Proceedings of ASME 2010 4th International Conference on Energy Sustainability
ES2010
May 17-22, 2010 Phoenix, Arizona, USA

ES2010-90381

DRAFT

A METHODOLOGY FOR COMPREHENSIVE CHARACTERIZATION OF ERRORS IN WIND POWER FORECASTING

Mark F. Bielecki
MS Candidate,
Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

Jason J. Kemper
MS Candidate,
Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

Thomas L. Acker
Professor of Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

Abstract

Wind power forecasting will play a more important role in electrical system planning with the greater wind penetrations of the coming decades. Wind will comprise a larger percentage of the generation mix, and as a result forecasting errors may have more significant effects on balancing operations. The natural uncertainties associated with wind along with limitations in numerical weather prediction (NWP) models lead to these forecasting errors, which play a considerable role in the impacts and costs of utility-scale wind integration. The premise of this project was to examine errors between the actual and commercially forecasted power production data from a typical wind power plant in the Northwestern United States. An exhaustive statistical characterization of the forecast behavior and error trends was undertaken, which allowed the most important metrics for describing wind power forecast errors to be identified. This paper presents only the metrics chosen as most significant. While basic information about wind forecast accuracy such as the mean absolute error (MAE) is valuable, a more detailed description is useful for system operators or in wind integration studies. System planners have expressed major concern in the area of forecast performance during large wind ramping events. For such reasons, this methodology included the development of a comprehensive ramp identification algorithm to select significant ramp events from the data record, and particular attention was paid to the error analysis during these events. The algorithm allows user input to select ramps of any desired magnitude, and also performs correlation analysis between forecasted ramp events and actual ramp events that coincide within a desired timing window. From this procedure, an investigation of the magnitude and phase of forecast errors was conducted for various forecast horizons. The metrics found to be of most importance for error characterization were selected based on overall impacts, and were ranked in order of significance. These metrics included: mean absolute error, root mean square error, average step change value, standard deviation of step changes, mean bias levels, correlation coefficient of power values, mean temporal bias of ramp events, and others. While these metrics were selected and the methodology was developed for a single dataset, the entire process can be applied generally to any wind power and forecast time series. The implications for such a process include use for generating a synthetic wind power forecast for wind integration studies that will reproduce the same error trends as those found in a real forecast.

Introduction

Although state-of-the-art wind power forecasts are not completely accurate, they have been shown to reduce the overall integration costs associated with utility-scale wind power as well as give system planners a better idea of the variability and uncertainty that will be faced by adding wind to the system [1,2]. Forecasting will undoubtedly play a more important role with the greater wind penetrations of the coming decades, and as a result forecasting errors may lead to larger consequences related to under or over-production when wind comprises a larger percentage of the generation mix. Many contemporary wind integration studies cite the need for accurate forecasting to accomplish successful large-scale wind integration [3,4,5].

Wind power forecasts are based on wind speed forecasts. The wind speed forecasts are typically generated from numerical weather prediction (NWP) models, which use a set of boundary and initial conditions as inputs to physics-based equations that propagate fluid behavior forward in space and time to generate predicted values. Limitations in the ability to numerically solve these complex equations along with inherent inaccuracies in the initial and boundary conditions lead to errors in wind speed forecasting. Once the speed has been forecasted, wind turbine power curves and other wind power plant information are used to formulate the wind power forecasts, frequently using complex and proprietary algorithms intended to decrease forecast error. The cubic relation between wind speed and wind power leads to the non-linear nature of wind turbine power curves [6]. Therefore, the errors in wind speed will be exacerbated by this non-linear relation, leading to increased errors in wind power predictions as discussed in [7].

As utilities and independent system operators (ISOs) plan to integrate more wind into their generation mixes, they seek a better understanding of how the variability and uncertainty added by wind will impact the balancing operations for the grid. The errors associated with wind power prediction can lead to significant challenges for system planners and operators. Wind integration studies are generally conducted prior to the wind development phase to assess these impacts. The studies often rely on simulated wind power and wind power forecasts for proposed development sites, as a lack of historical wind measurements exists for many rural areas [8]. A thorough understanding of the error characteristics that will be encountered in actuality is imperative when performing comprehensive wind integration studies to ensure that any simulated or synthetic data reproduces the same error trends. Such validations were carried out in [9] for the California Energy Commission's Intermittency Analysis Project, and in [4] for the California Independent System Operator's study on meeting RPS standards of 20% renewables. Extensive validations of simulated data have also been conducted for the Western Wind and Solar Integration Study (WWSIS) by NREL [10]. These contemporary examples have demonstrated a demand for a methodical approach to wind power forecast error description.

At present time, there is no universal means of evaluating forecast performance, although some efforts to standardize have been proposed [11]. Several metrics are commonly presented, yet no individual metric offers a complete description of error tendencies. The difference between the actual power value and the forecast power value at any given time is called the forecast error as shown by Equation 1[1]. The mean and standard deviation of the forecast error can be used to further describe the error spread, and will be discussed later in this report.

$$forecast\ error = P_f(t) - P_a(t) \tag{1}$$

Two other common metrics for evaluating the forecast performance for a time series with n elements are the mean absolute error (MAE) as shown by Equation 2, and the root mean square of the error (RMSE) as shown by Equation 3.

$$MAE = \frac{1}{n}\sum_1^n \left|P_a - P_f\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_1^n (P_a - P_f)^2} \tag{3}$$

An advantage to the MAE is that it gives more insight about the average magnitude of the errors over an entire dataset without the effect of cancelling positive and negative errors that occurs when the forecast error is simply averaged. However, this advantage is gained with the sacrifice of error directionality, which can be important when large amounts of wind are integrated into the grid system. Operators would like to know whether the wind component is being under or over-predicted. The RMSE can be another useful metric for evaluating wind forecast errors because it intrinsically places more weight on larger error terms, which are often of most interest to system planners. Again, this comes at the expense of losing error directionality. It should be noted that if the average of the forecast error, or the "mean bias" is zero, then the RMSE and the standard deviation of the forecast error would be equivalent.

[1] Where P_a represents actual power value and P_f represents forecast power value.

Although each of the above metrics has its place in evaluating errors in wind power prediction, none offers a complete description of forecast performance. Consequently, the objective of this work was to thoroughly examine errors between the actual and commercially forecasted power production data from a typical wind power plant and to identify a set of metrics that can be used to completely describe error trends. Data from an operating wind power plant in the Northwestern United States was used as the basis for the study, and considered representative of a state-of-the-art professional wind forecast. An exhaustive statistical characterization of the forecast behavior and error trends was undertaken with the most significant results presented in this paper. Special attention was given to forecast horizons that are important to grid system planning, as well as large wind ramping events which can be especially challenging for operational concerns.

Methodology

An hourly time series from three recent years (2004-2006) of actual wind power production and simultaneous forecast data from an operating wind power plant were used for analysis. The state-of-the-art forecast dataset was created by a commercial forecast provider using proprietary methods, and included hourly forecast data for 2004-2006. For each hour during this period, a wind power forecast value was provided for the following 140 hours. A "forecast horizon" is defined as the number of hours ahead of operation for which the forecast is intended (e.g. 1-hr horizon corresponds the predicted value of wind 1 hour from the hour of operation). Forecast values are seldom generated in real time, so there is generally a cut-off time several hours ahead of the operation hour during which they are actually generated. The forecast horizons presented in this study were chosen for their significance to unit scheduling concerns. For example, forecast horizons of 1 to several hours ahead of real time coincide with load following and dispatch optimization planning, while the day-ahead horizon coincides with unit-commitment planning [12]. The day-ahead forecast is not the same as a 24-hr forecast horizon in this context. Instead, the day-ahead forecast set used for this analysis consisted of hourly averages of forecast power that were created at 6AM for the following day (i.e. the day-ahead forecast for Tuesday consisted of 24 power values and was generated at 6AM on Monday). The actual and forecast datasets were synchronized in time, and any missing sections were ignored.

The methodology used in this study for characterizing the forecast errors consisted of blending traditional forecast performance evaluation tactics with ideas that had not been previously encountered. Trends in each individual dataset were examined first, and subsequent characterization of forecasting error was focused on faults in magnitude prediction as well as faults in the timing. Special emphasis was given to these magnitude and phase errors during large wind ramping events. The following discussion presents the key findings to an exhaustive statistical analysis that was performed on the available data. An extended presentation of this topic can be found in [13].

Ramp Identification

The wind industry has expressed major concern about forecasting errors during large ramp events, as these provide the greatest challenges for operators and carry the risks associated with resource planning. Thus, a ramp rate analysis was conducted including a statistical characterization of forecast errors during and near periods identified as significant ramp. Once the ramps were identified, a multi-faceted analysis of the correlation between actual ramp events and forecasted ramp events was conducted.

The available time series' of data for this project were limited to hourly resolution. The step change in power production from one hour to the next is equivalent to an hourly ramping rate (MW/hr), and is referred to as the "hourly delta" (or just delta) value for the remainder of this report. Equation 4 shows how the delta values were calculated. System planners are interested in the delta metric (particularly large values) as it provides insight into necessary balancing measures that might be needed to accommodate wind.

$$\Delta = P_{t+1} - P_t \tag{4}$$

Hourly delta values alone may not always be sufficient to capture the difficulties associated with larger ramping events. Therefore, a major component of this project was the development of a meticulous procedure for identifying entire ramp events that span multiple hourly timesteps. This procedure was written in MATLAB® and consists of a two-step moving average technique that allows for the definitive beginning and ending of a ramp to be specified. A total of five input parameters are used to search for desired ramp rates, ramp durations, and threshold values that capture the beginning and end of the event. A moving average is computed from the hourly delta values in either the actual or forecast time series. The moving average method was used to reduce the "noise", or eliminate smaller false ramps in the positive or negative direction that are actually part of a larger or sustained ramping

event in the opposite direction. The algorithm can search for ramps of any user-defined magnitude that occur over a selected period of time. Additionally, the algorithm performs correlation analysis between forecasted and actual ramp event starts that coincide within a desired timing window. It also determines if ramps were forecasted to occur in the correct or opposing direction of those that actually occurred, which is valuable information that must be captured in a synthetic forecast should the trend exist in state-of-the-art forecasts. This procedure has been named the Ramp Identification Algorithm (RIA), and was used throughout this project to identify the largest wind ramping events present in the actual power and forecast datasets. A more complete description of this algorithm is presented in [13].

Statistical Characterization

Subsequent statistical analysis was contingent on gaining complete confidence in the ability to specify entire ramp events. The statistical characterization process of this methodology was primarily focused on the following objectives:

- Trends were identified in power production levels and hourly delta values for both the actual data series and forecast data series.
- The forecast error was assessed using traditional metrics of MAE, RMSE etc., and a beta probability density function was fit to these forecast errors.
- Intensive examination of error trends during large ramp events selected by RIA was conducted. This included correlation analysis between actual and forecast ramps, and phase analysis of errors in ramp forecasting.
- All of the above steps were evaluated at significant forecast horizons.

Results and Discussion

Characteristics of the power production and hourly delta distributions from the actual and forecast data are presented. Additionally, forecast error characteristics are presented along with a beta probability density function that fits the results. Next, trends in the MAE, RMSE, and σ_Δ are presented during wind ramping events. Finally, temporal correlation between actual and forecast ramp events is presented. Most of these results are presented as a function of the forecast horizon.

A few of the methods used to quantify trends in the actual and commercially forecasted wind power time series are shown in Figure 1, Figure 2, and Figure 3. Histograms and duration curves were made for the distributions of power production values and distributions of delta values over the entire three-year period. Figure 1 presents the distribution of power production levels (over the entire dataset) as a percentage of plant capacity from both the actual wind plant and several forecast horizons from the commercial data. The commercial forecast reasonably matched the actual trends for very short forecast horizons, yet the forecast predicted very few high output values at even four hours out. In the day-ahead situation, less overall variability was predicted as demonstrated by the large zero output bin, yet some high output values were predicted.

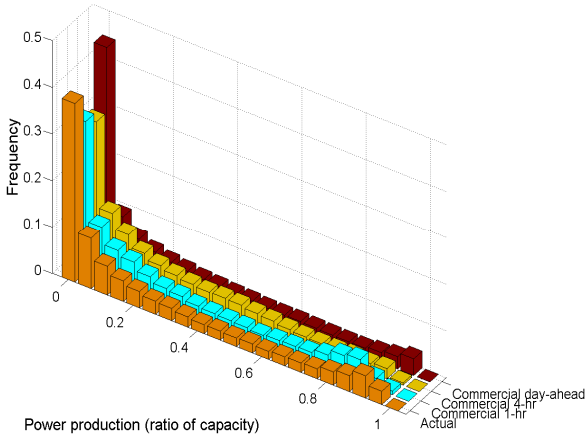

Figure 1: Distribution of power production from actual wind plant and commercial forecasts for 1-hr, 4-hr, and day-ahead horizons.

The duration curve shown by Figure 2 and the histogram in Figure 3 present the distribution of hourly step changes in power production, or hourly deltas for the actual power values. The distribution appears somewhat normal in the histogram, but the circle on the duration curve suggests that the hourly delta values from the actual wind plant are greater than or equal to zero approximately 54% of the time. System planners will take this to mean that the wind plant will be ramping up more often than not, and as shown by Figure 3, the majority or hourly deltas will be relatively small compared to the overall plant capacity. These figures serve as examples of useful methods to evaluate and compare two wind power time series. Matching distributions from the forecast data series and further implications of these plots are discussed in [13].

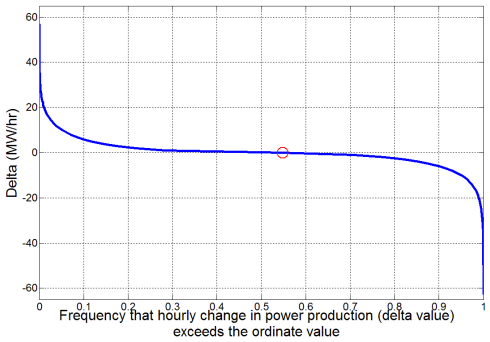

Figure 2: Duration curve of hourly step changes (deltas) in power production from actual wind power dataset. Circle indicates frequency that step changes exceed 0 MW/hr. Plant capacity is in excess of 60 MW.

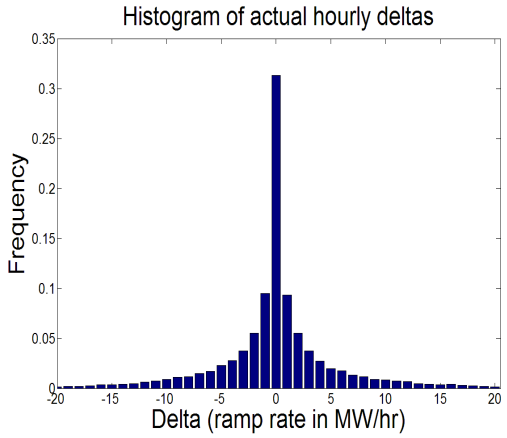

Figure 3: Histogram of delta values for actual power dataset indicating distribution of hour-to-hour changes in plant output. Plant capacity is in excess of 60 MW.

Traditional Error Metrics

The mathematical difference between forecasted power and actual power output at any given time is commonly called the forecast error (see Equation 1). The mean bias is obtained by taking the average of the forecast error values over the entire dataset, as shown by Equation 5.

$$Mean\ bias = \frac{1}{n}\sum_{i=1}^{n}(P_a - P_f) \qquad (5)$$

The standard deviation of forecast error, $\sigma_{f.e.}$ is given by taking the standard deviation of the values calculated by Equation 1. Figure 4 shows several of the traditional metrics used to evaluate forecast performance, including: the mean and standard deviation of the forecast error, the MAE given by Equation 2, and the RMSE given by Equation 3 (all versus the forecast horizon for the entire three years of data). Thus, each point plotted in the figure represents an average of 3-years of hourly forecasted and actual power data from the specified forecast horizon (roughly 26,000 data points). As demonstrated by the black line with '+' markers on the lower portion of the plot, there was an overall negative bias in the forecast errors, becoming increasingly negative as the forecast horizon increased, with a near-linear decrease during horizons 1 - 8. This negative bias of the forecast errors for each hour could be "by design" for this particular forecast, as the recipients of a commercial forecast often prefer that the wind power production be under-predicted. The MAE is shown by the line that contains no markers. The MAE increased in a near-linear fashion during forecast horizons 1-8, and less extreme afterward.

The $\sigma_{f.e}$ and the RMSE for each horizon are shown in Figure 4 by the upper two lines, with open and closed-circle markers, respectively. These metrics also grew in a linear fashion during the same horizons as the mean bias and MAE. The $\sigma_{f.e}$ metric for each hour is an important parameter to understand: it is an indicator of the variability of the hourly forecast error about its mean. Thus, the mean bias was nearly zero for the 1-hour forecast horizon, but the standard deviation of this bias was nearly 10%. This latter metric is arguably of more value to an electrical system planner/operator to aid in understanding the potential variability of errors associated with predicting the wind. The similarities between the RMSE and $\sigma_{f.e}$ are explained in [11], and recall that they would be equivalent if the mean bias was zero.

The distinct "elbow" feature seen in all four of these metrics is not a surprise. It is well-known that state-of-the-art forecast providers use proprietary techniques (often statistical) to decrease forecasting errors during the first several hours ahead of the operational hour, after which a transition to NWP predictions is made. The important things to gather from these results are the variability and dependence of each metric with the forecast horizon.

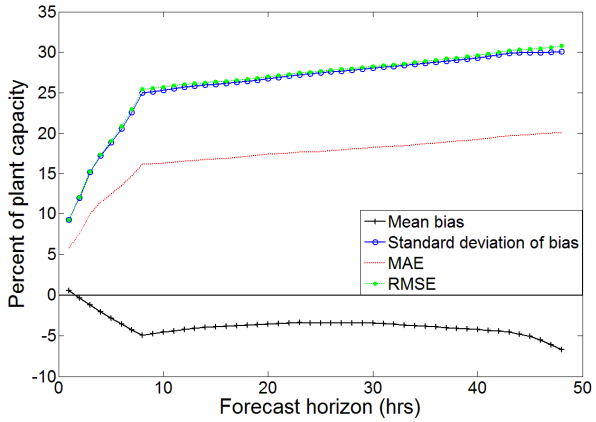

Figure 4: MAE, RMSE, Mean bias, and standard deviation of forecast error vs. forecast horizon.

Probability Density Function of Forecast Error

Historically, wind forecasting errors were presumed to follow a normal distribution. It was thought that there was roughly equal chance of over and under-prediction. This trend is generally true for wind velocity forecasting. However, it has been shown that the actual distribution of wind power errors more closely follows the Beta probability density function [14, 15]. The results of this project were in agreement with their conclusions, as can be seen by Figure 5 which presents the day-ahead forecast horizon. The normalized error bias was plotted against an equation-based beta function with calculated parameters as presented in [14], as well as the MATLAB® function called "betafit". The beta distribution is appropriate for wind power applications because it is typically bound between 0 and 1, consistent with normalized errors found in wind power forecasting. The forecast errors for other horizons also fit the beta distribution. This result is important for system planning concerns in that it allows error probabilities to be estimated based on anticipated production levels [15].

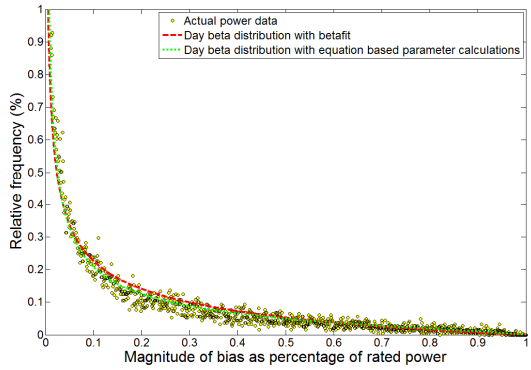

Figure 5: Distribution of forecast errors from available data along with two forms of a beta distribution.

Correlation coefficient of power values

The correlation coefficient, R, between forecast and actual power values was also investigated for various forecast horizons, two of which are shown by Figure 6 and Figure 7. The clustering shown in Figure 6 for the hour-ahead forecast around the linear trend line demonstrates more accurate forecast results than the scattered values in the 8-hr horizon as shown by Figure 7. The correlation coefficient can be extracted from these scatter plots, and the metric gives insight into forecast structure and can be used to validate synthetic forecast data. Figure 8 shows the correlation coefficient between the actual and forecast power values, vs. the forecast horizon. The R-value drops dramatically during forecast horizons 1-8, and then continues to drop at a lesser rate after that. These results should be compared to those in Figure 4 as the interesting behavior during the 1-8 hour horizons are captured in both.

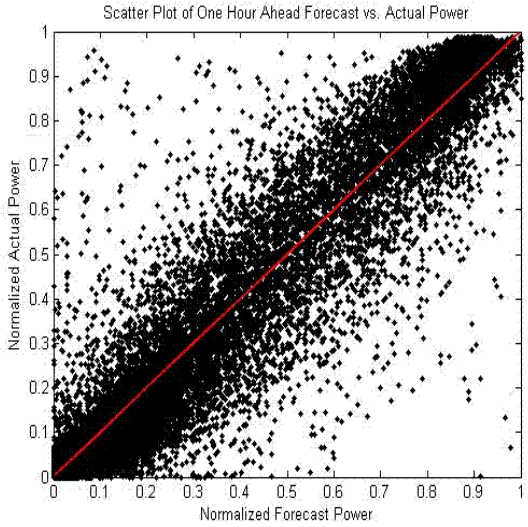

Figure 6: Scatter plot of power values for actual and forecast data for 1-hr forecast horizon.

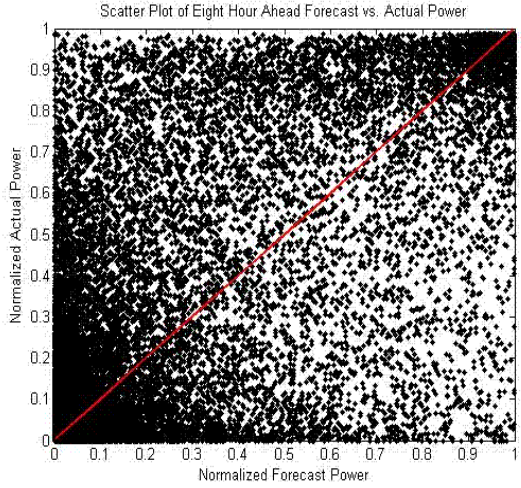

Figure 7: Scatter plot of power values for actual and forecast data for the 8-hr forecast horizon.

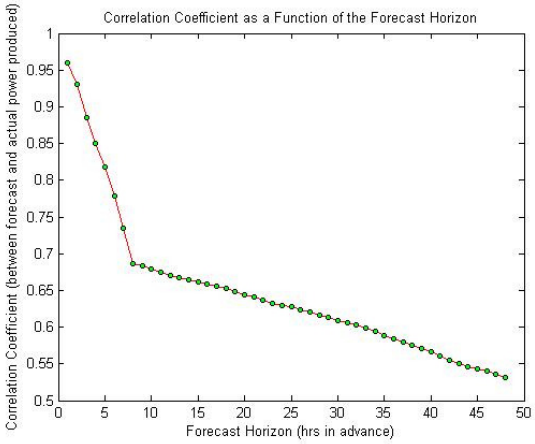

Figure 8: Correlation coefficient between actual and forecast power production vs. forecast horizon.

Ramp Analysis

The results of error characterization near wind ramping events are presented first as they relate to only the hourly delta values, and second as they relate to the multi-hour ramp events chosen by RIA.

Dependence of Errors on Delta Values

Recall that the hourly delta value simply gives the change in production from one hour to the next (see Equation 4). Figure 9 shows the number of hourly deltas from both the actual and forecasted data that exceeded 10 percent of the plant capacity. For forecast horizons of 3 hours and beyond, the forecast data contained far fewer deltas of these magnitudes than did the actual, suggesting less variability in forecasted power at longer time horizons. The lessened variability of the forecast data will impact some of the results shown later. It is interesting to note that a drastic drop in the number of deltas occurs during the same initial forecast horizons that showed notable results in Figure 4 and Figure 8. These results together show that a decrease in forecast power variability corresponds to increased forecasting errors.

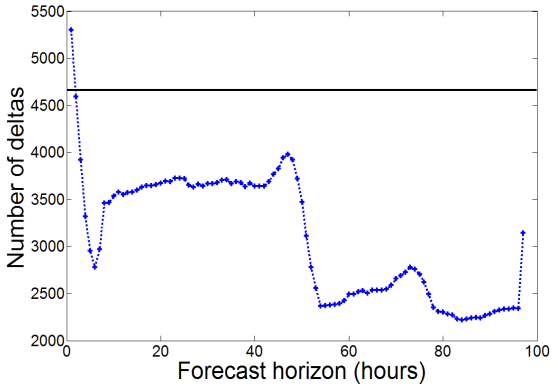

Figure 9: Number of deltas (hourly step changes in power production) > 10% capacity at various forecast horizons. The horizontal line represents the number in the actual dataset, and the other line represents forecast data.

As mentioned previously, major challenges and expenses experienced in balancing the electrical system often occur when the wind is ramping up or down. Figure 10 presents the mean absolute error in the 1-hour wind power forecast horizon as a function of the hourly delta value (the dashed line). The overall MAE (independent of delta values) is also shown for comparison. The abscissa in this plot gives the delta value (hourly step change) as a percentage of plant capacity. As shown, the mean absolute error increases as a function of the delta value. The solid line shows the frequency that delta values of each magnitude occur, and note that there are very few instances of which the hourly deltas are large. The interesting result is that although large delta values seldom occurred, the MAE during those times was much larger than when delta values were small. This figure not only demonstrates that forecast accuracy decreases inversely with the magnitude of the hourly delta, it also shows that an overall MAE value does not offer complete description of forecast error.

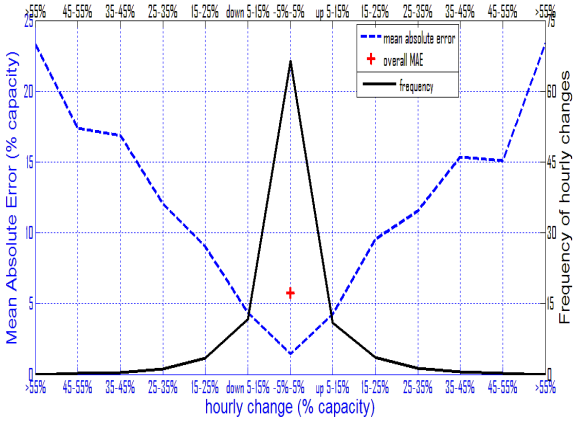

Figure 10: Frequency of occurrence and MAE during various sizes of step changes in actual hourly power production. Both up and down steps are included, as well as the overall MAE. 1-hr forecast horizon data are shown.

Treatment of Large Wind Ramping Events

Results for the multi-hourly ramping events chosen by RIA showed increased MAE during ramp events. Additionally, the predicted mean and standard deviation of the delta values *during* these ramp events became less-accurate at longer forecast horizons. The following analysis is based upon a two-hour rolling average of the deltas used to pick out the largest ramp events of the entire data set. Various rates of change were used to select the top 100 and 900 largest ramp events[2], with up ramps kept separate from down ramps. Although these extreme events made up a small percentage of total ramp events during the three-year period of the data, they will be of interest to system planners due to the challenges involved with load and production balancing. Figure 11 shows the mean delta value during the top 100 largest ramp events as chosen by RIA from the actual dataset. It is important to notice that predicted values were relatively accurate for the 1-hr forecast horizon, and not so accurate for hours 3-8. In fact, the average forecasted hourly delta value for both up and down ramps became closer to zero during these

[2] The top 900 ramp events occurred when the wind plant production level changed by approximately ±35% of total capacity in a two-hour timeframe. The top 100 ramp events occurred when the production level changed by approximately ±66% of total capacity in a two-hour timeframe. See [13] for further discussion.

horizons, confirming the results of Figure 9 that there is less variability in the forecast dataset. Figure 12 shows the standard deviation of the hourly delta values, or σ_Δ, during the same 100 largest ramp events. The forecasted σ_Δ values again were smaller for horizons 3-8, suggesting less variability in forecast power values.

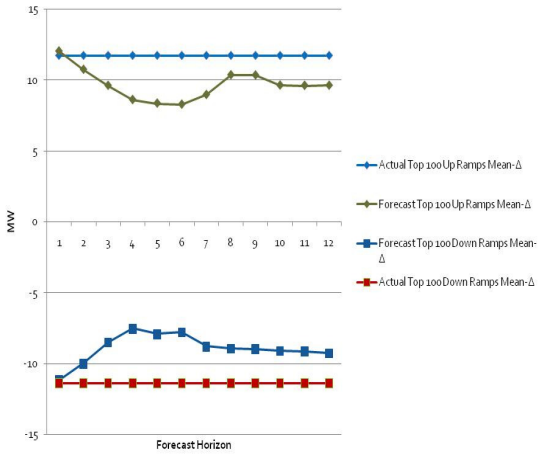

Figure 11: Mean delta value during top 100 largest ramp events for actual dataset.

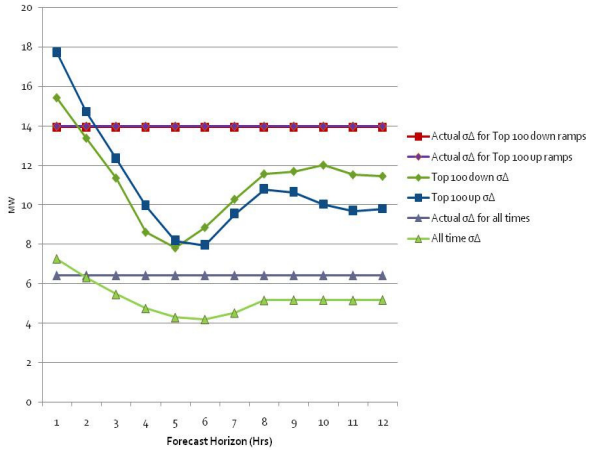

Figure 12: Standard deviation of delta value during top 100 largest ramp events for actual dataset.

A larger number of ramp events were chosen for further MAE analysis. Figure 13 demonstrates the increased MAE during the top 900 ramp events, which confirms the concerns of system operators and planners. During all forecast horizons, the MAE was greater during times defined as either up or down ramps in the actual power dataset. Additionally, it can be seen that the MAE tended to grow as a function of forecast horizon, with a notable steep increase in the 1-3 hour-ahead horizon. The MAE during ramps was found to be 2-5 times as large as during times not defined as ramps.

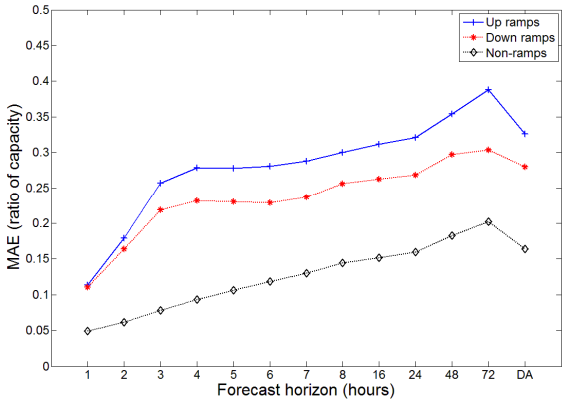

Figure 13: MAE during ramps and non-ramps for various forecast horizons (top 900 ramps)

Phase Error Analysis

The MAE has traditionally been used to quantify error during ramp events, but it does not offer a complete description of forecast accuracy. To extend the forecast error characterization, the magnitude component of the error was complemented by a phase error component that captures the correlation of actual and forecasted ramps in time. The first goal was to determine if trends existed in whether forecasted ramps lead or lagged the actual wind ramp events. Figure 14 and Figure 15 demonstrate the correlation analysis between actual and forecasted ramp events. Figure 14 shows the frequency of correctly forecasted start times as a function of forecast horizon, meaning that the forecasted ramp events began during the same hour as the actual ramp events, and were of similar magnitude. Because up and down ramps were kept separate, the overall frequency of correctly forecasted ramps can be obtained by adding the ordinate values of both lines in Figure 14, (e.g. at 1-hr horizon approximately 90% of ramp start times were correctly forecasted). Figure 15 shows the frequency of correlated forecast ramp events that were either leading or lagging actual events in time. This figure shows only those forecast ramps events that began within a timing window of ±4 hours of the actual ramp, and does not include any ramps that were forecasted to begin at exactly the correct hour. Therefore, of the ramps forecasted correctly within the ±4-hour timing window (but not on the exact hour), Figure 15 shows that an overwhelming majority of both up and down ramp events were forecasted to lag, or occur later than the actual ramp events. During similar hours of forecast horizon, the phase accuracy of predicting ramp events drops off dramatically in Figure 14. This finding is certainly valuable to system planners and could lead to adjustments in forecast interpretation.

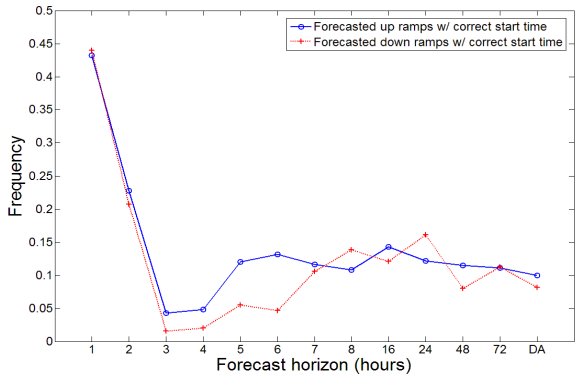

Figure 14: Frequency of correctly forecasted up and down ramps versus forecast horizon. Forecasted ramps were of similar size to actual ramps.

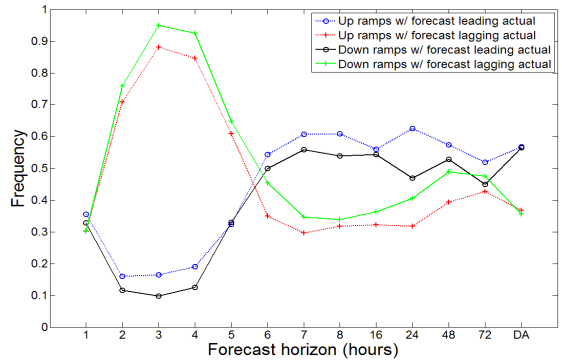

Figure 15: Frequency of up and down ramp starts leading or lagging actual ramp starts. Only forecasted ramps correlated within ± 4 hour window of actual ramps are included.

Phase errors for ramp events were quantified in terms of hours. Figure 16 presents the mean temporal bias (MTB) of the top 900 ramp events. The MTB indicates the average number of hours by which the actual ramp event starts were missed by the forecast, for various forecast horizons. The hour-ahead horizon had a near-zero MTB, suggesting that the majority of ramp start times were correctly forecasted in this short horizon. Hours 2-5 show a negative MTB, indicating that forecasted ramp starts were lagging, or occurring after the actual ramp starts (confirmed by Figure 15). At hour six, the average value of the phase error for predicted ramps makes a transition from lagging to leading the actual ramps. The ramp identification algorithm used a 4-hour window on either side of the actual ramp starts to search for forecasted ramps of similar magnitude. The percentage of actual ramps that were correctly forecast (refer again to Figure 14) can also be obtained from the algorithm that produced the MTB.

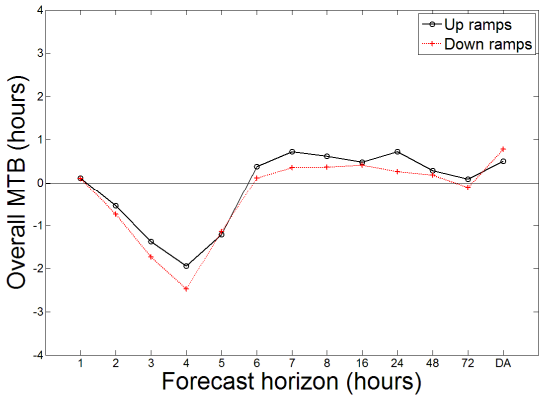

Figure 16: Mean temporal bias for top 900 largest actual ramp events that also have a corresponding forecasted ramp of similar magnitude within a ± 4 hr window.

There are two additional items of significance contained in Figure 16. First of all, there is a trend toward a worse MTB during hours two through four. Forecasted start times for both up and down ramps become progressively worse, with ramp events being missed by almost three hours (on average) in the worst case. The second item of significance is the appearance of a small MTB for hours six and beyond. This does not actually mean that ramp forecasting becomes better at longer horizons; recall from Figure 9 that there are far fewer forecasted ramp events during these time horizons. Since the figure only includes data from actual ramps that also have a forecasted ramp within the ±4 hour window, it is not unreasonable to expect skewed results during the forecast horizons that contain so few ramp events to begin with.

Conclusions

A rigorous statistical characterization of the wind power forecast errors was conducted, with only the most significant results presented in this paper. A more thorough presentation of all results is given in [13]. A number of analytical techniques have been presented, resulting in a process for comprehensively characterizing errors in wind power forecasting. Many of the most interesting findings occurred in the forecast horizons of 1-8 hours, when commercial forecast providers use proprietary methods to modify NWP models. The change from proprietary blending to NWP predictions at the approximately 8-hour forecast horizon is evident in the transitions seen in most figures of this report. The results and discussion above demonstrate that a relatively small number of statistical parameters can be used to adequately describe forecast error characteristics and capture both the trends and variability of the expected errors. This methodology contains characteristics of the actual and forecast

datasets by themselves, the mean bias levels computed from the raw differences between actual and forecast datasets, and the correlation and phase errors of large ramping events. These significant parameters are summarized in Table 1, and are ranked in order of most to least significant. It is important to note that there will be one set of these parameters for each forecast horizon. The intensive investigation of the magnitude and phase error trends near and during ramp events allows for a more complete understanding of the challenges that will be faced in reality. An overall MAE, RMSE, or simple bias metric does not accomplish this.

Table 1: Selected statistical parameters for error characterization

MAE (ramps)
MAE (non-ramps)
RMSE
$\sigma_{f.e}$
R
Mean Δ
σ_{Δ}
Mean Bias
MTB
% correct ramp starts

In addition to these metrics, the distributions of actual and forecasted power production levels and delta values should be investigated, along with the distribution of bias values.

Although the results presented here apply only to one particular power and forecast couple, the algorithm for ramp identification, statistical characterization, and important parameter selection has been developed and could be applied to any set of wind power and forecast time series. The techniques presented here could be used to verify simulated wind power data, and further implications include the creation of a computationally inexpensive synthetic forecast that is formulated by reproducing the statistical trends and significant error characteristics seen in an appropriate real forecast. This would be valuable for future wind integration studies. These implications are discussed further in [16].

Acknowledgements

Special thanks to the National Renewable Energy Laboratory for support of this work under subcontract XXL-7-77283-01.

References

[1] Zack, J. (2005) *Overview of Wind Energy Generation Forecasting*, AWS TrueWind, report to NYSERDA and NYISO.

[2] Piwko, R., et. al. (2005) *The Effects of Integrating Wind Power on Transmission System Planning, Reliability, and Operations*, GE Energy Consulting, prepared for NYSERDA.

[3] Lindenberg, S., Smith, B., O'Dell, K., DeMeo, D. (2008) *20% Wind Energy by 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply*, DOE/GO-102008-2567.

[4] Loutan, C., et al. (2007) *Integration of Renewable Resources*, California Independent System Operator.

[5] EWEA (2005) *Large Scale Integration of Wind Energy in the European Power Supply: analysis, issues, and recommendations*, A report by the European Wind Energy Association.

[6] Manwell, J.F., et. al. (2002) Wind Energy Explained, West Sussex, England: John Wiley & Sons Ltd.

[7] Lange, M. (2005) *On the uncertainty of wind power predictions – Analysis of the forecast accuracy and statistical distribution of errors*, J. Sol. Energy Eng,. Vol 127, pp 177-184.

[8] Söder, L. (1993) *Modeling of Wind Power Forecast Uncertainty*, Proceedings of the European Community Wind Energy Conference, 8 - 12 March 1993, Lübeck-Travemünde, pp. 786 - 789.

[9] Brower, M., (AWS Truewind, LLC) (2007) *Intermittency Analysis Project: Characterizing New Wind Resources in California*, California Energy Commission, PIER Renewable Energy Technologies. CEC-500-2007-XXX

[10] *Western Wind and Solar Integration Study*, working web site by the National Renewable Energy Laboratory, http://wind.nrel.gov/public/WWIS/

[11] Madsen, H., et. al. (2004) *A Protocol for Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models*, Project ANEMOS

[12] Dragoon, K., Milligan, M. (2003) *Assessing Wind Integration Costs with Dispatch Models: A Case Study of PacifiCorp*, NREL/CP-500-34022, National Renewable Energy Laboratory, Golde, CO.

[13] Bielecki, M. (2010) *Characterization of Errors in Wind Power Forecasting*, Master's Thesis, Northern Arizona University.

[14] Bludszuweit, H., Dominguez-Navarro, J. (2008) *Statistical Analysis of Wind Power Forecast Error*, IEEE Transactions on Power Systems, Vol. 23, No. 3

[15] Bofinger, S., Luig, A., Beyer, H., *Qualification of wind power forecasts*, University of Applied Sciences Magdeburg-Stendal, Dept. of Electrical Engineering

[16] Kemper, J., et al. (2010) *Modeling of Wind Power Production Forecast Errors for Wind Integration Studies*, Proc. Of ASME 2010 4th International Conference on Energy Sustainability.

