STATISTICAL CHARACTERIZATION OF ERRORS IN WIND POWER FORECASTING

Mark F. Bielecki
Graduate Research Assistant

Jason J. Kemper
Graduate Research Assistant

T. L. Acker
Professor, Mechanical Engineering

NORTHERN ARIZONA UNIVERSITY

Sustainable Energy Solutions

1. Intro to Wind Power Forecasting

Commercial wind power forecasts are created with a blend of physics-based and statistical techniques.

Model Output Statistics (MOS) and complex proprietary algorithms are used to formulate final wind power forecasts and minimize forecasting errors.

Errors associated with predicting the wind speed can be exacerbated when wind speed values are converted to wind power values due to the nonlinear nature of turbine power curves.

As global wind penetration levels increase, an understanding of forecast error characteristics can aid load-serving entities (LSEs) with accommodating the impacts associated with the variability and uncertainty of wind.

Large-scale wind integration studies often utilize simulated wind power and forecast data, which must be validated to ensure study results are realistic.

There is currently no industry standard to quantify errors in wind power forecasting

This project was completed as part of a master's thesis in mechanical engineering.

2. Study Methodology

A combination of traditional and new forecast performance evaluation criteria were used to characterize the errors in wind power forecasting.

Available data:
Three-year time series (2004-2006) of hourly wind power data from an operating wind power plant and simultaneous hourly forecast values generated for the same wind power plant by state-of-the-art commercial forecast provider. Distributions in both datasets were assessed prior to error analysis for data validation (example by Figure 1)

Particular focus was given to forecast errors during wind power ramp events, due to an increased interest by the power industry.

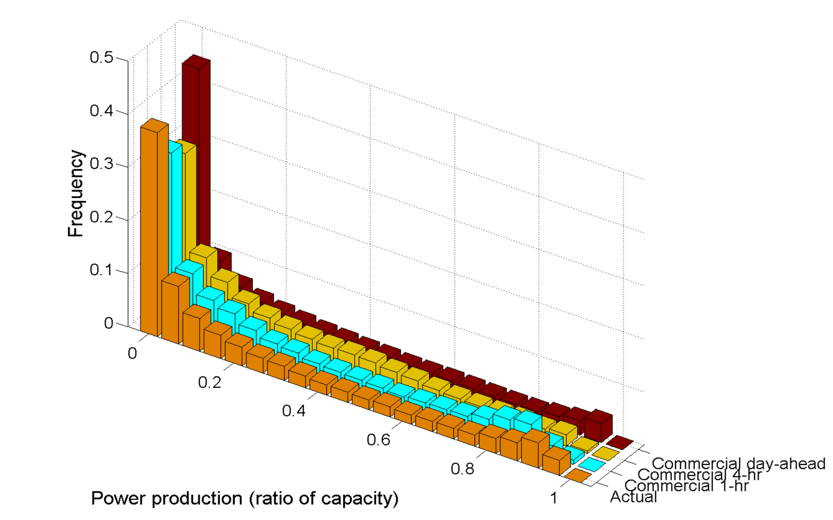

Figure 1: Distribution of power production from actual wind plant and commercial forecasts for 1-hr, 4-hr, and day-ahead horizons.

3. Traditional Error Analysis Methods

Wind power forecast errors can be quantified by several metrics, none of which offers a complete description of forecast performance. Values are typically presented as a function of forecast horizon, or the number of hours ahead of real-time.

The root mean square error (RMSE) and mean absolute error (MAE) are two of the most common.

The mean bias is defined as the average of the difference between forecast and actual power values for a given forecast horizon.

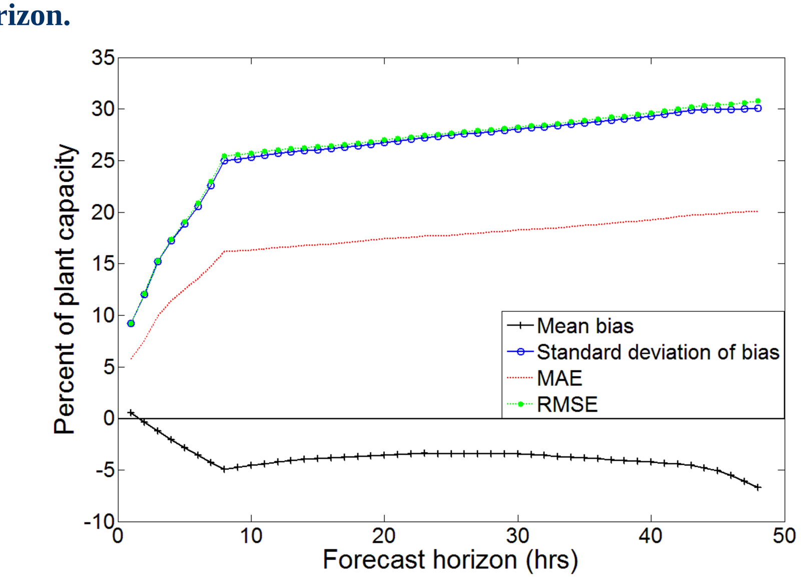

Figure 2: MAE, RMSE, mean and standard deviation of forecast error vs. forecast horizon. Each point represents average of data from all 3 years of actual and commercially forecast sets.

4. Ramp Identification Algorithm

Recently, there has been increased interest in the area of ramp event forecasting in the wind industry. Large wind ramping events can lead to significant challenges for electrical system operators. The potential for sizeable errors in wind power forecasts occurs during large ramp events due to the difficulty in predicting weather front timing and regional wind effects. Sufficient reserves must be available to accommodate these types of errors, which can lead to increased wind integration costs.

A portion of this project was dedicated to the formulation of a Ramp Identification Algorithm (called RIA) to select ramp events of a desired magnitude and duration from the actual or forecasted wind power time series. The algorithm can also perform correlation analysis between forecasted and actual ramp events that occur within a desired timing window (e.g. ± 4 hours of each other).

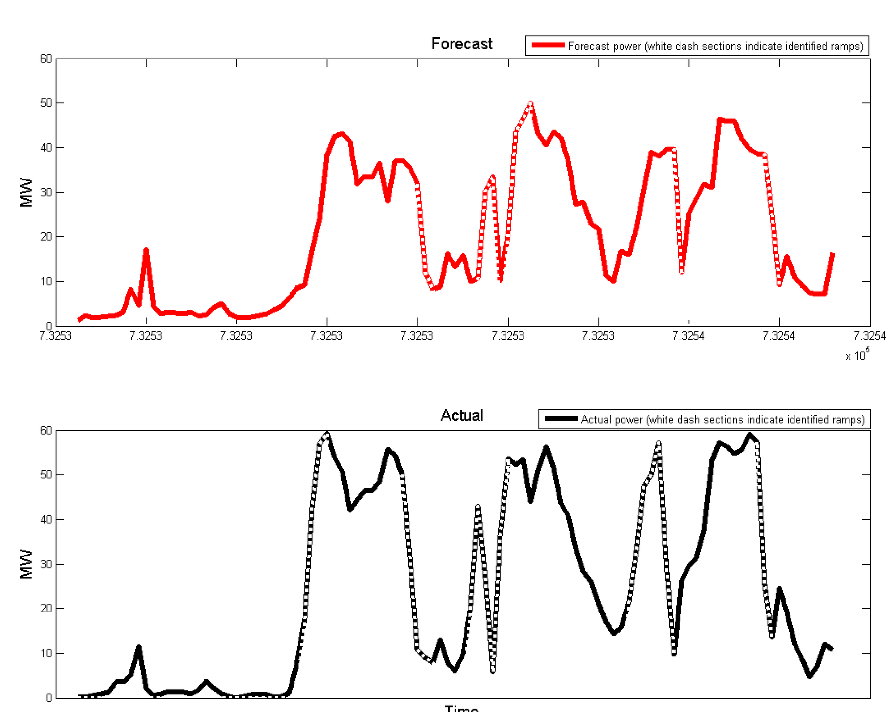

Figure 3: example of periods defined as ramps by the RIA for corresponding times of actual and forecast wind power time series

The algorithm consists of a two-step moving average technique that allows for the definitive beginning and ending of a ramp to be specified (see Figure 3). A total of five input parameters are selected by the user 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 (step change) values in either the actual or forecast time series.

Due to the hourly nature of the data, ramp rates are expressed in MW/hr, and all ramps are at least 1 hour in duration. The algorithm could identify shorter ramp events with finer resolution data.

As an example, the top 900 largest ramp events of the three-year series occurred when the wind plant production level changed by approximately ±35% of total capacity in a two-hour timeframe. The top 100 largest ramp events occurred when the production level changed by approximately ±66% of total capacity in a two-hour timeframe. See Bielecki (2010) for further discussion.

5. Forecast Errors During Wind Ramp Events

Once identified, detailed attention was paid to forecast error patterns during and near large ramp events.

Forecasting errors were evaluated first as a function of the hourly step change in power production (or hourly ramp rate) of the actual wind power plant. MAE increased as a function of hourly ramp rate (both up and down) as seen in the dashed blue line of Figure 4, despite fewer occurrences of large hourly ramp rates (as seen by solid black line). The overall MAE is shown by the red '+' sign, demonstrating that a single MAE metric is not sufficient for complete forecast error description.

When entire ramp events (often multi-hourly) were identified by the RIA (as opposed to single hourly ramp rates alone), it can be seen in Figure 5 that the average MAE was also greater during both up and down ramps than during times not defined as ramps (confirming system operator concerns).

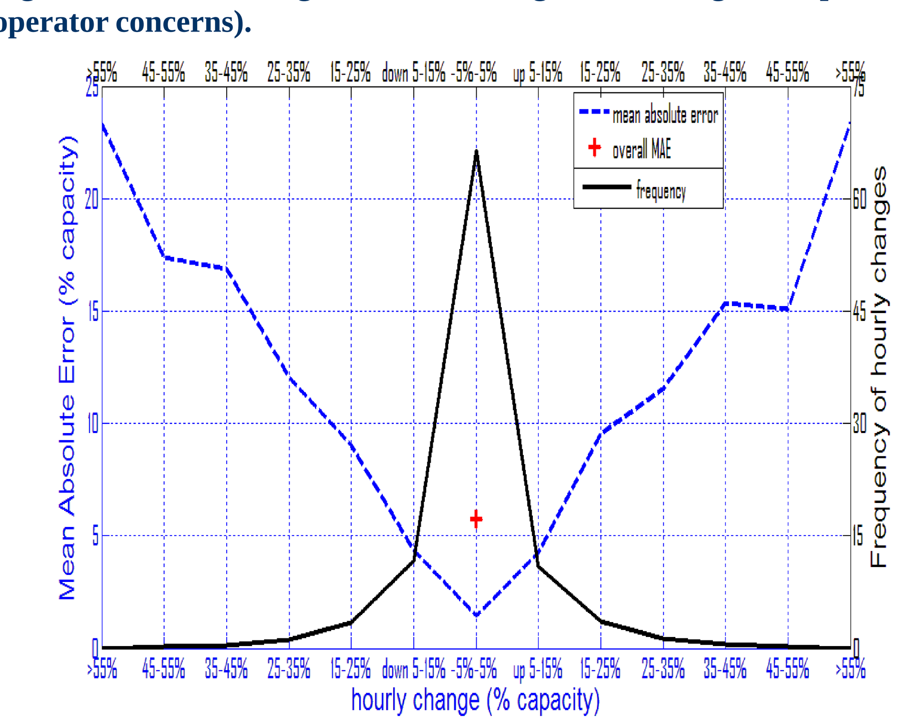

Figure 4: Frequency of occurrence and MAE during various sizes of step changes in actual hourly power production. Both up and down steps are included, as well as the overall MAE. 1-hr forecast horizon data are shown.

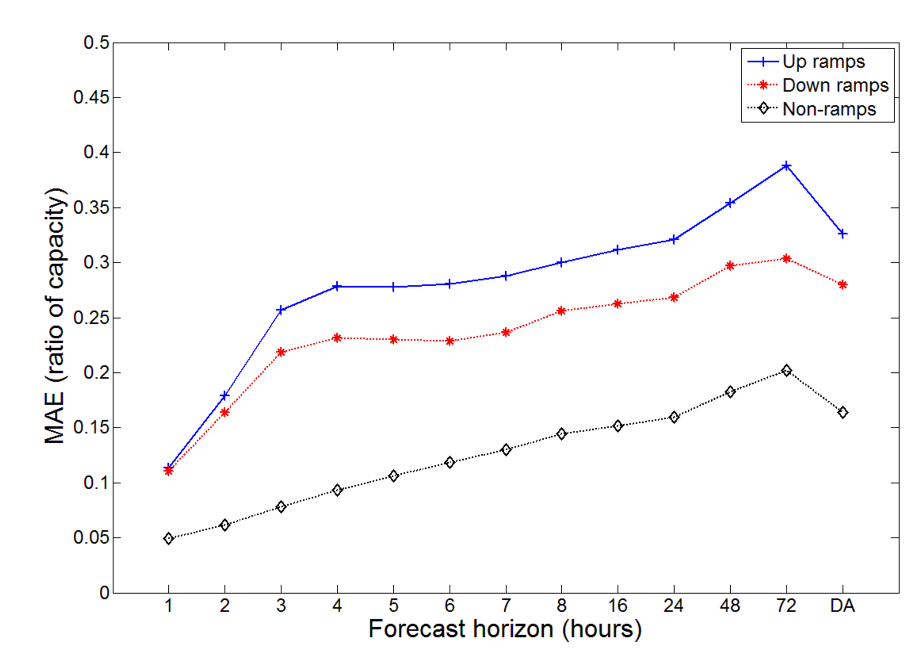

Figure 5: MAE during ramps and non-ramps for various forecast horizons (top 900 ramps)

6. Temporal Error Analysis

With respect to ramp forecasting, magnitude error metrics such as the forecast error (or bias defined as $P_{forecast}-P_{actual}$), MAE, and RMSE may not offer complete error description. Ramps may forecasted correctly in size, but timing can be missed (refer to Figure 6)

Characterization of temporal errors in ramp events can be an effective tool for additional forecast performance evaluation.

All ramp comparison presented here consists of forecasted and actual ramps of similar size occurring within ±4 hours of each other. Figure 7 shows the frequency of ramps correctly forecasted in both time and size.

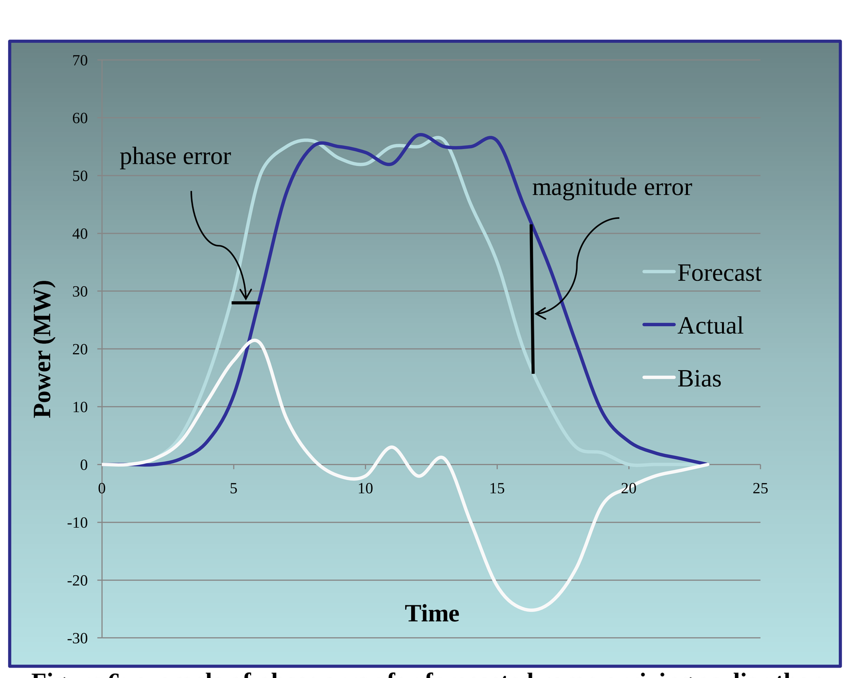

Figure 6: example of phase error for forecasted ramp arriving earlier than actual ramp of similar magnitude .

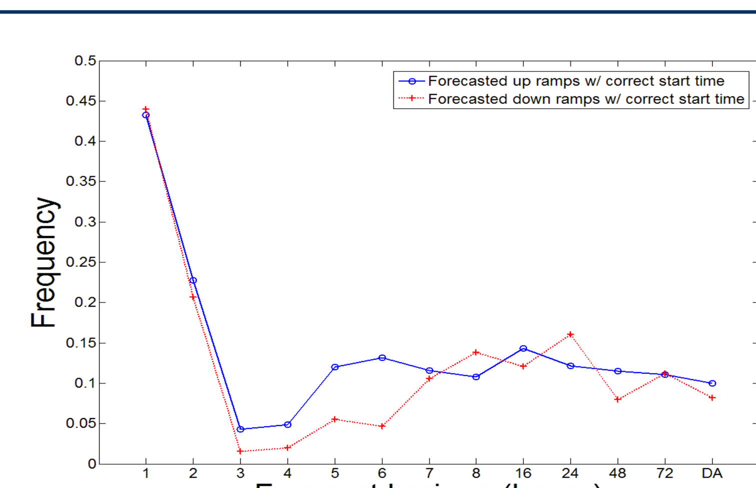

Figure 7: Frequency of correctly forecasted up and down ramps versus forecast horizon. Forecasted ramps were of similar size to actual ramps.

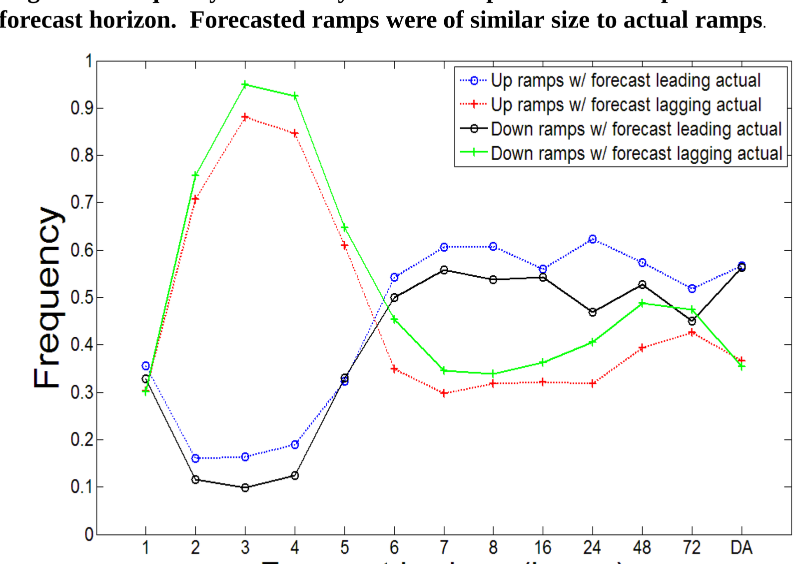

Figure 8: Frequency of up and down ramp starts leading or lagging actual ramp starts. Only forecasted ramps correlated within ± 4 hour window of actual ramps are included, correctly forecasted ramps are not.

Figure 8 shows the frequency of forecasted ramps that were either leading or lagging actual ramp events in time (if correlated in ± 4 hour window). For the first several forecast horizon hours, both up and down forecasted ramps were lagging behind, or occurring after the actual ramp events.

Figure 9 shows the mean temporal bias, or average number of hours by which the forecasted ramp event start times missed the actual ramp start (e.g. at the 4-hour forecast horizon, the start times for both up and down forecasted ramps occurred about 2 hours later than the actual ramp events).

Taken together, these plots provide valuable information to understand forecast performance. Consider the 3-hour forecast horizon: roughly 85-95% of the forecasted ramps (of the magnitude specified by the RIA) were lagging the actual ramps (Figure 8) by an average amount of about 1.5 hours (Figure 9)

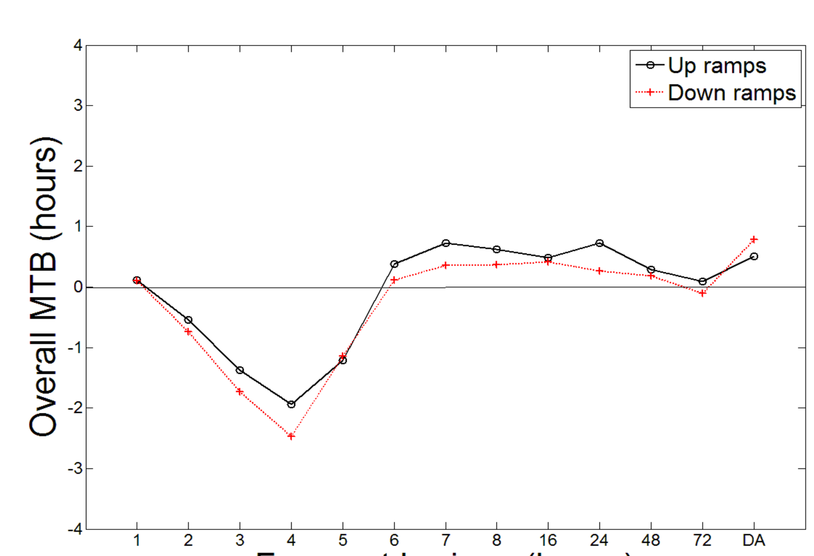

Figure 9: Mean temporal bias for top 900 largest actual ramp events that also have a corresponding forecasted ramp of similar magnitude within a ± 4 hr window.

7. Conclusion

A rigorous statistical characterization of wind power forecast errors was conducted for simultaneous actual and commercially forecasted wind power data from an operating wind power plant. Many of the most interesting findings occurred in forecast horizons of 1-8 hours when forecast providers use proprietary methods to modify NWP models.

The results presented demonstrate that a relatively small number of statistical parameters can be used to describe forecast errors and capture both the trends and variability of the expected errors.

It is important to note that the results presented here apply only to a single power and forecast couple. However, the methodology and metrics presented could be applied to any set of corresponding wind power and forecast data.

The techniques presented here could be used to verify simulated wind power data or evaluate a synthetic forecast that is formulated by reproducing the statistical trends and significant error characteristics seen in an appropriate real forecast (refer to Kemper (2010). The metrics presented are all of importance to electrical system planners and operators seeking to integrate wind power.

References:

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

Kemper, J. (2010) *Applications and Modeling of Wind Power Production Forecast Errors Produced from Meso-Scale Simulations*, Master's Thesis, Northern Arizona University.

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