.1016/j.eneco.2009.10.018>
2. Gonzalez-Aparicio I, Zucker A. Impact of wind power uncertainty forecasting on the market integration of wind energy in Spain. *Appl Energy.* 2015;159:334-349. <https://doi.org/10.1016/j.apenergy.2015.08.104>
3. Mc Garrigle EV, Leahy PG. Quantifying the value of improved wind energy forecasts in a pool-based electricity market. *Renew Energy.* 2015;80:517-524. <https://doi.org/10.1016/j.renene.2015.02.023>
4. Brancucci Martinez-Anido C, Hodge B. 2014. Impact of utility-scale distributed wind on transmission-level system operations. National Renewable Energy Laboratory (NREL), Technical Report No. NREL/TP-5D00-61824.
5. Hodge B, Florita A, Sharp J, Margulis M, McCreavy D 2015. The value of improved short-term wind power forecasting. National Renewable Energy Laboratory (NREL), Technical Report No. NREL/TP-5D00-63175.
6. Wilczak JM, Finley C, Freedman J, et al. The Wind Forecast Improvement Project (WFIP): a public-private partnership addressing wind energy forecast needs. *Bull Am Meteorol Soc.* 2015;19(10):1699-1718. <https://doi.org/10.1175/BAMS-D-14-00107.1>
7. Wilczak J, Bianco L, Olson J et al. 2014. The Wind Forecast Improvement Project (WFIP): a public/private partnership for improving short term wind energy forecasts and quantifying the benefits of utility operations. NOAA Final Technical Report to DOE, award number DE-EE0003080, 162 pp. [Available online at <http://energy.gov/sites/prod/files/2014/05/f15/wfipandnoaafinalreport.pdf>.]
8. Ancell BC, Kashawlic E, Schroeder JL. Evaluation of wind forecasts and observation impacts from variational and ensemble data assimilation for wind energy applications. *Month Wea Rev.* 2015;143(8):3230-3245. <https://doi.org/10.1175/MWR-D-15-0001.1>
9. Wu W-S, Parrish DF, Purser RJ. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon Weather Rev.* 2002;130(12):2905-2916. [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2)
10. Benjamin SG, Weygandt SS, Brown JM, et al. A North American hourly assimilation and model forecast cycle: the Rapid Refresh. *Mon Weather Rev.* 2016;144(4):1669-1694. <https://doi.org/10.1175/MWR-D-15-0242.1>
11. Kleist DT, Parrish DF, Derber JC, Treadon R, Wu W-S, Lord S. Introduction of the GSI into the NCEP global data assimilation system. *Wea Forecast.* 2009a;24(6):1691-1705. <https://doi.org/10.1175/2009WAF2222201.1>
12. Langland RH, Baker NL. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus.* 2004;56A(3):189-201. <https://doi.org/10.1111/j.1600-0870.2004.00056.x>
13. Zhu Y, Gelaro R. Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon Wea Rev.* 2008;136(1):335-351. <https://doi.org/10.1175/MWR3525.1>
14. Sommer M, Weissmann M. Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus A: Dyn Meteorol Oceanogr.* 2016;68(1):27885. <https://doi.org/10.3402/tellusa.v68.27885>
15. Kalnay E, Ota Y, Miyoshi T, Liu J. A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus.* 2012;64A:18462. <https://doi.org/10.3402/tellusa.v64i0.18462>
16. Hotta D, Chen T-C, Kalnay E, Ota Y, Miyoshi T. Proactive QC: a fully flow-dependent quality control scheme based on EFSO. *Month Wea Rev.* 2017;145(8):3331-3354. <https://doi.org/10.1175/MWR-D-16-0290.1>
17. Storm B, Dudhia J, Basu S, Swift A, Giammanco I. Evaluation of the Weather Research and Forecasting model on forecasting low-level jets: implications for wind energy. *Wind Energy.* 2009;12(1):81-90. <https://doi.org/10.1002/we.288>
18. Banta RM, Newsom RK, Lundquist JK, Pichugina YL, Coulter RL, Mahrt L. Nocturnal low-level jet characteristics over Kansas during CASES-99. *Bound-Layer Meteorol.* 2002;105(2):221-252. <https://doi.org/10.1023/A:1019992330866>
19. Mirocha JD, Simpson MD, Fast JD, Berg LK, Baskett RL. Investigation of boundary-layer wind predictions during nocturnal low-level jet events using the Weather Research and Forecasting model. *Wind Energy.* 2016;19(4):739-762. <https://doi.org/10.1002/we.1862>
20. Djalalova IV, Olson J, Carley J, et al. The POWER experiment: impact of assimilation of a network of coastal wind profiling radars on simulating offshore winds in and above the wind turbine layer. *Weather Forecast.* 2016;31(4):1071-1091. <https://doi.org/10.1175/WAF-D-15-0104.1>

How to cite this article: Wilczak JM, Olson JB, Djalalova I, et al. Data assimilation impact of in situ and remote sensing meteorological observations on wind power forecasts during the first Wind Forecast Improvement Project (WFIP). *Wind Energy.* 2019;22:932-944. <https://doi.org/10.1002/we.2332>

APPENDIX A

WIND SPEED TO POWER CONVERSION EQUATION

AWS Truepower (now UL Renewables) provided a set of normalized power values (between 0 and 1) for wind speed values at 1-ms^{-1} intervals between 3 and 25 ms^{-1} , based on the average of several power curves provided by various wind turbine manufactures for IEC Class 2 turbines. A seventh-order polynomial was then fit to the data points between 3 to 16 ms^{-1} (Figure A1), such that

for $3 \leq S \leq 16$,

$$\text{Power} = C_0 + C_1S + C_2S^2 + C_3S^3 + C_4S^4 + C_5S^5 + C_6S^6 + C_7S^7, \quad (\text{A1})$$

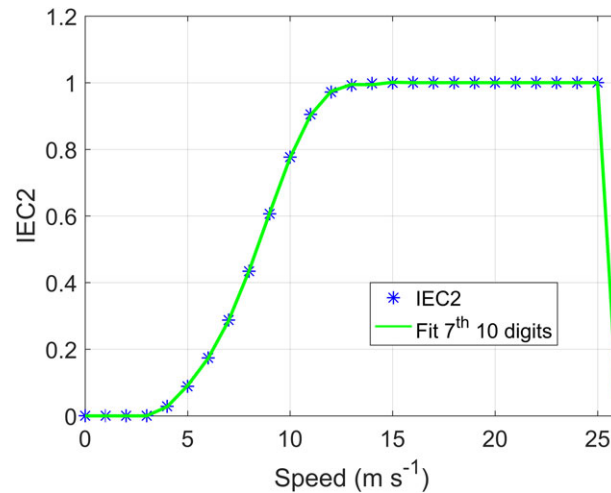


FIGURE A1 Normalized power curve values for an average of several IEC Class 2 wind turbines (solid points), and a seventh-order polynomial fit to those data points (green curve) [Colour figure can be viewed at wileyonlinelibrary.com]

where

$$\begin{aligned} C_0 &= 2.4820513186, \\ C_1 &= -2.6915377682, \\ C_2 &= 1.1681347571, \\ C_3 &= -0.2650969882, \\ C_4 &= 0.0342920032, \\ C_5 &= -0.0024951705, \\ C_6 &= 0.0000946197, \\ C_7 &= -0.0000014530. \end{aligned}$$

For $16 < S \leq 25$, the power was set to unity, while for $S < 3$, it was set to zero. Also, if Equation A1 ever gives a value greater than 1, the power is set to 1, and if it gives a value less than 0, the power is set to 0.

Normalized power differences between this procedure and the original set of power-wind speed data points was never greater than 0.004. Using less than 10 digits significantly increased the error of the fit (for example with nine digits the maximum power difference increased to 0.03) while using 11 or 15 digits changed the results by less than 0.0005.