# STATISTICAL CHARACTERIZATION OF ERRORS IN WIND POWER FORECASTING

By Mark F. Bielecki

A Thesis
Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science
in Engineering

Northern Arizona University

May 2010

Approved:

___________________________
Tom Acker, Ph. D., Chair

___________________________
William Auberle, Ph. D.

___________________________
Ernesto Penado, Ph. D.

ABSTRACT

STATISTICAL CHARACTERIZATION OF ERRORS IN WIND POWER FORECASTING

Mark Bielecki

Wind power forecasting will play a more important role in electrical system planning with the greater wind penetrations of the coming decades. Wind will most likely 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. This thesis project was designed 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.

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 a rudimentary (and perhaps subjective) order of significance. These metrics included: mean absolute error, root mean square error, average magnitude of step changes, 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.

## Acknowledgements

I would like to extend a special thanks to Dr. Tom Acker for his extensive guidance and motivation during this project. I am also grateful to the National Renewable Energy Laboratory for support of this work under subcontract XXL-7-77283-01. Brian Parsons, Michael Milligan, and Yih-Huei Wan of NREL have provided valuable feedback and assistance for this work.

## Table of contents

## List of Tables

## List of Figures

**List of Acronyms**

| | |
|---|---|
| ARMA | Autoregressive moving average |
| CAISO | California Independent System Operator |
| DOE | Department of Energy |
| GCPUD | Grant County Public Utility District |
| ISO | Independent system operator |
| LSE | Load serving entity |
| MAE | Mean Absolute Error |
| MOS | Model Output Statistics |
| MTB | Mean Temporal Bias |
| MW | megawatt |
| NREL | National Renewable Energy Laboratory |
| NWP | Numerical Weather Prediction |
| NYISO | New York Independent System Operator |
| NYSERDA | New York State Energy Research and Development Authority |
| pdf | probability density function |
| RIA | Ramp Identification Algorithm |
| RMSE | Root Mean Square Error |
| RPS | Renewable Portfolio Standard |
| TSD | Temporal Standard Deviation |
| WWSIS | Western Wind and Solar Integration Study |

## Introduction/Project Objectives

The wind industry is a growing sector in the global energy market, as it contributes a domestic, clean, and water-free source of electricity generation. International efforts to stabilize electricity prices, reduce carbon emissions, and diversify energy portfolios have lead to the transformation of wind from a small-scale and often experimental topic to a significant component of the power industry. However, due to its dependence upon meteorological phenomenon, there are a unique set of concerns that must be addressed in order to successfully integrate wind power into the existing and future electrical grid system. As Load Serving Entities (LSEs) such as electrical 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 variable nature of wind refers to the inherent fluctuations in magnitude and direction, while uncertainty refers to the unpredictability of wind in its exact timing, location, and strength. Taken together, these characteristics add complexity to wind power development.

In the next several decades, not only will the net installed amounts of wind generating capacity increase, but wind will most likely constitute a larger percentage of the electrical generation mix. As a result, wind forecasting errors may have more significant effects on balancing operations because they will no longer be simply absorbed into load forecasting errors. In 2008, the U.S. Department of Energy (DOE) published an ambitious, national-scale wind integration study to detail the feasibility of generating 20% of the U.S. electricity demand with wind power by 2030 [Lindenberg et al. 2008]. The study found that the total installed wind capacity in the U.S. would have to increase nearly tenfold from the 2009 levels of 35,000 megawatts (MW)[1] to over 300,000 MW to accomplish the 20% goals.

[1] As reported on the Wind Powering America website: http://www.windpoweringamerica.gov

Penetration levels in some electrical systems would reach incredible levels, and the impacts and costs associated with these levels are in early stages of being well-understood.

Wind power forecasting comprises a sizeable portion of the wind industry, and a number of entities exist to provide advanced wind power forecasts to interested parties. Although state-of-the-art wind power forecasts are not completely accurate, they can be powerful tools for electrical system planning and 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 [Piwko et al. 2005, Zack 2005]. Many contemporary wind integration studies have cited the need for accurate forecasting to accomplish successful large-scale wind integration [Lindenberg et al. 2008, Loutan et al. 2007, EWEA 2005 ].

The errors associated with wind power prediction can lead to significant challenges for system planners and operators. Some means exist to quantify the impacts associated with the variability and uncertainty of wind, which can be used by electrical system planners to act accordingly in order to ensure that demand and supply of electricity are always balanced. Wind integration studies are generally conducted prior to the wind project development phase to assess these impacts. The two-part study conducted by the New York Independent System Operator (NYISO) and the New York State Energy Research and Development Authority (NYSERDA) along with General Electric Energy Consulting is considered one of the most comprehensive of these types of studies [Piwko et al. 2004, Piwko et al. 2005]. EnerNex Corporation has also assembled several U.S. wind integration studies [EnerNex 2004, EnerNex 2007, Zavadil 2006]. The studies often rely on meso-scale simulations of wind power and wind power forecasts for proposed development sites, as a lack of real historical wind measurements exists for many rural areas where wind projects are typically located. The simulations are created by taking historical wind measurements from the nearest possible location (perhaps hundreds of miles away) and using them as boundary conditions for meteorological tools that approximate what the wind would have been at the project location. These simulations are computationally intensive and financially expensive.

Simulated data representing wind power or wind power forecasts must be validated to ensure that they contain characteristics inherent of real outputs from operational wind power plants. Examples of some types of these data validations were carried out in Brower (2007) for the California Energy Commission's Intermittency Analysis Project, and in Loutan et al. (2007) for the California Independent System Operator's study on meeting Renewable Portfolio Standards (RPS) of 20% renewable generation. Extensive validations of simulated data have also been conducted for the Western Wind and Solar Integration Study (WWSIS) carried out by the National Renewable Energy Laboratory (NREL) and General Electric Consulting [Piwko et al. 2010]. These contemporary examples have demonstrated a demand for a methodical approach to wind power time series description. Aside from confirming that simulated wind displays the same traits as actual measured wind, it is also necessary to investigate the structure of errors in wind power forecasting. A thorough understanding of the forecast 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.

This thesis seeks to perform a comprehensive statistical characterization of wind power forecasting errors from a typical wind power plant and associated state-of-the-art professional forecast. The methodology and set of analytical tools presented here can serve as a systematic means to classify forecast performance, and it should therefore be emphasized that the entire process and techniques developed during the course of this project are more important than the results themselves. Accurate state-of-the art forecasts can alleviate many of the worries associated with wind integration, but it is not reasonable to expect forecasting to be perfected given the nature of the wind resource. The purpose here is to quantify and classify real forecasting errors from a single dataset, and to gain an understanding of which characterization metrics will be of interest for any forecast error analysis from any dataset. Commercial forecasts are computationally expensive to produce, and simpler synthetic forecasts may have the potential to afford the same attributes at a lesser cost. The process for error characterization presented in this thesis can be incorporated into integration studies to ensure that any simulated or synthetic data reproduces the same error trends seen in real applications.

At present time, there is no universal means of evaluating forecast performance, although some efforts to develop standardized criteria have been proposed [Madsen et al. 2004]. Statistical analysis of wind power forecasting errors has also been conducted on several levels [Bludszuweit 2008, Bofinger 2002, Lange 2005, Milligan 2003]. Several metrics are commonly presented, yet no individual metric offers a complete description of error tendencies. This thesis seeks to develop the process for complete characterization of forecast error trends, and to identify a set of parameters that can be used to adequately describe forecast structure and performance. Special attention was given to forecast horizons that are important to grid system planning. Although not all of these things will be important during real-time grid operations, the procedure and performance evaluation metrics can be used for integration studies and the identification of error trends may aid in system planning and risk assessment for real and proposed wind projects.

Recently, the power industry has had a growing interest in wind forecasting during large wind ramp events. A wind ramping event occurs when the output from a wind power plant increases or decreases drastically in a short amount of time causing significant increases or decreases in the amount of power being delivered to the grid. Ramping events present significant challenges to electrical system planners and operators. They are particularly difficult to forecast due to uncertainties in weather front timing and regional weather patterns. For these reasons, forecast errors can be substantial near ramp events, leading at one extreme to a greater need for expensive ancillary services to meet demand, and at the other extreme the need for wind curtailment or other generator shut downs. A significant portion of this thesis was dedicated to formulating a Ramp Identification Algorithm to select ramp events from the data, and intensive error analysis was conducted during those events. A multi-faceted approach to quantify magnitude and temporal or phase errors, in addition to a correlation analysis between actual and forecasted ramp events is presented here.

The methodical approach developed during this project began with the investigation of the structure and patterns found in actual and forecast wind power datasets. Traditional forecast error assessment methods were then performed, along with some new techniques involving magnitude and phase error trends during large wind ramping events. The resulting parameters were then combined into a single analytical tool consisting of important metrics that can be used to evaluate forecast performance. Although the results presented in this thesis apply only to one particular wind power and forecast couple, the repeatable process offered during this report for ramp identification, statistical characterization, and important parameter analysis 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 evaluate a 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.

## Relevance to Engineering Community

This project will make an important contribution to the engineering community by providing a new perspective to wind power forecasting error characterization. Wind behavior is dictated by large-scale atmospheric fluid dynamics. The conversion of wind energy into electrical power involves a number of mechanical and electrical concepts that affect the end result concerns of the grid operators and electricity consumers.  Wind integration studies are conducted to understand these impacts, and the methodology developed throughout this thesis project has direct applications for performing these studies and perhaps in describing forecast errors in the control room environment.

# Chapter 1: Summary of contemporary forecast structure, integration impacts, and accuracy assessment.

## 1.1 Literature review of relevant topics

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

A persistence-based approach was used to generate forecast datasets from wind power time series. Forecast error distributions were investigated and found to sometimes have high kurtosis values, leading to the suggestion that the beta pdf is an appropriate fit to wind power forecast errors.

- Bofinger, S., Luig, A., Beyer, H., (2002) *Qualification of wind power forecasts*, University of Applied Sciences Magdeburg-Stendal, Dept. of Electrical Engineering

Statistical distribution of forecast errors was investigated from the German PREVIENTO model when applied to regional wind power. Neural network postprocessing was used to reduce bias prior to fitting a probability density function. The beta function was found to be a reasonable pdf for errors within a 5% significance level using a chi-squared test.

- 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

Part of a large study to assess the impacts on operational, economic, and reliability concerns that integrating large amounts of wind into the California Bulk Power System will have. Simulated wind plant data and forecast data were generated to assess 22 GW of existing and potential new wind development sites. Characteristics of the simulated power were compared to those of actual power outputs for validation purposes.

- Ernst, B. et al. *Predicting the Wind*, IEEE Power & Energy Magazine, (2007). "Special Issue on Wind Integration: Driving Technology, Policy, and Economics", *IEEE Power & Energy Special Issue,* Volume 5, Number 6, November/December.

Several models and methodologies pertaining to wind forecasting for utility operations and planning are presented. Contemporary developments in numerical weather prediction models and wind power forecasting tools are discussed, including the use of ensemble forecasts which combine several tools to generate a single forecast with the goal of minimizing overall forecast errors.

- Lange, M., *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, 2005.

This study demonstrated the effects that the non-linear power curve had on amplifying the errors in wind speed predictions. The errors in wind power were shown to increases by factors as large as 2.6 times those seen in wind speed predictions. Some phase error analysis was conducted, but not relating to ramp correlation between actual and forecast time series.

- Milligan, M. and Schwartz, M.N. et.al. 2003. *Statistical Wind Power Forecasting for U.S. Wind Farms*, NREL/CP-500-35087, National Renewable Energy Laboratory, Golden, CO, November.

Provides assessment of improvements gained over persistence forecasting in the short-term by applying statistical time-series models to wind speed or power data alone. Autoregressive moving average (ARMA) models were used for comparison, and the 1-6 hour forecast horizon was considered. Various training periods were explored to capture annual variability in wind output.

- Piwko, R., Xinggang, B., Clark, K., et al. (2004). *The Effects of Integrating Wind Power On Transmission System Planning, Reliability, and Operations, Report on Phase 1: Preliminary Overall Reliability Assessment,* Prepared for the New York State Energy Research and Development Authority, by General Electric's Power Systems Energy Consulting, Schenectady, NY.

This comprehensive study was conducted to determine impacts of integrating large wind penetrations into the NYISO. Provides reliability assessment of integrating wind penetrations up to 10% of peak load on the New York State Bulk Power System (some study zones had levels as high as 36% peak load). It was found that additional load following would be needed to accommodate the 3300 MW of wind, but unit commitment measures would remain the same. Several statistical techniques were used.

- Piwko, R., Xinggang, B., Clark, K., et al. (2005). *The Effects of Integrating Wind Power On Transmission System Planning, Reliability, and Operations, Report on Phase 2: System Performance Evaluation*, Prepared for the New York State Energy Research and Development Authority, by General Electric's Power Systems Energy Consulting, Schenectady, NY.

Phase two of this study focused on system performance impacts of wind integration. Market structure, economic dispatch, and wind power forecast performance were discussed.

- Söder, L., 2004, *Simulation of Wind Speed Forecast Errors for Operation Planning of Multi-Area Power Systems*, 8th International Conference on Probabilistic Methods Applied to Power Systems, 12 - 16 September 2004, Iowa State University, United States.

ARMA procedures were used to develop a method to simulate wind speed outcomes from results based on available forecasts. It was assumed that a correlation exists between forecast errors from different regions.

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

This report gives a broad overview of wind power forecasting, including forecast development, short-term, and long-term forecast generation. Forecast evaluation criteria are presented, along with benefits of adding commercial forecasts to wind integration.

## 1.2 System impacts of wind integration

In order to fully understand the implications of characterizing wind power forecast errors, it is necessary to discuss the impacts that wind integration has on the electrical grid system. In contrast to most conventional generating units (e.g. coal, nuclear, gas, hydro etc.), wind power is generally taken to be a non-dispatchable, meaning that it cannot simply be started, adjusted, or kept at a certain level on demand[2]. For this reason, the electrical demand is taken on a "load net wind", or simply "net load" basis, meaning that the wind power contribution is treated as negative load and subtracted from the overall demand signal. This method effectively allows the variability of wind to be absorbed into the inherent variability of load by itself. However, large penetration levels of wind may add substantial amounts of variability to the system.

Electrical grid system operations are typically divided into three time periods of interest: regulation, load following, and unit commitment. Regulation is generally used to describe the timeframe of seconds to minutes, during which the small fluctuations in load must be balanced by small fluctuations in electrical generation. Load following refers to the scale of minutes to hours, during which larger trends in load fluctuation such as transitions from off-peak to on-peak energy usage may

[2] Although wind output can be reduced, or curtailed if needed.

occur. The unit commitment horizon refers to the hour to several hours ahead and beyond that generating units must be scheduled for use. Commitment is often based on market prices, resource purchases centered around price speculation, and outage schedules for specific generators. Load serving entities in the U.S. generally schedule their generation based on stability and economic factors, usually with a base load covered by coal, nuclear, and large hydro if available. System operators balance fluctuations in load and generation by using flexible resources such as gas turbines, hydroelectric generators, and other means. Stringent guidelines are in place by regulatory agencies requiring additional generation capacity, called ancillary services, to be available to cover variations in demand or failures of large generators or transmission lines.

The variable nature of wind as well as the mechanical nature and inertia associated with wind turbine technology relate to the interrelation between wind power variability and electrical system operations during the timeframes of interest. Figure 1 shows wind power behavior on timescales relative to electrical system planning and operations. The analysis presented in this thesis pertains primarily to the load following and unit commitment timeframes, and as a result the time series variability is mostly due to fluctuations in the output of the wind power plant (or wind farm) and regional weather behavior. The impacts of wind integration on system operations are generally characterized by estimating the overall integration cost, added levels of ancillary services for regulation and load following, and possible increases in reserve requirements. A summary of recent U.S. case studies involving quantified wind integration costs is shown in Table 1.

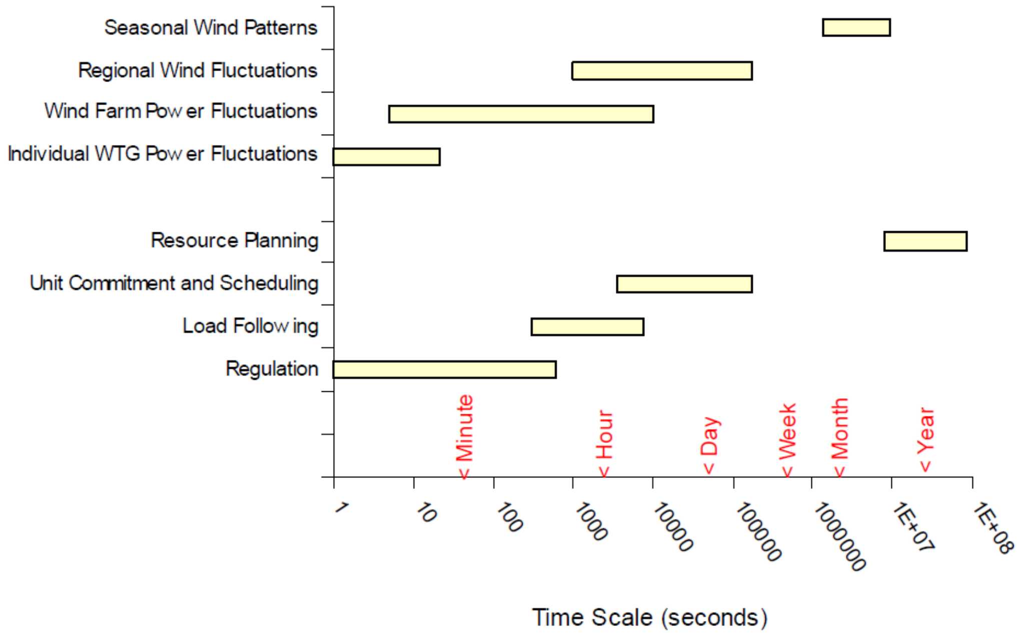


**Figure 1: Comparison of timeframes of interest to electrical system operations and associated wind power variability. (Source: Piwko 2004)**

**Table 1: Impacts and costs of several recent U.S. wind integration case studies. (Source: Wiser 2009)**

| Date | Study | Wind Capacity Penetration | Integration Cost ($/MWh) Regulation | Load Following | Unit Commit. | Gas Supply | TOTAL |
|---|---|---|---|---|---|---|---|
| 2003 | Xcel-UWIG | 3.5% | 0 | 0.41 | 1.44 | na | 1.85 |
| 2003 | We Energies | 29% | 1.02 | 0.15 | 1.75 | na | 2.92 |
| 2004 | Xcel-MNDOC | 15% | 0.23 | na | 4.37 | na | 4.60 |
| 2005 | PacifiCorp-2004 | 11% | 0 | 1.48 | 3.16 | na | 4.64 |
| 2006 | Calif. (multi-year)* | 4% | 0.45 | trace | trace | na | 0.45 |
| 2006 | Xcel-PSCo | 15% | 0.20 | na | 3.32 | 1.45 | 4.97 |
| 2006 | MN-MISO** | 31% | na | na | na | na | 4.41 |
| 2007 | Puget Sound Energy | 12% | na | na | na | na | 6.94 |
| 2007 | Arizona Pub. Service | 15% | 0.37 | 2.65 | 1.06 | na | 4.08 |
| 2007 | Avista Utilities | 30% | 1.43 | 4.40 | 3.00 | na | 8.84 |
| 2007 | Idaho Power | 20% | na | na | na | na | 7.92 |
| 2007 | PacifiCorp-2007 | 18% | na | 1.10 | 4.00 | na | 5.10 |
| 2008 | Xcel-PSCo*** | 20% | na | na | na | na | 8.56 |

* Regulation costs represent 3-year average.
** Highest over 3-year evaluation period.
*** This integration cost reflects a $10/MMBtu natural gas price scenario. This cost is much higher than the integration cost calculated for Xcel-PSCo in 2006, in large measure due to the higher natural gas price: had the gas price from the 2006 study been used in the 2008 study, the integration cost would drop to $5.13/MWh.

Wind power adds additional variability and uncertainty to the generation mix. The goal is to minimize the amount of additional reserves that will be required to accommodate wind. State-of-the-art wind power forecasts provide a means to mitigate some of the complications associated with adding wind to the system, but the errors in these forecasts can themselves be a source of added complexity. If a forecast over-predicts the wind, it means that the wind component of the planned generation mix will under-produce if system planners rely on the forecast. Likewise, under-predicting the wind can lead to over-generation. Over-predicting the wind can lead to situations where gaps between generation and load must be met by using expensive generation reserves. In these cases, load and generation balancing will usually be accomplished by using gas turbines, unless the control area possesses hydro generation capabilities that are flexible enough to meet demand. For these reasons, wind integration studies often seek to quantify the additional amount of generating reserves that will be needed to accommodate various wind penetration levels [Zavadil 2006, EnerNex 2007].

Under-predicting the wind may result in more serious consequences with regards to system planning and real-time operations. Production levels of large-scale fossil and nuclear fueled generating units are not easily ramped down, and although such is possible for combustion turbines (CTs) and combined cycle turbines (CCTs), utilities purchase gas well ahead of time and typically do not have on-site storage for excess gas that is not needed. Therefore, it becomes financially disadvantageous to shut these units down. If the control area contains hydro generation, overproduction of wind can lead to hydro spillage which is an inefficient use of the resource and could possibly have detrimental effects on the environment. The current solution to under-prediction is often to curtail the wind, which means that turbine blades are pitched so as to quickly reduce power output levels. This course of action leads to wasted potential for wind energy production, and it will become more frequent as wind penetration levels increase unless forecasting is perfected or market structure is changed to allow shorter-term transactions.

## 1.3 How wind power forecasts are made

Several types of wind power forecasts are commonly used in the industry, with selection often based on a cost vs. benefit tradeoff. Most forecasts can be categorized as either probabilistic or deterministic. Probabilistic forecasts provide probabilities of certain wind power outputs, often times including a range of uncertainties. Deterministic forecasts seek to predict an exact output value. Typically, a forecast value for a wind power plant or region is given for multiple forecast horizons, or the time ahead of real-time for which the forecast is made. For example, an hour-ahead forecast value gives the predicted wind power for one hour ahead of real time. However, there is typically a closure window or deadline by which a forecast must be provided due to the energy market structure. The closure window is usually on the scale of 1-4 hours, meaning that the "hour-ahead" forecast was actually created 1-5 hours ahead of real time.

For industry purposes, during longer forecast horizons the simplest approach is to assume that the wind will equal its known average for the region. This approach is often called a climatological forecast method and can be reported as an annual average, seasonal average, or whatever is most applicable. When the forecast horizon becomes shorter, such as hourly or sub-hourly, a persistence approach is the simplest. Persistence forecasting predicts that the future output will be equal to the current output, and therefore must be continuously updated. The accuracy of a persistence forecast drops off rather dramatically as the time horizon increases.

State-of-the-art forecasts are created from computationally intensive meteorological models that seek to solve physics-based equations in high-resolution representations of the atmosphere and topographical terrain features. These are called meso-scale simulations. The wind power forecasts are based on wind speed forecasts, which are generated from the numerical weather prediction (NWP) models usually produced by governmental atmospheric agencies. As inputs to the physical equations, these models use a set of boundary and initial conditions that are typically measured from meteorological stations (possibly meteorological towers installed at the sites themselves) and radiosonde-equipped weather balloons. The inputs are periodically updated and the behavior is propagated in space and time while solutions are generated for wind velocity, temperature, and pressure at grid points of the atmospheric model. These governing physical equations have yet to be solved analytically, and limitations in the ability to numerically solve these models lead to errors in wind speed forecasting.

Once the wind velocity forecast has been created, the results are used to formulate a wind power forecast. Wind speed predictions are scaled to turbine hub heights, and often refined to account for local conditions. Analytically, wind power varies as the cube of wind speed, and is directly proportional to the area swept by the turbine rotor [Manwell et al. 2002]. However, for the purposes of converting wind speed predictions to wind power predictions for a given site, the appropriate wind turbine power curves are used to formulate the wind power values in lieu of the analytical equations. Figure 2 shows a comparison between the theoretical power in the wind and the power that is actually captured by a typical wind turbine. The yellow line demonstrates the analytical relationship between wind speed (abscissa) and wind power (ordinate), and the blue line represents the same relation as given by a typical wind turbine power curve. Wind turbine power curves are created from empirical testing of turbine models and present a non-linear relation between the input of wind speed and the output of wind power. Therefore, the errors in wind speed can be exacerbated by this non-linear relation when the turbine is operating in Region II of the power curve shown in Figure 2, leading to increased errors in wind power predictions[3]. Various efficiency losses and mechanical characteristics of individual turbine models lead to the deviations from theoretical power conversion equations.

[3] The effects of wind speed prediction errors on wind power predictions will depend on the location within the turbine power curve (e.g. flat sections of the power curve such as Region III in Figure 2 during which the turbine is operating at rated power may lead to compressed errors).

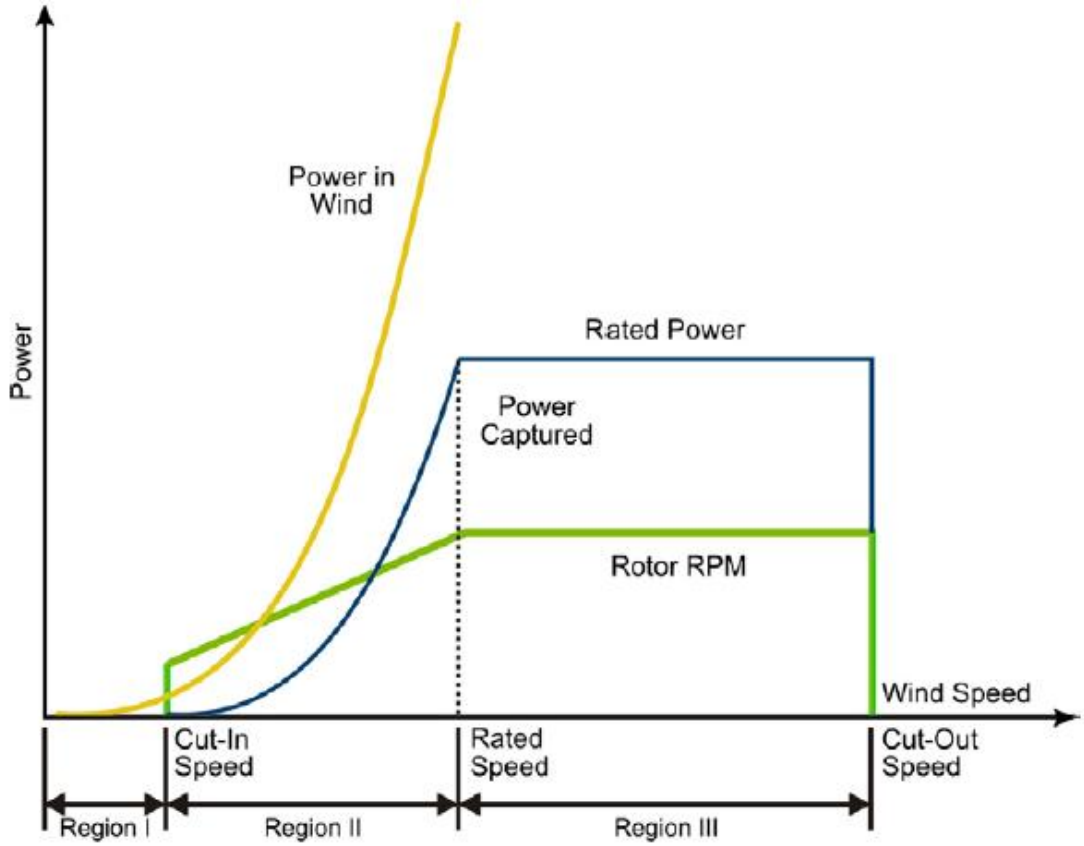


**Figure 2: Comparison between theoretical power in the wind and typical wind turbine power curve. (Source: Lindenberg 2008)**

Commercial forecast providers employ a number of proprietary methods to hone the accuracy of their products. They may start with output data from the NWP models, and tailor it to the specific topographical and wind turbine characteristics of the actual wind power plant site. They may then apply statistical blending tactics to alter the outputs and decrease biases that are inherent to the modeling process. These tactics are known as Model Output Statistics (MOS). The statistical blending is typically applied during shorter forecast horizons which may also be weighted toward persistence prediction. As the horizon increases and persistence accuracy falls off, statistical blending alone may be used until becoming less effective. This can occur between hours 4 and 8, after which the providers may rely heavily on the NWP results [Zack 2005]. A flowchart outlining the generalized process used by commercial forecast providers to create modern wind power forecasts is shown in Figure 3.

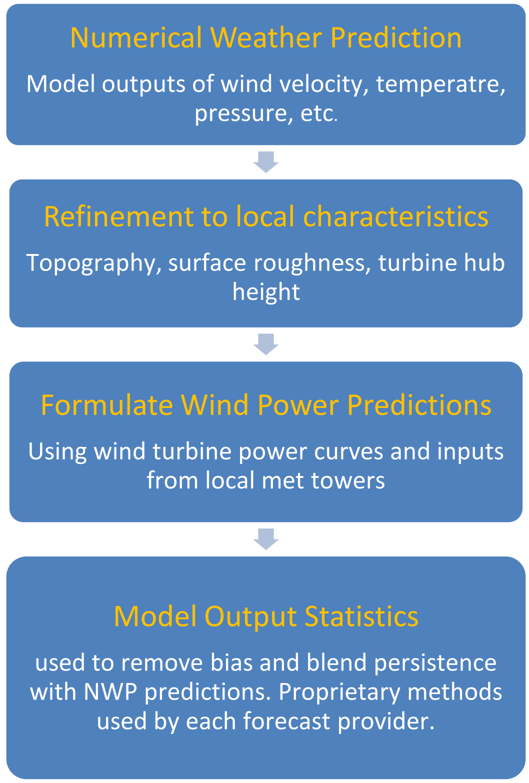


**Figure 3: Flowchart of process used to create commercial wind power forecasts**

Although the methods used by each can be quite different, commercial forecast providers share the goal of minimizing forecast error during all time horizons. The periods defined as regulation, load following, and unit commitment are most significant to electrical system planners and operators. With regards to wind power forecasting, several studies have shown persistence methods to be sufficiently accurate during regulation and the shorter end of load following timeframes [Zack 2005]. However, for the longer end of load following, and certainly unit commitment horizons, the more advanced forecasting methods are needed.

The above discussion pertains to real wind power forecasts that are created from actual meteorological measurements. Complete forecasts can be made well after the time period for which they apply by incorporating historical weather data. This type of real forecast is called a "backcast" or "hindcast". Although generated for times in the past, backcasts are still considered real, state-of-the-art forecasts because they incorporate meteorological data and use the same advanced techniques as real-time forecasts. It is also possible to create a "synthetic" forecast by using a variety of mathematical methods, such as an autoregressive moving average [Milligan 2003, Söder 2004]. Synthetic forecasts require a fraction of the computing power (and hence cost) or real forecasts, and they can incorporate a variety of statistical measures to simulate behavior of real forecasts. Forecast error characterization has broad implications for synthetic forecast analysis.

## 1.4 Traditional methodologies to analyze forecast accuracy

Forecast accuracy has typically been evaluated by using a small number of simple statistical metrics. The forecast error is defined as the difference between forecast and actual power values at any point in time as shown in Equation 1.

$$\boldsymbol{forecast\ error} = \boldsymbol{P_f(t) - P_a(t)} \qquad \textbf{1}$$

This metric is used to gauge instantaneous performance and can be averaged over the entire time series of n-values to obtain the mean bias, given by Equation 2.

$$\boldsymbol{mean\ bias} = \boldsymbol{\frac{1}{n}\sum_{i=1}^{n}(P_f - P_a)} \qquad \textbf{2}$$

The mean bias gives the average value by which the predicted power differs from the actual power over an entire dataset. Therefore, this metric gives insight as to whether the wind tends to be over or under-predicted. The standard deviation of the forecast error, or $\sigma_{f.e.}$ is given by Equation 3. This metric is found by taking the traditional standard deviation of the values calculated by Equation 1, where $x_i$ represents the $i^{th}$ component of the time series of forecast errors[4].

$$\boldsymbol{\sigma_{f.e.}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad \mathbf{3}$$

The standard deviation of the forecast error indicates the variability of the hourly forecast error about its mean. This metric is important to electrical system planners and operators to aid in understanding the potential variability of errors associated with predicting the wind.

The other two metrics commonly used to evaluate forecast performance for a time series with n elements are the mean absolute error (MAE), and the root mean square of the error (RMSE). The mean absolute error is obtained by taking the absolute value of all bias values, and then taking the mean as shown by Equation 4.

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{1}^{n}\left|P_f - P_a\right| \qquad \mathbf{4}$$

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 might occur with a simple mean bias metric. However, this advantaged 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, particularly during wind ramp events when errors can mean the difference between needing to increase or reduce system output.

[4] Note that the forecast error metric will be a time series of values corresponding to each predicted and actual power value, and the mean bias and $\sigma_{f.e}$ will be single values calculated from the time series of forecast errors.

The RMSE is given by:

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(P_f - P_a)^2} \qquad \textbf{5}$$

The RMSE can be a preferred metric for evaluating wind forecast errors because it intrinsically places more weight on larger error terms due to the squared nature of the subtraction term. Larger error terms are often of most interest to system planners but again this comes at the expense of specifying error directionality.

Each of the above metrics has its place in evaluating errors in wind power prediction, although it is generally agreed upon that there is no single metric that will completely describe forecast performance. A combination of metrics is needed and each may be important for unique reasons.

# Chapter 2: Study methodology

The methodology used throughout the course of this thesis can be adapted as a methodical approach to wind power forecast performance evaluation. This chapter outlines the major components to this approach, beginning with data organization and individual wind power time series analysis. An in-depth description of the Ramp Identification Algorithm (RIA) is then presented. Finally, the process used for assessing forecast errors is described for all times and for use during large ramp events, including an introduction to the phase error characterization.

## 2.1 Available data and organization of data

An hourly time series from 2004-2006 of actual wind power production and simultaneous forecast data from an operating wind power plant were used for analysis. The datasets were synchronized in time, and missing values were omitted. A meticulous organization effort was undertaken to account for shifts related to daylight savings, the leap year status of 2004, and any other abnormal findings to ensure that the two time series were properly aligned prior to analysis.

### Power production data

The actual wind power data used for this project were from the Grant County Public Utility District (GCPUD) in Washington State. The dataset consisted of a 2-second resolution time series of wind production from January 1, 2004 through November 30, 2006. The raw data came from GCPUD's share of the total output of the Nine Canyon Wind Project, which at the time consisted of 49, 1.3 MW turbines for a total capacity of 63.7 MW. The power data signal came from a connection bus showing only GCPUD's share of the wind power, and were then scaled up to reflect the total plant output.

The 2-second resolution raw data was converted to both 10-minute and hourly averages using MATLAB®[5]. Each 10-minute or hourly average was created based on the number of valid data points within that period, so periods with some missing data were averaged over fewer points. Some of the actual data had values of -99000. These points corresponded to daylight savings hours and a few other seemingly random times. Such values were flagged for removal during analysis. Additionally, some power values were slightly negative, described by a term known as "parasitic loads". This occurs when the wind is not blowing at all at the wind power plant, but turbine electronics and control systems remain in operation and use some amount of power from the grid. These negative values were set to zero for analysis purposes due to the fact that forecasts will not predict negative values. All other values were left intact resulting in an hourly actual power time series containing 25,560 data points, which is exactly equivalent to the number of hours during the 35-month period used (including the extra 24 hours for the leap year in 2004).

**Forecast data**

The matching hourly wind power forecast data for GCPUD were obtained from a commercial forecast provider for the years 2004-2006[6]. The forecast used for this thesis was deterministic in nature, meaning that a power value was given for each point in time, as opposed to a probability of certain power levels. The state-of-the-art forecast dataset was created using proprietary methods, and will be referred to as the "commercial" or simply forecast dataset. The raw data was provided as a set

[5] The 10-minute averages of actual wind power were not used for most of the analysis presented in this thesis due to the hourly resolution of forecast data, however they were used to test and validate the ramp identification algorithm discussed in Section 2.3.

[6] Although the available data from the commercial forecast provider contained one additional month (December, 2006), that month was omitted for all analysis to maintain seasonal trending.

consisting of one file for each hour of the three years. Within each of these files was a forecast for 144 hours (6 days) ahead of the hour specified by the file name. Data was provided for "low", "mid", and "high" estimates of power production. These values appeared to be related by a near-linear scaling factor, and the mid-levels were used for this entire thesis. These data were used to generate time series for hours 1-96 ahead of each hour of the actual power time series. Having only hourly averaged forecast data limits the scope of the results to the load following and unit commitment timeframes.

In addition to analyzing forecasts based on a specific hour ahead, a "day-ahead" forecast was generated from the commercial forecast data. A day-ahead forecast is not synonymous with a 24-hour ahead forecast. Day-ahead forecasts are commonly used in the industry to allow system planners to make energy market transactions and commit generating units for the following day. There are two common methods for providing a day-ahead forecast. The first is simply to estimate the amount of energy that will be delivered over the entire day and report it as a single block with no resolution on specific wind events or timing. The second method and the one used for this thesis involves creating an hourly forecast for the following day that must be completed at the time of the closure window used for unit commitment concerns. This type of day-ahead forecast is based on the forecast for the following day's wind power production as predicted at 6AM[7] on the day before. Therefore, the wind power production at midnight is predicted 18 hours in advance (created at 6 AM). The forecast horizon for each hour following midnight then spans from 18-41 hours in advance to reach the final hour of the next day (11PM-midnight). The eighteen hours between 6AM and midnight is known as the forecast window closure period and may vary in utility applications. When the day-ahead wind power production forecast is synchronized with a day-ahead load forecast, the LSEs can then trade blocks of energy based on their expected needs. Therefore, the quality of this type of forecast is of significant economic interest to LSEs. Hence, for the

[7] The 6AM closure window was chosen for this project because it is common in the industry, although some markets may allow for a shorter window.

remainder of this report, the day-ahead forecast 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).

For each of the 25,560 hourly data points of actual wind power, there were 97 corresponding forecast data points (hours 1-96 and day-ahead). This resulted in a forecast set with roughly 2.5 million data points. Not all forecast horizons were used for analysis, as most efforts were focused on forecast horizons less than 12 hours.

An example section of overlaid data from the actual power time series and 1-hr, 4-hr, and 12-hr forecast data is shown in Figure 4. The hour-ahead matched the actual data quite well during this time. The 4-hr matched the overall trend, yet with considerable under-prediction. The 12-hr forecast showed little resemblance to the actual during this section.

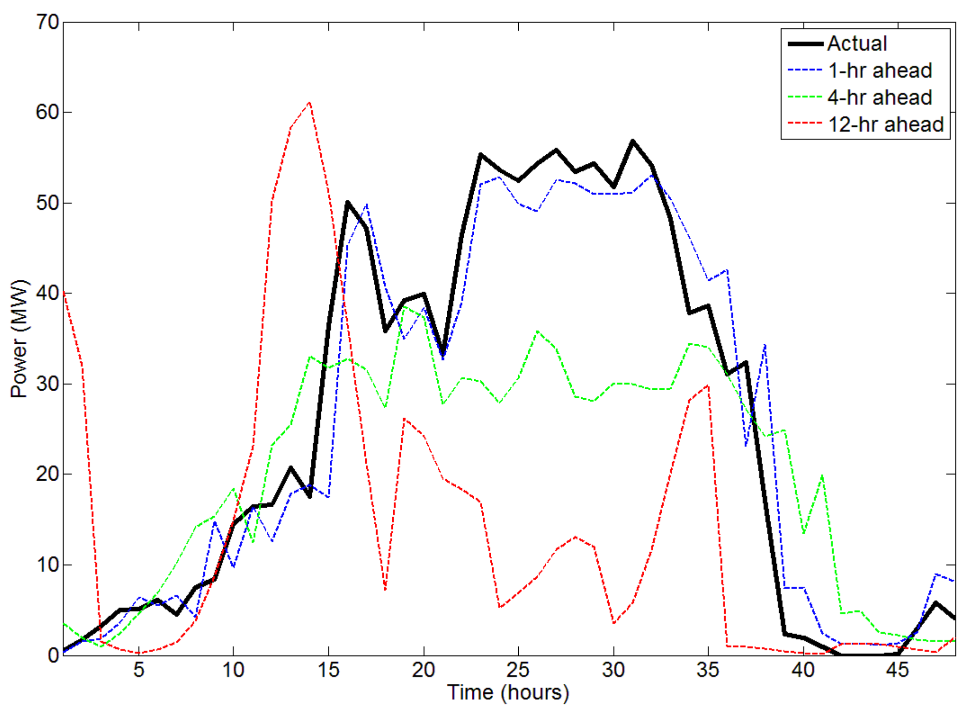


**Figure 4: Example section of time series comparing actual power data with matching 1-hr, 4-hr, and 12-hr forecast data.**

## 2.2 Assessment of actual and forecast datasets

As a preliminary stage for assessing the relation between actual and forecast data, the characteristics of the datasets by themselves were investigated. The structure of the actual wind power time series was used as a baseline to which the structure of each hour of the forecast wind power time series was compared. Differences between the patterns and trends that make up the structure of the actual and forecast time series are effectively the cause of errors, and understanding them aided in the process of identifying significant metrics for evaluating forecast performance. It is not necessary for the structure of a synthetic forecast that matches the real error trends to mimic every attribute of the commercial forecast, but the analytical tools presented below serve as means to compare two time series of data.

Recall that the available time series of data for this project were limited to hourly resolution. Two important features of any wind power time series are the distribution of hourly power production levels and the distribution of hourly step changes in power production. The power production level of a wind power plant can vary from zero to the rated capacity of the plant[8]. By dividing the power output levels into a number of bins, it is possible to obtain a distribution showing the frequency that the wind power plant is operating at certain levels.

An hourly step change in power production is the difference between power values from one hour to the next. This metric is also commonly referred to as the hourly ramp rate or hourly delta value (as it will be referred to during the majority of this report), with units expressed in MW/hr, as shown by Equation 6.

$$\Delta = \boldsymbol{P_{t+1}} - \boldsymbol{P_t} \qquad \mathbf{6}$$

[8] Values slightly less than zero are possible when no wind is present and the electronic control systems for the turbines are using some power from the grid.

The hourly delta values are used to quantify wind ramping events and time series variability. For example, an hourly delta that exceeds a certain percentage of a wind power plant capacity may be considered a "significant ramp". This type of analysis is commonly used in the industry to determine variability impacts of wind [Piwko 2004, EnerNex 2004]. Delta values were tabulated for the actual and forecast datasets, and the average and standard deviation of the delta values were also calculated.

Information given by the distribution of power and delta values is valuable to system planners with regards to understanding the overall variability of the wind generation component. These two techniques serve as examples of useful methods to evaluate and compare two wind power time series.

Power production levels, delta distributions, and error characteristics can be categorized on a diurnal, seasonal, or annual basis if desired. This is especially useful in regions or systems that are known to have strong patterns on these time scales.

### 2.3 Ramp Identification Algorithm

Analysis of trends in hourly delta values may not always be sufficient to capture the difficulties associated with larger ramping events that span multiple hourly timesteps. The delta approach does have the benefits of being simple and well-understood (especially pertaining to data with coarse temporal resolution), and planners and operators are frequently interested in looking for large positive or negative changes in power production over short time periods. Although extreme ramp events are rare, they are of significant interest to system planners with regards to reserve requirements. A single large hourly delta value gives no information regarding previous or future values, and large ramping events spanning multiple timesteps present significant challenges for system operators. Even less information about large ramp events may be obtained from considering step changes with finer resolutin data.

Currently, there is no industry standard for classifying wind power ramp events. Efforts have typically been focused on step change analysis alone. A major component of this thesis project was the development of a meticulous procedure for identifying entire ramp events. This procedure consists of a two-step moving average technique that allows for the definitive beginning and ending of a ramp to be specified. Once the hourly delta values have been calculated, a rolling average of them is taken when identifying a ramp event as opposed to searching only for single delta values of a given magnitude. For the hourly resolution of the data used for this thesis, a two-hour rolling average was appropriate to identify ramp events. Therefore, a time series was genearted by taking a two-hour rolling average of the hourly delta values (e.g. by averaging the current and previous delta values). An average over more data points would be appropriate with finer resolution data. 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. Additionally, the algorithm performs a check to ensure that neighboring data points are consecutive in time in order to avoid “false” ramps that might occur as a result of a missing data point. This procedure has been named the Ramp Identification Algorithm, or RIA.

There are essentially 5 input parameters required for the RIA. The first parameter has been called “mrate”, for moving average rate of change. The mrate input specifies a minimum rate of change (ramp rate or delta value) in actual or forecasted power production that the algorithm will look for. The algorithm will search for all rates of change greater than or equal to the mrate input. Therefore, a larger value of mrate corresponds to a greater change in power over a set amount of time. This is analogous to steeper ramp events. The algorithm searches for both positive and negative ramps (known respectively as up and down ramps) of the same rate, which corresponds to positive and negative mrate values. Up and down ramps were kept separate for most of this project, as each type has unique significance to electrical system planners and operators.

The second parameter of the ramp identification algorithm was called “bdur” for beginning duration. Bdur specifies the time period over which the moving average is computed. For example, a bdur value of two was most commonly used for the hourly data, which resulted in a moving average of the deltas over two hours (two data points). Hence, each value in the moving average set is an average of the delta (step change) from the current time and previous time. The algorithm will only identify a ramp specified by the mrate value if it is sustained for a time period less than or equal to the moving average period given by bdur. For example, if an mrate of 20 MW/hr is combined with a bdur of 2 hours, then any period for which two consecutive, hourly output levels average 20 MW/hr will be defined as a ramp (e.g. [40,0], [20,20], [30,10], [-30,-10] etc.). Single occurrences of delta values greater than or equal to 20 MW/hr (or less than or equal to -20 MW/hr) will also be included.

The combination of a moving average over the desired time period specified by bdur and a selected mrate value are enough to identify the bulk of a ramp event in the same way as the simple step change technique used by many in the wind industry. However, this combination does not allow for a definitive beginning and end to an entire ramp event to be established. As an example, consider a case where the wind slowly beings to ramp up, then becomes more rapid, and again slowly levels off at the end. An approach where the ramp was specified only by a large rate of change would miss the tail ends of the event. The third, fourth, and fifth parameters in the ramp identification algorithm pertain to a second moving average that is used to identify these points in time. The second moving average is computed from the same delta values as the first moving average. A second rate of change, called “mrate2” was used to specify bounds to the rate of change needed to officially declare the beginning or ending of a ramp event. The third parameter, known as “bdur2” specifies the time period over which to compute the second moving average at the beginning of the ramp. The fourth parameter, known as “edur2” is used to specify the time period over which to compute the second moving average at the end of the ramp. For all further analysis, an mrate2 of zero was used. This means that a sign change in deltas (from positive to negative or vice versa) was needed to specify the beginning or end of a ramp event unless the time period specified by bdur2 or edur2 was reached. The RIA input parameters and their purposes are summarized in Table 2.

**Table 2: RIA Input parameters and their purposes.**

| Parameter | Purpose |
|---|---|
| **mrate** | minimum rate of change to identify bulk of ramp event |
| **bdur** | time period over which to compute rolling average |
| **mrate2** | bounds on rate of change for beginning and end of ramp |
| **bdur2** | time period over which to compute second moving average at beginning of ramp |
| **edur2** | time period over which to compute second moving average at end of ramp |

For the majority of the analysis presented in this thesis, the values of bdur, bdur2, and edur2 were held constant at 2,1,1, respectively. Because the datasets were of hourly resolution, it was decided that a two-hour rolling average was appropriate to identify the desired rates of change in power production and that a one-hour period before and after the main portion of the ramp was sufficient to identify beginning and end points for ramps. Load following and optimization planning occurs on the timescale of one to several hours, and the inherent variability of wind makes it impractical to take a rolling average over too many hours.

Because of the constraints on the tail ends of ramp events, the primary parameter varied during this project was the mrate value. By holding all others constant, a changing mrate allowed for ramps of different sizes to be extracted for analysis. When an mrate of 11 MW/hr was used, the RIA would extract ramp events that were equivalent to or exceeded ±22 MW, or approximately ±35% plant capacity over the two-hour rolling average timeframe (given the bdur value of 2). When an mrate of 21 MW/hr was used, the RIA would extract ramp events that were equivalent to or exceeded ±42 MW, or approximately ±66% plant capacity over the two-hour rolling average timeframe. The mrates of 11 MW/hr and 21 MW/hr were used most frequently during this analysis, and resulted in the respective top 900 and 100 largest ramp events to be identified. Essentially, out of the entire three-year dataset, there were only 100 instances during which the actual wind power plant output changed by ±66% plant capacity in a two-hour timeframe. Although seldom, these events would be of extreme interest to system planners and operators with regards to balancing concerns and ancillary service planning.

The series of plots below illustrate the RIA in action. The 10-minute averages of actual power were used to create these plots for visualization purposes. For both Figure 5 and Figure 6, an mrate value of 5 MW/10min was used to search for the ramp and an mrate2 of 0 was used to define the beginning and ending. In Figure 5 the bdur value of 3 indicated that a moving average of deltas over 3 time steps (30 min) was used to identify the ramps, and a bdur2 and edur2 of 1 indicated that any change delta value greater than or less than mrate2 (hence any sign change) would begin or end the ramp. This explains the black middle section in Figure 5 that was not identified as a ramp.

Figure 6 demonstrates the full scope of the RIA capabilities. An mrate of 5 MW/10min was again used, but the bdur parameter was changed to 6, indicating that a rolling average would be computed over six time steps (60 min). In addition, the bdur2 and edur2 parameters were changed from 1 to 3, indicating that a delta sign change within 3 time steps was acceptable, so long as the mrate value was maintained. By doing this, the entire event is identified as a single up ramp, despite the small segment during which the power slightly ramps down during the middle (not identified in Figure 5). Also note that the tail at the very beginning of the ramp is captured this time.

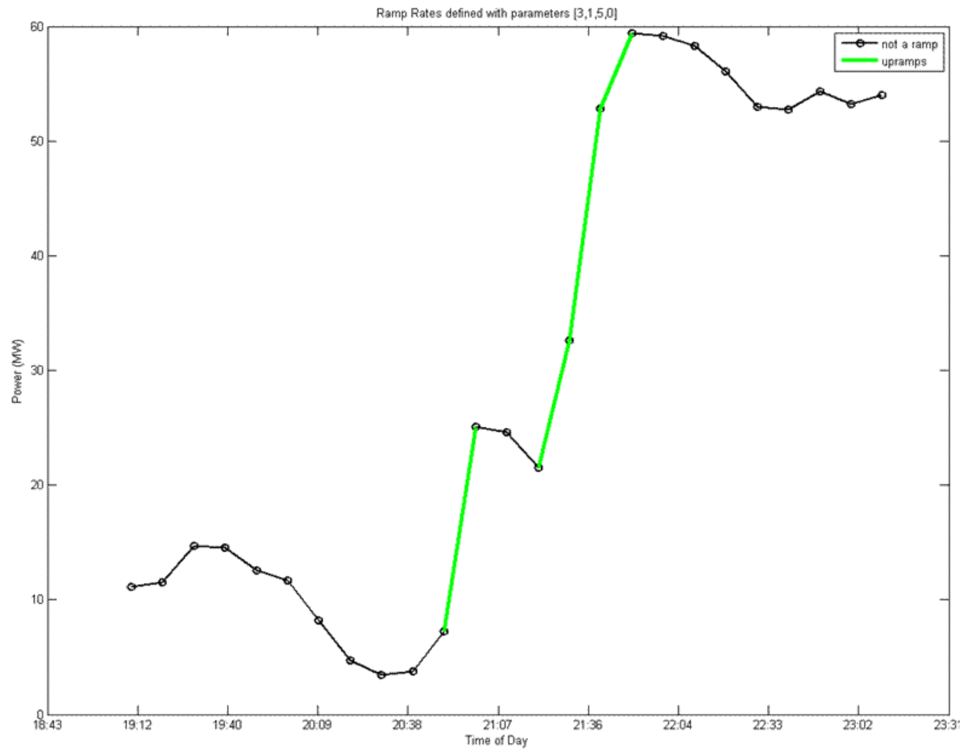


**Figure 5: Demonstration of the RIA using mrate of 5 MW/10min, bdur of 3 (30 min), bdur2 and edur2 of 1 (10 min), and mrate2 of 0.**

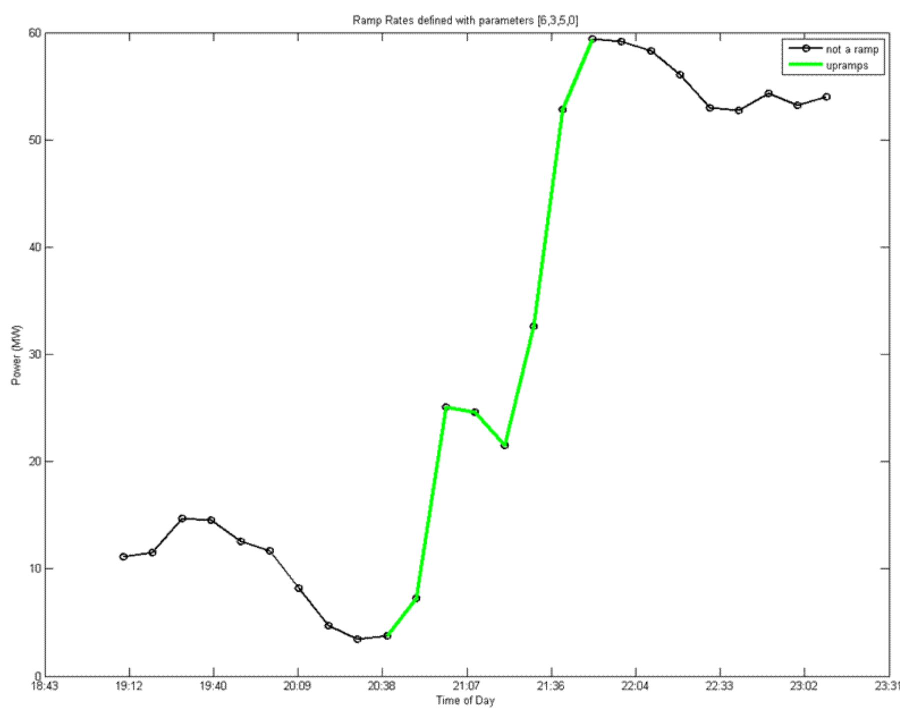


**Figure 6: Demonstration of the RIA using mrate of 5 MW/10min, bdur of 6 (30 min), bdur2 and edur2 of 3 (30 min), and mrate2 of 0.**

A critical point to make is that a major strength of the ramp identification algorithm is an added amount of flexibility that can be used to capture ramp events of interest. This technique can be tailored to datasets with various time resolutions and adjusted to work on ramping timeframes that are relevant to the stakeholder. The input parameters can be adjusted to appropriately reflect the resolution of the data and the desires of the user. Such adjustments allow for ramp events of any size to be identified during time periods that are crucial to regulation, load following, and unit commitment horizons. The algorithm was tested extensively and seemed to perform better with 10-minute data, which is the resolution commonly used for simulations and wind integration studies. This is primarily due to the fact that the tail ends of ramp events occur on small scales, and the overall wind can vary considerably in the course of an hour.

The RIA can be applied equally to the actual or forecast data series. Upon identifying ramp events from both time series, the RIA can also perform a temporal correlation analysis to assess the accuracy in the forecast's ability to predict ramp events. This correlation analysis is completed by identifying the start times of actual ramp events (although any time within the ramp could be used), and searching for forecasted ramps that occur within a user-inputted timing window around the actual ramp event. Figure 7 shows a time-synchronized portion of both the actual and forecast datasets with the dashed sections representing sections identified as ramp events with the particular input parameters used. For this thesis, a ±4-hour timing window was used to search for forecasted ramp events that correlated with actual ramp events of similar size. The 4-hour window was chosen because it is appropriate when addressing load following concerns of system operators. Weather fronts, which are a common cause of large ramp events, are often missed by a couple of hours and additional generating reserves may be needed to accommodate the large forecasting errors that occur as a result. The RIA could also be used to search for correlated ramps of different sizes, if for example the forecast underestimated the actual ramp size by some amount but accurately predicted the timing. The MATLAB® code for the RIA is provided in Appendix A.

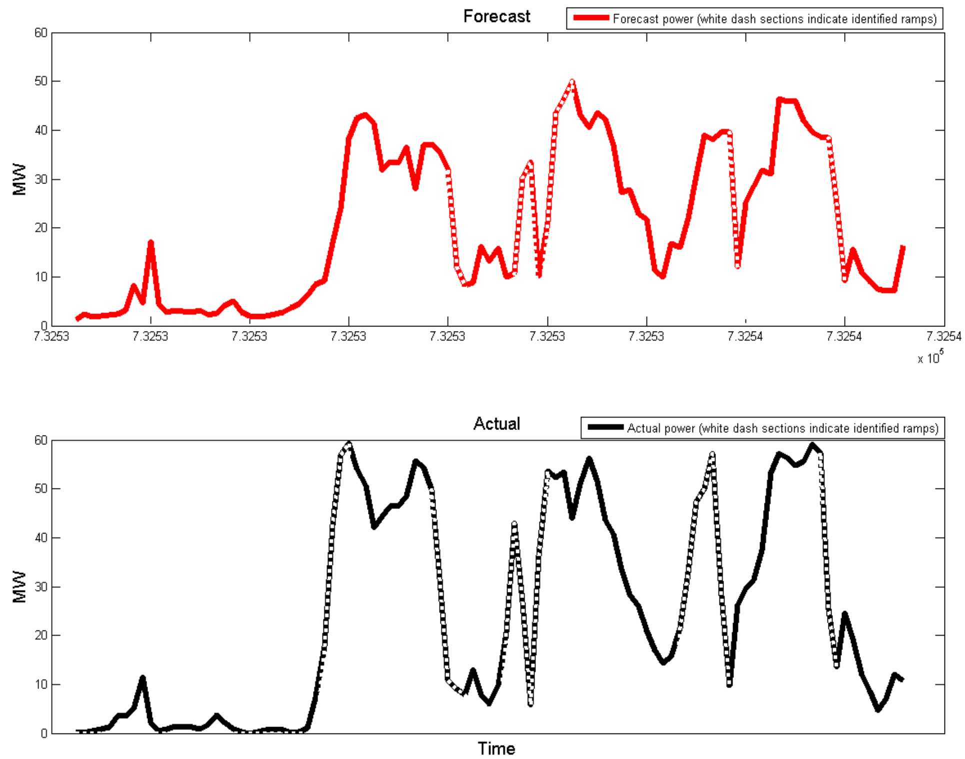


**Figure 7: Example of ramp events identified by the Ramp Identification Algorithm from a matching section of the actual wind power and forecast wind power data. Dashed sections indicate sections identified as ramps for the particular criteria used for this run of the RIA.**

## 2.4 Process for general error assessment

The underlying theme of this thesis project was to develop a methodical process for characterizing errors in wind power forecasting and to select a descriptive set of metrics that will thoroughly evaluate the performance of a forecast. In order to accomplish this, an exhaustive investigation of forecasting errors was needed to identify trends associated with or impacts resulting from various metrics. The metrics discussed below were evaluated at various forecast horizons, in order to demonstrate how the structure of the commercial forecast changes with respect to forecast horizon. The purpose of this section is to present an overview of the tactics used for error assessment, and all results and discussion are presented in Chapter 4 of this report.

### Traditional error analysis

The methodology used in this study for characterizing the forecast errors consisted of using both traditional forecast performance evaluation tactics along with new evaluation criteria. The traditional metrics of mean bias, mean absolute error, and root mean square of the error were computed as described in Section 1.3. The standard deviation of the forecast error was also computed. These metrics were evaluated at many different forecast horizons for the purpose of demonstrating traditional means of characterizing forecast performance.

### Probability density function

An additional step was taken to fit a probability density function (pdf) to the forecast error values. Doing this allows the probability of forecast error values falling within certain ranges to be estimated based on the anticipated power production levels of the actual wind power plant. In the past, wind power forecasting errors were often presumed to follow a normal distribution, with roughly equal chance of over and under-prediction. This trend is generally true for wind velocity forecasting. However it has been shown in the literature that the actual distribution of wind power errors more closely follows the beta probability density function [Bludszuweit 2008, Bofinger 2002]. Therefore, the beta pdf was the only distribution applied to the data for this project. The beta pdf can be described by two shape parameters. The $\alpha$-parameter is based on the mean and variance ($\sigma^2$) of the forecast error. The $\beta$-parameter is based on the $\alpha$-parameter and the mean of the forecast error.

The normalized error values from the available data were plotted against an equation-based beta function with calculated parameters as presented in Bludszuweit (2008), as well as the MATLAB® function called "betafit". The beta pdf as defined in Bludszuweit (2008) is given in Equation 7.

$$\boldsymbol{B(\alpha,\beta) = \int_0^1 P^{\alpha-1} \cdot (1-P)^{\beta-1} dP} \qquad \mathbf{7}$$

$$\alpha = \frac{(1-\mu)\cdot\mu^2}{\sigma^2} - \mu$$

$$\beta = \frac{1-\mu}{\mu} \cdot \alpha$$

The beta distribution is appropriate for wind power applications because it is bound between zero and one. This is consistent with the magnitude (absolute value) of the normalized errors found in wind power forecasting. For example, one would not predict the output of a wind power plant to be less than zero or greater than the capacity of the plant. Therefore, when normalized to the capacity, the absolute value of the forecasting errors falls between zero and one.

### Dependence on power production levels

Forecast error values were further assessed for dependence on the power production level of the wind power plant. Important insight may be gained by understanding how the accuracy of forecasting changes with the level at which the actual wind power plant is producing. The beta pdf can allow the forecast error magnitudes to be estimated based on power production levels, but it does not determine if the forecast errors will be positive or negative. Distributions were created to investigate error directionality and size as a function of power production levels.

### Correlation coefficient

The correlation coefficient, R, between actual and forecast power levels was computed for various forecast horizons. The correlation coefficient is the square root of the coefficient of determination, which provides an idea of how well a predicted value of a time series matches the actual value. The formula for R is given by Equation 8, taken from Ott (2001).

$$\boldsymbol{R_{x,y}} = \frac{\Sigma\left(\boldsymbol{x_i} - \bar{x}\right)\left(\mathbf{y_i} - \bar{y}\right)}{\sqrt{\boldsymbol{S_{xx}S_{yy}}}} \qquad \mathbf{8}$$

$$S_{xx} = \sum_i \left(x_i - \bar{x}\right)^2 \quad S_{yy} = \sum_i \left(y_i - \bar{y}\right)^2$$

The x-values represent forecast data, and the y-values represent actual. It is commonly used in regression analysis to determine the linear dependence of one variable on another, or in this case the dependence of actual wind power output on the predicted value. The coefficient of determination falls between zero and one, indicating a percentage by which the error for prediction of the time series is reduced from simply predicting the mean. A value of R or $R^2$ of 1 would indicate a perfect forecast, or a 100% reduction in error from predicting the mean. As discussed earlier, a climatological forecast predicts that the wind power output will always equal its mean value for the region. Therefore by applying the coefficient of determination, it is possible to determine how much the commercial forecast improves upon the climatological forecast.

**Hourly ramp rate analysis**

Forecast error assessment during wind ramping events was conducted using two methodologies. The first method included investigation of errors for a dependence exclusively on the hourly ramp rate, or delta value as given by Equation 6. The mean and standard deviations of the hourly delta values for the actual data and various forecast horizons were calculated. Due to the hourly resolution of the data, this type of investigation was warranted with regards to load following concerns, however it does not constitute an exhaustive procedure for characterizing errors during entire ramp events. Therefore, the second method included an intensive error analysis during larger ramp events selected by the RIA, which is discussed further in Section 2.5 of this report.

## 2.5 Error analysis during large wind ramping events selected by the RIA

The next component of statistical characterization involved comparison between actual and forecasted ramp events as selected by the RIA. The wind industry has expressed particular concern for forecasting error during large ramp events. At smaller penetration levels, these errors can be easily accommodated by available reserves. However, the consequences for larger penetration levels become more significant. It is generally accepted that geographic diversity of several wind power plants will decrease the variability of wind that is absorbed into the electrical grid system [Wan 2004, Wan 2005, Wan 2009]. But with the scale of individual projects expected to substantially increase, these individual wind sites may still be subject to large ramping events with sizes that have yet to be encountered.

An intensive examination of error behavior during and near large ramping events was conducted. This component of the study was multi-faceted, and involved the application of traditional forecast performance evaluation techniques during ramps events, as well as a phase analysis that will be presented later.

The RIA was used to extract the largest ramping events from the actual wind power dataset. A multitude of mrate values were used to select ramps of various sizes, and these ramps were tracked by their place in time. Most of the analysis was conducted using the top 100 and top 900 largest ramping events of the entire three-year time series of data. The top 900 ramp events occurred when the wind plant production level changed by approximately ±35% of total capacity in a two-hour timeframe, and the top 100 ramp events occurred when the production level changed by approximately ±66% of total capacity in a two-hour timeframe[9]. These values were chosen for their applicability to system operations.

After these ramps were identified, the error characteristics were investigated during and near the times defined as actual ramp events. This allowed the forecast performance evaluation to be separated into times defined as ramps and non-ramps. Traditional metrics such as the MAE, mean bias, and $\sigma_{f.e.}$ were used to compare forecast accuracy during the times defined as ramps and those defined as non-ramps. Additionally, forecast accuracy was assessed as a function of ramp event size. Up ramps were kept separate from down ramps for all analysis, as each has its unique implications for grid integration.

## 2.6 Phase error process

It is well-known that wind power forecasts often under or over-predict ramp events in magnitude, but they can also miss the time of the beginning, ending, and duration of the events. This tends to occur as storm fronts pass through a region and during times of pressure changes in the atmosphere. Wind ramp events occur in all sizes and durations, and may occur many times within the course of a single day. An important realization is that even when a ramp event is properly forecasted in magnitude, a fault in the timing of the ramp arrival will have drastic effects on the overall magnitude error characteristics such as MAE and mean bias. The difference between phase error and magnitude error is demonstrated in the Figure 8 below. The general shape of the forecasted ramp event closely matches that of the actual, which suggests that magnitude and ramp duration were accurately predicted. However, a temporal error in the arrival time of the ramp leads to a large bias that will affect the overall MAE and RMSE (although in this case the mean bias may be nearly zero). Clearly, the traditional forecast error metrics do not fully describe this event. Any insight on trends in phase errors will be valuable for system planning

[9] As discussed in Section 2.3

and may also provide a key validation point for synthetic forecasts used in wind integration studies.  The only previous work encountered that involved phase error analysis for wind power forecasting was in Lange (2005), where phase errors were briefly mentioned "in a statistical sense in terms of the cross-correlation not as a well-defined phase shift between two time series".

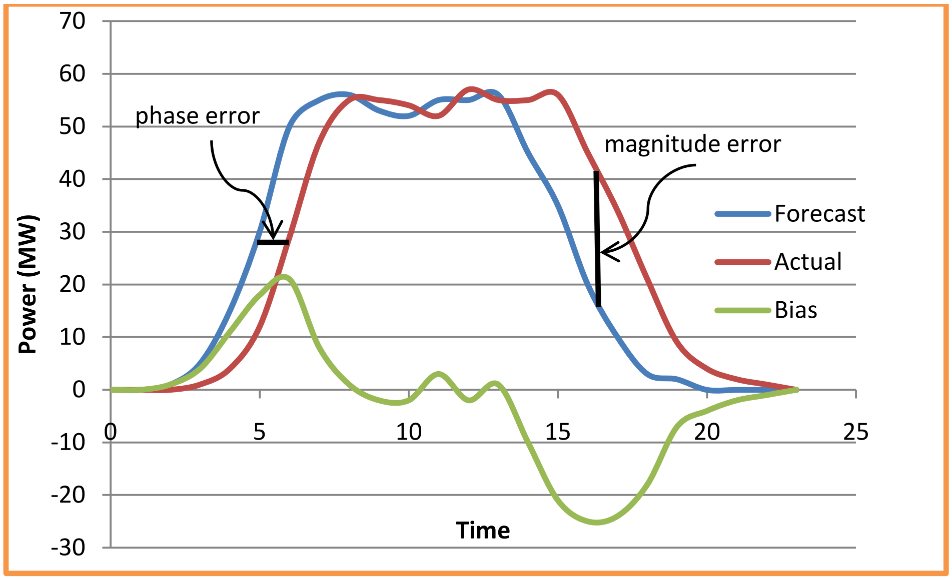


**Figure 8: Illustration of phase error in forecasting for a ramp event. The black lines indicate separation between actual and forecast data, the green bias line is the obtained from subtracting actual from forecast at each point in time (representing magnitude error).**

To begin the process of temporal error analysis, a time-synchronized comparison was undertaken to identify predicted ramp events that corresponded to actual ramp events. Therefore, forecasted ramps of similar size were identified within a four-hour window on either side of actual ramps.  The RIA will identify the time at which the actual ramp event begins, and then search for forecasted ramp events of user-chosen magnitude that also begin within the timing window.  The window of four-hours was chosen for its significance to utility system planning, although any period of time can be input into the RIA program.  Also, the RIA searched for forecasted ramps of similar magnitude to the actual (e.g. using same input parameters as described in Table 2).

The following contingency table summarizes the possible outcomes of the RIA correlation analysis between forecast and actual ramp events. The algorithm identifies up and down ramp events in the commercial forecast time series that correctly or incorrectly match the direction of actual ramp events. Forecasted ramps occurring in the opposite direction as the actual can lead to extremely large forecasting errors, and they may be of more interest to planners than the instances when the forecast merely under-predicts or does not predict a ramp. Consequently, a synthetic forecast must contain these cases of extreme errors if such are found in commercial forecasts.

**Table 3: Contingency table for correlation analysis between forecast and actual ramp events.**

| | Actual Up | Actual Down |
|---|---|---|
| Forecast Up | x | x |
| Forecast Down | x | x |

A few points of discussion arise from this temporal correlation analysis. First of all, the ramp events are identified by the user input for the mrate value in the RIA program[10]. The RIA will identify any ramp that contains an average mrate value greater than or equal to the specified limit and occurring over the time period specified by bdur. For the correlation analysis to be performed, the identification algorithm is run first on the actual data and subsequently on the forecast data. This allows the user to choose different mrate values if desired. For example, it is possible for a forecasted ramp event to miss *both* the magnitude and timing of the actual event. Therefore, the user can search for forecasted ramps of slightly less magnitude than the actual within the same timing window as opposed to only searching for ramps of the same magnitude. A ramp of slightly smaller magnitude and reasonable timing does not necessarily mean the forecast is blown, and error

[10] The mrate value specifies the moving average of the rate of change in power production that will be identified in RIA, as discussed in Section 2.3 of this report.

assessment may be of value in these cases. This is especially true if the overall mean bias of the forecast is negative (as will be shown in the results of this project). Trial runs were conducted for this thesis that included correlation analysis between forecasted ramp events that were at least 80% of the magnitude of the actual events and some results are presented in Appendix B. However, all subsequent analysis presented for phase error analysis of ramp events contains only those with the same magnitudes. The levels can be adjusted as desired during further use of these procedures.

## Chapter 3: Document for Journal Publication

This section contains an abbreviated version of this thesis project intended for professional publication. The example provide here will be published in the Proceedings of the American Society of Mechanical Engineers 2010 4$^{th}$ International Conference on Energy Sustainability.

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

**ES2010-90381**

# 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 most likely 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 considered by the authors to be 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 a rudimentary (and perhaps subjective) order of significance. These metrics included: mean absolute error, root mean square error, average magnitude of step changes, 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 greater consequences 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].

Commercial forecasts are a synergistic blend of statistical and physics-based approaches. 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 can be exacerbated by this non-linear relation, leading to increased[1] errors in wind power predictions as discussed in [7].

As load-serving entities (LSEs) 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 Renewable Portfolio Standards (RPS) of 20% renewable generation. 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 develop standardized criteria 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[2]. 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.

$$forecast\ error = P_f(t) - P_a(t) \qquad (1)$$

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.

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

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

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.

Although each of the above metrics has its place in evaluating errors in wind power prediction, none offers a complete description of forecast performance.

[1] The effects of wind speed prediction errors on wind power predictions will depend on the location within the turbine power curve (e.g. flat sections of the power curve may lead to compressed errors).

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

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. Forecast data for 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[3]. 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 using both traditional forecast performance evaluation tactics along with new evaluation criteria. Trends in each individual dataset were examined first, and subsequent characterization of forecasting error was focused on magnitude and timing faults. 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 utility and wind industries have 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 ramp events. Once the ramps were identified, a multi-faceted analysis of the correlation between actual ramp events and forecasted ramp events was conducted. It should be noted that the commercial forecast used in this project was not optimized to predict ramp events, but the identification and error analysis of such are still valid.

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

[3] It should be noted that all professional forecasts are different, and may be optimized for specific purposes.

event in the opposite direction. The algorithm can search for ramps of any user-defined magnitude that occur over a selected period of time (1-hr or greater for the hourly data used in this project, although shorter ramps could be identified with finer resolution data). 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. 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.

In order 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[4], 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. 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:

- Distributions were investigated in power production levels and hourly delta values for both the actual 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 patterns during large ramp events selected by the 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.

[4] 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.

## 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 $\sigma_\Delta$ (standard deviation of delta values) 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 the structure of 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 still predicted.

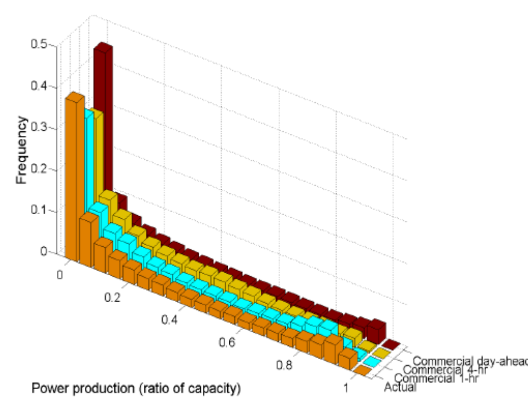


**Figure 1: Distribution of power production from actual wind plant and commercial forecasts for 1-hr, 4-hr, and day-ahead horizons.**

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 symmetrical 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 ramping down, and the average down ramp is slightly larger than the average up ramp. 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].

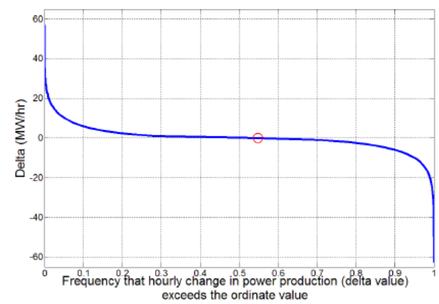


**Figure 2: Duration curve of hourly step changes (deltas) in power production from actual wind power dataset. Circle indicates frequency that step changes exceed 0 MW/hr. Plant capacity is in excess of 60 MW.**

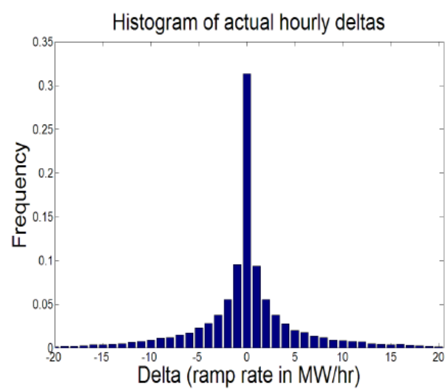


**Figure 3: Histogram of delta values for actual power dataset indicating distribution of hour-to-hour changes in plant output. Plant capacity is in excess of 60 MW.**

*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.

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

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 could be the result of a bias in the NWP models that was not completely removed by the forecast provider. 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 it should be noted that if the "mean bias" is zero, then the RMSE and the standard deviation of the forecast error are equivalent.

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 a proprietary combination of techniques (including statistical) to decrease forecasting errors during the first several hours ahead of the operational hour, after which a transition weighted more heavily on NWP predictions is made. The important things to gather from these results are the variability and dependence of each metric on the forecast horizon.

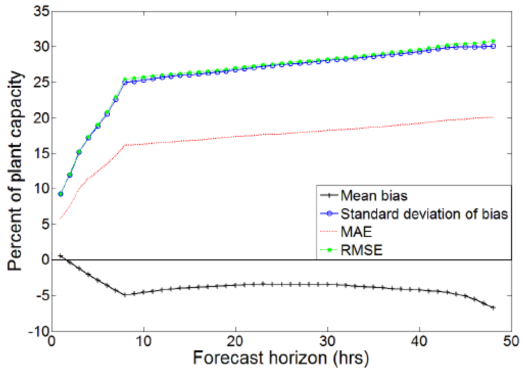


**Figure 4: MAE, RMSE, Mean bias, and standard deviation of forecast error vs. forecast horizon.**

*Probability Density Function of Forecast Error*

Wind forecasting errors are often presumed to follow a normal distribution, with 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 (dotted line in Figure 5) as presented in [14], as well as the MATLAB® function called "betafit" (dashed line in Figure 5). The beta distribution is appropriate for wind power applications because it is typically bound between 0 and 1, consistent with the magnitude (absolute value) of the 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].

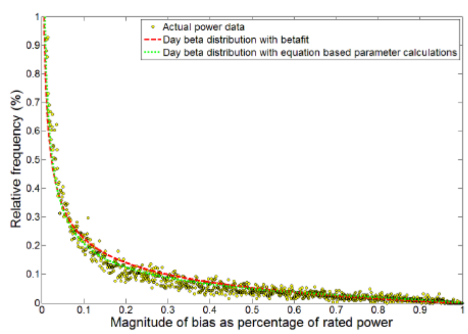


**Figure 5: Distribution of forecast errors from available data along with two forms of a beta distribution.**

*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 inferred 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 more gradually 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.

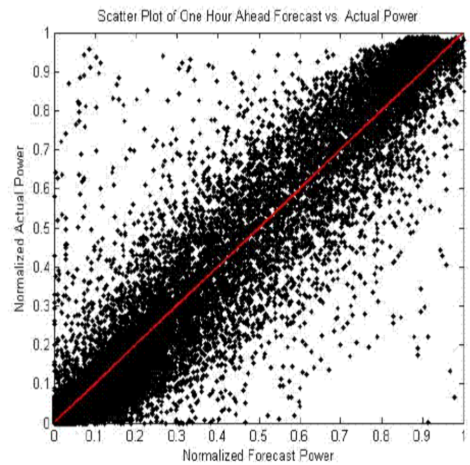


**Figure 6: Scatter plot of power values for actual and forecast data for 1-hr forecast horizon.**

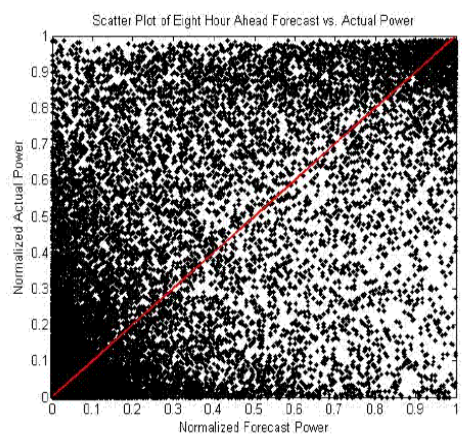


**Figure 7: Scatter plot of power values for actual and forecast data for the 8-hr forecast horizon.**

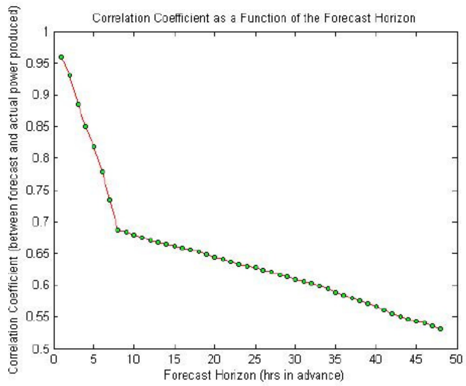


**Figure 8: Correlation coefficient between actual and forecast power production vs. forecast horizon.**

*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 the RIA. It should again be noted that the commercial forecast used was not optimized for ramp prediction.

*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. 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.

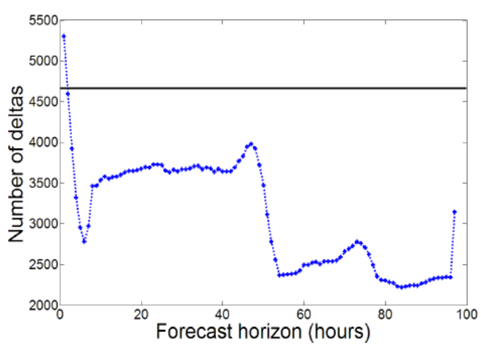


**Figure 9: Number of deltas (hourly step changes in power production) > 10% capacity at various forecast horizons. The horizontal line represents the number in the actual dataset, and the other line represents forecast data.**

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.

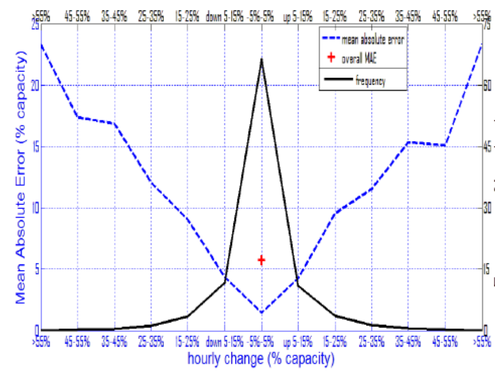


**Figure 10: Frequency of occurrence and MAE during various sizes of step changes in actual hourly power production. Both up and down steps are included, as well as the overall MAE. 1-hr forecast horizon data are shown.**

*Treatment of Large Wind Ramping Events*

Results for the multi-hourly ramping events chosen by the 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. Figure 11 shows the mean delta value during the top 100 largest ramp events as chosen by the RIA from the actual dataset. It is important to notice that predicted values were close to the actual delta values for the 1-hr forecast horizon, but were dissimilar for hours 3-8. In fact, the average forecasted hourly delta value for both up and down ramps became closer to zero during the first 4-6 hours of lead time before reversing direction and eventually leveling off at a forecast horizon of 8 hours (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 $\sigma_\Delta$, during the same 100 largest ramp events. The forecasted $\sigma_\Delta$ values again were smaller for horizons 3-8, suggesting less variability in forecast power values.

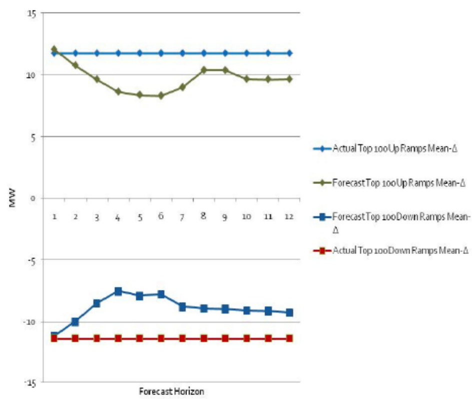


**Figure 11: Mean delta values during top 100 largest ramp events for actual dataset.**

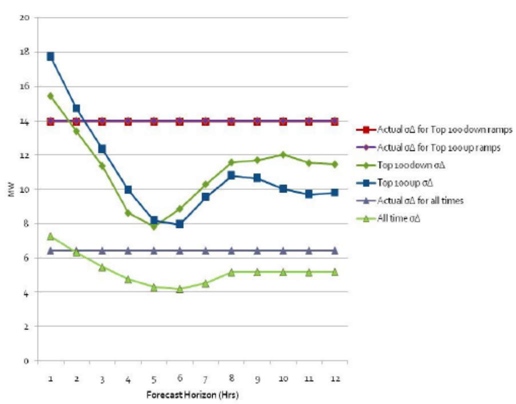


**Figure 12: Standard deviation of delta values during top 100 largest ramp events for actual dataset.**

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.

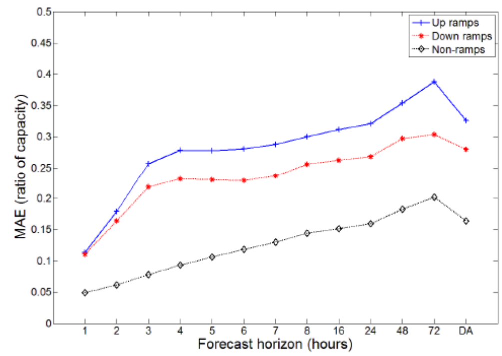


**Figure 13: MAE during ramps and non-ramps for various forecast horizons (top 900 ramps)**

*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 patterns 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[5], 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

[5] The RIA is also capable of finding forecasted ramp events that have no associated actual ramp.

interpretation as well as providing a strong argument for reducing forecast closure windows.

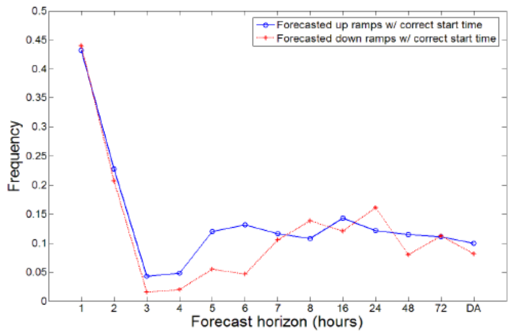


**Figure 14: Frequency of correctly forecasted up and down ramps versus forecast horizon. Forecasted ramps were of similar size to actual ramps.**

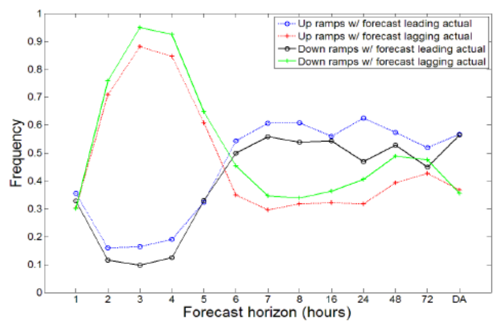


**Figure 15: Frequency of up and down ramp starts leading or lagging actual ramp starts. Only forecasted ramps correlated within ± 4 hour window of actual ramps are included.**

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.

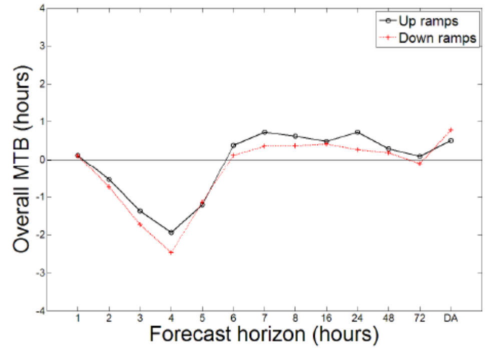


**Figure 16: Mean temporal bias for top 900 largest actual ramp events that also have a corresponding forecasted ramp of similar magnitude within a ± 4 hr window.**

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 positive MTB for hours six and beyond. This does not actually mean that ramp forecasting becomes better at longer horizons. Since Figure 16 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 (recall from Figure 9 that there are far fewer forecasted ramp events during these time horizons).

## 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 reproducible 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 more reliance on 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 rudimentary order of most to least significant (based on the opinion of the authors). 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 the system control rooms. 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 $\Delta$ |
| $\sigma_{\Delta}$ |
| 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 errors.

Although the results presented here apply only to one particular power and forecast couple, the repeatable process offered during this report for ramp identification, statistical characterization, and important parameter analysis 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 evaluation of a 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.

# Chapter 4: Results and Discussion

Section 1.3 and all of Chapter 2 of this report outlined the process followed for forecast error characterization during this thesis project. In this chapter, the results are presented and discussion is provided in the same order as the techniques were introduced earlier.

## 4.1 Characteristics of individual datasets

The distribution of actual power production levels throughout the entire three-year span of data is shown in Figure 9 with bin sizes of 5% plant capacity. The bimodal distribution of power production levels is characteristic of many wind power plants. Production levels were fairly low for the majority of the time as indicated by the larger mode on the left side, and the second mode on the far right indicates that a significant amount of time was spent near full capacity. Intermediate levels were less common.

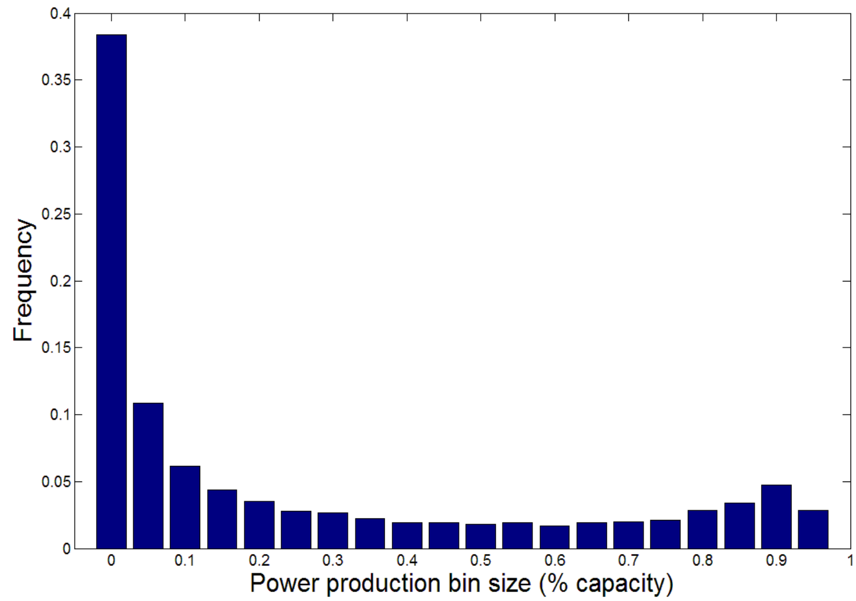


**Figure 9: Distribution of wind power production levels for 2004-2006 GCPUD data with bin sizes of 5% capacity.**

The distribution of hourly step changes in power production from the three years of actual data is shown in Figure 10. The distribution is fairly symmetrical, and offers valuable information about the variability of the wind power plant. The hourly step changes in power production were generally small, with about one third of them falling in the smallest delta bin spanning from -0.5 MW to 0.5 MW. Additionally, there were very few large step changes, therefore the abscissa in Figure 10 is limited to ± 20 MW, or approximately one third of the 63.7 MW total plant capacity.

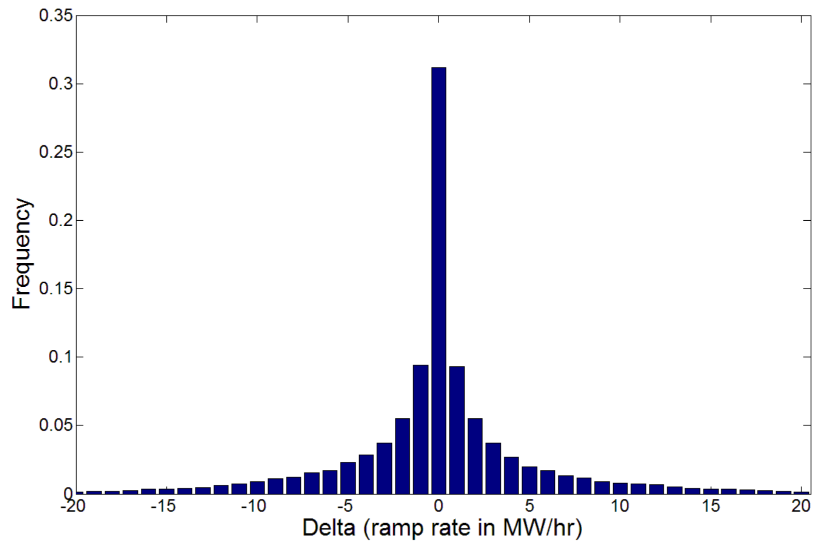


**Figure 10: Histogram of delta values for actual power production for 2004-2006 GCPUD data, indicating distribution of hour-to-hour changes in plant output. Plant capacity was in excess of 60 MW.**

The accompanying duration curve shown in Figure 11 offers another perspective of the hourly step change distribution. The red circle denotes the location of step changes equaling zero, which means that the hourly delta values from the actual wind plant were greater than or equal to zero approximately 54% of the time.

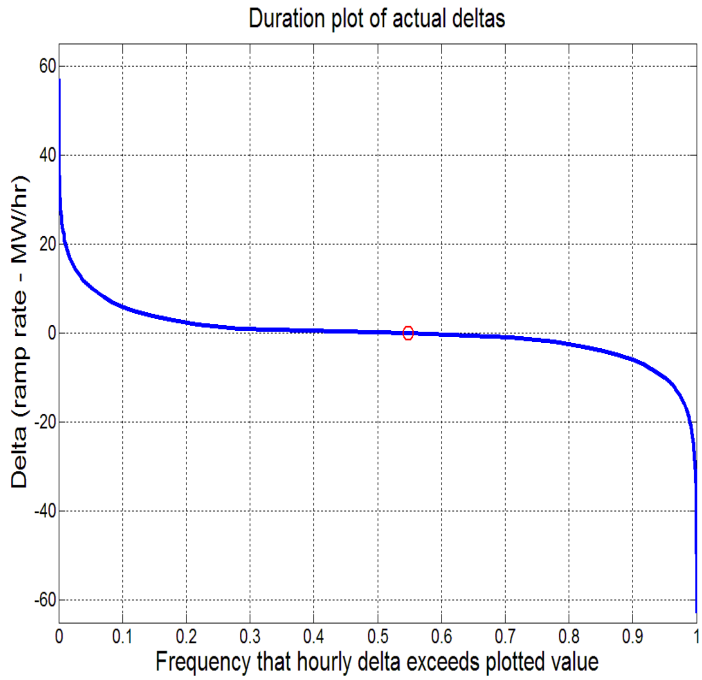


**Figure 11: Duration curve of hourly step changes (deltas) in wind power production for 2004-2006 GCPUD data. The red circle indicates the frequency that step changes exceed 0 (y-intercept).**

The same metrics are presented below for the commercial forecast time series. Several forecast time horizons are offered to demonstrate forecast behavior over time. Figure 12 demonstrates the effects of forecast horizon on the distribution of predicted power levels. As might be expected, the 1-hour forecast did a reasonable job matching the bimodal shape and magnitudes of the actual power data. The second mode is lost during the 4-hour horizon, suggesting the possibility of a more conservative statistical blending to the forecast creation during this horizon. The second mode returns in the day-ahead horizon, yet the frequency of power predicted to occur in the smallest bin increases substantially.

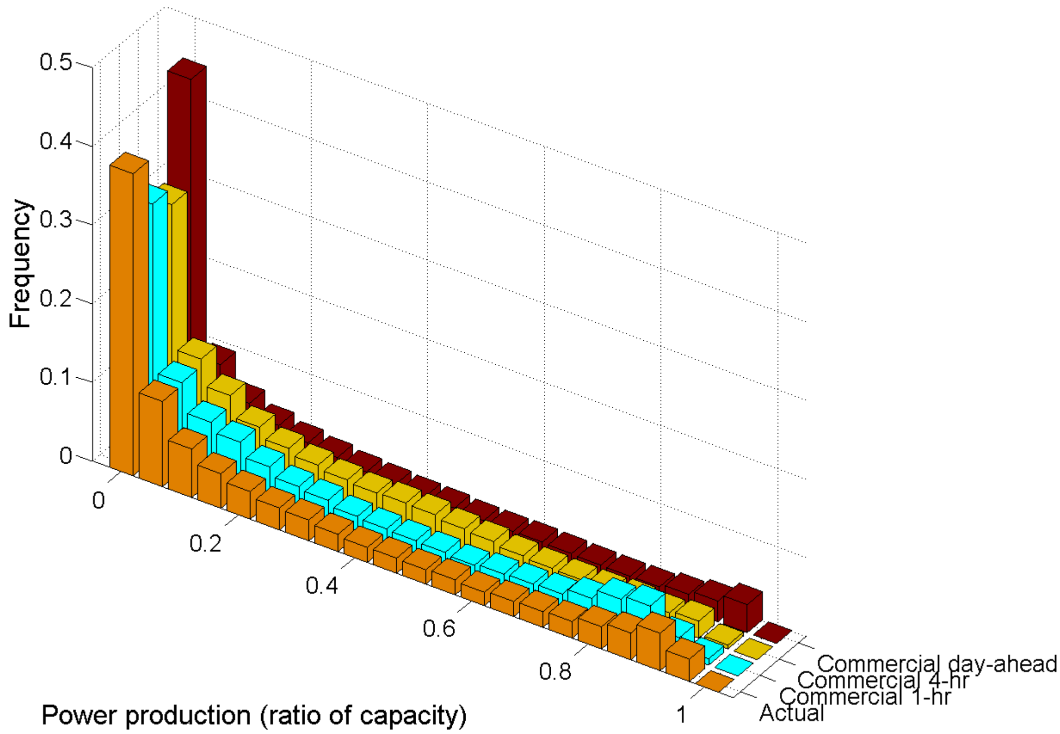


**Figure 12: Distribution of power production levels from actual wind plant and commercial forecasts for 1-hr, 4-hr. and day-ahead horizons. Bin sizes are 5% plant capacity.**

Hourly step changes in forecasted power production are presented in Figure 13 for the same horizons. Again, the 1-hour horizon was reasonably similar to the actual distribution in Figure 10. The subsequent horizons remain fairly symmetrical, but an increased frequency of near-zero step changes can be seen for the 8-hour and day-ahead horizons, indicating less predicted variability at these times.

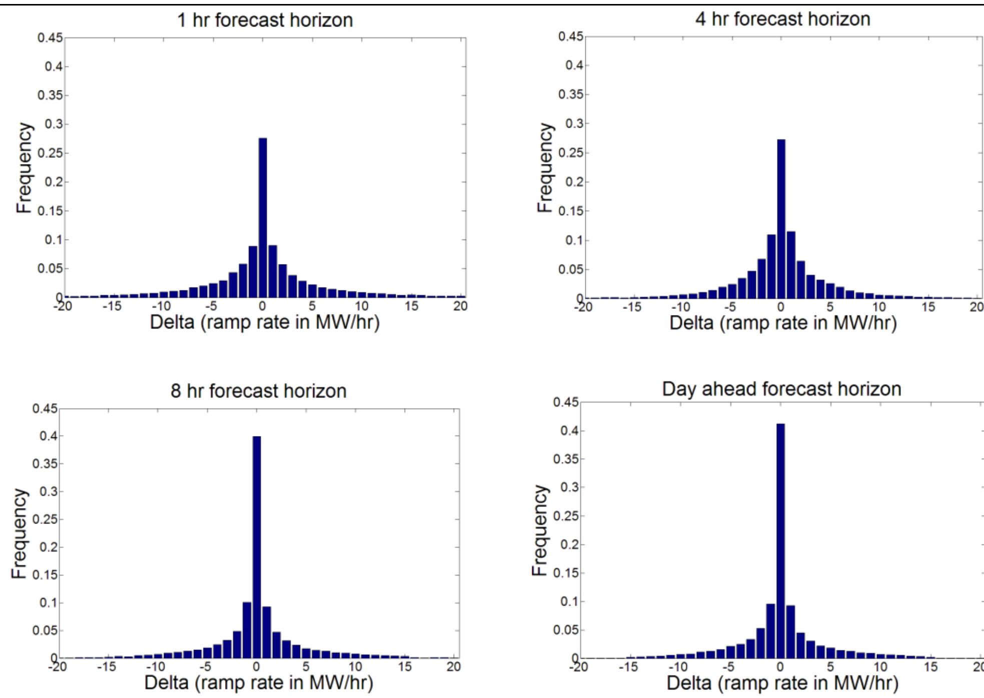


**Figure 13: Distribution of hourly step changes in commercially forecasted power production for various forecast horizons. Total plant capacity was in excess of 60 MW.**

Further discussion of these figures as they relate to forecast error characterization will be presented later in this report.

Table 4 shows the number of hourly delta values of certain sizes. Counts from the actual data are compared with two horizons of forecast data. The number of deltas (as percentage of nameplate capacity) from the hour-ahead matched the number of actual deltas much more closely than did the 4-hr horizon. In fact, the lower delta-count of the 4-hr horizon, particularly of large sizes, indicate that the forecast had much less overall variability during the later horizon. These types of delta tabulation provide a way to quantify the variability in both the actual and forecast time series. Results from additional forecast horizons are shown in Appendix C.

**Table 4: Tabulation of hourly delta values as percentage of nameplate capacity. Actual data shown, as well as hour-ahead and 4-hr ahead forecast values.**

| 1-hr step change (>= % of nameplate) | GC Actual | | | Forecast (HA) | | | Forecast (4-hr) | |
|---|---|---|---|---|---|---|---|---|
| | up | down | | up | down | | up | down |
| 90 | 0 | 0 | | 0 | 0 | | 0 | 0 |
| 80 | 2 | 0 | | 4 | 0 | | 0 | 0 |
| 70 | 9 | 3 | | 11 | 8 | | 0 | 0 |
| 60 | 18 | 10 | | 35 | 27 | | 0 | 0 |
| 50 | 41 | 45 | | 85 | 70 | | 2 | 0 |
| 40 | 110 | 112 | | 197 | 173 | | 22 | 13 |
| 30 | 301 | 289 | | 470 | 410 | | 101 | 72 |
| 20 | 846 | 782 | | 1040 | 982 | | 392 | 363 |
| 10 | 2315 | 2364 | | 2633 | 2670 | | 1673 | 1648 |
| 0 | 1982 | | | 1643 | | | 168 | |

The mean delta value versus forecast horizon is shown in Figure 14. Although the forecast values fluctuate a bit, they remain on the same order of magnitude as the actual mean delta and near zero. More information is given by Figure 15, which shows the standard deviation of delta values, or σΔ versus forecast horizon. This figures shows a decreasing σΔ, and hence decreasing forecast variability for horizons 1-6, with slight relaxation for hours 7 and 8 where the metric stabilizes.

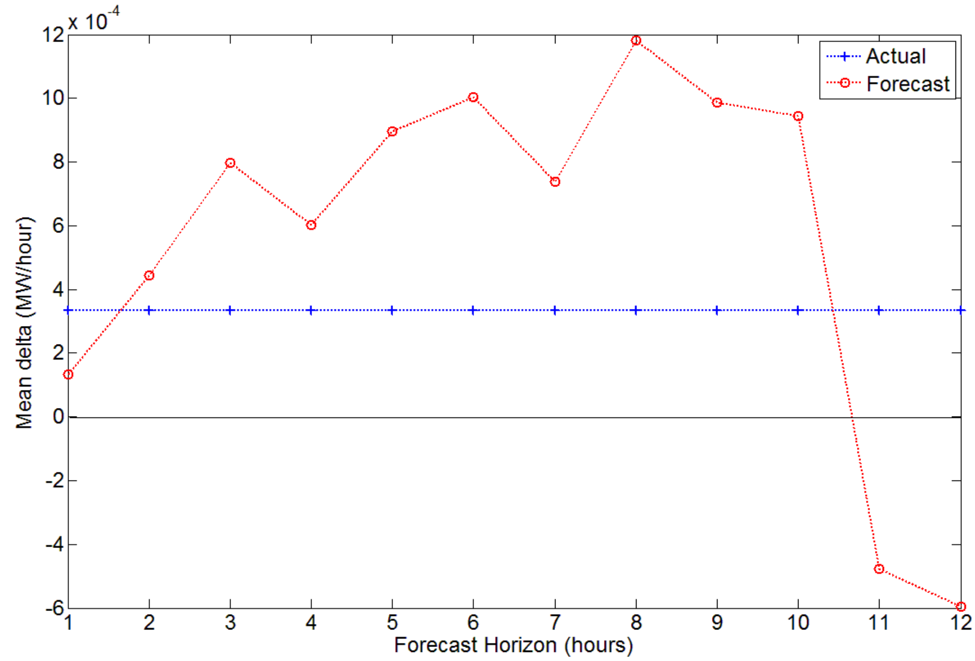


**Figure 14: Mean hourly delta vs. forecast horizon.**

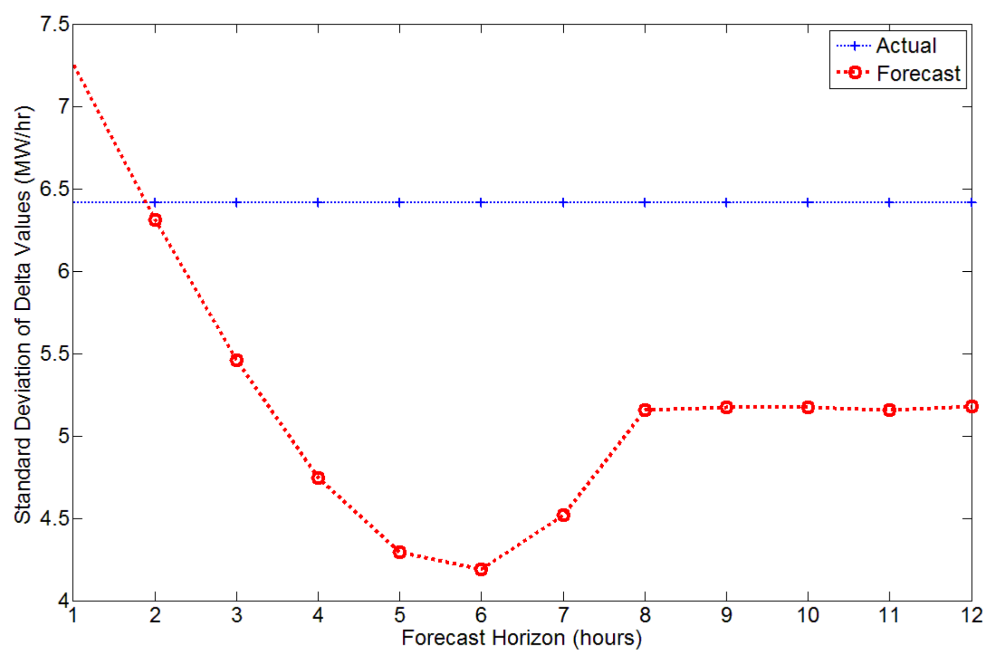


**Figure 15: Standard deviation of hourly delta values (σΔ) vs. forecast horizon.**

## 4.2 Standard characteristics of errors

Results for the traditional metrics for evaluating forecast performance are shown in Figure 16. This includes the mean bias, standard deviation of the bias, MAE, and RMSE for various forecast horizons. Equations for these metrics were provided in Section 1.4 of this report. Each point plotted in Figure 16 represents an average of 3 years of hourly and forecasted data from the specified horizon (roughly 26,000 data points). There was an overall negative mean bias between the actual and commercial data for all time horizons except the hour ahead, as shown by the black line with '+' markers. The mean bias became increasingly negative in a near-linear fashion from hours 1-8, and subsequently showed a slight lessening. Occasionally, bias levels are intentionally added by forecast providers, but most of the time are the result of original bias from the NWP models that was not completely removed during the blending process used by the forecast providers. The standard deviation of the bias, or $\sigma_{f.e.}$, grew in a linear fashion during the same horizons of 1-8, and then continued to grow in a less-extreme fashion after. The interplay between the mean and standard deviation of the forecasting error becomes apparent at this point. For example, at the 1-hour forecast horizon, the mean bias was found to be about 1% of plant capacity, yet the standard deviation was about 9%. This result suggests that although the positive and negative forecast error values nearly cancelled each other out for the 1-hour horizon, there was still a relatively considerable amount of variability in the forecast errors as given by the standard deviation value.

Both the MAE and RMSE metrics shown respectively by the red and green lines in Figure 16 displayed steep increases for horizons 1-8, and continued to increase at less-extreme rates afterward. These evaluation metrics demonstrate that forecast performance becomes less-accurate as the forecast horizon increases.

A point of clarification should be made regarding the similarities of the RMSE and standard deviation of forecast error lines in Figure 16. These mathematical similarities are explained in Madsen (2004), and it should be noted that the two metrics would be identical if the mean bias was zero. This can also be seen by comparing Equations 3 and 5 of this report.

The distinct "elbow" feature seen in all four metrics of Figure 16 is not a surprise. It is well-known that state-of-the-art forecast providers use a proprietary combination of techniques (including statistical) to decrease forecasting errors during the first several hours ahead of the operational hour, after which a transition weighted more heavily on NWP predictions is made. Many of these techniques (which include model output statistics) are centered around removing original bias that is inherent in NWP models. The important points to gather from these results are the variability and dependence of each metric on the forecast horizon.

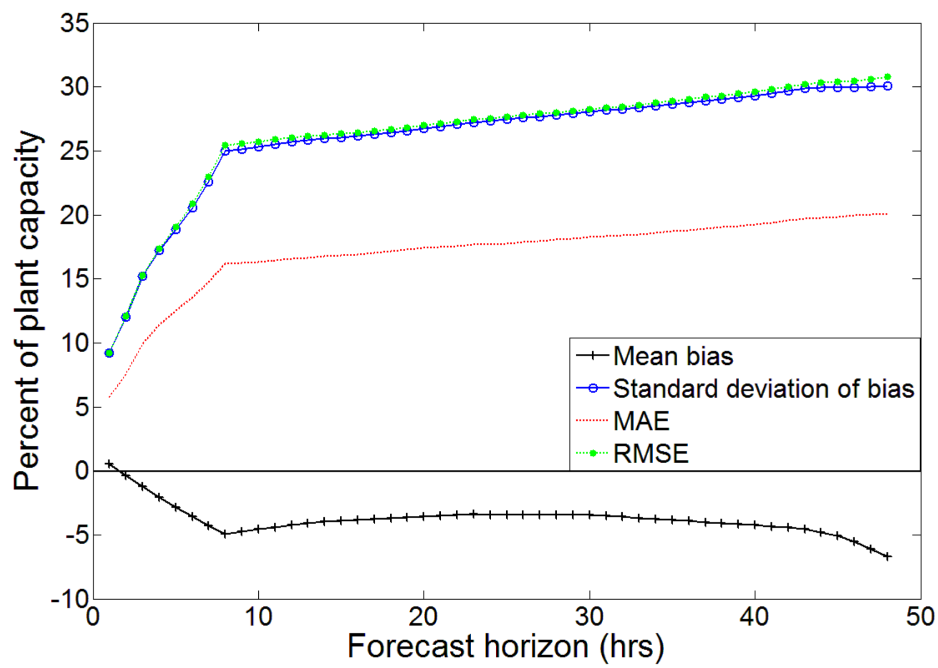


**Figure 16: MAE, RMSE, Mean bias, and standard deviation of forecast error ($\sigma_{f.e.}$) vs. forecast horizon.**

### 4.3 Probability Density Function

Results for the distribution of forecast errors matched a beta probability density function fairly well, in agreement with Bludszuweit (2008) and Bofinger (2002).

Figure 17 shows the normalized forecast errors for the day-ahead forecast horizon plotted against the equation-based beta function with calculated parameters (dashed line) as well as the MATLAB® betafit function (dotted line). The respective α and β parameters for the equation-based beta function were 0.544 and 8.93, and were 0.806 and 12.74 for the MATLAB® betafit function. The errors for other forecast horizons also fit the beta distribution.

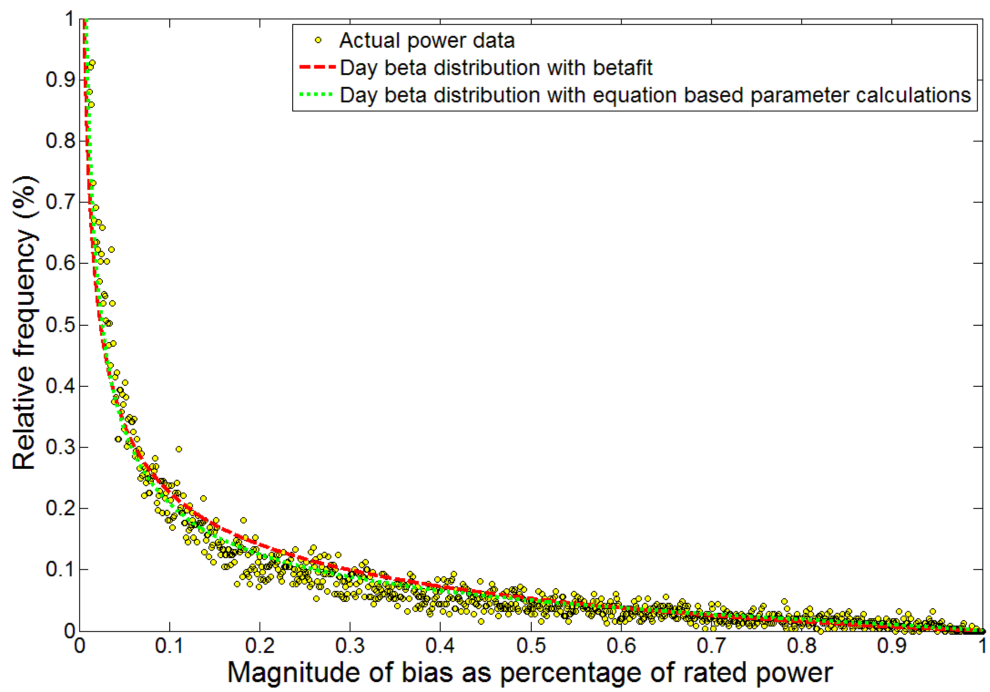


**Figure 17: Distribution of forecast errors from actual data along with two forms of beta probability density function. The equation-based form can be found in Bludszuweit and the betafit form is a built-in function of MATLAB®**

A cumulative beta function was also used to demonstrate the distribution of forecast error values from the actual data and the equation-based beta distribution as shown in Figure 18. Increments of 10% of power production levels were used as shown by each pair of red and blue lines in Figure 18. The cumulative beta is used to predict the frequency of actual power production bins based on the occurrence of forecasted power bin sizes. The distribution matched empirical values well when forecast power bin sizes were at mid-levels, but the actual power seemed to less-accurately match the beta distribution during times when the forecast was predicting low levels (lines on left portion of Figure 18) or very high levels (lines on right portion of Figure 18). These results are similar to those found in Bofinger (2002) where values less than 10% capacity failed the Chi Squared test at a 5% significance level, indicating that the cumulative beta was not a good fit for the data. Referring back to Figure 12, it can be seen that in reality the actual wind power plant spent a significant amount of time in the lowest production bin, suggestion that the cumulative beta may not be a sufficient means to predict actual power bin probabilities based on forecast bin sizes.

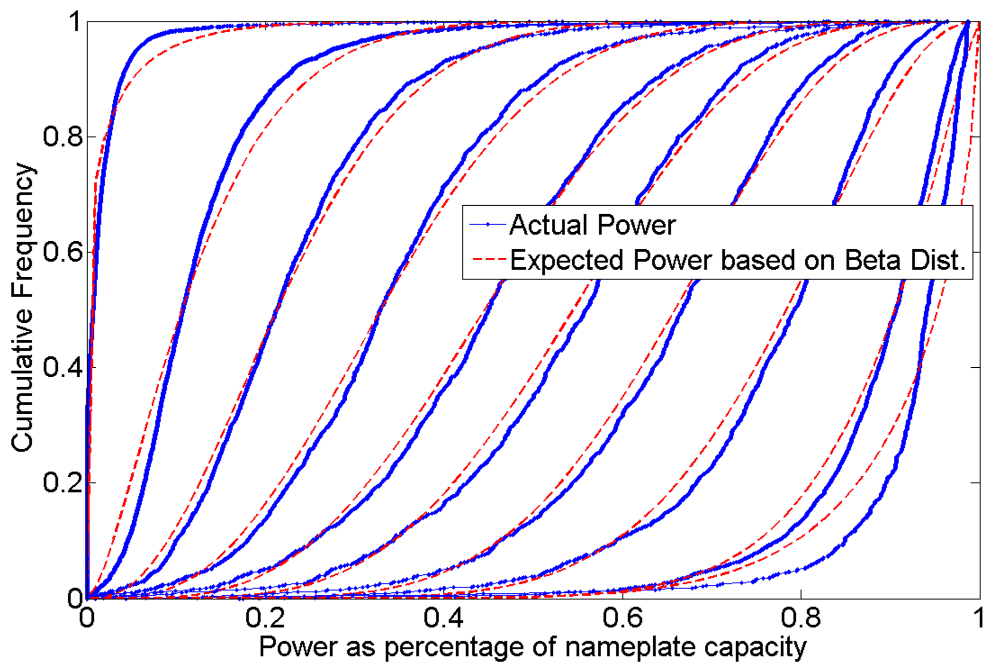


**Figure 18: Cumulative probability of forecast errors occurring for various levels of power production. Actual forecast errors are plotted against equation-based beta pdf.**

### 4.4 Dependence on power production levels

The beta pdf discussed in the previous section provides a means to relate normalized error probabilities to the power production levels, but insight as to whether the forecasting errors are positive or negative is also of value. Trends in forecasting error, including error direction, that depend on power production levels of a wind site may be of interest for integration studies and system planning. Wind power plants spend most of the time producing at lower levels of capacity, and therefore accurate forecasts should predict the same. It is important to know how forecasting error may change with actual production levels, and especially if there are consistent levels of production that seem to correspond to the largest forecasting errors.

The series of histograms presented in Figure 19 and Figure 20 show the mean bias, or average forecast error as a function of power production levels. The color of each bar represents the power production bin size, or the level at which the wind plant was producing. Each grouping of bars represents the bin sizes of bias levels, as indicated by the abscissa labels. For example, the center grouping of colored bars surrounded by the red box in Figure 19 indicates the frequency that the mean bias was between -5 MW and 5 MW for various power production levels (specified by the color of each bar). Each bar was normalized by the number of values used to make it, for ease of visualization.

Histograms are shown for forecast horizons of one through eight hours to show evolution over time. The benefits of this type of plot include the ability to visualize the directionality of the forecasting error. As demonstrated by Figure 16 above, there is a negative mean bias for all forecast horizons except hour ahead. This is confirmed in Figure 19 and Figure 20 by the larger groupings of colored bars with negative bias levels (or to the left of center). Additionally, Figure 16 shows that the magnitude of this bias grows in a linear fashion for horizons one through eight. These observations are also validated by Figure 19 and Figure 20 below, but an extra piece of information that can be obtained is that the negative bias tends to grow in the low power production bins, as indicated by the or leftmost colored bars in each grouping.

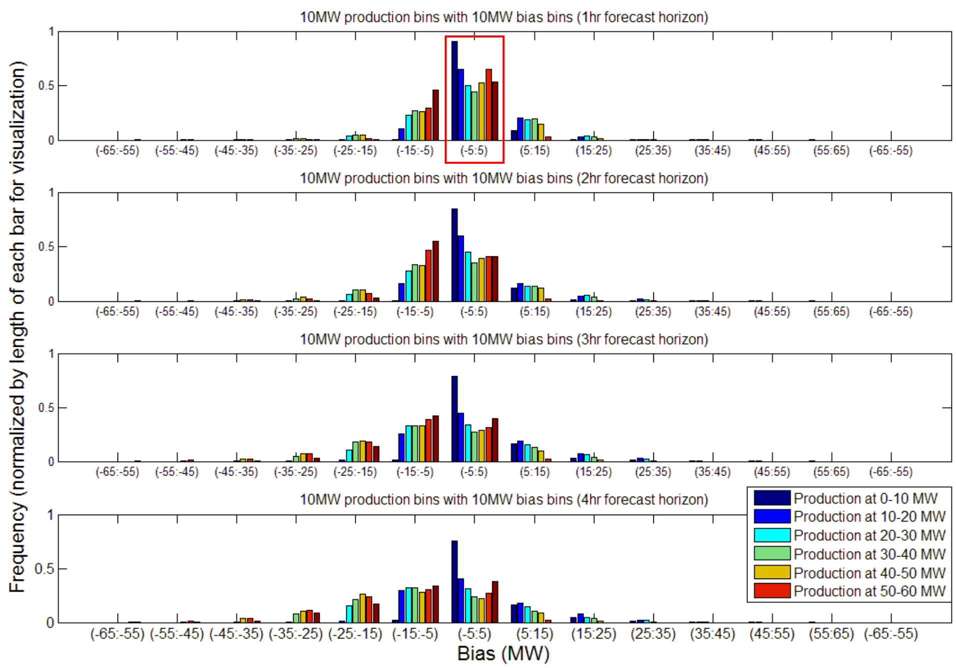


**Figure 19: Forecast error (or bias) levels as a function of wind power production levels for forecast horizons 1-4 hours. Total capacity was in excess of 60 MW.**

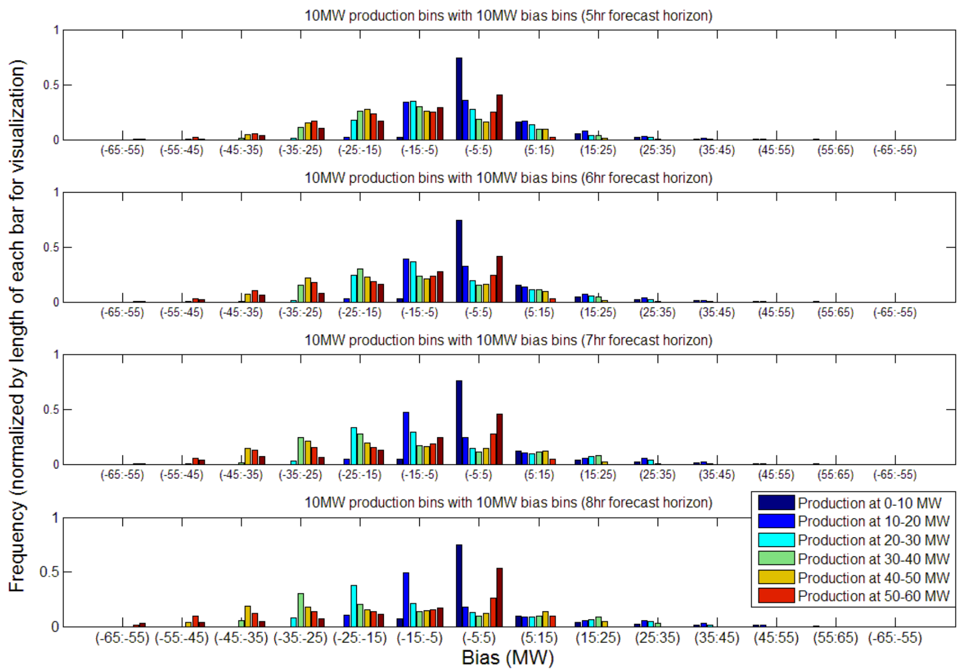


**Figure 20: Forecast error (or bias) levels as a function of wind power production levels for forecast horizons 5-8 hours. Total capacity was in excess of 60 MW.**

## 4.5 Correlation Coefficients

Scatter plots were created to show the correlation between actual and forecasted power values. Figure 21 and Figure 22 show the results for the 1-hour and 8-hour forecast horizons, respectively. Each black dot represents the ratio of normalized actual power to normalized forecast power for a given hour of the time series. The red trend lines were added to show what a perfect forecast would look like, i.e. if the ratio of actual to forecasted power were always one. The clustering of dots in Figure 21 about the trend line indicates a fairly accurate forecast for the 1-hour horizon, and the more widespread results of Figure 22 show less accuracy.

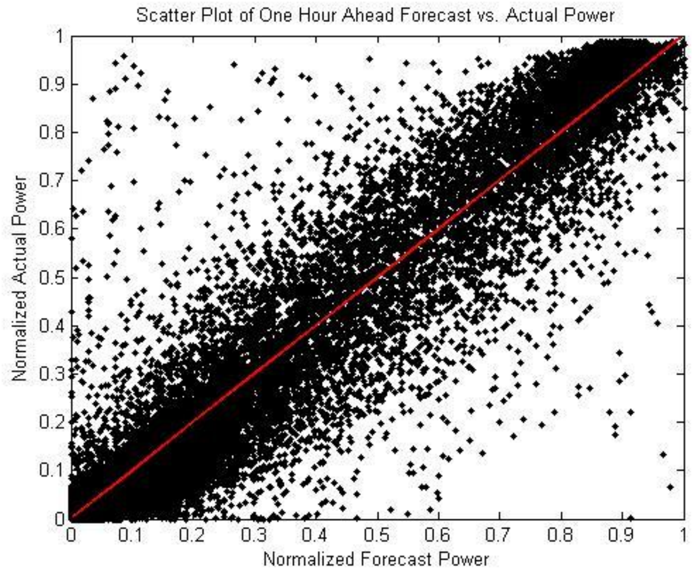


**Figure 21: Scatter plot of power values for actual and forecast data during 1-hour forecast horizon.**

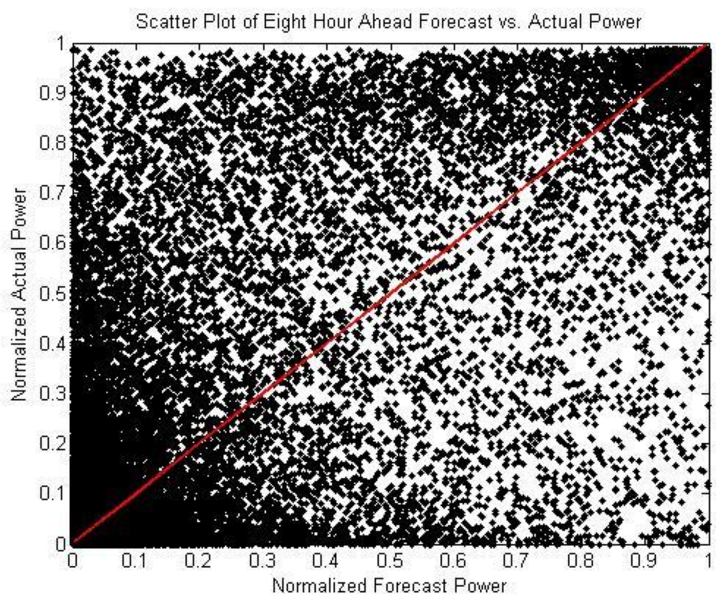


**Figure 22: Scatter plot of power values for actual and forecast data during 8-hour forecast horizon.**

The correlation is related to the information contained in these scatter plots (see Equation 8), and the metric gives insight into how the forecast structure depends on forecast horizon. Figure 23 shows the correlation coefficient vs. the forecast horizon. Recall that the correlation coefficient is approximately the percentage amount by which the prediction error is reduced from simply predicting the mean. Therefore, during the 1-hour horizon the commercial forecast shows about a 96% reduction in prediction errors than would have been the case if the mean value of actual wind power was used as the prediction. The R-value dropped dramatically during forecast horizons 1-8, and continued to drop more gradually after that. These results should be compared to those in Figure 16 as the interesting behavior in early horizons is captured in both.

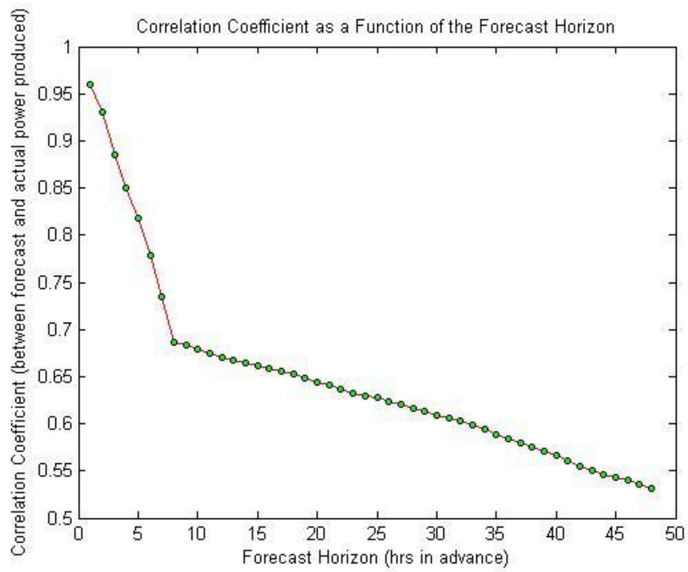


**Figure 23: Correlation coefficient between actual and forecast power production vs. forecast horizon.**

## 4.6 Dependence on rate of change in power production levels

Figure 24 shows the number of hourly deltas from both the actual and forecasted data that exceeded 10 percent of the total plant capacity. The number of deltas of this size from the actual power data set is shown by the black solid line. The number of forecasted deltas for each horizon are shown by the blue "plus" signs, connected by a dotted line. For horizons of three hours and beyond, the forecast data contained far fewer deltas of these magnitudes than did the actual, suggesting less variability in the forecasted power at longer horizons (also refer to Figure 15). Although there was less variability in the forecasted power itself, there was more variability in the size of the forecast errors at longer horizons as shown by the standard deviation of the forecasting errors in Figure 16. The lessened variability of the forecast data will impact some results present later in this chapter. Note that drastic behavior in the number of deltas occured during the forecast horizons 1-8 hours, corresponding to the same horizons as many of the plots already shown.

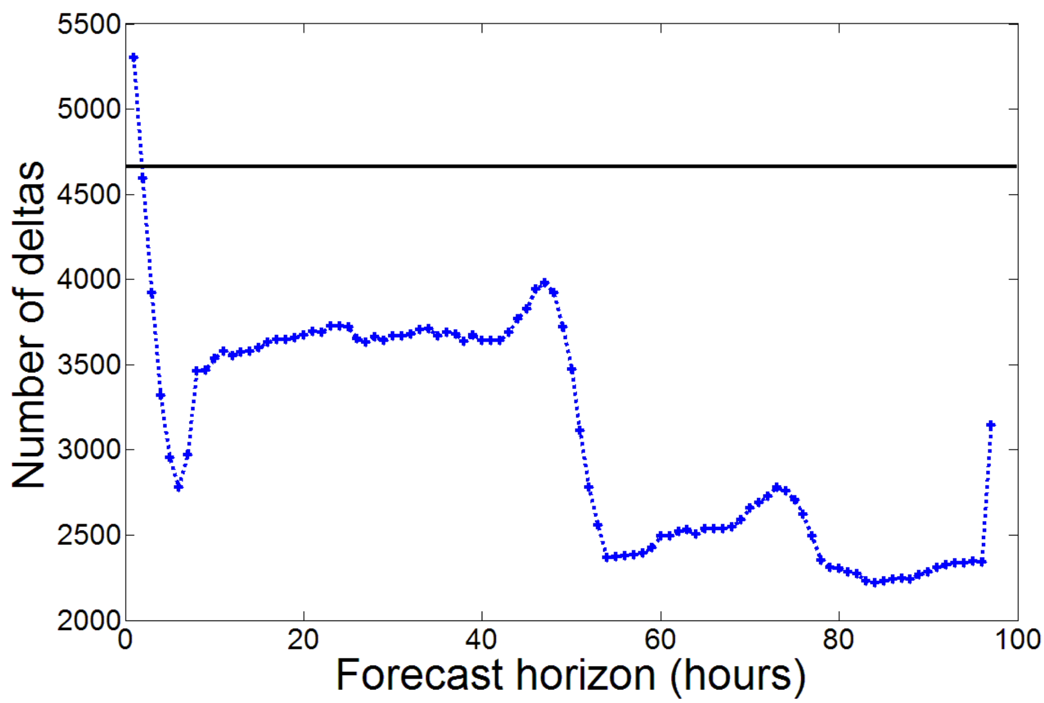


**Figure 24: Number of deltas (hourly step changes in power production) greater than 10% plant capacity at various forecast horizons. The horizontal line represents the number in the actual dataset, and the blue points represent forecast values.**

The dependence of forecast error averages on ramp rates for the 1-hour forecast horizon is shown in Figure 25. The delta values were sorted into 1 MW/hr bin sizes as shown by the abscissa value of each plus sign data point. The ordinate value of each point gives the average of all forecast error values that fall within the respective delta bin sizes (or the mean bias versus delta value). The red line shows a linear fit to all data points. The placement of the trend line shows that in general, the average forecast error was negative when the wind power plant was ramping down (delta < 0), and the average forecasting error was positive when the wind power plant was ramping up (delta > 0).

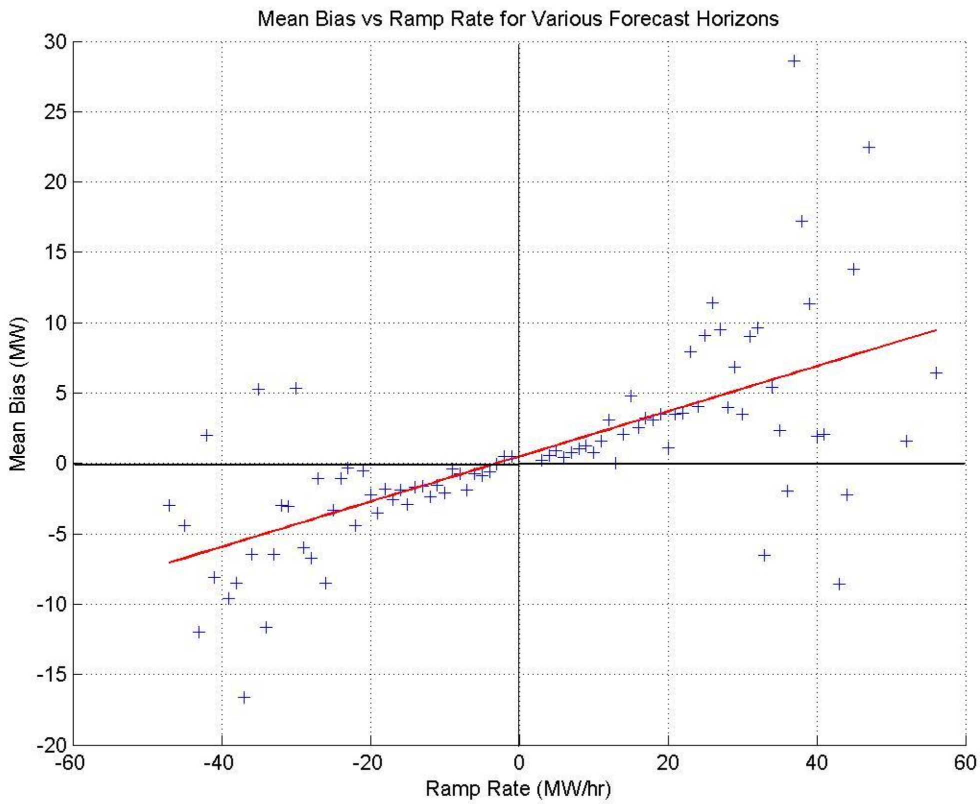


**Figure 25: Mean bias as a function of hourly ramp rate in actual wind power data for the 1-hr forecast horizon. Red line is a linear fit to all plus sign data points**

Figure 26 shows the mean bias as a function of the delta value, for various forecast horizons. Each line in this figure is a trend line of scatter plot values for the respective forecast horizon. In general, the forecast tended to under-predict the wind when the actual wind power plant was ramping down (denoted by the negative bias corresponding to negative ramp rates in the bottom left quadrant of Figure 26), and tended to over-predict the wind when the actual power plant was ramping up (denoted by the positive bias corresponding to positive ramp rates in the upper right quadrant of Figure 26). The trends are less-drastic for the short forecast horizons of 1 and 2-hours, suggesting increased accuracy.  Keep in mind that Figure 26 presents the trend lines only, and scattered values for each horizon were not necessarily bound to the quadrants shown. The results shown in Figure 26 may be affected by the decreased forecast variability at longer forecast horizons as shown by Figure 24. Because there were effectively fewer forecasted deltas of large sizes during the later horizons, the mean bias would be expected to be larger because the forecast ramps would not adequately match the actual ramps.  Also recall that there was an overall negative mean bias for all forecast horizons other than the hour-ahead (Figure 16).

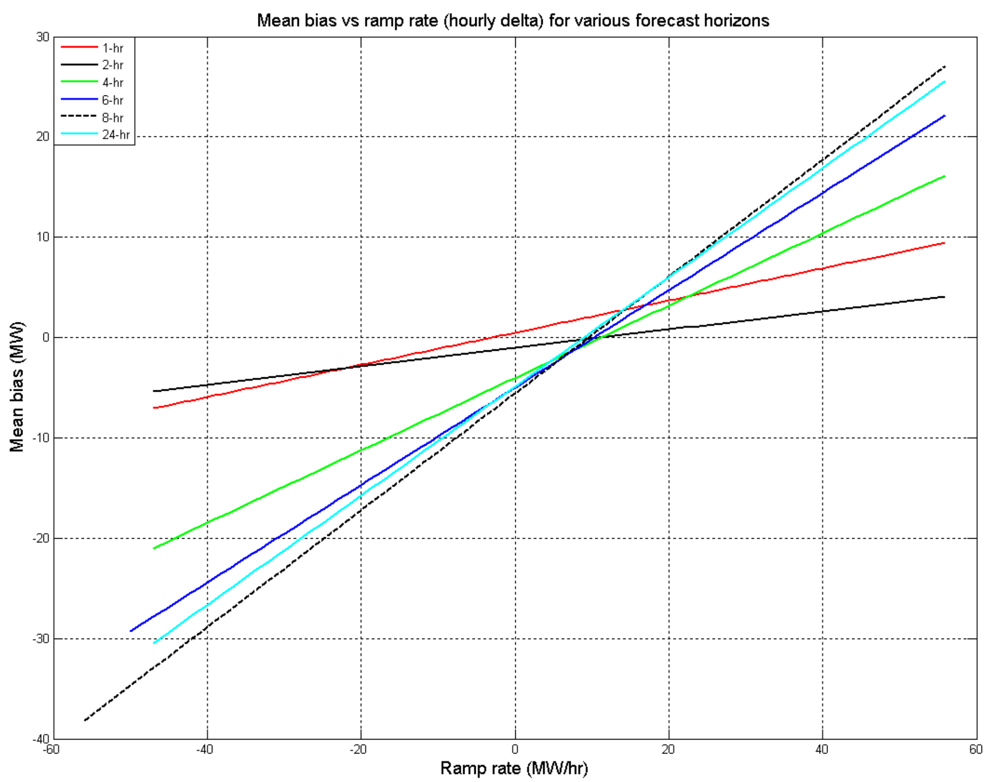


**Figure 26: Mean bias (or average value of forecasting errors) versus hourly ramp rate (delta value) for various forecast horizons.**

Figure 27 presents the mean absolute error (blue dashed line) for the 1-hour wind power forecast horizon as a function of the hourly delta value.  The overall MAE (independent of delta values) is also shown for comparison (red plus sign).  The abscissa in this plot gives the delta value (hourly step change) as a percentage of plant capacity. As shown by the blue dashed line, the MAE increases with the magnitude of the delta value, both positive and negative. The solid black line shows the frequency that delta values of each magnitude occurred, and note that there are very few instances of which the hourly deltas were 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.

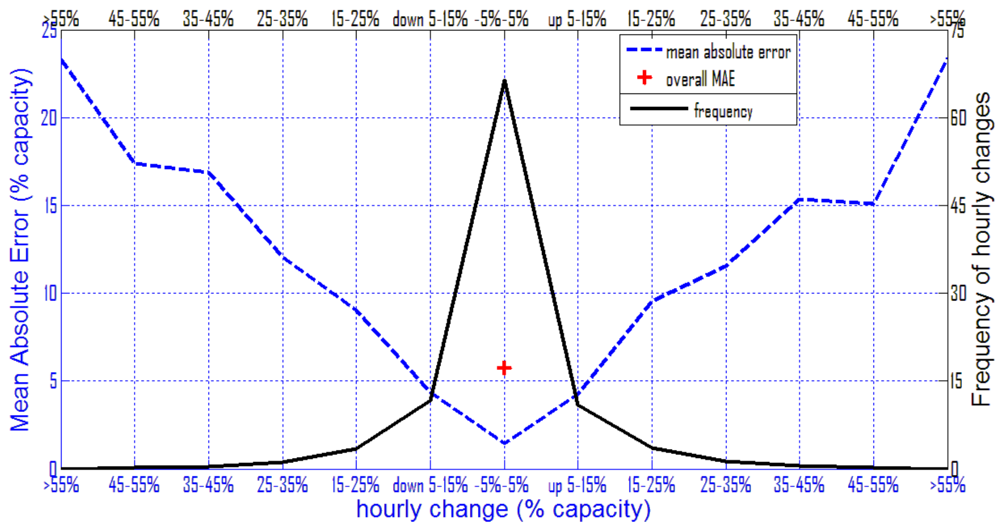


**Figure 27: Frequency of occurrence and MAE as a function of normalized hourly step changes (deltas) in power production. Both up and down deltas are included, as well as the overall MAE. 1-hr forecast horizon data are shown.**

Taken together, Figure 26 and Figure 27 demonstrate that forecast accuracy tends to decrease inversely with the magnitude of the ramp rate, whether positive or negative. It also shows that the overall MAE value given by the red plus sign does not offer complete description of forecast accuracy. The instantaneous MAE may be lower than the overall MAE for the majority of the time, yet the overall MAE may be highly influenced by the outlying errors associated with larger ramping rates.

## 4.7 Characteristics of errors during ramp events selected by the RIA

Error analysis during large ramp events selected by the RIA began with investigating the number of ramps and mean ramp duration of ramps versus the forecast horizon. The RIA parameters were set to select the top 900 actual ramp events[11], and to compare forecasted ramps of similar size within the ±4-hour timing window. A two-hour moving average of ramp rates was used, an mrate of 11 MW/hr resulted in approximately 900 ramp events for the entire 35-month dataset. This is roughly equivalent to 900 occurrences during which the power increased or decreased by at least 22 MW (or about 35% total capacity) in a two-hour timeframe. This size was used both because it identified a sufficient number of ramp events to perform error analysis, and also because production changes of one third capacity in two hours are of interest to system planners due to the balancing actions that may be necessary. Figure 28 shows the number of forecasted up and down ramps compared to the number of actual ramps when all RIA parameters were equivalent. The forecast dataset contained far more large ramps during the hour-ahead, and steadily decreased for horizons 1-5 before increasing and stabilizing again near the actual numbers. These results are in agreement with Figure 24, considering they are only indicative of the largest ramps.

[11] As described in Section 2.3

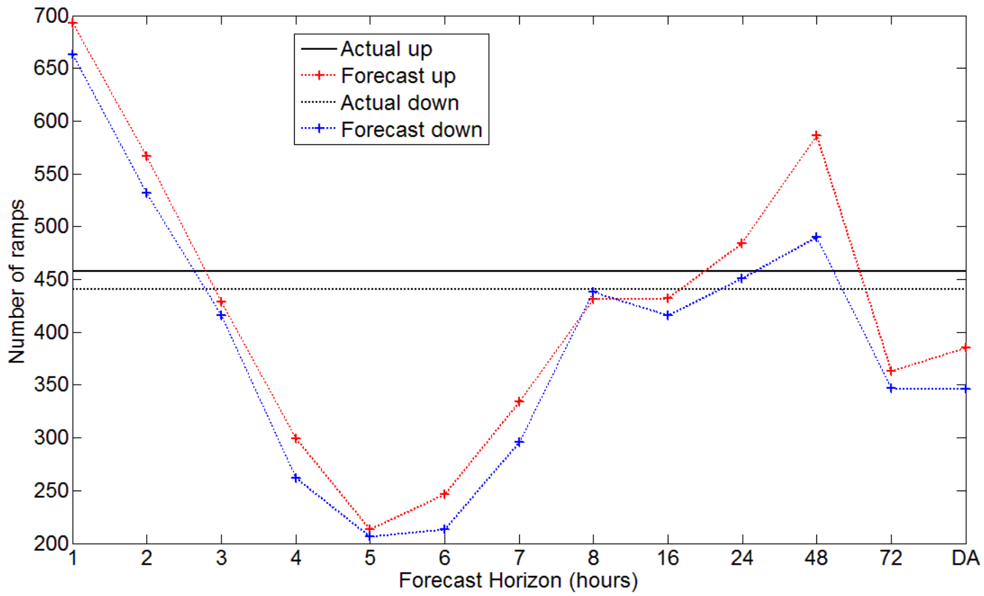


**Figure 28: Number of actual ramps and forecast ramps vs. forecast horizon. The RIA parameters were chosen to select the top 900 actual ramp events.**

Figure 29 shows the mean duration (in hours) of the actual and forecast ramps when the RIA was set to select the top 900 actual ramps. The forecasted mean ramp duration was fairly accurate for horizons 1-4, after which it increased significantly.

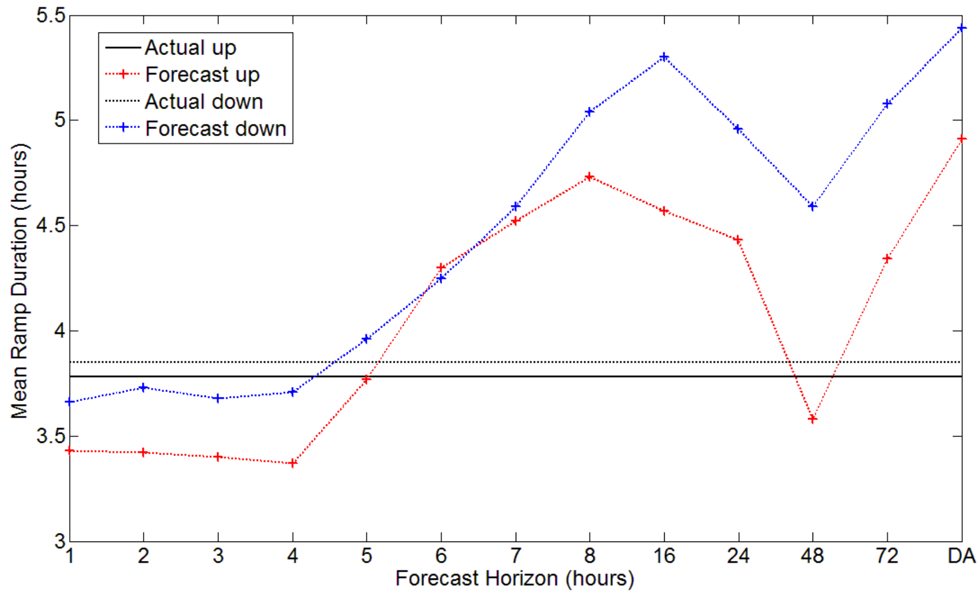


**Figure 29: Mean duration of actual ramps and forecast ramps vs. forecast horizon. The RIA parameters were chosen to select the top 900 actual ramp events.**

The detailed treatment of error analysis during ramp events that were selected by the ramp identification algorithm revealed decreased performance of the commercial forecast during these events, however it should be re-stated that the commercial forecast used for this project was not optimized to predict ramp events. Figure 30 below presents the mean absolute error during times defined as ramp events in the actual GCPUD dataset.

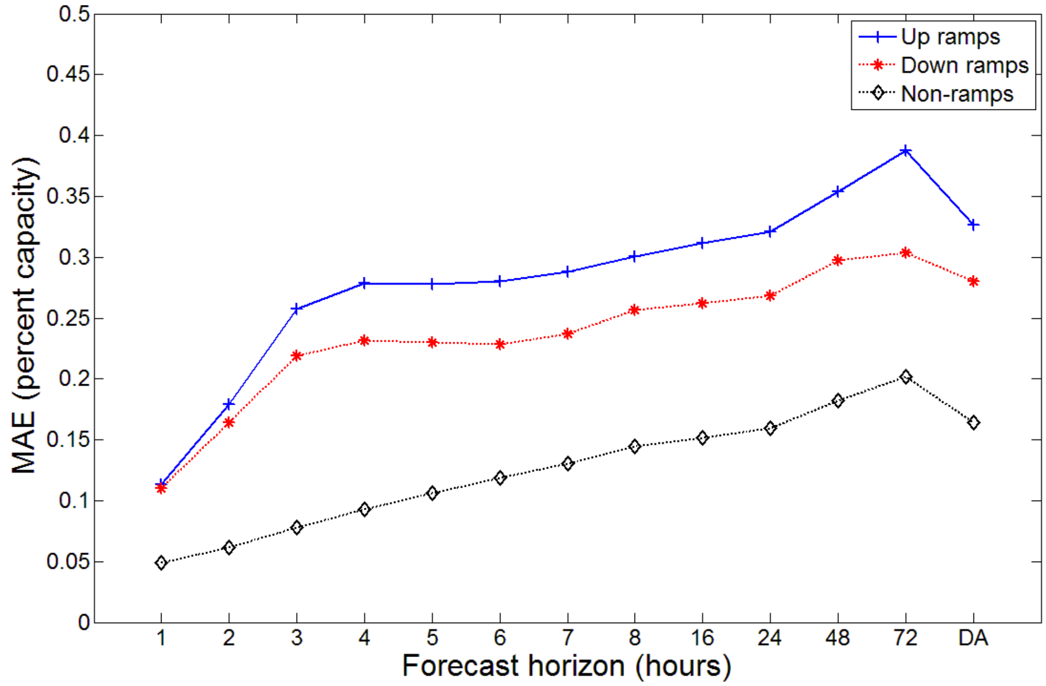


**Figure 30: Mean absolute error during times defined as ramp events by the Ramp Identification Algorithm. The top 900 largest ramp events in the 2004-2006 actual data are included, as well as the MAE during all other times not defined as ramps.**

Figure 30 was generated by computing the MAE during all hours defined as ramp events when an mrate of 11 MW/hr was used, for various forecast horizons. This figure demonstrates the increased forecasting error during ramp events, which confirms the concerns of system operators and planners. During all forecast horizons, there was a greater MAE 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, and the MAE during up ramps is also consistently larger than during down ramps.

The standard deviation of the forecast error, or $\sigma_{f.e.}$, was also computed during the top 900 ramp events for various forecast horizons, as shown in Figure 31. This figure demonstrates that there was consistently more variability in the size of forecasting errors during ramp events than during other times.

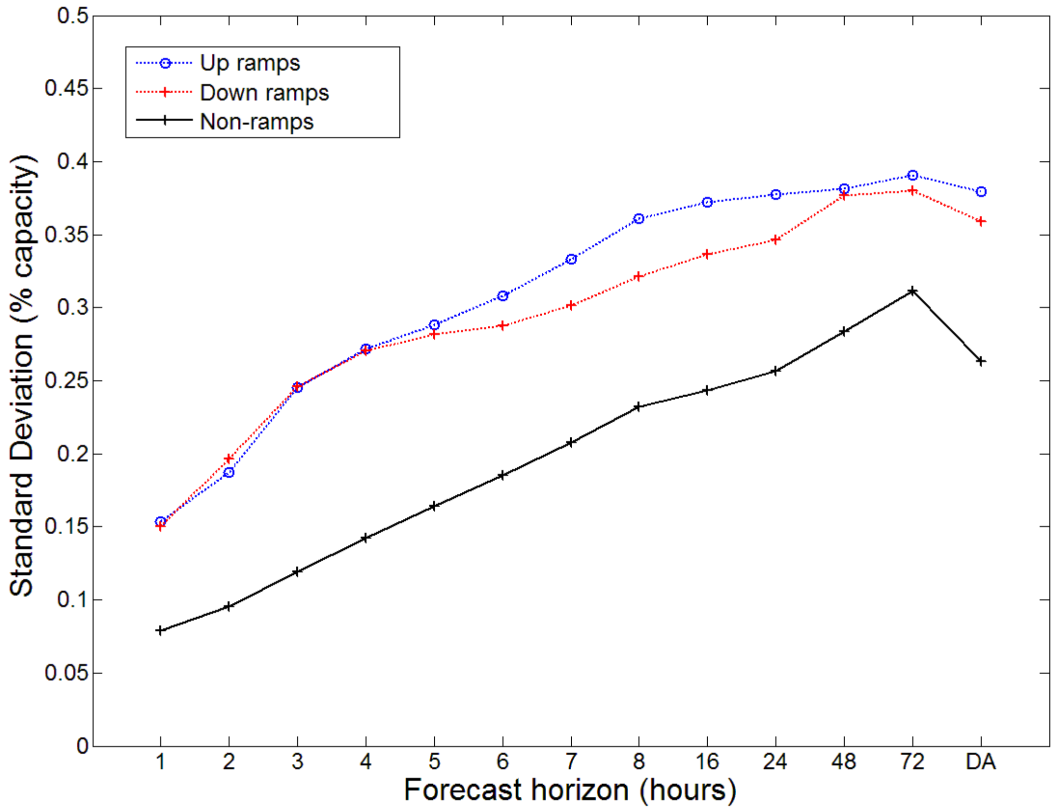


**Figure 31: Standard deviation of bias during times defined as ramp events by the Ramp Identification Algorithm. The top 900 largest ramp events in the 2004-2006 actual data are included, as well as the SD during all other times not defined as ramps.**

The MAE as a function of ramp event size was also investigated. Larger mrate threshold values resulted in larger, and hence fewer ramp events. Figure 32 shows the MAE during ramps and non-ramps as a function of mrate size (and hence ramp event size). The overall MAE, which includes all times is also shown by the green reference line in Figure 32. As shown by the blue line in Figure 32, the MAE is worse for larger ramp events, and considerably worse than the overall MAE for all ramps. The MAE during non-ramps (black line in Figure 32) eventually approaches the overall MAE due to the fact that there are very few ramps at all of the sizes captured by the largest mrate values.

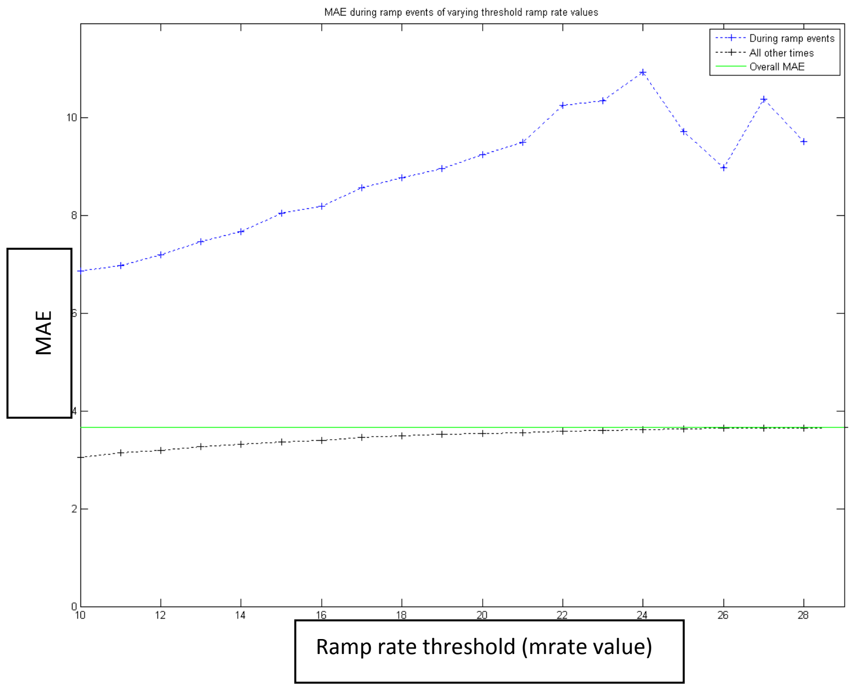


**Figure 32: MAE during ramps and non-ramps vs. size of ramp event.**

The dependence of MAE and $\sigma_{f.e.}$ on ramp size is presented in Table 5. The left side of the table gives the MAE and $\sigma_{f.e.}$ during up ramps, down ramps, and all other times (e.g. those not defined as ramp events by the RIA). The right side of the table gives the same metrics, only each was also computed for the hour before and after the ramp event. It was decided to investigate these error characteristics during one hour before and one hour after each ramp in attempt to include errors from possible phase shifts in predicted ramps. Mrate values ranged from 5-25 MW/hr[12]. Nearly every value in Table 5 increases with the size of mrate, demonstrating that both MAE and $\sigma_{f.e.}$ (indicating variability in forecast error) increased with the size of ramp events.

[12] Bdur = 2, bdur2 = 1, and edur2 = 1, as discussed in Section 2.3.

**Table 5: Forecast performance during ramp events of various sizes. Larger mrate values correspond to larger (and fewer) ramps. Stats are included for times defined as ramps by the RIA, and for the case of 1 hour before and after the ramp (in attempt to include any phase shift).**

| Ramps only | | | 1-hr horizon | | | |
|---|---|---|---|---|---|---|
| mrate | MAE (up) (MW) | MAE (down) (MW) | MAE (other) (MW) | SD (up) (MW) | SD (down) (MW) | SD (other) (MW) |
| **5** | 6.22 | 5.71 | 2.52 | 8.52 | 8.07 | 4.16 |
| **10** | 7.14 | 6.86 | 3.06 | 9.70 | 9.34 | 4.91 |
| **15** | 8.34 | 8.16 | 3.37 | 11.60 | 10.80 | 5.31 |
| **20** | 9.57 | 9.18 | 3.54 | 13.49 | 12.08 | 5.61 |
| **25** | 10.02 | 9.83 | 3.64 | 14.65 | 13.59 | 5.80 |
| | **MAE ALL** | **3.6578** | | **SD All** | **5.8519** | |

| Ramps ± 1hr | | | 1-hr horizon | | | |
|---|---|---|---|---|---|---|
| mrate | MAE (up) (MW) | MAE (down) (MW) | MAE (other) (MW) | SD (up) (MW) | SD (down) (MW) | SD (other) (MW) |
| **5** | 6.01 | 5.37 | 3.37 | 8.37 | 7.73 | 5.27 |
| **10** | 6.73 | 6.24 | 3.56 | 9.38 | 8.70 | 5.79 |
| **15** | 7.61 | 7.17 | 3.62 | 11.00 | 9.99 | 5.83 |
| **20** | 8.38 | 7.82 | 3.65 | 12.55 | 10.90 | 5.82 |
| **25** | 7.84 | 8.24 | 3.67 | 12.05 | 11.62 | 5.86 |

The mean and standard deviation of delta values can also be calculated for only the times selected as ramp events by the RIA. For example, Figure 33 and Figure 34 show these metrics during the top 100 ramp events as selected by the RIA from the actual wind power time series. Recall that an mrate of 21 MW/hr was used to obtain these ramps with a rolling average of deltas taken over two hours. Therefore, it makes sense that the mean delta from the actual time series during these times would be approximately 10.5 MW/hr, or 21/2. The mean deltas for the forecast time series during these times again showed less-variability. A similar trend is seen between actual and forecast $\sigma\Delta$ values shown in Figure 34. A greater variability is expected during large ramp events, as ramp events are essentially defined as variability in power output.

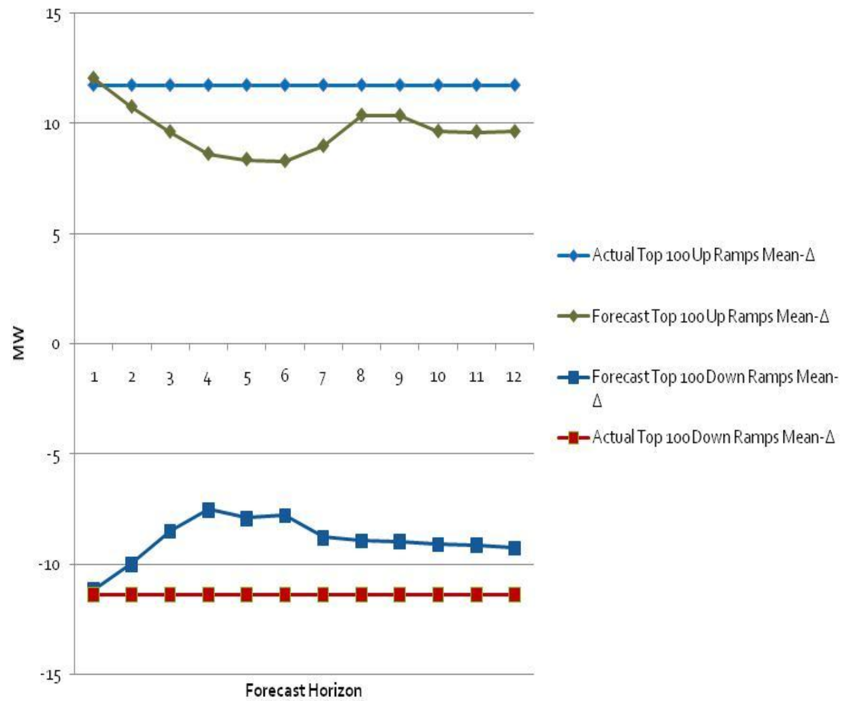


**Figure 33: Mean hourly delta vs. forecast horizon during top 100 ramp events of actual power time series.**

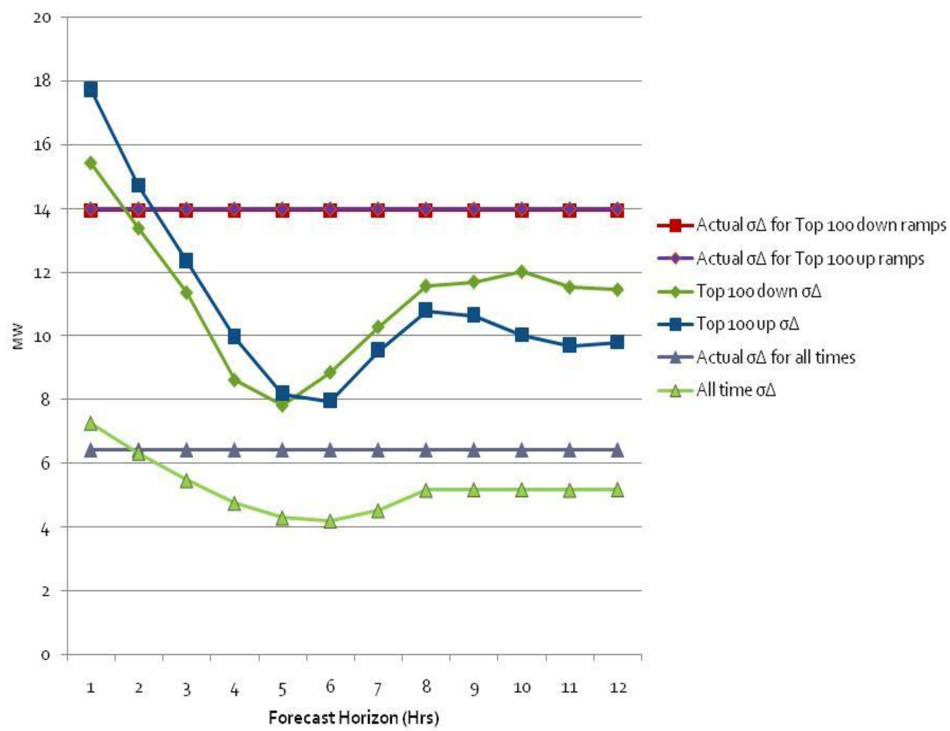


**Figure 34: Standard deviation of hourly delta vs. forecast horizon during top 100 ramp events of actual power time series.**

## 4.8 Ramp event phase error analysis

Recall from Section 2.3 that in addition to selecting ramp events of desired sizes and durations, the Ramp Identification Algorithm is capable of performing magnitude and temporal correlation analysis between actual and forecast ramp events. The motivation behind this was to establish a means to characterize the phase errors in ramp forecasting. The RIA can be used to search for actual and forecasted ramps that are correlated within a user-inputted timing window. Since the algorithm specifies a definitive start and end time for a ramp event, either of these (or any time between) can be chosen to perform the correlation. For all of the results discussed in this section, temporal correlation analysis was based on the start times of actual and forecasted ramp events of similar magnitude that occurred within a window of ±4 hours of each other. The top 900 actual ramp events were used for this analysis, meaning that the results below are representative of forecasted ramp events similar in size that occurred within ±4 hours of each of the top 900 actual ramp events[13].

The template used to keep track of ramp correlation statistics and phase error characteristics is presented in Table 6. The total number of actual ramps is shown, along with the number of them that did not have a correlated forecast ramp within the ±4 hour window. The number of forecast ramps that both lead and lagged behind actual ramps in time are shown, as well as several metrics (to be discussed later in this section) describing other properties of timing errors in forecasted ramps. The average ramp durations are shown, along with the directionality of actual and forecasted ramp events described by the contingency tool in Table 3. The lower section of Table 6 gives the number of forecasted ramp event starts that fell within the specified hourly increments before, after, or at the correct start times of the actual ramps.

[13] The top 900 actual ramp events were chosen for this analysis for the fact that system operators are generally concerned with the larger ramp events only, which are sufficiently captured by the top 900 for this dataset. The analysis could be repeated with any combination of actual and forecasted ramp sizes, with temporal correlation windows of any size.

A new metric was developed during this thesis to quantify the phase errors for ramp events. This metric was called the mean temporal bias (MTB), which indicates the average amount of time by which the actual ramp events were missed by the forecast. For the purposes of this project, the MTB will be reported in hours although any temporal units could be used provided sufficient resolution of the data. The term "temporal standard deviation" (TSD) is also used to denote the standard deviation of the phase errors in ramp forecasting (also in hours for this project).

**Table 6: Template used to track statistics for ramp correlations and phase error analysis.**

**2-hr Forecast Comparison**

**RIA Input Parameters = [11,2,1,1,0]**

| | Total Ramps | Uncorrelated With Forecast* | Forecast Ramp Start Leads | Mean Bias (leads) | σ lead | Forecast Ramp Start Correct | Mean Temporal Bias | σ all | Forecast Ramp Start Lags | Mean Bias lags | σ lags |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Up Ramps** | 458 | 71 | 62 | 2.53 | 1.18 | 88 | -0.54 | 1.53 | 274 | -1.41 | 0.57 |
| **Down Ramps** | 441 | 69 | 43 | 2.72 | 1.14 | 77 | -0.74 | 1.46 | 282 | -1.47 | 0.58 |

*Search limit for correlated forecasts = 4 hours

| | Total Ramps | Mean Duration | Standard Deviation of Duration |
|---|---|---|---|
| **Up Ramp Actual** | 458.00 | 3.78 | 1.44 |
| **Up Ramp Forecast** | 567.00 | 3.42 | 1.51 |
| **Down Ramp Actual** | 441.00 | 3.85 | 1.68 |
| **Down Ramp Forecast** | 532.00 | 3.73 | 1.66 |

| | Mean Temporal Bias (hours) |
|---|---|
| **Actual up; Forecast up** | -0.54 |
| **Actual down; Forecast down** | -0.74 |
| **Actual up; Forecast down** | -0.74 |
| **Actual down; Forecast up** | 0.86 |

Forecast Lead(+) Lag(-)

| | Total Correlated Ramps | 4hrs | 3hrs | 2hrs | 1hr | Forecast Start Correct | -1hr | -2hrs | -3hrs | -4hrs |
|---|---|---|---|---|---|---|---|---|---|---|
| **Actual up; Forecast up** | 424 | 19 | 11 | 16 | 16 | 88 | 172 | 93 | 8 | 1 |
| **Actual down; Forecast down** | 402 | 14 | 12 | 8 | 9 | 77 | 160 | 114 | 6 | 2 |
| **Actual up; Forecast down** | 27 | 1 | 1 | 4 | 6 | 3 | 0 | 0 | 7 | 5 |
| **Actual down; Forecast up** | 21 | 2 | 3 | 6 | 4 | 1 | 0 | 2 | 1 | 2 |

Table 7 presents the template used to track several metrics for characterizing magnitude and phase errors of ramp events. Three different mrate values were used, and one of these tables was created for each forecast horizon. Most values in Table 7 are given as percentages of the wind power plant capacity, and were generated from the raw template values of Table 6.

**Table 7: Template of metrics relating to temporal and magnitude errors during ramp events.**

| Temporal error stats for ramp events of various sizes (2-hr horizon) | | | |
|---|---|---|---|
| **(mrate)** | **21.50** | **16.00** | **11.00** |
| Total up ramps | 58 | 188 | 458 |
| Total down ramps | 42 | 172 | 441 |
| Uncorrelated up | 0.33 | 0.21 | 0.16 |
| Uncorrelated down | 0.29 | 0.24 | 0.16 |
| Correct start up * | 0.26 | 0.18 | 0.21 |
| Correct start down* | 0.17 | 0.18 | 0.19 |
| Up leads * | 0.03 | 0.07 | 0.15 |
| Up lags * | 0.72 | 0.75 | 0.65 |
| Down leads * | 0.00 | 0.04 | 0.11 |
| Down lags * | 0.83 | 0.79 | 0.70 |
| * % of correlated ramps (subject to rounding) | | | |
| **Up Ramps** | | | |
| Mean Temporal Bias | 0.92 | 0.83 | -0.54 |
| MTB (leads) | 2.00 | 2.91 | 2.53 |
| TSD (leads) | 0.00 | 1.14 | 1.18 |
| MTB(lags) | -1.36 | -1.37 | -1.41 |
| TSD (lags) | 0.56 | 0.57 | 0.57 |
| **Down Ramps** | | | |
| Mean Temporal Bias | 1.20 | 0.98 | -0.74 |
| MTB (leads) | NaN | 3.20 | 2.72 |
| TSD (leads) | NaN | 1.10 | 1.14 |
| MTB(lags) | -1.44 | -1.39 | -1.47 |
| TSD (lags) | 0.51 | 0.51 | 0.58 |
| **Error Stats** | | | |
| Mean Bias up | -0.17 | -0.16 | -0.13 |
| Mean Bias down | 0.19 | 0.14 | 0.10 |
| Mean Bias non-ramps | 0.00 | 0.00 | 0.00 |
| MAE up | 0.23 | 0.20 | 0.18 |
| MAE down | 0.25 | 0.20 | 0.16 |
| MAE other | 0.07 | 0.07 | 0.06 |
| SD up | 0.26 | 0.21 | 0.19 |
| SD down | 0.26 | 0.22 | 0.20 |
| SD other | 0.12 | 0.11 | 0.10 |
| MAE up +- 1 | 0.19 | 0.18 | 0.16 |
| MAE down +- 1 | 0.20 | 0.17 | 0.15 |
| MAE other +- 1 | 0.08 | 0.07 | 0.07 |
| SD up +- 1 | 0.26 | 0.22 | 0.19 |
| SD down +- 1 | 0.26 | 0.23 | 0.19 |
| SD other +- 1 | 0.12 | 0.12 | 0.12 |

Upon creating a series of tables similar to Table 6 and Table 7 for several forecast horizons of interest, it was possible to investigate the dependence of many of these metrics on forecast horizon thus allowing the commercial forecast structure to be better-understood. Figure 35 shows the frequency of correctly forecasted ramp event start times for various forecast horizons. This technique effectively represents the phase accuracy of the commercial forecast with respect to ramp prediction. Up ramps were kept separate from down ramps, and only the actual ramps that had a correlated forecasted ramp within the ±4 hour timing window were used for these results. For the hour-ahead forecast horizon, roughly 40% of both up and down ramps were predicted to begin at the correct hour by the commercial forecast, compared to approximately 20% for the 2-hour horizon. Clearly, the accuracy of predicting ramp event start times drops off dramatically during the first three hours of forecast horizon, and becomes somewhat more stable after.

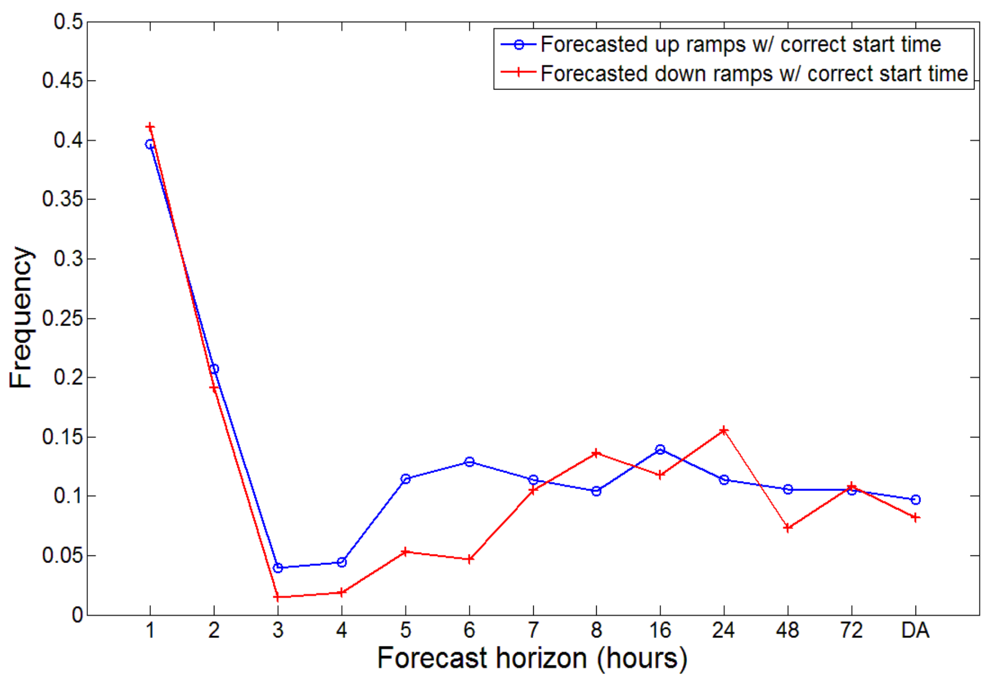


**Figure 35: Frequency of correctly forecasted up and down ramps versus forecast horizon (represents phase accuracy of commercial forecast). These data include only actual ramps that also had a forecasted ramp within the ±4 hour timing window.**

Of the ramp events that were not forecasted to begin on the correct hour, yet still fell within the ±4 hour window, it was important to investigate whether the forecasted ramps were leading (occurring before) or lagging (occurring after) the actual ramps in time. Figure 36 shows the frequency of correlated up and down ramps that were either leading or lagging the actual ramps. For the hour-ahead forecast horizon, both up and down predicted ramps showed relatively equal frequencies of leading or lagging the actual ramps (roughly 28-32%). This is not a surprising result, recalling from Figure 35 that a good number of ramps were accurately forecasted during the hour-ahead horizon, leading one to suspect the remaining ramps to be somewhat normally distributed in time about the actual ramps. During forecast horizons of 2-6 hours, the overwhelming majority of both up and down forecasted ramps lagged behind the actual ramps, during the same hours that the phase accuracy dropped off dramatically in Figure 35. The frequencies tend to switch and somewhat stabilize after the six-hour horizon.

Also note that for any forecast horizon, the values in Figure 35 and Figure 36 for either up or down ramps should add to 1, since all are based on the number of correlated ramp events. For example, consider up ramps for the 3-hour forecast horizon. Adding the blue line in Figure 35 with both the blue and red lines of Figure 36 gives (approximately),0 .05+0.8+0.15=1.

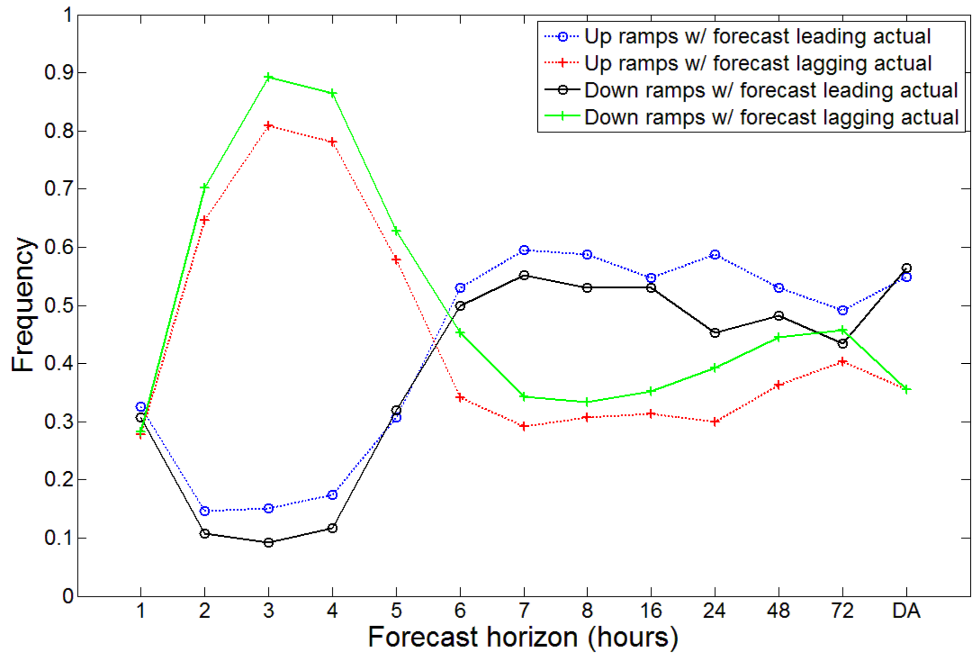


**Figure 36: Frequency of up and down ramp starts leading or lagging actual ramp starts in time. This plot excludes ramps (out of the top 900) that had no correlated forecast ramp at all.**

Significant information regarding the temporal accuracy of predicted ramp events was provided by the techniques shown in Figure 35 and Figure 36. In addition to these analytical results, another important attribute of error characterization during ramp events is to quantify the size of the phase errors. Figure 37 presents the mean temporal bias of the top 900 ramp events. The ordinate values in this plot give the MTB, with negative values indicating that the forecasted ramp was lagging the actual ramp in time, and positive values indicating a prediction that leads in time. The hour-ahead forecast horizon had a near-zero MTB, suggesting that the distribution of forecast ramp phase errors was somewhat normally distributed (and also reiterating the phase accuracy in the hour-ahead horizon. Hours 2-4 show a near-linear decrease in the MTB, suggesting that predicted ramps tended to lag behind the actual ramps. Consider the 4-hour forecast horizon as an example. On average, predicted up ramps tended to lag behind the actual up ramps by about two hours, while predicted down ramps tended to occur about 2.5 hours later than the actual down ramps. At hour six, the average value of the phase error for predicted ramps makes a transition from lagging to leading the actual ramps, confirming the results of Figure 36.

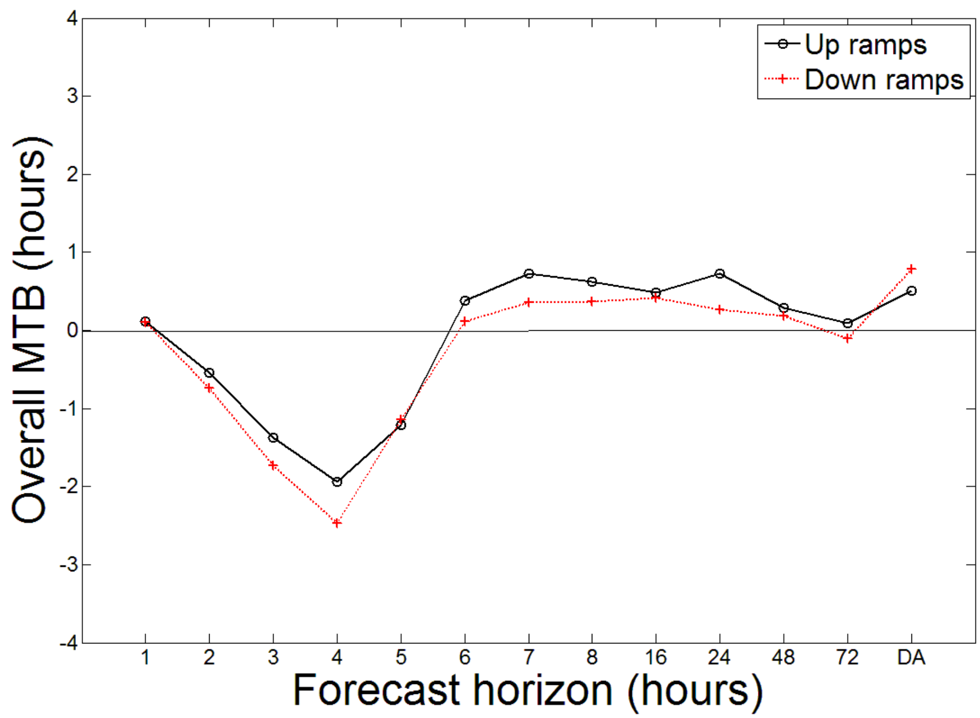


**Figure 37: Mean temporal bias vs. forecast horizon for top 900 largest actual ramp events that also have a corresponding forecasted ramp of similar magnitude within a ±4 hr window.**

An additional item of significance contained in Figure 37 is the appearance of a small positive (and hence near-zero) MTB for hours six and beyond. This could mean that ramp forecasting becomes better at these longer horizons, which would be supported by the results of Figure 28 that show a fairly accurate ramp count during these same horizons. However, Figure 29 shows decreased accuracy in mean ramp duration during later horizons, and referring back to Figure 24 it is evident that the forecast dataset contains less overall variability than the actual dataset during later horizons. The important point to take away is that all of these metrics must be taken together, as none by itself gives complete information about the relation between the two time series. For electrical system planning and operational concerns, the first several hours of the forecast horizon will be most significant with regards to regulation concerns.

The mean temporal bias can also be used to separately quantify leading and lagging times of ramp events. Figure 38 shows the MTB for both the leading and lagging predicted ramps. The amount of time by which lagging ramps follow the actual ramps increases for horizons 1-5. The leading ramps stabilize more quickly at hour two with an MTB remaining near 2.5 hours. The results of Figure 37 were obtained by averaging the leading with the lagging (taking into account their respective frequencies).

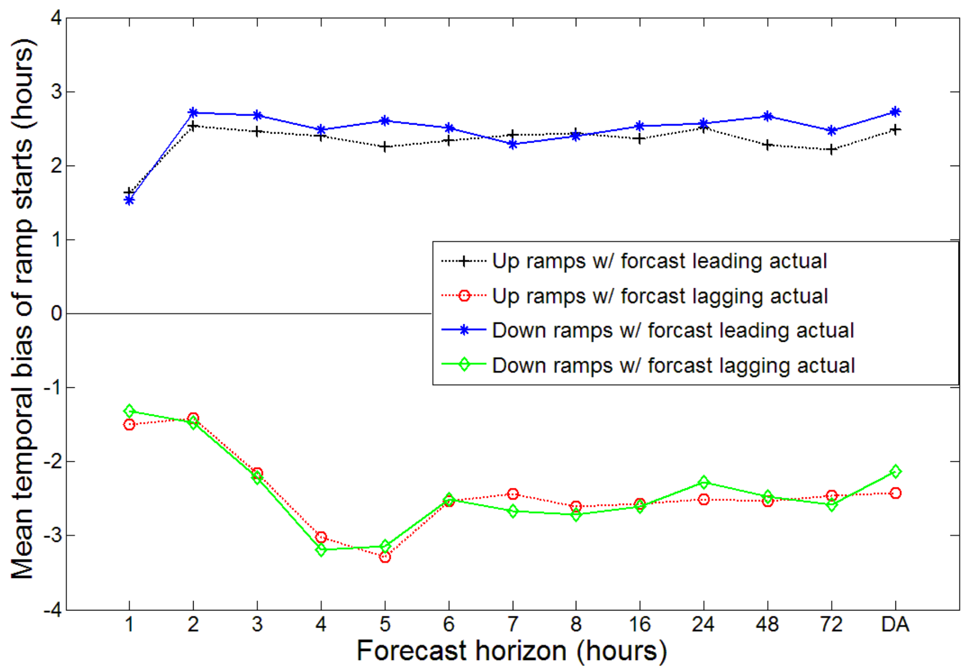


**Figure 38: Mean temporal bias vs. forecast horizon with leading and lagging ramps kept separate.**

## Chapter 5: Selection of metrics to evaluate forecast performance

The goal for incorporating wind power into the electrical generation mix is to gain an understanding of the variability and uncertainty associated with the natural resource. Wind power forecasting errors are an unavoidable component of wind integration, and a thorough grasp on common error trends can aid with the process. There is no single metric or evaluation technique that encompasses all of the complexities associated with the properties of variability, uncertainty, and forecasting errors. The above discussion demonstrates a rigorous characterization of the trends and errors seen in a typical wind power forecast created by a state-of-the-art forecast provider. The purpose of this section is to assemble a relatively small number of evaluation criteria from above into a single tool that can be used to evaluate forecast performance and characterize the patterns seen in forecast composition and error behaviors.

Each of the techniques presented in this thesis provides a means to characterize either the structure of a commercial forecast or the associated prediction errors. When used together as part of a methodical approach to analyzing any pair of wind power and forecast time series, these metrics can be sufficient to characterize the forecasting errors. The list of significant metrics to form a comprehensive evaluation tool is summarized in Table 8.

**Table 8: Selected statistical parameters for forecast performance evaluation.**

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

The mean Δ and σΔ provide a way to quantify the variability of wind power. In addition to these two metrics, the variability of wind power can also be analyzed by creating the distributions of power production levels and rates of change of power production given by the delta values. These were demonstrated by a variety of plots in Section 4.1 as well as Figure 14, and have also been demonstrated in Piwko (2004), EnerNex (2004), and other studies. Tabulations of various-sized delta counts can also be made as demonstrated by Table 4.

The uncertainty component of wind is more difficult to quantify by itself, but may perhaps best be described by a combination of magnitude and phase analysis of the errors. Intensive examination of ramping events allows each of these approaches to be taken. The number of ramps of similar size in each time series should be counted (ramps identified by the RIA and tabulated in the template shown in Table 6). Upon doing this, the percentage of correct ramp starts coupled with the mean temporal bias allows the phase errors to be quantified for each forecast horizon. Further assessment of the MAE during ramps and non-ramps provides information relating to the magnitude errors during ramp events.

Overall forecast accuracy can be captured by the MAE, RMSE, correlation coefficient, and mean bias. The standard deviation of the forecast errors gives more information regarding the distribution of error sizes. The beta pdf can also be applied to the instantaneous forecasting errors for confirmation of distribution trends. These metrics can be tested by creating the trend plots and histograms relating the mean bias to the power production level or to the hourly ramp rate (refer to Figure 19, Figure 20, and Figure 25).

It is important to note that one set of these parameters will be needed for each forecast horizon. Also, commercial forecasts can be optimized to meet specific customer criteria such as achieving a desired MAE. The above metrics can be tailored to address these concerns if needed.

## Chapter 6: Implications, uses and continued research

Implications for characterizing wind power forecasting errors are primarily concentrated in the area of wind integration studies, however there are also many components that are relevant to the planning and operation of electrical systems containing actual wind power plants.

Wind integration studies may incorporate either a state-of-the-art real forecast along with actual wind power data (from existing regional wind power plants) to perform impact assessment, or a combination of simulated data can be used. For a new development site, the wind power data must be simulated using NWP methods along with turbine power curves, regional inputs from met towers or other devices, and model output statistics. State-of-the-art wind power forecasts are computationally and financially expensive to produce. For this reason, simulated or synthetic wind power forecasts can be incorporated into wind integration studies. These forecasts can be created using ARMA methods [Milligan 2003] or other techniques. Synthetic forecasts may be based on statistical properties of the wind power time series as opposed to real meteorological data. Therefore, if a simulated forecast is used in a wind integration study, it is imperative that it possesses the patterns and characteristics that will occur in an actual forecast so that the impacts can be properly assessed. This includes the replication of forecast error trends.

The methodology developed in this thesis for forecast error characterization can be used to validate synthetic forecast data. Although the actual results presented in this thesis apply only to the commercial forecast and actual wind power data used, the evaluation tools can be applied to any matching forecast and power datasets. For example, consider a region that contains operational wind power plants as well as proposals to expand the size of them or to construct new power plants. If the errors from a single, regional plant can be characterized using the methods developed in this thesis, they can subsequently be applied to integration studies carried out within the region.

In addition to being used to validate a simulated forecast or to evaluate forecast performance in general, the process developed in this thesis could be used as part of a procedure to create a rapid synthetic forecast. For example, a synthetic forecast that reproduces the same error patterns that are commonly found in a real forecast may be sufficient for use in wind integration studies. This type of synthetic time series could be generated from the statistical error properties themselves. It could be optimized to meet the desired criteria for the various metrics used for evaluation. For example, a forecast could be created with specified bounds to the MAE during ramps and non-ramps, or to achieve a specific MTB for ramps during a particular forecast horizon. This topic is the basis for Kemper (2010), and serves as a continuation of this thesis project.

One final implication of uses for this thesis is further development and use of the Ramp Identification Algorithm. There is no industry standard for identifying entire wind power ramping events, and the RIA could serve as a foundation for a standardization tool. It can be used on data of any temporal resolution and optimized to search for ramps of desired size and duration, as well as perform correlation analysis between two time series.

## Chapter 7: Conclusion

A rigorous statistical characterization of the wind power forecast errors was conducted. 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 more reliance on NWP predictions at the approximately 8-hour forecast horizon is evident in the transitions seen in most figures of this report, notably the classic elbow feature of Figure 16. 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 8.

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 the system control rooms. An overall MAE, RMSE, or simple bias metric does not accomplish this. 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 errors which can be compared to the beta pdf.

Although the results presented here apply only to one particular power and forecast couple, the repeatable process offered during this report for ramp identification, statistical characterization, and important parameter analysis 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 evaluation of a 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 Kemper (2010).

## References


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

Bofinger, S., Luig, A., Beyer, H., (2002) *Qualification of wind power forecasts*, University of Applied Sciences Magdeburg-Stendal, Dept. of Electrical Engineering.

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

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, Golden, CO.

EnerNex Corp. and Windlogics Inc. (2004) "Xcel Energy and the Minnesota Department of Commerce, Wind Integration Study – Final Report," http://www.uwig.org/XcelMNDOCStudyReport.pdf.

EnerNex Corp. (2007) "Avista Corporation Wind Integration Study Final Report," http://www.uwig.org/AvistaWindIntegrationStudy.pdf.

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.

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

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.

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.

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

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

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

Milligan, M., Schwartz, M., Wan, Y, (2003) *Statistical Wind Power Forecasting for U.S. Wind Farms,* National Renewable Energy Laboratory, NREL/CP-500-35087.

Ott, R., Longnecker, M. (2001). An Introduction to Statistical Methods and Data Analysis, fifth edition, Pacific Grove, CA, USA: Wadsworth Group

Piwko, R., Boukarim, G., Clark, K., et al. (2004). "The Effects of Integrating Wind Power On Transmission System Planning, Reliability, and Operations, Report on Phase 1: Preliminary Overall Reliability Assessment," Prepared for the New York State Energy Research and Development Authority, by General Electric's Power Systems Energy Consulting, Schenectady, NY.

Piwko, R., Xinggang, B., Clark, K., et al. (2005). "The Effects of Integrating Wind Power On Transmission System Planning, Reliability, and Operations, Report on Phase 2: System Performance Evaluation," Prepared for the New York State Energy Research and Development Authority, by General Electric's Power Systems Energy Consulting, Schenectady, NY.

Piwko, R., Clark, K., Freeman, L., Jordan, G., Miller, N., (2010). "Western Wind and Solar Integration Study," *NREL Subcontract Report.* Available at http://wind.nrel.gov/public/WWSIS/

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.

Söder, L., 2004, "Simulation of Wind Speed Forecast Errors for Operation Planning of Multi-Area Power Systems", 8th International Conference on Probabilistic Methods Applied to Power Systems, 12 - 16 September 2004, Iowa State University, United States.

Wan, Y., (2004). *Wind power plant behaviors: analyses of long-term wind power data*, National Renewable Energy Laboratory, Technical Report, NREL/TP-500-36551. Available: http://www.nrel.gov/docs/fy04osti/36551.pdf

Wan, Y., (2005). *A Primer on Wind Power for Utility Applications*, National Renewable Energy Laboratory, Technical Report, NREL/TP-500-36230, August. Available: http://www.nrel.gov/docs/fy05osti/36230.pdf

Wan, Y., (2009). *Summary Report of Wind Farm Data*, National Renewable Energy Laboritory, Technical Report, NREL/TP-500-4438, May. Available: http://www.nrel.gov/docs/fy09osti/44348.pdf

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

Wind Powering America: http://www.windpoweringamerica.gov

Wiser, R., Bolinger, M., (2009). “2008 Wind Technologies Market Report”, produced for the U. S. Department of Energy by NREL, DOE/GO-102009-2868, July. Available at http://www.windpoweringamerica.gov/pdfs/2008_annual_wind_market_report.pdf

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

Zavadil, R. (2006). “WAPA Wind Integration Study,” EnerNex Corporation, Knowville, TN.

## Appendix A: MATLAB Code for Ramp Identification Algorithm

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%            Ramp Analysis for GC 10-min Data
%%%%%%%
%%%%%%
%%%%%%%
%%%%%%          Creates Result array with:
%%%%%%%
%%%%%% [datenum|power|raw ramp rate|1st movAVG|movAVG ramps|2nd MovAVG|
%%
%%%%%%      2nd MovAVG ramps| Ramps as % capacity]
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%

 %requires iscons.m function

%clear
%count=[];
maeR=[];
maenR=[];
maeB=[];
maeG=[];
maeA1=[];
maeB1=[];
maeG1=[];
cR=[];
sdR=[];
sdnR=[];
sdG=[];
sdG1=[];
sdA1=[];
sdB1=[];
rmseR=[];
rmsenR=[];
Aerror=[];
tot_error=[];



%for q=2:1:28   %f=.1:.1:1=.1:.05:.75 %:

plant=63.7;           %specifies wind farm size (MW)
rate = 5;            %specifies desired MW/change in time ramp rate
mrate =21;            %specifies desired ramp rate of moving average
mrate2 = 0;           %specifies desired ramp rate of ending moving
average
duration = 5;         %specifies desired ramp duration to search for
capacity = .5;         %specifies ramp as percent of capacity
period=3;              % specifies length of time for ramp of size
'capacity' to occur
```

```
percent = 80;        %specifies ramp as percent of current production

bdur=2;
edur=2;
bdur2=1;
edur2=1;

%power = xlsread('U:\Thesis\Wind Thesis Mark B Jason K S-
Drive\Analysis\ten_min1');   %inputs production data for ramp
identification
%Alternately, save a matlab array and name it power...
%power = ten_min_new;

%dat=xlsread('S:\groups\RERC\Wind Thesis Mark B Jason K S-
Drive\Analysis\error_day');
dat=xlsread('I:\Documents\Thesis\Work from NREL\From NREL\Thesis\Wind
Thesis Mark B Jason K S-Drive\Analysis\error_4hr');
%syn_dat = xlsread('S:\groups\RERC\MBielecki\Thesis\Work from NREL\From
NREL\Thesis\synthetic_stats','sheet3','j1:j24989');
%dat=dat1;
power = [dat(:,1),dat(:,3)];%,dat(:,2),dat(:,4)];
%power=[dat(:,1),syn_dat(:,1)];


dates=[datenum('01-mar-2004 00:00:00'):1/144:datenum('31-mar-2004
23:59:59')]';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%       Compute moving average for ramp identification
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%

% a = 1;                          %starting coeff for standard diff
formula
% b=[1/3 1/3 1/3];        %uses current data point plus previous 4 to
avg
% movAVG = filter(b,a,power(:,2));
% %plot(movAVG(5:500,1));



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%   Generate basic array of ramps by subtracting previous
%%%%%%%
%%%%%%   hour prodution from current production.
%%%%%%%
%%%%%%   Also compute 1st and 2nd moving averages
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%

ramps = dat(2:length(dat),1);
```

```
ramps(:,2) = dat(2:end,3)-dat(1:end-1,3); %actual deltas
ramps(:,3) = dat(2:end,2)-dat(1:end-1,2); %fcast deltas

%compute 1st moving average
a = 1;                              %starting coeff for standard diff
formula
b=repmat(1/bdur,1,bdur);            %uses current data point plus previous
4 to avg
movAVG = filter(b,a,ramps(:,2));
movAVG=[ramps(:,1),movAVG];

%compute 2nd moving average
a = 1;                              %starting coeff for standard diff
formula
b=repmat(1/bdur2,1,bdur2);            %uses current data point plus
previous 4 to avg
movAVG2 = filter(b,a,ramps(:,2));
movAVG2=[ramps(:,1),movAVG2];

%compute 3rd moving average
a = 1;                              %starting coeff for standard diff
formula
b=repmat(1/edur2,1,edur2);            %uses current data point plus
previous 4 to avg
movAVG3 = filter(b,a,ramps(:,2));
movAVG3=[ramps(:,1),movAVG3];



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%   Plot power w raw ramps and 1st mov avg
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% figure
%
[AX,H1,H2]=plotyy(power(36:108,1),ramps(36:108,2),power(37:109,1),power
(37:109,2));%,'-o','markersize', 3,'linewidth',2)
% %datetick('x','mm-dd-yyyy','keeplimits');
% hold on
% set(get(AX(1),'Ylabel'),'String','Ramp Rate (MW/10-min)');
% set(get(AX(2),'Ylabel'),'String','Power Production (MW)');
% %set(H1,'Color','-bo');
% %set(H2,'Color','p');
% set(AX(1),'YColor','k');
% set(H1,'linewidth',2);
%
% %set(H1,'LineStyle','--');
% %hold on
% %set(H1,'LineStyle','o','markersize',3);
% hold on
% plot(ramps(36:108,1),movAVG(36:108,2),'-ro','markersize',
3,'linewidth',2);
% hold on
% %plot(power(36:108,2),power(36:108,3),'-ko','markersize',2)
% %xlim([ramps(36,1),ramps(108,1)]);
```

```
%
%
% %datetick('x','mm-dd-yyyy','keeplimits');
% title('10-min Ramp vs Moving Average Ramp');ylabel('ramp rate
(MW/10min)');
% legend('actual power','10-min ramp', 'moving average ramp (20 min)');
%



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%        To identify ramps by specified rate
%%%%%%%
%%%%%%        (> operator must be changed for down ramps)
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%

Rloc = [];      %establishes array with location of rate ramps
nRloc=[];       %location of down ramps

Q=[find(ramps(:,2) >= rate & ramps(:,2) <= plant)];  %location of up
ramps within ramps
Rloc=[ramps(Q,:)];           %array of up ramps
nQ=[find(ramps(:,2) <= -rate & ramps(:,2) >= -plant)];   %location of
down ramps within ramps
nRloc=[ramps(nQ,:)];         %array of down ramps



mRloc=[];
mnRloc=[];

mQ=[find(movAVG(:,2) >= mrate & movAVG(:,2) < plant)];    %location of
up ramps within movAvg
mRloc=[movAVG(mQ,:)];           %array of up ramps
mnQ=[find(movAVG(:,2) <= -mrate & movAVG(:,2) > -plant)];   %location
of down ramps within movAvg
mnRloc=[movAVG(mnQ,:)];         %array of down ramps


Result=[];
Result=[power(2:length(power),:),ramps(:,2),movAVG(:,2)];
Result(:,5)=0;
Result=[Result,movAVG2(:,2)];
Result(:,7:9)=0;
Result = [Result,movAVG3(:,2)];

%%%%%%%%%%%%%%%%%%%%%%%% Filter times when gaps in data
occur%%%%%%%%%%%%%%

t = round(Result(:,1)*24); %Remove 6 for hourly data
t(:,2) = iscons(t(:,1)); %requires iscons.m function
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%           To Plot power w/ all rates of change
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%


% figure
% plot(Result(1:150,2));hold on; plot(Result(1:150,3),'r');hold on;
plot(Result(1:150,4),'k');hold on; plot(Result(1:150,6),'g')
% legend('Power','10-min rate of change','20-min moving average','2nd
moving average');




%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%     Place 1 or 2 in Result col 5 & 7  for ramps that
%%%%%%%
%%%%%%     begin when a threshold (mrate) is exceeded and
%%%%%%%
%%%%%%     continue until another threshold (mrate2) is exceeded.
%%%%%%%
%%%%%%     Col 9 is a sum of col 5 and 7. Thus if there are
%%%%%%%
%%%%%%     consecutive occurences of 3, ramp rate may be poorly
%%%%%%%
%%%%%%     described here by this algorithm.
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%

j = 0;

%Loop to identify upramps which = 1 if true (in Col 5)


for k = 1:length(Result)-6
    if Result(k,4)>mrate && Result(k-bdur+bdur2,6)>mrate && Result(k-
1,4)<mrate && t(k,2)>0
        Result(k-bdur:k-bdur+bdur2,5) = 1;
        while Result(k+j-bdur+edur2+1,10)>mrate2;
            Result(k+j-bdur+2,5) =1; j=j+1;
            % Result(k+j-bdur+edur2,5) =1; j=j+1;
        end
        j=0;
    elseif Result(k,4)>mrate && Result(k-bdur+bdur2+1,6)>mrate &&
Result(k-1,4)<mrate && t(k,2)>0
        Result(k-bdur+1:k-bdur+bdur2+1,5) =1;
        while Result(k+j-bdur+edur2+2,10)>mrate2;
            Result(k+j-bdur+3,5) =1; j=j+1;
            %Result(k+j-bdur+edur2+1,5) =1; j=j+1;
```

```
        end
        j=0;
    elseif Result(k,4)>mrate && Result(k-bdur+bdur2+2,6)>mrate &&
Result(k-1,4)<mrate && t(k,2)>0
        Result(k-bdur+2:k-bdur+bdur2+2,5) =1;
        while Result(k+j-bdur+edur2+3,10)>mrate2;
            Result(k+j-bdur+4,5) =1; j=j+1;
            %Result(k+j-bdur+edur2+2,5) =1; j=j+1;
        end
        j=0;
    elseif Result(k,4)>mrate && Result(k-bdur+bdur2+3,6)>mrate &&
Result(k-1,4)<mrate && t(k,2)>0
        Result(k-bdur+3:k-bdur+bdur2+3,5) =1;
        while Result(k+j-bdur+edur2+4,10)>mrate2;
            Result(k+j-bdur+5,5) =1; j=j+1;
            %Result(k+j-bdur+edur2+3,5) =1; j=j+1;
        end
        j=0;
    elseif Result(k,4)>mrate && Result(k-bdur+bdur2+4,6)>mrate &&
Result(k-1,4)<mrate && t(k,2)>0
        Result(k-bdur+4:k-bdur+bdur2+4,5) =1;
        while Result(k+j-bdur+edur2+5,10)>mrate2;
            Result(k+j-bdur+6,5) =1; j=j+1;
            %Result(k+j-bdur+edur2+4,5) =1; j=j+1;
        end
        j=0;
    elseif Result(k,4)>mrate && Result(k-bdur+bdur2+5,6)>mrate &&
Result(k-1,4)<mrate && t(k,2)>0
        Result(k-bdur+5:k-bdur+bdur2+5,5) =1;
        while Result(k+j-bdur+edur2+6,10)>mrate2;
            Result(k+j-bdur+7,5) =1; j=j+1;
            %Result(k+j-bdur+edur2+5,5) =1; j=j+1;
        end
        j=0;
    end
end

%Repeat above for downramps which = 2 if true (in Col 7)
for k = 1:length(Result)-6
    if Result(k,4)<-mrate && Result(k-bdur+bdur2,6)<-mrate && Result(k-
1,4)>-mrate && t(k,2)>0
        Result(k-bdur:k-bdur+bdur2,7) = 2;
        while Result(k+j-bdur+edur2+1,10)<-mrate2;
            Result(k+j-bdur+2,7) =2; j=j+1;
            %Result(k+j-bdur+edur2,7) =2; j=j+1;
        end
        j=0;
    elseif Result(k,4)<-mrate && Result(k-bdur+bdur2+1,6)<-mrate &&
Result(k-1,4)>-mrate && t(k,2)>0
        Result(k-bdur+1:k-bdur+bdur2+1,7) =2;
        while Result(k+j-bdur+edur2+2,10)<-mrate2;
            Result(k+j-bdur+3,7) =2; j=j+1;
            %Result(k+j-bdur+bdur2+1,7) =2; j=j+1;
        end
        j=0;
    elseif Result(k,4)<-mrate && Result(k-bdur+bdur2+2,6)<-mrate &&
Result(k-1,4)>-mrate && t(k,2)>0
```

```
        Result(k-bdur+2:k-bdur+bdur2+2,7) =2;
        while Result(k+j-bdur+edur2+3,10)<-mrate2;
            Result(k+j-bdur+4,7) =2; j=j+1;
            %Result(k+j-bdur+bdur2+2,7) =2; j=j+1;
        end
        j=0;
    elseif Result(k,4)<-mrate && Result(k-bdur+bdur2+3,6)<-mrate &&
Result(k-1,4)>-mrate && t(k,2)>0
        Result(k-bdur+3:k-bdur+bdur2+3,7) =2;
        while Result(k+j-bdur+edur2+4,10)<-mrate2;
            Result(k+j-bdur+5,7) =2; j=j+1;
            %Result(k+j-bdur+bdur2+3,7) =2; j=j+1;
        end
        j=0;
    elseif Result(k,4)<-mrate && Result(k-bdur+bdur2+4,6)<-mrate &&
Result(k-1,4)>-mrate && t(k,2)>0
        Result(k-bdur+4:k-bdur+bdur2+4,7) =2;
        while Result(k+j-bdur+edur2+5,10)<-mrate2;
            Result(k+j-bdur+6,7) =2; j=j+1;
            %Result(k+j-bdur+bdur2+4,7) =2; j=j+1;
        end
        j=0;
    elseif Result(k,4)<-mrate && Result(k-bdur+bdur2+5,6)<-mrate &&
Result(k-1,4)>-mrate && t(k,2)>0
        Result(k-bdur+5:k-bdur+bdur2+5,7) =2;
        while Result(k+j-bdur+edur2+6,10)<-mrate2;
            Result(k+j-bdur+7,7) =2; j=j+1;
            %Result(k+j-bdur+bdur2+5,7) =2; j=j+1;
        end
        j=0;
    end
end

Result(:,9) = sum(Result(:,5)+Result(:,7),2);


%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
% %%%%%%        To identify ramps by specified duration
%%%%%%%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%
% Z=[];        %location of up ramps (to look for several in a row)
%
% Z=find(Result(:,7)>0);
%
% for i = 1:length(Z)-1
%     if Z(i+1)-Z(i)==1
%         Z(i,2)=1;
%         Z(i+1,2)=1;
%     else Z(i,2)=0;
%     end
% end
```

```
%
%
% nZ=[];        %location of down ramps (to look for several in a row)
% nZ=find(Result(:,7)<0);
%
% for i = 1:length(nZ)-1
%     if nZ(i+1)-nZ(i)==1
%         nZ(i,2)=1;
%         nZ(i+1,2)=1;
%     else nZ(i,2)=0;
%     end
% end



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%        To identify ramps by percentage of capacity
%%%%%%%
%%%%%%        Puts 1 in Result Col 8 for beginnin of ramp
%%%%%%%
%%%%%%        that exceeds % of capacity in 30 min
%%%%%%%
%%%%%%        Uses simple step change of size 'period'
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%


Result(:,8)=0;

% for k=1:length(Result(:,2))-(period+1)
%     if Result(k+period,2)-Result(k,2) >= capacity*plant  %&&
Result(k,2)>=0 && Result(k+1,2)>=0 && Result(k+2,2)>=0 &&
Result(k+3,2)>=0 && Result(k-1,2)>=0
%        Result(k:k+period,8)=1;
%     end
% end
%
% for k=1:length(Result(:,2))-(period+1)
%     if Result(k+period,2)-Result(k,2) <= -capacity*plant  %&&
Result(k,2)>=0 && Result(k+1,2)>=0 && Result(k+2,2)>=0 &&
Result(k+3,2)>=0 && Result(k-1,2)>=0
%        Result(k:k+period,8)=-1;
%     end
% end


for k=1:length(Result(:,2))-(1)
    if Result(k+1,2)-Result(k,2) >= capacity*plant  &&
round(Result(k+1,1)-Result(k,1))==0 %&& Result(k+1,2)>=0 &&
Result(k+2,2)>=0 && Result(k+3,2)>=0 && Result(k-1,2)>=0
       Result(k:k+1,8)=1;
    end
end
```

```
for k=1:length(Result(:,2))-(1)
    if Result(k+1,2)-Result(k,2) <= -capacity*plant  &&
round(Result(k+1,1)-Result(k,1))==0 %&& Result(k+1,2)>=0 &&
Result(k+2,2)>=0 && Result(k+3,2)>=0 && Result(k-1,2)>=0
        Result(k:k+1,8)=-1;
    end
end


Result(:,11)=dat(2:end,2);  %forecast
Result(:,12)=dat(2:end,4);  %bias

%Result(:,11)=power(2:end,3);
%Result(:,12)=power(2:end,4);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
%%%%%%              Exceedence and Frequency plots
%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%


% production=[-63.5:1:63.5];      %characterize all ramps
% production1=[-63:1:64]';
% T=[histc(ramps(:,2),production)];  %***** %hist of ramp rate
distribution
% Tnorm=[T/sum(T)];
% %Tnew=[1;Tnew(1:end-1)];
%
% % figure
% % %subplot(2,1,1)
% % plot(production1,Tnorm)
% % set(gca,'xlim',[-45,45]);
% % title('Ramp frequency as function of ramp rate');
% % xlabel('Ramp Rate (MW/hr)');
% % ylabel('Frequency');
% % grid on
% % %
% Tnew=[];
% Tnew(1,1)=Tnorm(1)/sum(Tnorm);
% for i = 2:length(Tnorm)
%     Tnew=[Tnew;Tnew(i-1)+(Tnorm(i)/sum(Tnorm))];
% end
% Tproduction=[-1*production];
%
%
% Tnew=1-Tnew;
% %
% %
% %
% % figure
% % %subplot(2,1,2)
```

```
% % plot(Tnew(1:128,:),production1(1:128,:))
% % hold on;
% %
plot(Tnew(find(production1(:,1)==0),1),production1(find(production1(:,1
)==0),1),'ro');
% % set(gca,'xlim',[0,1]);
% % set(gca,'ylim',[-65,65]);
% % title('Duration plot of actual deltas','fontsize',14);
% % xlabel('Frequency that hourly delta exceeds plotted
value','fontsize',14);
% % ylabel('Duration (ramp rate - MW/hr)','fontsize',14);
% % grid on
% % % %
% % figure;
% % bar(production',Tnorm)
% % set(gca,'xlim',[-20,20])
% % %set(gca,'xtick',-21:5:19)
% % %set(gca,'xticklabel',{'-20','-15','-10','-
5','0','5','10','15','20'})
% % title('Histogram of actual hourly deltas','fontsize',14);
% % xlabel('delta (ramp rate - MW/hr)','fontsize',14);
% % ylabel('Frequency','fontsize',14);
%
% %
% %
% % T=[T,production'];
% % Tnorm=[Tnorm,production'];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%


production=[-63.5:1:63.5];       %characterize all ramps
production1=[-63:1:64]';
production2=[0:64/20:64];
production3=[-63.5:5:63.5];
T=[histc(ramps(:,2),production)];  %***** %hist of act ramp rate
distribution
Tnorm=[T/sum(T)];
Tf=[histc(ramps(:,3),production)];  %***** %hist of fcst ramp rate
distribution
Tfnorm=[Tf/sum(Tf)];
J=[histc(dat(:,2),production2)]; %fcast power bins
Ja=[histc(dat(:,3),production2)]; %actual power bins
J=[J/sum(J)];
Ja=[Ja/sum(Ja)];
E=[histc(dat(:,4),production)];  %error power bins
Enorm=[E/sum(E)];


%Tnew=[1;Tnew(1:end-1)];

% figure
% %subplot(2,1,1)
% plot(production1,Tnorm)
% set(gca,'xlim',[-45,45]);
```

```
% title('Ramp frequency as function of ramp rate');
% xlabel('Ramp Rate (MW/hr)');
% ylabel('Frequency');
% grid on
% %
Tnew=[];
Tnew(1,1)=Tnorm(1)/sum(Tnorm);
for i = 2:length(Tnorm)
    Tnew=[Tnew;Tnew(i-1)+(Tnorm(i)/sum(Tnorm))];
end
Tproduction=[-1*production];
Tnew=1-Tnew;

Tfnew=[];
Tfnew(1,1)=Tfnorm(1)/sum(Tfnorm);
for i = 2:length(Tfnorm)
    Tfnew=[Tfnew;Tfnew(i-1)+(Tfnorm(i)/sum(Tfnorm))];
end
Tfproduction=[-1*production];
Tfnew=1-Tfnew;
%
%
%
figure
%subplot(2,1,2)
plot(Tnew(1:128,:),production1(1:128,:),'linewidth',2)
hold on;
plot(Tnew(find(production1(:,1)==0),1),production1(find(production1(:,1
)==0),1),'ro','markersize',8,'linewidth',2);
set(gca,'xlim',[0,1]);
set(gca,'ylim',[-65,65]);
title('Duration plot of actual deltas','fontsize',18);
xlabel('Frequency that hourly delta exceeds plotted
value','fontsize',18);
ylabel('Duration (ramp rate in MW/hr)','fontsize',18);
grid on

figure
%subplot(2,1,2)
plot(Tfnew(1:128,:),production1(1:128,:),'linewidth',2)
hold on;
plot(Tfnew(find(production1(:,1)==0),1),production1(find(production1(:,
1)==0),1),'ro','markersize',8,'linewidth',2);
set(gca,'xlim',[0,1]);
set(gca,'ylim',[-65,65]);
title('Duration plot of forecast deltas (DA)','fontsize',18);
xlabel('Frequency that hourly delta exceeds plotted
value','fontsize',18);
ylabel('Duration (ramp rate in MW/hr)','fontsize',18);
grid on


% %
figure;
bar(production',Tnorm)
set(gca,'xlim',[-20.5,20])
```

```
set(gca,'xtick',-20.5:5:19.5)
set(gca,'xticklabel',{'-20','-15','-10','-5','0','5','10','15','20'})
title('Histogram of actual hourly deltas','fontsize',18);
xlabel('Delta (ramp rate in MW/hr)','fontsize',18);
ylabel('Frequency','fontsize',18);

figure;
bar(production',Tfnorm)
set(gca,'xlim',[-20.5,20])
set(gca,'xtick',-20.5:5:19.5)
set(gca,'xticklabel',{'-20','-15','-10','-5','0','5','10','15','20'})
title('Histogram of forecast hourly deltas (DA)','fontsize',18);
xlabel('Delta (ramp rate in MW/hr)','fontsize',18);
ylabel('Frequency','fontsize',18);

figure;
bar([0:1/20:1],J);
set(gca,'xlim',[-.05,1])
title('Distribution of forecasted power (DA)','fontsize',18)
xlabel('Forecast power bin size (% capacity)','fontsize',18)
ylabel('Frequency','fontsize',18)

figure
bar([0:1/20:1],Ja);
set(gca,'xlim',[-.05,1])
title('Distribution of actual power','fontsize',18)
xlabel('Actual power bin size (% capacity)','fontsize',18)
ylabel('Frequency','fontsize',18)

figure;
bar(production',Enorm)
set(gca,'xlim',[-60,60])
title('Histogram of forecast error values (DA)','fontsize',18);
xlabel('Forecast error (MW)','fontsize',18);
ylabel('Frequency','fontsize',18);



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%          STATISTICS          %%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


Result(:,13)=0;
for i = 2:length(Result)-1
    if Result(i,5)==1 && Result(i-1,5)~=1
        Result(i,13)=1;
    elseif Result(i,5)==1 && Result(i+1,5)~=1
        Result(i,13)=2;
    elseif Result(i,7)==2 && Result(i-1,7)~=2
        Result(i,13)=3;
    elseif Result(i,7)==2 && Result(i+1,7)~=2
        Result(i,13)=4;
    end
end
```

```
x1=[find(Result(:,13)==1)];  %beg upramp
x2=[find(Result(:,13)==2)];  %end upramp
x3=[find(Result(:,13)==3)];  %beg downramp
x4=[find(Result(:,13)==4)];  %end downramp

x=[];A=[];A1=[];
x=[find(Result(:,5)==1)];       %picks out up ramps for stats
A=[Result(x,12)];
A1=[A;Result(x1-1,12);Result(x2+1,12)];  %grabs up ramps +- 1 hr


y=[];B=[];B1=[];
y=[find(Result(:,7)==2)];       %picks out down ramps for stats
B=[Result(y,12)];
B1=[B;Result(x3-1,12);Result(x4+1,12)];

z=[];C=[];                      %picks out ramps from capacity col
z=[find(Result(:,8)~=0)];
C=[Result(z,12)];

g=[]; G=[];                     %stats for all times other than ramps
g=[find(Result(:,9)==0)];
G=[Result(g,12)];

Get=[find(Result(g,12));find(Result(x1-
1,12));find(Result(x2+1,12));find(Result(x3-
1,12));find(Result(x4+1,12))];
G1=[Result(Get,12)];

 maeR=[maeR;mrate,mean(abs(A))];             %MAE of up ramps
 maenR=[maenR;mrate,mean(abs(B))];           %MAE of down ramps
 maeB=[maeB;mrate,mean(abs(Result(:,12)))];  %MAE of all error
 cR=[cR;capacity,mean(abs(C))];
 maeG=[maeG;mrate,mean(abs(G))];
 maeA1=[maeA1;mrate,mean(abs(A1))];
 maeB1=[maeB1;mrate,mean(abs(B1))];
 maeG1=[maeG1;mrate,mean(abs(G1))];
 mean_biasA=mean(A);    %mean bias of all upramps
 mean_biasB=mean(B);    %mean bias of all downramps
 mean_biasG=mean(G);    %mean bias of non-ramp times
 table=[mean_biasA;mean_biasB;mean_biasG];

 sdR=[sdR;mrate,std(A)];
 sdnR=[sdnR;mrate,std(B)];
 sdG=[sdG;mrate,std(G)];
 sdA1=[sdA1;mrate,std(A1)];
 sdB1=[sdB1;mrate,std(B1)];
 sdG1=[sdG1;mrate,std(G1)];

 Aa=[];
 for i=1:length(A)
     Aa(i)=[A(i)*A(i)];
 end
 Bb=[];
```

```
 for j=1:length(B)
     Bb(j)=[B(j)*B(j)];
 end
 rmseR=[rmseR;mrate,sqrt(mean(Aa))];
 rmsenR=[rmsenR;mrate,sqrt(mean(Bb))];


%end
Aerror=[maeR,maenR(:,2),maeG(:,2),sdR(:,2),sdnR(:,2),sdG(:,2)];
Aerror_plus=[maeA1,maeB1(:,2),maeG1(:,2),sdA1(:,2),sdB1(:,2),sdG1(:,2)]
;
%tot_error=[tot_error;error]
AMAEall=mean(abs(Result(:,12)));
Asdall=std(Result(:,12));

maeR11=maeR/plant;
maenR11=maenR/plant;
maeG11=maeG/plant;
maeB11=maeB/plant;

% figure;
% plot(maeR(:,1),maeR11(:,2),'b+:');
% hold on;
% plot(maenR(:,1),maenR11(:,2),'k+:');
% hold on;
% plot(maeG(:,1),maeG11(:,2),'r+:')
% hold on;
% plot(maeB(:,1),maeB11(:,2),'g');
%
%
% xlabel('ramp rate threshold (mrate value in MW/hr)','fontsize',14)
% ylabel('MAE (% capacity)','fontsize',14)
% title('MAE as percentage of capacity during ramp events of varying
threshold ramp rate values','fontsize',14);
% legend('During up ramp events','During down ramp events','During non-
ramp times','All times')
% set(gca, 'ylim', [0,.4])
%count=[count;length(find(Result(:,13)==1 | Result(:,13)==3))];




%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%       Correlation Analysis Between       %%%%%%%%%%%
%%%%%%%%%%%%%%           Actual and Forecast Ramps      %%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



%This program can be run on the array, Result, to find the number and
locations
%of upramps and downramps.

urows = find(Result(:,5));
uc = iscons(urows);
ustart = find(uc<1);
```

```
usrows = urows(ustart);
udates = datestr(Result(usrows,1));
ulength = ustart(2:end)-ustart(1:end-1);
ulength = [ulength;length(urows)-ustart(end)+1];

drows = find(Result(:,7));
dc = iscons(drows);
dstart = find(dc<1);
dsrows = drows(dstart);
ddates = datestr(Result(dsrows,1));
dlength = dstart(2:end)-dstart(1:end-1);
dlength = [dlength;length(drows)-dstart(end)+1];

%Additional code for second Result array comparison

urows2 = find(Result2(:,5));
uc2 = iscons(urows2);
ustart2 = find(uc2<1);
usrows2 = urows2(ustart2);
udates2 = datestr(Result2(usrows2,1));
ulength2 = ustart2(2:end)-ustart2(1:end-1);
ulength2 = [ulength2;length(urows2)-ustart2(end)+1];

drows2 = find(Result2(:,7));
dc2 = iscons(drows2);
dstart2 = find(dc2<1);
dsrows2 = drows2(dstart2);
ddates2 = datestr(Result2(dsrows2,1));
dlength2 = dstart2(2:end)-dstart2(1:end-1);
dlength2 = [dlength2;length(drows2)-dstart2(end)+1];

limit = 4; %limit of time periods to correlate nearby ramps.
unear=[];%(:,2) = usrows;
uscrew=[];%(:,2) = usrows;
for i = 1:length(usrows)
    x = find(usrows2<= usrows(i) + limit & usrows2>= usrows(i) -
limit);
    xx = find(dsrows2<= usrows(i) + limit & dsrows2>= usrows(i) -
limit);
    if isempty(x)
        unear=[unear;0,usrows(i)];
        if isempty(xx)
            uscrew = [uscrew;0,usrows(i)];
        else
            uscrew = [uscrew;xx,repmat(usrows(i),length(xx),1)];
        end
    else
        unear = [unear;x,repmat(usrows(i),length(x),1)];
        uscrew = [uscrew;0,usrows(i)];
    end
end
unear(:,3) = 99;
%unear(find(unear(:,1)),3) = usrows(find(unear(:,1)))-
usrows2(unear(find(unear(:,1)>0)));
unear(find(unear(:,1)),3) = unear(find(unear(:,1)),2) -
usrows2(unear(find(unear(:,1)>0)));
```

```
uscrew(:,3) = 99;
%uscrew(find(uscrew(:,1)),3) = usrows(find(uscrew(:,1)))-
dsrows2(uscrew(find(uscrew(:,1)>0)));
uscrew(find(uscrew(:,1)),3) = uscrew(find(uscrew(:,1)),2)-
dsrows2(uscrew(find(uscrew(:,1)>0)));


dnear=[];%(:,2) = dsrows;
dscrew=[];%(:,2) = dsrows;
for i = 1:length(dsrows)
    x = find(dsrows2<= dsrows(i) + limit & dsrows2>= dsrows(i) -
limit);
    xx = find(usrows2<= dsrows(i) + limit & usrows2>= dsrows(i) -
limit);
    if isempty(x)
        dnear = [dnear;0,dsrows(i)];
        if isempty(xx)
            dscrew = [dscrew;0,dsrows(i)];
        else
            dscrew = [dscrew;xx,repmat(dsrows(i),length(xx),1)];
        end
    else
        dnear = [dnear;x,repmat(dsrows(i),length(x),1)];
        dscrew = [dscrew;0,dsrows(i)];
    end
end
dnear(:,3) = 99;
%dnear(find(dnear(:,1)),3) = dsrows(find(dnear(:,1)))-
dsrows2(dnear(find(dnear(:,1)>0)));
dnear(find(dnear(:,1)),3) = dnear(find(dnear(:,1)),2)-
dsrows2(dnear(find(dnear(:,1)>0)));
dscrew(:,3) = 99;
%dscrew(find(dscrew(:,1)),3) = dsrows(find(dscrew(:,1)))-
usrows2(dscrew(find(dscrew(:,1)>0)));
dscrew(find(dscrew(:,1)),3) = dscrew(find(dscrew(:,1)),2)-
usrows2(dscrew(find(dscrew(:,1)>0)));


% dnear = [];
% for i = 1:length(dsrows)
%     x = find(dsrows2<= dsrows(i) + limit & dsrows2>= dsrows(i) -
limit);
% %     if isempty(x)
% %         dnear = [dnear;0];
%     if length(x) ==1
%         dnear(i,1) = x;
%     end
% end
% dnear(:,2) = dsrows;
% dnear(:,3) = 99;
% dnear(find(dnear(:,1)),3) = dsrows(find(dnear(:,1)))-
dsrows2(dnear(find(dnear(:,1)>0)));

%unear and dnear are the resulting matrices describing the correlation
%between Result and Result2. Column 1 tells us which corresponding row
in
```

```
%usrows2 had a true ramp start near (within limit) the true ramp start
for
%usrows (0 if false). Column 2 is usrows (the rows of the true ramp
start from Result.
%Column 3 is the number of time periods the ramp start at farm 1 lagged
the ramp
%start at farm 2 (99 if false). So a positive value indicates farm1
lags
%farm2 and negative means farm1 lags farm2.

farm1_lag_upramp_count = length(find(unear(:,3)>0 & unear(:,3)<99));
f1_lag_4up = length(find(unear(:,3)>3 & unear(:,3)<5)); %hour style
f1_lag_3up = length(find(unear(:,3)>2 & unear(:,3)<4)); %hour style
f1_lag_2up = length(find(unear(:,3)>1 & unear(:,3)<3)); %hour style
f1_lag_1up = length(find(unear(:,3)>0 & unear(:,3)<2)); %hour style
% f1_lag_2up = length(find(unear(:,3)>6 & unear(:,3)<13)); %10-min
style
% f1_lag_1up = length(find(unear(:,3)>0 & unear(:,3)<7)); %10-min style
tied_upramp_starts = length(find(unear(:,3)<1 & unear(:,3)>-1));
farm1_lead_upramp_count = length(find(unear(:,3)<0));
f1_lead_4up = length(find(unear(:,3)>-5 & unear(:,3)<-3)); %hour style
f1_lead_3up = length(find(unear(:,3)>-4 & unear(:,3)<-2)); %hour style
f1_lead_2up = length(find(unear(:,3)>-3 & unear(:,3)<-1)); %hour style
f1_lead_1up = length(find(unear(:,3)>-2 & unear(:,3)<-0)); %hour style
% f1_lead_2up = length(find(unear(:,3)>-13 & unear(:,3)<-6)); %10-min
style
% f1_lead_1up = length(find(unear(:,3)>-7 & unear(:,3)<0)); %10-min
style
uncorrelated_upramps = length(find(unear(:,1)<1 & unear(:,1)>-1));

farm1_lag_upramp_count_screw = length(find(uscrew(:,3)>0 &
uscrew(:,3)<99));
f1_lag_4up_screw = length(find(uscrew(:,3)>3 & uscrew(:,3)<5)); %hour
style
f1_lag_3up_screw = length(find(uscrew(:,3)>2 & uscrew(:,3)<4)); %hour
style
f1_lag_2up_screw = length(find(uscrew(:,3)>1 & uscrew(:,3)<3)); %hour
style
f1_lag_1up_screw = length(find(uscrew(:,3)>0 & uscrew(:,3)<2)); %hour
style
% f1_lag_2up_screw = length(find(uscrew(:,3)>6 & uscrew(:,3)<13)); %10-
min style
% f1_lag_1up_screw = length(find(uscrew(:,3)>0 & uscrew(:,3)<7)); %10-
min style
tied_upramp_starts_screw = length(find(uscrew(:,3)<1 & uscrew(:,3)>-
1));
farm1_lead_upramp_count_screw = length(find(uscrew(:,3)<0));
f1_lead_4up_screw = length(find(uscrew(:,3)>-5 & uscrew(:,3)<-3));
%hour style
f1_lead_3up_screw = length(find(uscrew(:,3)>-4 & uscrew(:,3)<-2));
%hour style
f1_lead_2up_screw = length(find(uscrew(:,3)>-3 & uscrew(:,3)<-1));
%hour style
f1_lead_1up_screw = length(find(uscrew(:,3)>-2 & uscrew(:,3)<-0));
%hour style
```

```
% f1_lead_2up_screw = length(find(uscrew(:,3)>-13 & uscrew(:,3)<-6));
%10-min style
% f1_lead_1up_screw = length(find(uscrew(:,3)>-7 & uscrew(:,3)<0));
%10-min style
uncorrelated_upramps_screw = length(find(uscrew(:,1)<1 & uscrew(:,1)>-
1));


farm1_lag_downramp_count = length(find(dnear(:,3)>0 & dnear(:,3)<99));
f1_lag_4down = length(find(dnear(:,3)>3 & dnear(:,3)<5)); %hour style
f1_lag_3down = length(find(dnear(:,3)>2 & dnear(:,3)<4)); %hour style
f1_lag_2down = length(find(dnear(:,3)>1 & dnear(:,3)<3)); %hour style
f1_lag_1down = length(find(dnear(:,3)>0 & dnear(:,3)<2)); %hour style
% f1_lag_2down = length(find(dnear(:,3)>6 & dnear(:,3)<13));%10-min
style
% f1_lag_1down = length(find(dnear(:,3)>0 & dnear(:,3)<7));%10-min
style
tied_downramp_starts = length(find(dnear(:,3)<1 & dnear(:,3)>-1));
farm1_lead_downramp_count = length(find(dnear(:,3)<0));
f1_lead_4down = length(find(dnear(:,3)>-5 & dnear(:,3)<-3)); %hour
style
f1_lead_3down = length(find(dnear(:,3)>-4 & dnear(:,3)<-2)); %hour
style
f1_lead_2down = length(find(dnear(:,3)>-3 & dnear(:,3)<-1)); %hour
style
f1_lead_1down = length(find(dnear(:,3)>-2 & dnear(:,3)<-0)); %hour
style
% f1_lead_2down = length(find(dnear(:,3)>-13 & dnear(:,3)<-6));%10-min
style
% f1_lead_1down = length(find(dnear(:,3)>-7 & dnear(:,3)<0));%10-min
style
uncorrelated_downramps = length(find(dnear(:,1)<1 & dnear(:,1)>-1));

farm1_lag_downramp_count_screw = length(find(dscrew(:,3)>0 &
dscrew(:,3)<99));
f1_lag_4down_screw = length(find(dscrew(:,3)>3 & dscrew(:,3)<5)); %hour
style
f1_lag_3down_screw = length(find(dscrew(:,3)>2 & dscrew(:,3)<4)); %hour
style
f1_lag_2down_screw = length(find(dscrew(:,3)>1 & dscrew(:,3)<3)); %hour
style
f1_lag_1down_screw = length(find(dscrew(:,3)>0 & dscrew(:,3)<2)); %hour
style
% f1_lag_2down_screw = length(find(dscrew(:,3)>6 &
dscrew(:,3)<13));%10-min style
% f1_lag_1down_screw = length(find(dscrew(:,3)>0 & dscrew(:,3)<7));%10-
min style
tied_downramp_starts_screw = length(find(dscrew(:,3)<1 & dscrew(:,3)>-
1));
farm1_lead_downramp_count_screw = length(find(dscrew(:,3)<0));
f1_lead_4down_screw = length(find(dscrew(:,3)>-5 & dscrew(:,3)<-3));
%hour style
f1_lead_3down_screw = length(find(dscrew(:,3)>-4 & dscrew(:,3)<-2));
%hour style
f1_lead_2down_screw = length(find(dscrew(:,3)>-3 & dscrew(:,3)<-1));
%hour style
```

```
f1_lead_1down_screw = length(find(dscrew(:,3)>-2 & dscrew(:,3)<-0));
%hour style
% f1_lead_2down_screw = length(find(dscrew(:,3)>-13 & dscrew(:,3)<-
6));%10-min style
% f1_lead_1down_screw = length(find(dscrew(:,3)>-7 &
dscrew(:,3)<0));%10-min style
uncorrelated_downramps_screw = length(find(dscrew(:,1)<1 &
dscrew(:,1)>-1));



Compare2 =
[length(ulength),mean(ulength),std(ulength);length(ulength2),mean(uleng
th2),std(ulength2);...

length(dlength),mean(dlength),std(dlength);length(dlength2),mean(dlengt
h2),std(dlength2)];

Compare3 = [length(unear)-
uncorrelated_upramps,f1_lag_4up,f1_lag_3up,f1_lag_2up,f1_lag_1up,...

tied_upramp_starts,f1_lead_1up,f1_lead_2up,f1_lead_3up,f1_lead_4up;...
    length(dnear)-
uncorrelated_downramps,f1_lag_4down,f1_lag_3down,f1_lag_2down,f1_lag_1d
own,...

tied_downramp_starts,f1_lead_1down,f1_lead_2down,f1_lead_3down,f1_lead_
4down];

Screw3 = [length(uscrew)-
uncorrelated_upramps_screw,f1_lag_4up_screw,f1_lag_3up_screw,f1_lag_2up
_screw,f1_lag_1up_screw,...

tied_upramp_starts_screw,f1_lead_1up_screw,f1_lead_2up_screw,f1_lead_3u
p_screw,f1_lead_4up_screw;...
    length(dscrew)-
uncorrelated_downramps_screw,f1_lag_4down_screw,f1_lag_3down_screw,f1_l
ag_2down_screw,f1_lag_1down_screw,...

tied_downramp_starts_screw,f1_lead_1down_screw,f1_lead_2down_screw,f1_l
ead_3down_screw,f1_lead_4down_screw];

compmat_up = [repmat(-4,Compare3(1,2),1);repmat(-
3,Compare3(1,3),1);repmat(-2,Compare3(1,4),1);...
    repmat(-
1,Compare3(1,5),1);repmat(0,Compare3(1,6),1);repmat(1,Compare3(1,7),1);
repmat(2,Compare3(1,8),1);...
    repmat(3,Compare3(1,9),1);repmat(4,Compare3(1,10),1)];

compmat_down = [repmat(-4,Compare3(2,2),1);repmat(-
3,Compare3(2,3),1);repmat(-2,Compare3(2,4),1);...
    repmat(-
1,Compare3(2,5),1);repmat(0,Compare3(2,6),1);repmat(1,Compare3(2,7),1);
repmat(2,Compare3(2,8),1);...
    repmat(3,Compare3(2,9),1);repmat(4,Compare3(2,10),1)];
% compmat_up = [repmat(-2,Compare3(1,2),1);repmat(-
1,Compare3(1,3),1);repmat(0,Compare3(1,4),1);...
```

```
%     repmat(1,Compare3(1,5),1);repmat(2,Compare3(1,6),1)];
% compmat_down = [repmat(-2,Compare3(2,2),1);repmat(-
1,Compare3(2,3),1);repmat(0,Compare3(2,4),1);...
%     repmat(1,Compare3(2,5),1);repmat(2,Compare3(2,6),1)];
screw_up = [repmat(-2,Screw3(1,2),1);repmat(-
1,Screw3(1,3),1);repmat(0,Screw3(1,4),1);...
    repmat(1,Screw3(1,5),1);repmat(2,Screw3(1,6),1)];

lag_bias_up = -mean(compmat_up(compmat_up<0));
lead_bias_up =-mean(compmat_up(compmat_up>0));
lag_sd_up = std(compmat_up(compmat_up<0));
lead_sd_up = std(compmat_up(compmat_up>0));
lag_bias_down = -mean(compmat_down(compmat_down<0));
lead_bias_down = -mean(compmat_down(compmat_down>0));
lag_sd_down = std(compmat_down(compmat_down<0));
lead_sd_down = std(compmat_down(compmat_down>0));
mean_bias_up = -mean(compmat_up);
sd_up = std(compmat_up);
mean_bias_down = -mean(compmat_down);
sd_down = std(compmat_down);

temp_bias_up = mean(unear(find(unear(:,3)<99),3));% *10;
temp_bias_down =  mean(dnear(find(dnear(:,3)<99),3));% *10;
temp_bias_up_screw = mean(uscrew(find(uscrew(:,3)<99),3));% *10;
temp_bias_down_screw =  mean(dscrew(find(dscrew(:,3)<99),3));% *10;

% Comparison =
[length(unear),uncorrelated_upramps,farm1_lag_upramp_count,mean_bias_up
,sd_up,lag_bias_up...
%
lag_sd_up,tied_upramp_starts,farm1_lead_upramp_count,lead_bias_up,lead_
sd_up;...
%
length(dnear),uncorrelated_downramps,farm1_lag_downramp_count,mean_bias
_down,sd_down,lag_bias_down...
%
lag_sd_down,tied_downramp_starts,farm1_lead_downramp_count,lead_bias_do
wn,lead_sd_down];
Comparison =
[length(ulength),uncorrelated_upramps,farm1_lag_upramp_count,lag_bias_u
p,...

lag_sd_up,tied_upramp_starts,mean_bias_up,sd_up,farm1_lead_upramp_count
,lead_bias_up,lead_sd_up;...

length(dlength),uncorrelated_downramps,farm1_lag_downramp_count,lag_bia
s_down,...

lag_sd_down,tied_downramp_starts,mean_bias_down,sd_down,farm1_lead_down
ramp_count,lead_bias_down,lead_sd_down];

Comparison_screw =
[temp_bias_up;temp_bias_down;temp_bias_up_screw;temp_bias_down_screw];

figure
bar3(Compare3(:,2:end))
```

```
xlabel('Load Lead(+) or Lag(-) Time (Hrs)')
set(gca,'XTickLabel',4:-1:-4)
set(gca,'YTickLabel',{'Up','Down'})
zlabel('# of Corresponding Ramp Events')
title('Actual Ramp Events Correlated with 7Hr Forecast Ramp Events (of
same direction) [2,1,1,11,0] [2,1,1,11,0]')

figure
bar3(Screw3(:,2:end))
xlabel('Load Lead(+) or Lag(-) Time (Hrs)')
set(gca,'XTickLabel',4:-1:-4)
set(gca,'YTickLabel',{'A Up F Down','A Down F Up'})
zlabel('# of Corresponding Ramp Events')
title('Actual Ramp Events Correlated with Opposing 7Hr Forecast Ramp
Events [2,1,1,11,0] [2,1,1,11,0]')
```