NREL
NATIONAL RENEWABLE ENERGY LABORATORY

# Wind Power Forecasting Error Distributions over Multiple Timescales

## Preprint

Bri-Mathias Hodge and Michael Milligan

*To be presented at the Power & Energy Society General Meeting*
*Detroit, Michigan*
*July 24-29, 2011*

**NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.**

**Conference Paper**
NREL/CP-5500-50614
March 2011

Contract No. DE-AC36-08GO28308

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information

P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

# Wind Power Forecasting Error Distributions over Multiple Timescales

Bri-Mathias Hodge, *Member, IEEE* and Michael Milligan, *Senior Member, IEEE*

*Abstract* — **Wind forecasting is an important consideration in integrating large amounts of wind power into the electricity grid. The wind power forecast error distribution assumed can have a large impact on the confidence intervals produced in wind power forecasting. In this work we examine the shape of the persistence model error distribution for ten different wind plants in the Electric Reliability Council of Texas (ERCOT) system over multiple timescales. Comparisons are made between the experimental distribution shape and that of the normal distribution. The shape of the distribution is found to change significantly with the length of the forecasting timescale. The Cauchy distribution is proposed as a model distribution for the forecast errors and model parameters are fitted. Finally, the differences in confidence intervals obtained using the Cauchy distribution and the normal distribution are compared.**

***Index Terms*—Wind power generation, wind energy, stochastic systems, error probability, forecasting**

## I. Introduction

With the increasing amounts of wind generation being added into the electricity system, the importance of being able to accurately forecast the wind power output for future time stages is also increasing. More accurate wind power forecasts can lead to economic efficiency in the unit commitment and dispatch process, as fewer reserves must be kept, or deployed, in order to compensate for changes in the wind power output. Although most wind power forecasts are currently given as point forecasts, a more useful approach involves producing interval forecasts. Interval forecasts can complement the point forecast by providing bounds on the expected future value with associated probabilities. This can be a very important consideration in the commitment and dispatching of generating units. For example, Fabbri et al. reported on the cost associated with supplemental regulation reserves required because of forecasting errors in the Spanish electricity market [1]. They found that the total annual cost of wind forecast errors for a single wind plant ranged from €15,000-18,000 per MW of installed capacity, depending on the forecasting time frame. They also observed that these costs decline with increasing installed capacity, demonstrating the benefits of geographical diversity in wind plant placement.

As a further illustration of the importance of wind forecasting errors, consider a system in which there is expected to be 1000 MW of total load with a point forecast of 100 MW of wind power for the next time period. Now we assume in this situation that the marginal producer is a 50 MW gas turbine producing at only 40% of capacity; that is 20 MW. Using only the point forecast we do not know whether the 30 MW of capacity available in the marginal gas turbine will be sufficient if the wind power forecast is inaccurate. If we use an interval forecast that specifies the 95% confidence lower bound for the wind power output to be 85 MW, the spinning reserve capacity in the marginal gas turbine should be sufficient to handle any up regulation. However, if the lower 95% confidence interval for the forecast is only 40 MW, the system operator would mostly likely bring more reserves online to handle a lower wind output situation.

The production of these forecast confidence intervals is often calculated using an assumed error distribution on the point forecast. The errors are often assumed to follow a normal distribution [2-5], though Weibull [6] and Beta [7] distributions have also been utilized. Lange studied the distribution of wind power forecast errors for timescales between 6 and 48 hours ahead with a particular focus on the additional errors created from converting wind speed forecasts created by numerical weather prediction to wind power output [8]. The study demonstrates that while the NWP errors are well represented by a Gaussian curve, the power forecast error distributions exhibit both skewness and excess kurtosis. Focken et al. investigated the smoothing of forecast errors for multiple wind plants at timescales from 6 to 48 hours ahead [9]. They determined that geographical diversity of wind plants and ensemble forecasts could reduce the forecast errors, with increased benefits for increased wind plant distance. Bludszuweit et al. examined the forecast error distribution of the persistence model and find that the distribution is "fat-tailed" and should not be modeled using the normal distribution [7]. They measure variable kurtosis values between different time scales (with a minimum of ten-minute averaged output data), and model the error distributions using a beta function. This function was then applied to the sizing of an energy storage system that will act to smooth wind power output.

Previous studies have examined the distribution of wind power forecast errors, but have tended to look at timescales of six hours or more. In this work we examine the forecast error distributions for a number of different timescales, ranging from one minute to three hours, demonstrating that the normal distribution is unsuitable to represent the error distributions at these smaller timescales. We then examine the differences in distribution shape with regard to timescale. A number of

The authors are with the National Renewable Energy Laboratory, Golden, CO 8041 USA (email: bri-mathias.hodge@nrel.gov, michael.milligan@nrel.gov)

different distribution types are then fitted to sample time series and a new distribution is found to outperform those reported in the literature. Finally this new distribution is applied to the problem of generating forecast error confidence intervals.

The remainder of the paper is organized as follows. In Section II the methods and data used in this study are detailed. Section III reports on the results of analyzing the forecast error over a number of timescales, and demonstrates the effectiveness of modeling the observed forecast errors with a Cauchy distribution. Conclusions are then drawn and future areas for examination outlined in Section IV.

## II. Methods and Data

In this section we describe some of the important methods utilized in the study. Section II-A contains information on the datasets analyzed. The persistence model is described in Section II-B, and Section II-C discusses the differences between a pair of methods for data aggregation. Background information on the statistical distributions discussed in the work is given in Section II-D.

### *A. Data Utilized*

In this study we have utilized data from the Electric Reliability Council of Texas (ERCOT) interconnection area in the United States. The dataset used is from the year 2009 and contains one minute average power output for ten different wind plants ranging in size from 30 MW to 215 MW. In addition, we have also included a series composed of the combined output of the ten wind plants with a total capacity of approximately 940 MW. This aggregated ERCOT-wide time series is very useful for helping to identify trends that may not be apparent in the individual wind plant data. Additionally, this time series is more representative of the patterns noticed at the Independent System Operator (ISO) level where the geographic diversity of a number of wind plants tends to smooth the data. Two main subsets of the data were used: a winter period and a summer period. The winter period consists of the months of January until April and the summer period is comprised of June until September.

### *B. Persistence Model Forecasting*

There have been a number of statistical techniques developed for the forecasting of time series data, but none are as widely applicable, and surprisingly effective, as the persistence model. While a simple approach, the persistence model is the baseline against which all other forecasting methods must be compared. Equation (1) shows the persistence model, where $\hat{P}$ is the forecast power output, $P(t)$ is the measured power output at time $t$, and $k$ is the forecast delay.

$$\hat{P}(t+k \mid t) = P(t) \qquad (1)$$

One issue with the use of the persistence model is the choice of the time delay between the observation and forecast. For longer forecast time periods, it is reasonable to assume that the previous point can be measured and relayed with sufficient time allowed to make the forecast for the next period. However, as the forecast period decreases the importance of this market closure delay increases. However, in this work we have assumed no delay in the forecast and thus every forecast is simply the output at the previous time point. It is important to note that the forecast errors have been standardized as a fraction of the wind plant installed capacity, creating forecast errors contained within the interval [-1,1].

### *C. Mean Values versus Point Values in Data Aggregation*

One of the goals of the current study is to examine the changes in forecast errors of the same data set over a number of different timescales. Since the basic wind power data is at the minute timescale, we need to aggregate the data set in order to examine forecast errors for five minute, fifteen minute, one hour and three hour time periods. There are two basic ways that one could turn the minute data into data a longer timescale; either using point values and simply removing values not coincident with the desired time steps, or use the mean of the values for the previous time period. As may be seen from Fig. 1 the point method produces Root Mean Squared Forecast Error (RMSFE) values that are substantially higher at all of the timescales under consideration for all of the individual wind plants as well as the aggregated output.

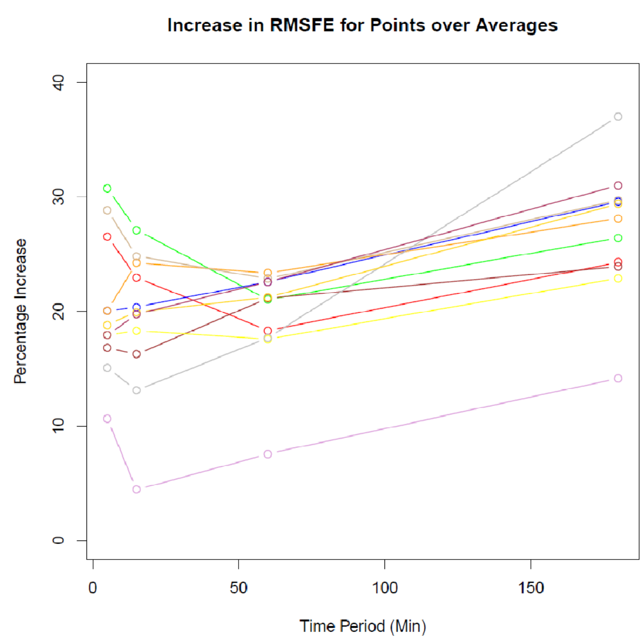


Fig. 1. Increase in RMSFE for point periods as a percentage of the RMSSFE found using the mean value period with the persistence model without a forecast delay. These values were obtained for the four month winter period.

Fig. 1 represents the values obtained using the persistence method with no time lag over a four month period starting at the beginning of the year. For this reason we have used the averaged minutely values over the previous time period in the subsequent analysis. For the sake of comparison the RMSFE values for each wind plant using the mean values at each time period length are shown in Fig. 2.

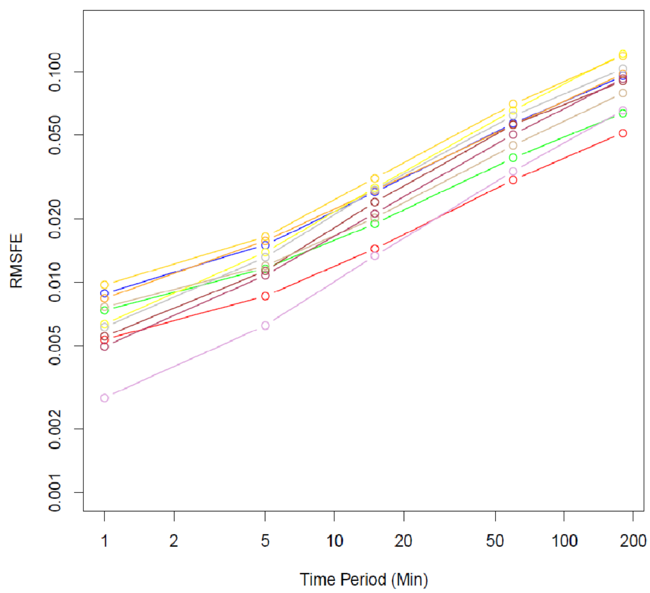


Fig. 2. RMSFE found using the mean values for each wind plant during the summer period. Note the log-log scale of the plot.

### D. Statistical Distributions

The probability density function is used to describe the range of values that a random variable can obtain, and the likelihood that a sample falls in a particular interval. In this work the primary focus is on two particular distributions: the normal (or Gaussian) distribution and the Cauchy-Lorentz Distribution. The beta distribution family, where the distribution can vary along the interval (0,1) based on the values of two shape parameters α and β, will also be discussed along with the Weibull distribution; which is a continuous distribution with two parameters relating to shape and scale, denoted by $k$ and λ respectively. Besides the standard deviation and mean, we will apply two other statistical measures to the characterization of the forecast error distributions. Skewness is a measure of the asymmetry of the probability distribution. This is demonstrated in Fig. 3 by the Weibull distribution that is slightly positively skewed. Skewness is the third standardized moment, as shown in (2).

$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{2}$$

where γ is the skewness, $X$ is the random variable, σ is the standard deviation and μ is the mean.

Kurtosis is a measure of the magnitude of the peak of the distribution, or conversely how fat-tailed the distribution is, and is the fourth standardized moment, defined in (3).

$$\kappa = \frac{E(\varepsilon^4)}{\sigma^4} \tag{3}$$

where κ is the kurtosis and ε is the normalized forecast error.

A distribution with a high kurtosis is known as leptokurtic and it describes a distribution where more of the variance is due to a lesser number of large deviations rather than the very frequent small deviations. The Cauchy distribution in Fig. 3 displays high kurtosis compared to the normal distribution, possessing a more pronounced peak, slimmer shoulders and longer tails. The difference between the kurtosis of a sample distribution and that of the normal distribution is known as the excess kurtosis. In the subsequent analysis the term kurtosis will be treated synonymously with excess kurtosis since we use the normal distribution as a baseline for the comparison of distributions.

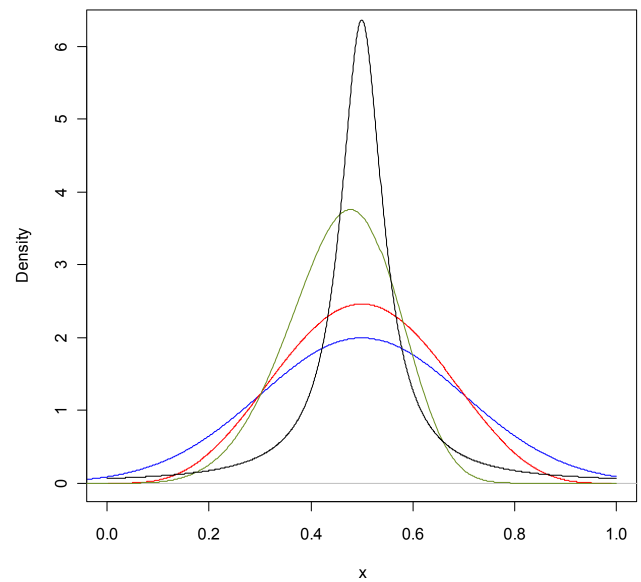


Fig. 3. Illustration of the different probability distributions on the x-interval [0,1]. Blue is the normal distribution with mean = 0.5 and standard deviation = 0.2. Red is the beta distribution with shape parameters α = β =5. The Weibull distribution is shown in green with $k$ = 5 and λ = 0.5. The Cauchy distribution is in black with $x_o$ = 0.5 and γ = 0.05.

## III. Results

Having established the importance of wind power forecasting error distributions and described the relevant statistical background, we now characterize the distributions for the datasets under consideration. In Sections III-A and III-B we examine the forecast error distribution shapes and use a pair of graphical techniques to compare the resulting distributions to the normal distribution. In Sections III-C and III-D we use a pair of numerical techniques to examine the forecast error distribution shapes. After ruling out the normal distribution as a valid model for the sample distributions, Section III-E endeavors to identify a more accurate model distribution. Finally, in Section III-F we apply the results of the new model to the problem of generating wind forecast error confidence intervals.

### A. Histograms

In order to examine the distributions of the persistence forecast errors, histograms of the data were first created and analyzed. The number of bins used exceeded the number needed according to Scott's rule [10] in all cases and a value of $n$ = 300 was found to work well over all of the timescales. An examination of the sample histograms provided in Fig. 4 and Fig. 5 show examples of leptokurtic distributions at two different timescales. The sample distributions have more pronounced peaks, steeper shoulders and fatter tails than the normal distributions generated using the sample means and standard deviations. Distributions are shown from both summer and winter four-month time periods.

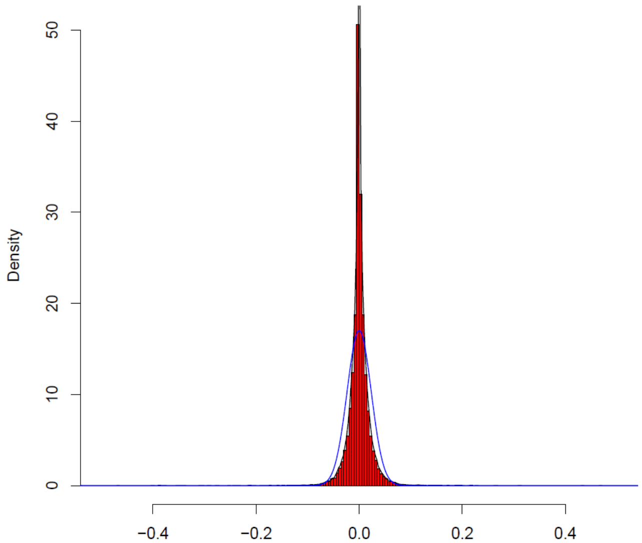


Fig. 4. Histogram of five minute persistence forecast error for wind plant #3 during the four month winter period. $\gamma$ = -2.14, $\kappa$ = 151.25. The blue line represents a normal distribution with the same mean and standard deviation. The black line is an approximate fit to the binned data.

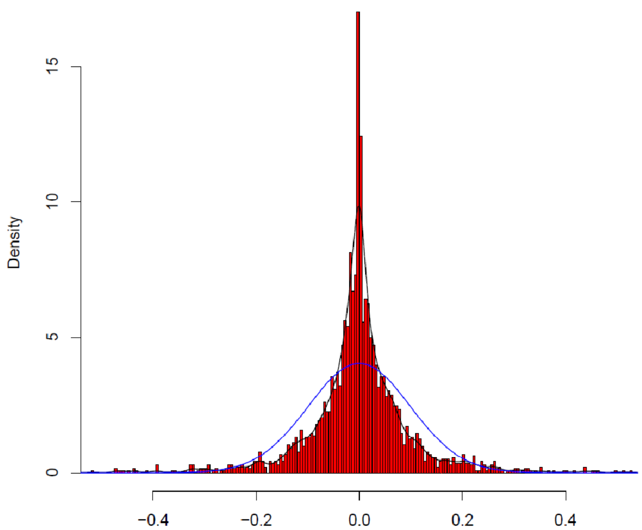


Fig. 5. Histogram of one hour persistence forecast error for wind plant #8 during the four month summer period. $\gamma$ = -0.02, $\kappa$ = 6.18. The blue line represents a normal distribution with the same mean and standard deviation. The black line is an approximate fit to the binned data.

*B. Normal Quantile-Quantile Plots*

The quantile-quantile (Q-Q) plot is a means by which two distributions can be graphically compared. Though the histograms shown in the previous section seem to indicate that the forecast error distributions are not normal over the timescales considered, the use of normal Q-Q plots can provide additional assurance. When the quantiles of the sampled quantity are compared with the quantiles of a normal distribution they should follow a linear pattern and approximately lie on the line y = x if the two distributions are similar.

The normal Q-Q plot in Fig. 6 shows the one minute forecast error distributions for wind plant #1. The non-normal nature of the distribution is immediately apparent from the non-linear nature of the quantile comparison, with significantly arched patterns suggesting a high degree of kurtosis, as confirmed by numerical calculation.

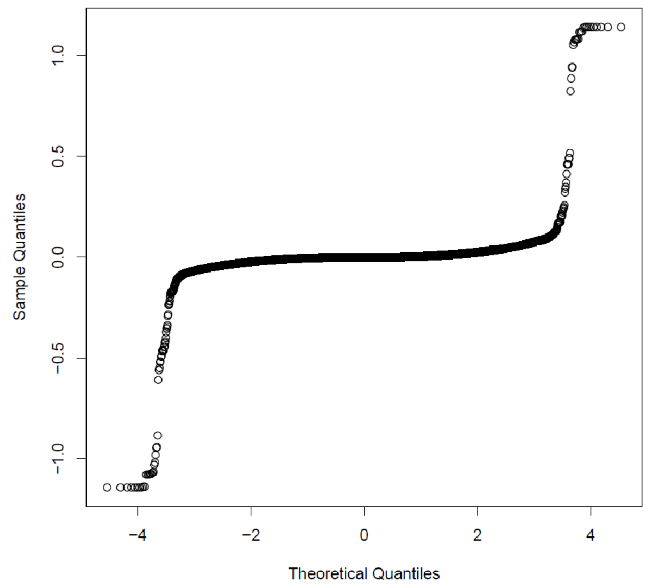


Fig. 6. Normal Quantile-Quantile plot for wind plant #1 – one minute persistence forecast error for the four month summer period. $\gamma$ = -0.45, $\kappa$ = 1717.81.

Fig. 7 shows a normal Q-Q plot of the one hour forecast error for wind plant #6. The plotted values are much more linear than those found in the previous plot, but still exhibit noticeable deviations. The distribution is also leptokurtic, as evidenced by the kurtosis value calculated.

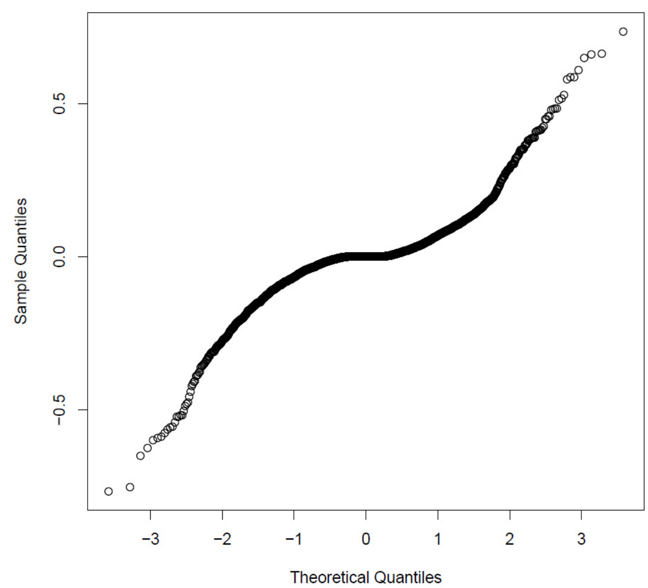


Fig. 7. Normal Quantile-Quantile plot for the wind plant #6 – one hour average persistence forecast error for the four month summer period. $\gamma$ = -0.08, $\kappa$ = 8.04.

Finally, Fig. 8 is an example of a normal Q-Q plot where the distribution is close to normal. The plotted points are very close to linear and the small kurtosis value confirms the visual assessment.

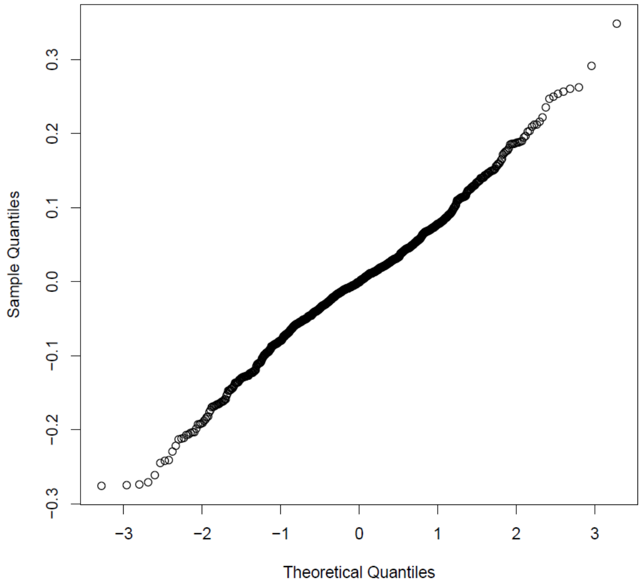


Fig. 8. Normal Quantile-Quantile plot for the combined ERCOT data – three hour average persistence forecast error for the four month summer period. $\gamma = 0.02$, $\kappa = 0.82$.

## C. *Kurtosis of the Forecast Errors*

One of the more interesting findings to come from examining the kurtosis values for the distribution of persistence forecasting errors over different timescales is the trend of decreasing kurtosis with larger timescales. As may be seen in Table I and Fig. 9, the kurtosis values have a strong decreasing trend with increasing timescale.

TABLE I
KURTOSIS VALUES FOR PERSISTENCE FORECAST ERROR WITH NO TIME LAG OVER A FOUR MONTH SUMMER PERIOD AT DIFFERENT TIME SCALES

| | 1 Minute | 5 Minute | 15 Minute | 1 Hour | 3 Hour |
|---|---|---|---|---|---|
| #1 | 1,717.89 | 328.43 | 163.59 | 39.50 | 25.07 |
| #2 | 210.06 | 62.61 | 34.09 | 8.24 | 3.42 |
| #3 | 46.61 | 42.17 | 29.05 | 9.33 | 2.34 |
| #4 | 122.11 | 51.94 | 30.63 | 7.61 | 2.23 |
| #5 | 595.13 | 69.88 | 31.87 | 9.30 | 2.65 |
| #6 | 900.03 | 137.24 | 34.84 | 8.04 | 2.16 |
| #7 | 79.91 | 27.79 | 19.69 | 5.16 | 1.72 |
| #8 | 50.72 | 51.31 | 28.08 | 6.18 | 1.13 |
| #9 | 119.07 | 37.23 | 18.41 | 6.23 | 1.99 |
| #10 | 328.18 | 81.80 | 38.93 | 5.14 | 1.39 |
| ERCOT | 149.12 | 16.32 | 7.85 | 2.51 | 0.82 |

It must be noted that the increasing timescale also corresponds to an increasing number of values used in the calculation of the time series values, due to the mean values used in the analysis. This trend can be interpreted to be a result of the central limit theorem because as the timescale increases so do the number of factors that may act on the wind power output values. This would also helps to explain why the aggregated ERCOT data's kurtosis values are far lower than simply the mean of the values for all of the individual wind plants in the summer data. However, the opposite effect is observed in the winter data, with the ERCOT forecast error distribution kurtosis values consistently larger than the mean values for the individual wind plants, as seen in the distinct purple line in Fig. 9.

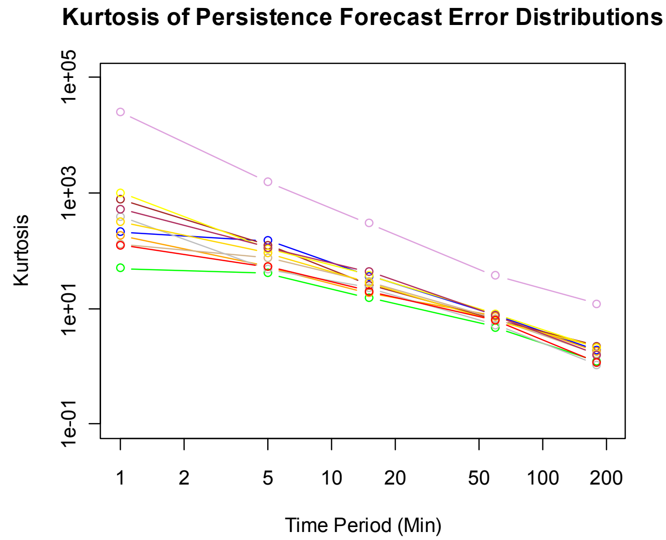


Fig. 9. Kurtosis values for persistence forecast error distributions for the winter time period. Note the log – log scale of the plot.

## D. *Skewness of the Forecast Errors*

In addition to the strength of the peak, the symmetry of the distributions is another very important aspect of the forecast error distribution characterization. The skewness values of the forecast error distributions for the various wind plants and timescales are shown in Table II. The trends in the data with respect to timescale are much weaker than in the kurtosis values; however, there is a weak trend toward more symmetric distributions with increasing time. In general, most of the skewness values are relatively minor, suggesting that the distribution of errors is close to symmetric. Similar to the kurtosis values for the summer period, the combined ERCOT data shows more Gaussian behavior than the individual wind plant mean skewness values alone would suggest.

TABLE II
SKEWNESS VALUES FOR PERSISTENCE FORECAST ERROR WITH NO TIME LAG OVER A FOUR MONTH SUMMER PERIOD AT DIFFERENT TIME SCALES

| | 1 Minute | 5 Minute | 15 Minute | 1 Hour | 3 Hour |
|---|---|---|---|---|---|
| #1 | -0.44 | -0.48 | -0.11 | 0.35 | 1.13 |
| #2 | -0.75 | -0.08 | 0.94 | 0.63 | 0.44 |
| #3 | 0.00 | -0.40 | 0.06 | 0.06 | 0.16 |
| #4 | -0.40 | 1.56 | 1.48 | 0.56 | 0.18 |
| #5 | 4.21 | 1.67 | 1.03 | 0.50 | 0.00 |
| #6 | -4.26 | -0.92 | 0.06 | -0.08 | -0.12 |
| #7 | -0.15 | 0.48 | 0.61 | 0.13 | -0.03 |
| #8 | 0.14 | -0.06 | -0.48 | -0.02 | 0.02 |
| #9 | -0.25 | -0.20 | 0.07 | -0.08 | -0.17 |
| #10 | -1.07 | 0.10 | 0.26 | -0.24 | 0.00 |
| ERCOT | 0.25 | 0.01 | -0.07 | 0.05 | 0.02 |

## E. *Distribution Fitting*

Having established that the normal distribution is a poor fit for the persistence forecast error distributions at the timescales under study, the next step in generating better forecasting intervals is to find a model that can more accurately represent the observed distributions. To accomplish this goal we have chosen three different model distributions that resemble those observed in the data. The Beta [7] and Weibull [6] distributions have been previously suggested in the literature, however, the Cauchy distribution appeared to be the most natural fit to the authors. In order to accommodate the fitting of the Beta and Weibull distributions, which are supported only on the intervals [0,1] and [0,∞) respectively, the forecast errors were converted from the [-1,1] interval to lie on the (0,1) interval. The distributions models were fit to the observed data using a maximum-likelihood optimization routine *fitdist* from the *MASS* package in the R statistical software environment [11].

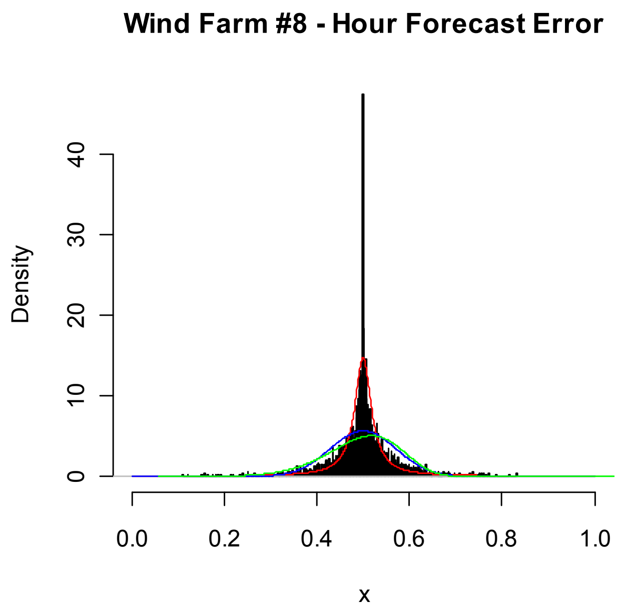


Fig. 10. Illustration of the fitted probability distributions on the x-interval [0,1] for wind plant #8 using hour average data for the winter time period. Blue is the beta distribution with shape parameters α = 25.2414 and β = 25.2692. The Weibull distribution is shown in green with $k$ = 7.1940 and λ = 0.5285. The Cauchy distribution is in red with $x_o$ = 0.4996 and γ = 0.0215.

The log-likelihood values for the fitted distributions indicate that the Cauchy distribution is a better fit for the forecast error distributions for a majority of the wind plants at all timescales. The Cauchy distribution fit was better than the Beta and Weibull distributions in 89% and 95% of the 55 cases, respectively. One such example is shown in Fig. 10, where the model distributions are compared to the histogram of the forecast error distribution for wind plant #8 at the hour timescale. In this case the Cauchy distribution represents a 20.24 % and 16.12% improvement over the Weibull and Beta distributions, respectively, in terms of the optimized log-likelihood values. For the ERCOT case at the 15 minute timescale the Cauchy distribution represents a 54.39 % and 36.19% improvement over the Weibull and Beta distributions, respectively, in terms of the optimized log-likelihood values.

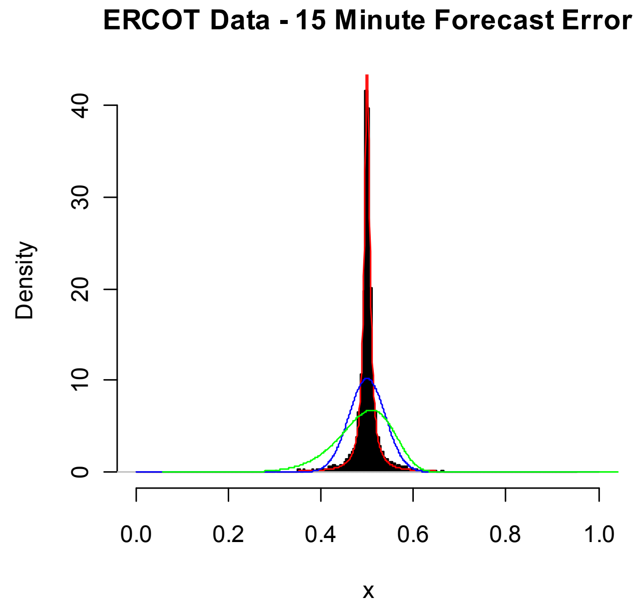


Fig. 11. Illustration of the fitted probability distributions on the x-interval [0,1] for combined ERCOT data using 15 minute average data for the winter time period. Blue is the beta distribution with shape parameters α = 82.0833 and β = 82.1398. The Weibull distribution is shown in green with $k$ = 9.4390 and λ = 0.5160. The Cauchy distribution is in red with $x_o$ = 0.4999 and γ = 0.0069.

### *F. Forecast Probability Intervals*

Thus far we have characterized the distribution of forecast errors for the persistence model over a number of timescales. We have also identified a distribution model that can represent the distribution of errors more accurately than the Gaussian distribution, the beta distribution and the Weibull distribution. But what are the practical implications of these findings on wind forecasting?

Fig. 12 shows the ERCOT data power forecast at 15 minute intervals along with 90% and 95% confidence intervals created using both a Cauchy distribution fit to the data and a normal distribution based on the data's mean and standard deviation values. The confidence intervals were created through 10,000 draws from each distribution. It is interesting to see that the Cauchy distribution produces a tighter interval for the 90% confidence values than the normal distribution. However, while the difference between the 90% and 95% confidence intervals for the normal distribution is relatively minor, there is a large difference for the Cauchy distribution. In fact, the 95% confidence interval for the Cauchy distribution is broader than for the normal distribution. This interesting result can be explained by examining the shapes of the respective distributions. While the Cauchy distribution produces more instances close to the mean than the normal distribution, it also has fat tails that indicate large deviations at the ends of the distribution, leading to narrow bounds for lower percentage confidence intervals and very wide bounds at high percentage confidence intervals. For the ERCOT 15 minute series shown in Figure 12, the 90% confidence interval band for the Cauchy distribution is 18.62% of total capacity, while for the normal distribution it is 25.32% of total capacity.

The 95% confidence interval spans 30.42% of the capacity for the normal distribution and 38.35% for the Cauchy distribution.

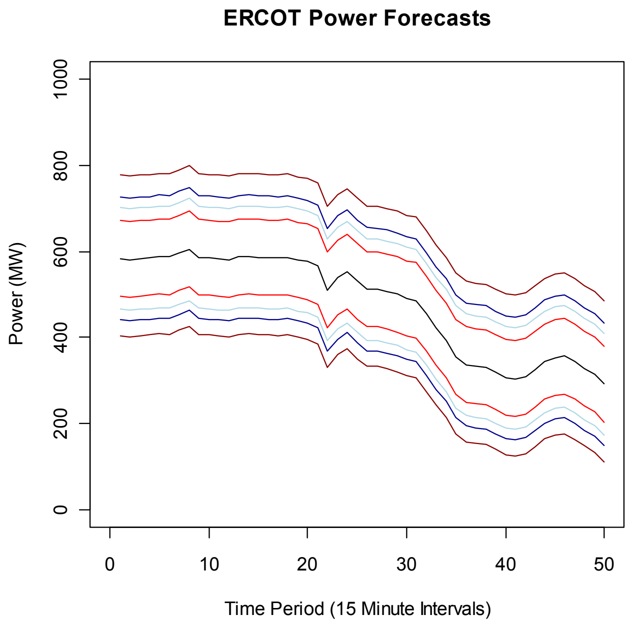


Fig.12. Plot of the wind power forecasts and associated forecast confidence intervals for the combined ERCOT output during a section of the winter period. The black line represents the power forecast. The normal distribution confidence intervals at 90% and 95% are shown in light blue and blue respectively. The Cauchy distribution confidence intervals at 90% and 95% are shown in red and dark red.

It is a slightly different story for the wind plant #5 hour series shown in Fig. 13, the result of a more platykurtic Cauchy distribution. The 90% confidence interval for the Cauchy distribution is nearly identical to the 95% confidence interval for the normal distribution. Once again the difference between the 90% and 95% confidence intervals for the Cauchy distribution is much wider than for the normal distribution due to the heavy tails. The Cauchy distribution 90% confidence interval is 36.86% of total capacity, while for the normal distribution it is 30.56% of total capacity. The 95% confidence interval for the normal distribution spans 36.39% of the capacity and 76.36% for the Cauchy distribution.

The results indicate that the normal distribution can overestimate the range of values for lower percentage confidence intervals, while underestimating the range for higher percentage confidence intervals. This difference can have important consequences in the determination of operating reserve levels necessary in systems with high wind penetration. More accurate modeling of the true wind power output confidence intervals with a Cauchy, or other, distribution can lead to economic savings through more efficient operating reserve allotment and operation. The confidence intervals chosen during operations may differ both by the balancing authority in question and the time of day. The current state of the system will guide the choice of confidence intervals and knowing the large marginal differences in the Cauchy distribution for confidence intervals above 90% can help inform the decision making process to ensure economic and reliable system operation.

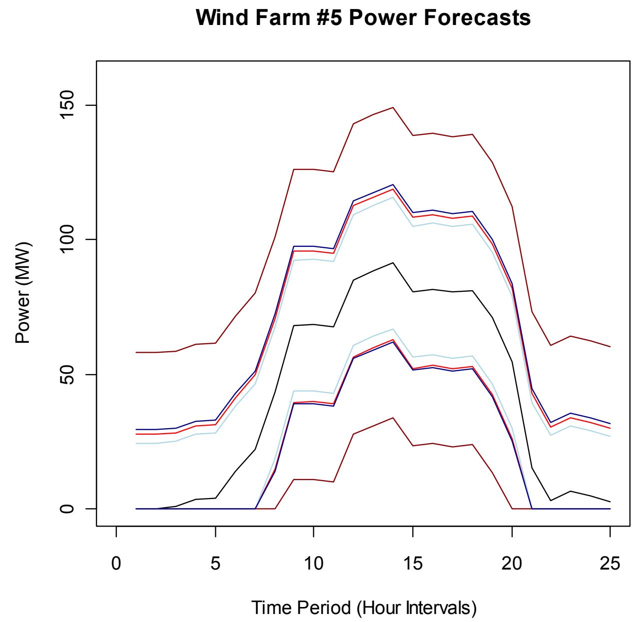


Fig. 13. Plot of the wind power forecasts and associated forecast confidence intervals for wind plant #5 output during a section of the winter period. The black line represents the power forecast. The normal distribution confidence intervals at 90% and 95% are shown in light blue and blue respectively. The Cauchy distribution confidence intervals at 90% and 95% are shown in red and dark red.

## IV. Conclusion

In this work we have examined the shape of wind power forecast error distributions through a statistical analysis. The distributions were found to differ greatly from the commonly assumed normal distribution, and the kurtosis of the distribution was found to vary with timescale. The Cauchy distribution has been proposed as a means of representing the forecast error distributions and was compared with some of the other distributions used in the literature. Finally, the effect that the new forecast error distribution model has on the calculation of wind power forecast confidence intervals was illustrated.

The current work suggests a number of directions for further examination. The analysis can be expanded to include more sources of wind power output data, as well as data at even smaller time scales. One possible caveat to the current work is that the wind power data sets are taken from the bus level and so they may include instances of wind curtailment that may skew the ends of the error distributions. The examination of the forecast error distributions at differing geographic scales would also be an interesting extension of the current work. The effect of forecasting lag within the persistence model is another area for further study. The examination of forecast error distributions for more sophisticated forecasting models, such as ARMA and ARIMA models, is also planned.

The current work also has applicability beyond the field of wind forecasting. The analysis found in this study could be used in the development of methods for the scaling down of wind power data. The quality of wind power output data is an important concern in wind integration studies. In cases where the wind data available is at a higher timescale than desired for the study, conditioning of the data is necessary in order to obtain data at the correct time scale. The findings from using the persistence model are especially relevant in this case because it represents what can be expected in the next time frame without any statistical transformation. The use of the Cauchy distribution to create data for the smaller scale time points in a scaling down effort could lead to the use of more representative wind power data in these studies.

## V. Acknowledgment

The authors gratefully acknowledge the contributions of Yih-huei Wan for supplying the data used in the study.

## VI. References


[1] A. Fabbri*, et al.*, "Assessment of the Cost Associated With Wind Generation Prediction Errors in a Liberalized Electricity Market," *IEEE Transactions on Power Systems,* vol. 20, August 2005.

[2] K. Methaprayoon*, et al.*, "An integration of ANN wind power estimation into UC considering the forecasting uncertainty," in *IEEE Industrial and Commercial Power Systems Technical Conference*, Saratoga Springs, NY, 2005.

[3] V. Pappala*, et al.*, "A Stochastic Model for the Optimal Operation of a Wind-Thermal Power System," *IEEE Transactions on Power Systems,* vol. 24, pp. 940 - 950, May 2009.

[4] E. Castronuovo and J. Lopes, "On the Optimization of the Daily Operation of a Wind-Hydro Power Plant," *IEEE Transactions on Power Systems,* vol. 19, pp. 1599 - 1606, August 2004.

[5] R. Doherty and M. O'Malley, "A New Approach to Quantify Reserve Demand in Systems With Significant Installed Wind Capacity," *IEEE Transactions on Power Systems,* vol. 20, pp. 587 - 595, May 2005.

[6] K. Dietrich*, et al.*, "Stochastic unit commitment considering uncertain wind production in an isolated system," in *4th Conference on Energy Economics and Technology*, Dresden, Germany, 2009.

[7] H. Bludzuweit*, et al.*, "Statistical Analysis of Wind Power Forecast Error," *IEEE Transactions on Power Systems,* vol. 23, pp. 983 - 991, August 2008.

[8] M. Lange, "On the Uncertainty of Wind Power Predictions - Analysis of the Forecast Accuracy and Statistical Distribution of Errors," *Journal of Solar Energy Engineering,* vol. 127, pp. 177-184, 2005.

[9] U. Focken*, et al.*, "Short-term prediction of the aggregated power output of wind farms - a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering and Industrial Aerodynamics,* vol. 90, pp. 231 - 246, 2002.

[10] D. Scott, "On optimal and data-based histograms," *Biometrika,* vol. 66, pp. 605-610, 1979.

[11] "R: A Language and Environment for Statistical Computing," ed. Vienna, Austria: R Foundation for Statistical Computing, 2010.


## VII. Biographies

**Bri-Mathias Hodge** (M'10) received the B.S. degree in chemical engineering from Carnegie Mellon University in 2004. He received a M.S. from the Process Design and Systems Engineering Laboratory of Åbo Akademi, Turku, Finland in 2005, and completed the Ph.D. in chemical engineering at Purdue University in 2010.

He is currently a post-doctoral researcher in the grid integration team for the National Renewable Energy Laboratory (NREL). His research interests include energy systems modeling, simulation and optimization.

**Michael Milligan** (M'98, SM'10) received the B.A. degree from Albion College, Albion, MI, and the M.A. and Ph.D. degrees from the University of Colorado, Boulder.

He is part of the grid integration team for the National Renewable Energy Laboratory (NREL), where he has worked on various aspects of electricity supply related to wind energy since 1992. He has worked on issues such as the ancillary service impacts of wind generation, the value of accurate wind forecasting, optimal selection of geographically dispersed wind power plants, modeling wind plant variability, and reliability contribution of wind power plants. He has authored or coauthored more than 90 reports and book chapters and has served on numerous technical review committees for wind integration studies around the United States.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** *(DD-MM-YYYY)*
March 2011

**2. REPORT TYPE**
Conference Paper

**3. DATES COVERED** *(From - To)*

**4. TITLE AND SUBTITLE**
Wind Power Forecasting Error Distributions over Multiple Timescales: Preprint

**5a. CONTRACT NUMBER**
DE-AC36-08GO28308

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
B.M. Hodge and M. Milligan

**5d. PROJECT NUMBER**
NREL/CP-5500-50614

**5e. TASK NUMBER**
WER9.5550

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401-3393

**8. PERFORMING ORGANIZATION REPORT NUMBER**
NREL/CP-5500-50614

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

**10. SPONSOR/MONITOR'S ACRONYM(S)**
NREL

**11. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**12. DISTRIBUTION AVAILABILITY STATEMENT**
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT** *(Maximum 200 Words)*
In this paper, we examine the shape of the persistence model error distribution for ten different wind plants in the ERCOT system over multiple timescales. Comparisons are made between the experimental distribution shape and that of the normal distribution.

**15. SUBJECT TERMS**
Wind power generation; wind energy; stochastic systems; error probability; forecasting

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL | | 19b. TELEPHONE NUMBER (*Include area code*) |

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39.18