An effective method for modeling wind power forecast uncertainty

Brandon Mauch*[,a,d], Jay Apt[a,b], Pedro M.S. Carvalho[d] and Mitchell J. Small[a,c]

[a]Department of Engineering and Public Policy Carnegie Mellon University; [b]Tepper School of Business, Carnegie Mellon University; [c]Department of Civil and Environmental Engineering, Carnegie Mellon University; [d]Department of Electrical and Computer Engineering, Instituto Superior Tecnico, Technical University of Lisbon

[*]Corresponding author, Email: bmauch@andrew.cmu.edu

Abstract

Wind forecasts are an important tool for electric system operators. Proper use of wind power forecasts to make operating decisions must account for the uncertainty associated with the forecast. Data from different regions in the USA with forecasts made by different vendors show the forecast error distribution is strongly dependent on the forecast level of wind power. At low wind forecast power, the forecasts tend to under-predict the actual wind power produced, whereas when the forecast is for high power, the forecast tends to over-predict the actual wind power. Most of the work in this field neglects the influence of wind forecast levels on wind forecast uncertainty and analyzes wind forecast errors as a whole. The few papers that account for this dependence, bin wind forecast data and fit parametric distributions to actual wind power in each bin. In the latter case, different parameters and possibly different distributions are estimated for each data bin. We present a method to model wind power forecast uncertainty as a single closed-form solution using a logit transformation of historical wind power forecast and actual wind power data. Once transformed, the data become close to jointly normally distributed. We show the process of calculating confidence intervals of wind power forecast errors using the jointly normally distributed logit transformed data. This method has the advantage of fitting the entire dataset with five parameters while also providing the ability to make calculations conditioned on the value of the wind power forecast.

Symbols

CI_α – Confidence interval at the level of α (0 to 100%)

cov( ) – Covariance; cov(A, B) = **E**[AB] – **E**[A]**E**[B], where **E** is the expectation value

e – Wind power forecast error

F – Wind power forecast

F^* – Logit transformed wind forecast

W – Actual wind power

W^* – Logit transformed actual wind power

Z_α – Number of standard deviations from mean covering α data in a normal distribution

f ( ) – Probability density function

ρ – Pearson's correlation coefficient

σ – Standard deviation

μ – Mean

1 Introduction

Wind power experienced substantial growth over the past decade in the U.S. and Europe. Installed capacity in the U.S. increased tenfold from 4.2 GW in 2001 to 47 GW in 2011 and now provides nearly 3% of total electrical energy [1]. European installed wind capacity increased fivefold in the same time period, from 17.3 GW to 94 GW, and produces over 6% of total electrical energy [2]. Since wind power is not fully dispatchable, wind forecasts are useful for planning and operations in electric power systems.

Electric system operators rely on wind power forecasts for decision making. For system operators, it is important not only that forecasts are accurate, but that the degree of inaccuracy is known. Operating reserves must be procured in advance to cover the uncertainty of wind power forecasts (and load forecasts). This not only applies to short term operation planning, but also to long term resource analysis. Here we present a model for wind power forecast uncertainty dependent on the wind power forecast based on historical wind forecast errors.

Broadly speaking, there are two approaches in the research literature to model wind forecast errors. One method models all wind power forecast errors as having been drawn from a single population with some known parametric probability distribution. Often a normal distribution is used under the assumption that the aggregation of many geographically diverse wind generators in a system justifies the application of the central limit theorem (e.g. [3], [4] and [5]). In other work, forecast errors were fit with a Cauchy distribution [6], a hyperbolic distribution [7], and a doubly truncated normal distribution [8].

Modeling all forecast errors with a single distribution makes the assumption that there is no dependence of forecast error on the forecast wind power level. Lange [9] transformed wind speed forecast errors into wind power forecast errors using the nonlinear relation of wind power to wind speed, which produced a distribution of power forecast errors more sharply peaked than the distribution of wind speed forecast errors. Power forecast errors were also shown to be skewed at the extreme values of the forecast range and symmetric near the center of the forecast range.

The other broad approach to modeling wind forecast errors is to condition forecast error distributions on the expected level of wind power. Neilsen et al. [10] determined wind power forecast confidence levels conditioned on the forecast wind power using quantile regression. In [11-13] wind data were binned by the wind power forecast values with Beta distributions fit to the observed wind power values associated with each bin of power forecasts. In [14] this method was extended to include the extreme value distribution. In each case, separate distribution parameters were required for each bin.

We present data showing that forecast error distributions for the largest USA wind regions are dependent on forecast wind values, and present a method to model wind forecast errors conditioned on the forecast value by applying a logit (or logistic) transformation to the wind forecast and actual wind power data. We find that the logit transformed variables can be reasonably modeled with a bivariate normal distribution, which is considerably easier to analyze than the original data. This method fits a single model to the entire set of data, meaning that only one set of parameters must be estimated, and produces smooth results without discontinuities that arise between bins.

Lau and McSharry [15] and Pinson [16] applied logit transformations to wind power time series to make them "more Gaussian" before fitting time series models for wind power forecasting. Both papers found that logit transformations of wind power data can be accurately modeled with normal distributions. They then fit statistical time series models to the transformed data to generate short term wind power forecasts from 1 to 48 hours. Here we utilize wind power forecasts independently generated by electric power system operators, show that these forecasts are also well described by the logit-normal model, and represent the observed and forecast wind power pairs as a bivariate logit-normal distribution in which the forecast error is implicitly included. This allows a characterization of the forecast error over the range of time scales for which these forecasts are provided.

This paper is organized as follows. In section 2 we briefly describe the data used in this study. Section 3 contains some discussion on the dependence of wind forecast uncertainty to wind forecast values. In section 4 we outline a method to model this uncertainty by applying a logit transformation to the wind forecast and wind power data. Results are shown for day-ahead wind power forecasts at the ISO level. We also present results using hour-ahead forecasts. Section 5 summarizes our conclusions.

2 Wind data used

We used wind forecast and actual wind power data from the Electric Reliability Council of Texas (ERCOT) and the Midwest Independent System Operator (MISO). ERCOT's territory covers most of the state of Texas while MISO covers most of the Midwestern portion of the U.S.

The ERCOT data included hourly wind forecast values for 1 to 48 hour look-ahead times covering the years 2009 and 2010. Included with the forecast data were hourly actual wind generation values and hourly estimates of what wind generation would have been if there were no curtailments. When analyzing wind power uncertainty, we used the estimated values of uncurtailed wind power since wind curtailments are not considered in wind power forecasts. ERCOT curtailed an estimated 17% of wind generation in 2009 and 10% of wind generation in 2010 [17, 18].

The estimates of uncurtailed hourly wind generation data were made for ERCOT by AWS Truepower based on actual wind generation, meteorological data and curtailment instructions sent from ERCOT to individual wind farms. The AWS analysis assumes that all curtailment instructions were followed and that wind turbine availability was known. In reality, it is not known how well wind farms followed curtailments instructions, and it was apparent that not all wind farms reported wind turbine availability status. Whenever our analysis below refers to "uncurtailed wind power" the AWS estimate is meant.

MISO day-ahead wind forecast and actual wind power data from February 2011 to May 2012 were obtained from the MISO website [19]. MISO wind forecasts are produced by Energy and Meteo GmbH [20]. During this period, wind curtailments in MISO were estimated to be 2 to 6% each month [21] which create minor problems with our analysis as will be shown in the next section. Further information on the data is presented in Appendix A.

3 Wind power forecast error characteristics

Wind power forecasts are provided for system operators of electric power networks in order to assist with decision making. Look-ahead times range from 5 minutes to several days. Wind forecasts for time periods up to 1 hour ahead provide information used in economic dispatch or real-time trading. Longer look-ahead times assist system operators in unit commitment decisions and provide wind farm operators with information for day-ahead market bids.

Decision makers who rely on wind power forecasts must not only prepare for the amount of wind power expected in the grid, they must also prepare for the chance that the forecast is wrong. Wind forecast uncertainty here is characterized by the distribution of wind forecast errors. We define wind forecast error (e) as the forecast value (F) minus the actual wind power (W) with all variables normalized by the installed wind capacity.

$$e = F - W \tag{1}$$

Modeling wind power uncertainty is challenging due both to the highly variable nature of wind speed over different time scales and to the non-linear and variable (for different turbines) transformation relating wind speed to wind power. This causes wind power forecast errors to be non-Gaussian in general and have much different distributions at low wind forecast values than high forecast values [9]. Figure 1 (a) shows a scatter plot of uncurtailed wind power levels plotted against wind power forecasts for ERCOT during the years 2009 and 2010. As indicated in the plot, points above the unit-slope line are under-forecast while points below the line are over-forecast. Forecast errors are plotted against forecast levels of wind power in Figure 1 (b). Since wind power is subtracted from forecasts, positive errors represent over-forecasts and negative errors represent under-forecasts. The three bottom plots in Figure 1 (c, d and e) show forecast error distributions at three different wind forecast levels. As evident in the plot, error distributions are skewed left near the low end of the forecast range and skewed right for high wind forecasts.

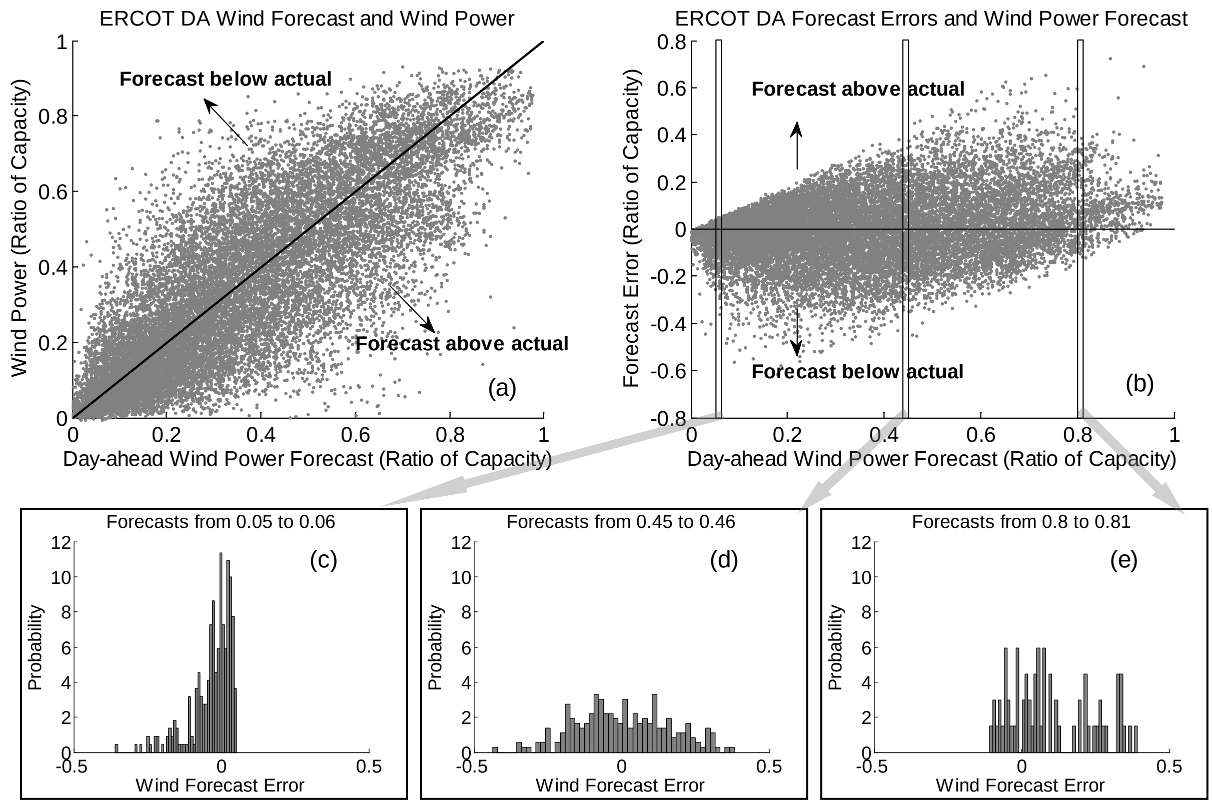

Figure 1: ERCOT Estimated uncurtailed wind power (a) and wind forecast errors (b) plotted against forecast wind power. Bottom plots show probability distributions of wind forecast errors corresponding to the three bins highlighted in the (b). Forecast value ranges in the bottom plots are (c) 0.05 to 0.06, (d) 0.45 to 0.46 and (e) 0.8 to 0.81. All values are shown as ratios of installed wind capacity.

Figure 2 shows the forecast bias calculated over the range of forecast values with a moving window of size 0.1 of the normalized wind power forecast. Unlike the work in [11 – 14], we did not attempt to remove the conditional forecast bias in our analysis. As a whole, the mean forecast error in the ERCOT data is 0.007 indicating the forecasts are unbiased, but when conditioned on the forecast value a bias emerges that is related to the forecast level of wind. When the predicted wind power is small, the actual power averages higher than the forecast, producing a negative bias. For high wind power predictions the actual wind power averages lower than forecast producing a positive bias. A similar bias pattern occurs in the MISO wind data and has been observed in wind forecasts for California [22] and Germany [12]. More discussion on the conditional bias is presented in Appendix C.

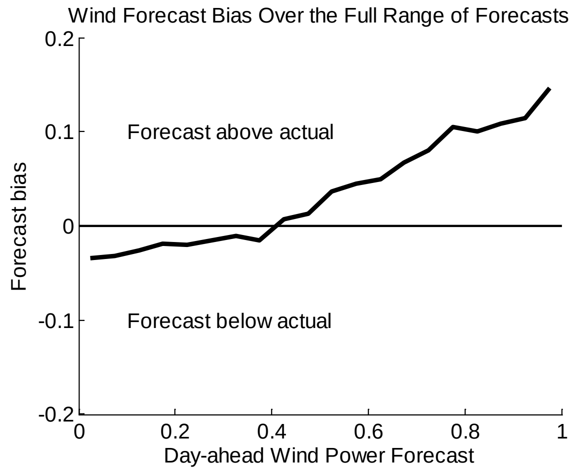

Figure 2: ERCOT forecast bias over the range of possible forecasts.

4 Logit transformation of wind data

4.1 Day-ahead wind power forecasts

We applied logit transforms to data in order to obtain a dataset that can be modeled with a normal distribution. Logit transforms are valid for data constrained between the values 0 and 1. Wind data normalized by installed wind capacity by definition must lie in the range from 0 to 1. Normalized wind data aggregated over an entire electric grid rarely register values equal to 0 or 1 making these data likely candidates for analysis with logit transforms. The underlying assumption in fitting logit transformed data to a normal distribution is that the forecast and wind power data fit a logit-normal distribution [23], as discussed in greater detail in Appendix B. If we define normalized wind forecast and wind power data as F and W, the transformed variables, F^* and W^* are defined in Equation 2. Once transformed, the variables can take values ranging from negative to positive infinity.

$$F^* = \ln\left(\frac{F}{1-F}\right) \qquad W^* = \ln\left(\frac{W}{1-W}\right) \tag{2}$$

The transformed variables were fit to a normal distribution with a density function

$$f(X^*) = \frac{1}{\sqrt{2\pi}\sigma_{X^*}} exp\left(-\frac{(X^*-\mu_{X^*})^2}{2\sigma_{X^*}^2}\right) \tag{3}$$

The symbols μ_{X^*} and σ_{X^*} are the mean and standard deviation of the transformed data. Figure 3 shows the distributions of F^* and W^* with a fitted normal distribution overlaid. As the graphs show, the normal distribution fits the transformed data well.

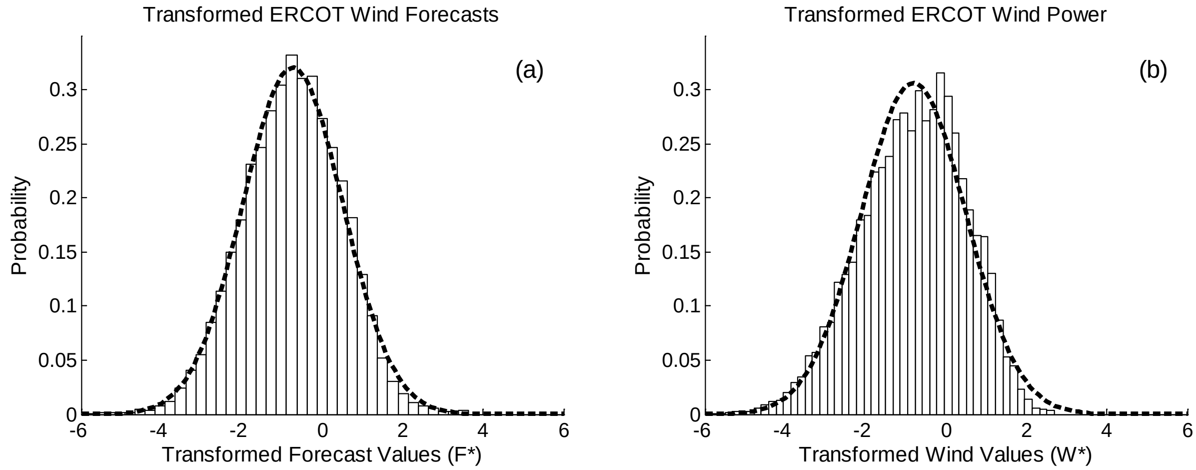

Figure 3: Relative frequency distributions (histogram bars) for the transformed wind power forecast (a) and wind power (b) in ERCOT with a fitted normal distribution (dashed line) overlaid. Analogous MISO data are shown in Figure 10.

The transformed variables are plotted against each other in Figure 4. The solid lines represent the contour of a bivariate normal distribution fit to the data.

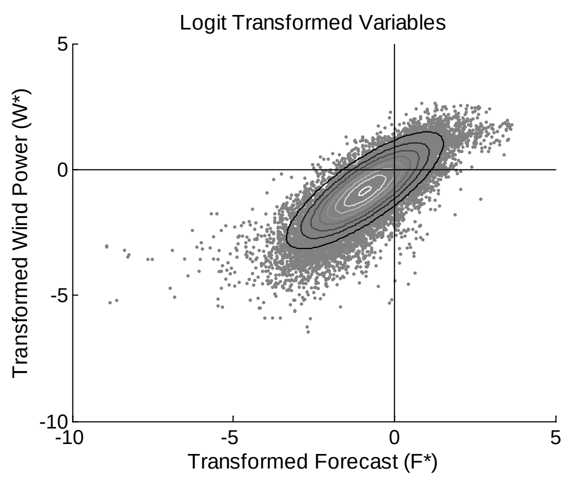

Figure 4: Logit transformed wind power data plotted against transformed wind forecast data. The data shown here are transformations of the data displayed in Figure 1 (a). Solid lines are contours of a fitted bivariate normal distribution with μ_{F^*} = -0.74, μ_{W^*} = -0.81, σ_{F^*} = 1.55, σ_{W^*} = 1.70 and ρ = 0.80

We modeled the transformed variables as jointly normally distributed with the bivariate normal distribution:

$$f(F^*,W^*) = \frac{1}{2\pi\sigma_{F^*}\sigma_{W^*}\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(F^*-\mu_{F^*})^2}{\sigma_{F^*}^2} + \frac{(W^*-\mu_{W^*})^2}{\sigma_{W^*}^2} - \frac{2\rho(F^*-\mu_{F^*})(W^*-\mu_{W^*})}{\sigma_{F^*}\sigma_{W^*}}\right]\right) \quad (4)$$

where $$\rho = \frac{cov(W^*, F^*)}{\sigma_{W^*}\sigma_{F^*}} \tag{5}$$

and cov(,) is the covariance.

Using the fitted normal distribution provides a convenient way to model wind forecast errors. For a particular forecast value, the transformed wind power is modeled with a normal distribution with the conditional probability density defined by.

$$f(W^*|F^*) = \frac{1}{\sqrt{2\pi}\sigma_{W^*|F^*}} exp\left(-\frac{(W^* - \mu_{W^*|F^*})^2}{2\sigma^2_{W^*|F^*}}\right) \tag{6}$$

where $$\mu_{W^*|F^*} = \mu_{W^*} + \frac{\rho\sigma_{W^*}}{\sigma_{F^*}}(F^* - \mu_{F^*}) \tag{7}$$

$$\sigma_{W^*|F^*} = \sigma_{W^*}\sqrt{1-\rho^2} \tag{8}$$

Given a wind power forecast value, a decision maker is interested in knowing a particular confidence interval for the expected wind power. A confidence interval of 95% ranges from the 2.5th percentile to the 97.5th percentile calculated from the inverse of the cumulative distribution function (*CDF*). In the case of a normal distribution, this calculation simplifies to a function of the conditional mean and standard deviation. The end values of a given confidence interval are calculated by

$$CI_\alpha\{W^*|F^*\} = \left[\mathbf{CDF}^{-1}_{W^*|F^*}\left(\frac{1-\alpha}{2}\right), \quad \mathbf{CDF}^{-1}_{W^*|F^*}\left(\frac{1+\alpha}{2}\right)\right] \tag{9}$$

$$= [\mu_{W^*|F^*} - Z_\alpha\, \sigma_{W^*|F^*}, \quad \mu_{W^*|F^*} + Z_\alpha\, \sigma_{W^*|F^*}] \tag{10}$$

The symbol α ranges between 0 and 1 and indicates the desired confidence interval level (i.e. 0.95 for a 95% confidence interval). The value of Z_α in Equation 10 is selected from standard normal distribution tables to produce a particular confidence interval, CI_α. Table 2 shows selected values of Z_α for desired confidence intervals.

Table 1: Z values for selected confidence intervals (CI).

α	70%	80%	85%	90%	95%
Z_α	1.047	1.282	1.440	1.645	1.960

Figure 5 shows the 95% normal distribution confidence interval of actual wind power in the transformed space conditioned on the forecast wind power. The solid line is the mean of W^* as a function of F^*, and the region within the dashed lines represents 1.96 standard deviations from the mean which contains 95% of the W^* values assuming a bivariate normal distribution.

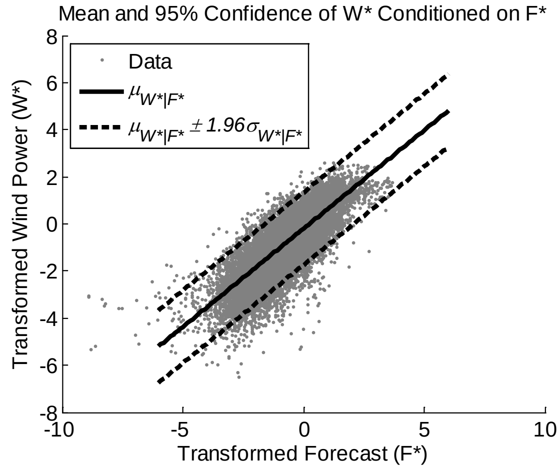

Figure 5: Transformed ERCOT data with conditional mean for W* and 95% confidence interval of W* as a function of F*.

A confidence interval in the transformed space is converted to the confidence interval in the original data space using Equation 11 for each variable. The result is displayed as solid lines in Figure 6.

$$F = \frac{1}{1+e^{-F^*}} \qquad\qquad W = \frac{1}{1+e^{-W^*}} \tag{11}$$

Once a confidence interval is determined for the observed wind power given a wind forecast, we can determine a confidence level for wind forecast errors by subtracting the confidence interval values of wind power from the wind forecast values. Figure7 shows a range of confidence intervals for actual wind power based on the day-ahead wind forecast and the confidence levels of wind forecast errors. As shown in Figure 7 (b), the confidence intervals are generally asymmetric and not centered on zero.

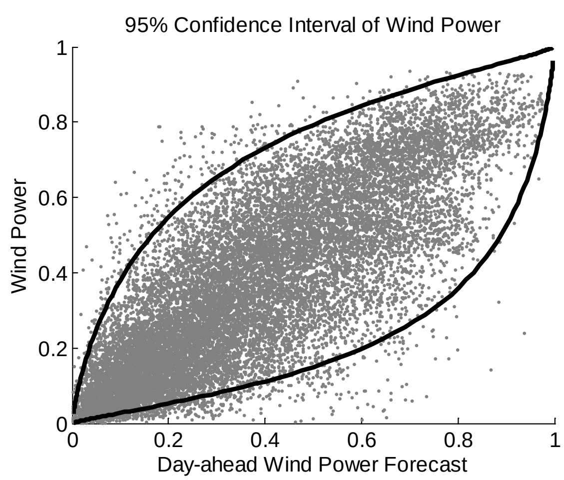

Figure 6: Ninety-five percent confidence interval for estimated uncurtailed ERCOT wind power in the original space plotted as a function of the day-ahead forecast level of wind power.

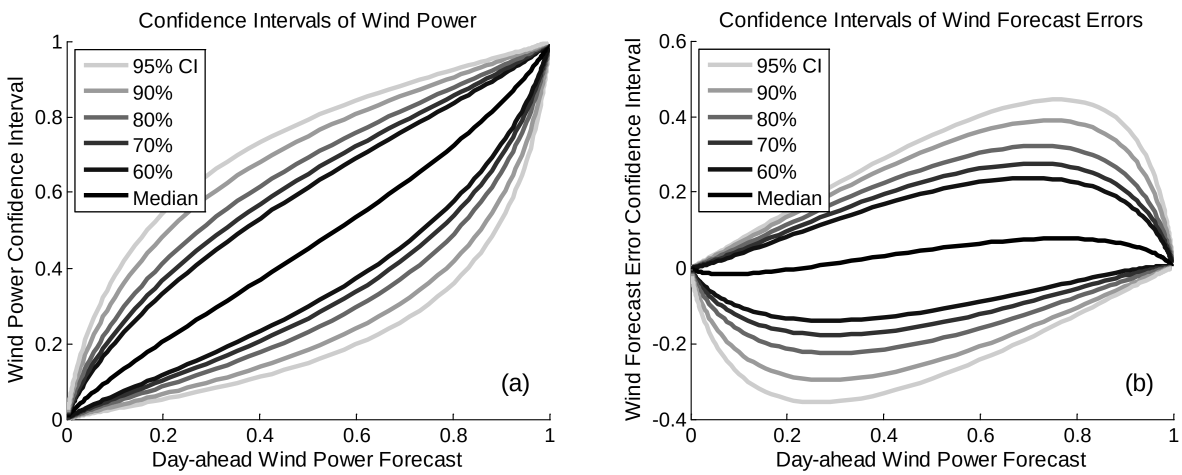

Figure 7: ERCOT wind power confidence levels (left plot) based on wind forecast level and wind error confidence levels (right plot) based on wind forecast level.

We plotted 95, 90 and 80% confidence intervals of wind forecast errors using the logit transform and compared these to the same confidence intervals calculated with a moving window of width 0.1 in Figure 8. Fitting a bivariate normal distribution to the logit transform provides a good closed-form representation of the wind forecast errors requiring a fit of only 5 parameters; μ_{F^*}, μ_{W^*}, σ_{F^*}, σ_{W^*} and ρ. In comparison, fitting a Beta distribution to 50 bins of wind data requires the estimation of 100 parameters (2 parameters for each bin) as was done in [11].

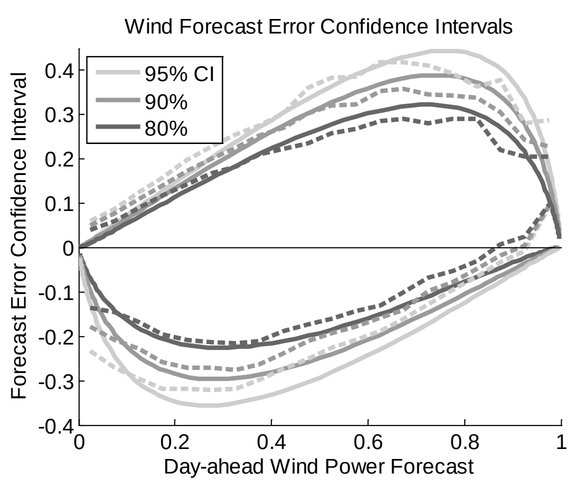

Figure 8: Confidence intervals conditioned on ERCOT wind forecast level. Solid lines are calculated using the logit transformation, and dashed lines are calculated by binning data according to the forecasts.

Figures 2-8 were created using day-ahead wind forecasts and estimated uncurtailed wind power data from ERCOT. We also modeled wind forecast errors with MISO wind data. The MISO data did not contain wind curtailment estimates so we used observed wind power with wind forecasts. As mentioned earlier, curtailments in MISO ranged from 2 to 6% of total wind energy generated, depending on the month. We feel that these levels of wind curtailments are low enough to provide meaningful results. A system operator analyzing wind forecast errors would likely have uncurtailed wind generation estimates available, and would be able to model the forecast errors more accurately.

Figure 9 (a) shows the actual wind power against the day-ahead forecast wind power for each hour of the sample period with all data normalized by the wind capacity. Note that wind power levels in MISO rarely reach 80% of installed capacity. This may be due to the fact that wind farms in MISO are distributed over a much wider and more geographically diverse region than ERCOT wind farms. Figure 9 (b) displays the same data after applying the logit transformation.

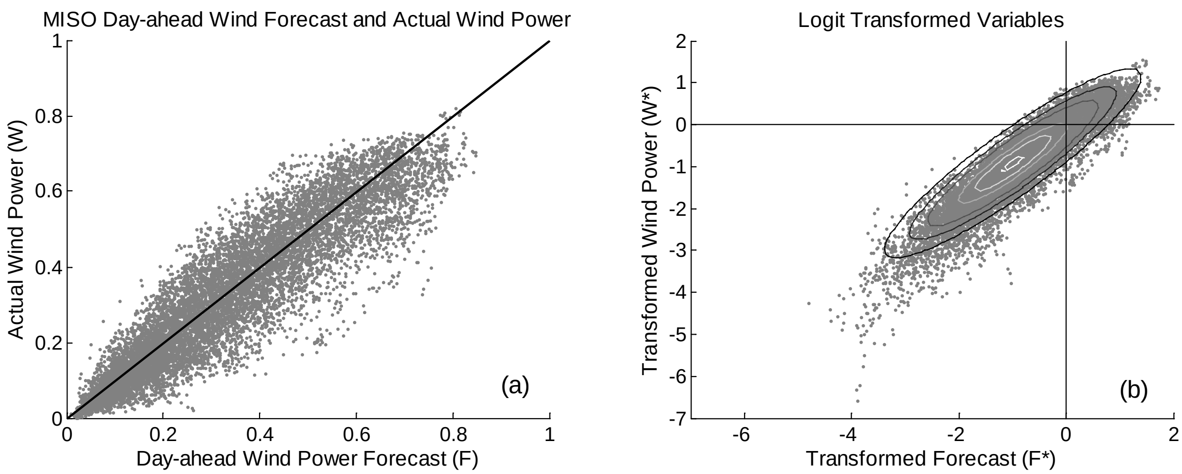

Figure 9: (a) Actual wind power in MISO plotted against the day-ahead forecast values and (b) the logit transformation of the same data. A contour (solid lines) of the fitted bivariate normal distribution is plotted over the data in (b) with μ_{F*} = -0.82, μ_{W*} = -0.89, σ_{F*} = 1.05, σ_{W*} = 1.18 and ρ = 0.92

In order to evaluate how normally distributed the transformed data are, the relative frequency distributions of the transformed data are displayed in Figure 10.

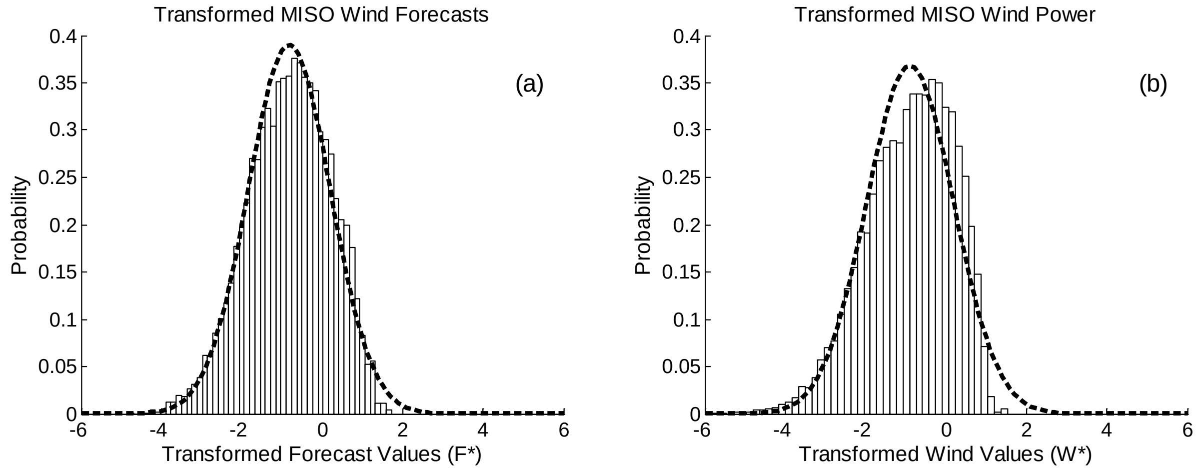

Figure 10: Relative frequency distributions (histogram bars) for the transformed wind power forecast (a) and wind power (b) in ERCOT with a fitted normal distribution (dashed line) overlaid.

The transformed actual wind power (W^*) distribution is skewed left. One likely reason for this skewness is the effect of wind curtailments on the data. Since wind power in MISO rarely exceeds 80% of installed capacity, verification of confidence intervals was impossible for forecasts near a normalized value of 1. Figure 11 shows the 95% confidence envelope for MISO wind errors calculated with the logit transformation (solid lines) and a moving window of width 0.1 (dashed lines).

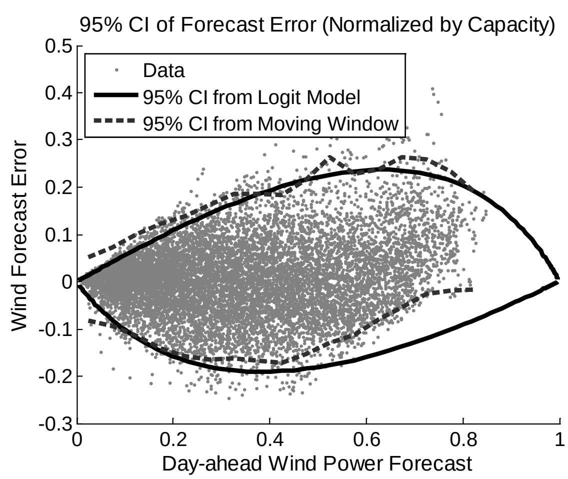

Figure 11: Wind forecast errors in MISO plotted against wind forecast values with 95% confidence intervals. Solid lines are calculated using the logit transformation, and dashed lines are calculated with a moving window of width 0.1.

4.2 Hour-ahead wind power forecasts

We applied the same analysis to ERCOT and MISO wind data using hour-ahead forecasts. Forecasts were created using the persistence method where the wind power forecast for one hour is equal to the actual wind power from the previous hour (Equation 12).

$$F(t) = W(t-1) \tag{12}$$

As in the previous section, we used the estimated uncurtailed values for wind power in ERCOT. Figure 12 (a) shows the actual wind power values plotted against the hour-ahead forecast wind power values. Hour-ahead wind forecast error confidence intervals are shown in Figure 12 (b) where dashed lines show intervals calculated with a moving window of width 0.1 and solid lines are derived from a logit transformation. Figure 13 shows the same plots for the MISO data.

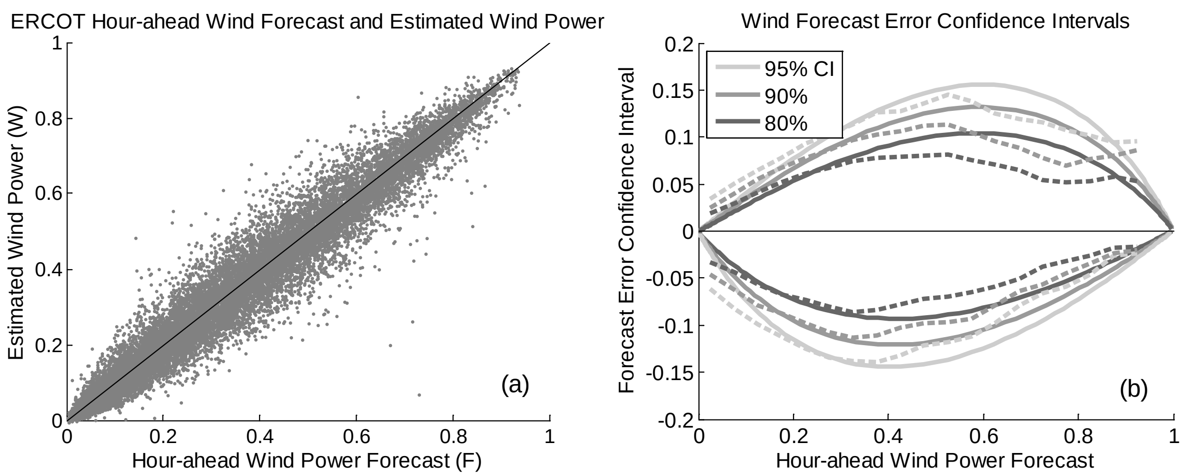

Figure 12: ERCOT estimated uncurtailed wind power plotted against the hour-ahead persistence forecasts (a), and confidence intervals for wind forecast errors (b). Solid lines are calculated using the logit transformation, and dashed lines are calculated with a moving window.

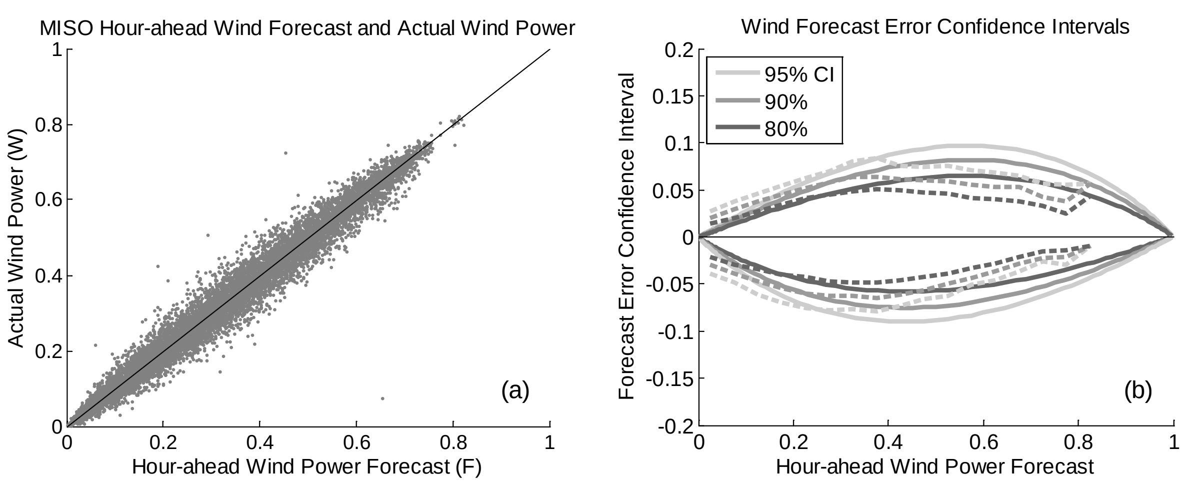

Figure 13: MISO wind power plotted against the hour-ahead persistence forecasts (a), and confidence intervals for wind forecast errors (b). Solid lines are calculated using the logit transformation, and dashed lines are calculated with a moving window.

The agreement between confidence intervals calculated with logit transforms and the moving window are similar to that obtained with the day-ahead forecast data. The scale of the vertical axis is smaller in Figures 12 (b) and 13 (b) than in the figures for the day-ahead results. Once again, verification of confidence intervals was impossible for forecasts near a normalized value of 1 with the MISO data since the maximum wind power observed was around 0.8. While the agreement between confidence intervals calculated with the logit transformation model and the moving window is not perfect, it is reasonably close. A logit transformation model is shown to be an effective tool to estimate uncertainty in wind power forecasts.

5 Conclusion

Observed wind power forecast error distributions are highly dependent on the forecast level of wind power. At low wind forecast power, the forecasts over-predict the actual wind power produced, whereas when the forecast is for high power, the forecast tends to under-predict the actual wind power.

Thus, forecast errors modeled with a single distribution do not provide adequate information for system operators. We presented a method to analyze wind forecast errors with a logit transformation. Transforming wind data with a logit transform in this manner is a straightforward method to determine the amount of uncertainty associated with wind forecasts using historical data. The advantage of this method is that transformed data can be accurately modeled with a bivariate normal distribution. This greatly simplifies the analysis since one set of parameters is estimated instead of multiple parameters for different forecast levels. Calculations of confidence intervals with this method use a model fit to the entire dataset while providing the ability to condition wind uncertainty on the wind forecast value for a given time period. We applied the logit transform to hourly data, but there is no reason that this method should not work well with different time scales.

Electric grid system operators can use this model in their respective decision making analysis. Proper confidence intervals of wind forecast errors for a given forecast level of wind power is essential for decision making. Uncertainty associated with wind power and load forecasts determine the amount of reserve requirements for reliable grid operation.

6 Acknowledgements

This work was supported in part by the Doris Duke Charitable Foundation, the Richard King Mellon Foundation, the Heinz Endowments, the Department of Energy National Energy Technology Laboratory, and the Electric Power Research Institute through the RenewElec project at Carnegie Mellon University. Support was also received from the Center for Climate and Energy Decision Making created through a cooperative agreement between the National Science Foundation (SES-0949710) and Carnegie Mellon University. Brandon Mauch was supported by a fellowship from the Portuguese Foundation for Science and Technology

(Fundação para a Ciência e a Tecnologia), number SFRH/BD/33764/2009. The authors would like to thank Dr. Mark Handschy for his helpful comments. They are also grateful for the data provided by AWS Truepower.

7 References:

[1] U.S. Department of Energy, Wind powering america: Installed U.S. wind capacity and wind project locations, (2012) Retrieved April 25, 2012.

[2] J. Wilkes, J. Moccia, M. Drangan, Wind in power: 2011 European wind statistics, European Wind Energy Association Technical Report, (2012) online: http://www.ewea.org/ fileadmin/ewea_documents/documents/publications/statistics/Stats_2011.pdf

[3] M.A. Ortega-Vazquez, D.S. Kirschen, Estimating the spinning reserve requirements in systems with significant wind power generation penetration, IEEE Transactions on Power Systems, Vol 24, Iss. 1, (2009) 114-124.

[4] F. Bouffard, F.D. Galiana, Stochastic security for operations planning with significant wind power generation, 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, (2008) 1-11.

[5] R. Doherty, M. O'Malley, A new approach to quantify reserve demand in systems with significant installed wind capacity, IEEE Transactions on Power Systems, Vol. 20, Iss. 2, (2005) 587- 595.

[6] B. Hodge, M. Milligan, Wind power forecasting error distributions over multiple timescales, 2011 IEEE Power and Energy Society General Meeting, (2011) 1-8.

[7] B. Hodge, A. Florita, K. Orwig, D. Lew, M. Milligan, A Comparison of wind power and load forecasting distributions, 2012 World Renewable Energy Forum, (2012) NREL/CP-5500-54384, http://www.nrel.gov/docs/fy12osti/54384.pdf

[8] Y.V. Makarov, C. Loutan, Jian Ma, P. de Mello, Operational impacts of wind generation on California power systems, IEEE Transactions on Power Systems, Vol. 24, Iss.2, (2009) 1039-1050.

[9] 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, (2005) 177–184.

[10] H.A. Nielsen, H. Madsen, T.S. Nielsen, 2006, Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts, Wind Energy, Vol. 9, (2006) 95–108.

[11] H. Bludszuweit, J.A. Dominguez-Navarro, A. Llombart, Statistical analysis of wind power forecast error, IEEE Transactions on Power Systems, Vol.23, Iss.3, (2008) 983-991.

[12] A. Luig, S. Bofinger, H.G. Beyer, Analysis of confidence intervals for the prediction of regional wind power output, Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, 2-6 June 2001, 725-728.

[13] A. Fabbri, T. Gomez San Roman, J. Rivier Abbad, V.H. Mendez Quezada, Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market, IEEE Transactions on Power Systems,Vol.20, Iss.3, (2005) 1440- 1446.
[14] A.T. Al-Awami, M.A. El-Sharkawi, Statistical characterization of wind power output for a given wind power forecast, North American Power Symposium (NAPS), (2009) 1-4.
[15] A. Lau, P. McSharry, Approaches for multi-step density forecasts with applications to aggregated wind power, The Annals of Applied Statistics, Vol. 4, Iss. 3, (2010) 1311–1341.
[16] P. Pinson, Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions, Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol 61, Iss. 4, (2012) 555-576.
[17] J. Rogers, S. Fink, K. Porter, Examples of wind energy curtailment practices, NREL Subcontract Report, (2010) NREL/SR-550-48737.
[18] R. Wiser, M. Bolinger, DOE 2010 Wind technologies market report, Department of Energy, (2011) DOE/GO-102011-3322.
[19] MISO, Day Ahead Wind Forecast, online: https://www.midwestiso.org/MarketsOperations/ RealTimeMarketData/ Pages/DayAheadWindForecast.aspx
[20] J. Kehler, H. Ming, M. McMullen, J. Blatchford, ISO perspective and experience with integrating wind power forecasts into operations, 2010 IEEE Power and Energy Society General Meeting,(2010) 1-5.
[21] MISO Reliability Subcommittee Report June 27, 2012. https://www.misoenergy.org/Library/Repository/Meeting%20Material/Stakeholder/RSC/2012/20120626/20120626%20RSC%20Item%2008%20Wind%20Curtailment%20Data.pdf
[22] Y.V. Makarov, R.T. Guttromson , Z. Huang, K. Subbarao, P.V. Etingov, B.B. Chakrabarti, J. Ma, Incorporating wind generation and load forecast uncertainties into power grid operations, Report PNNL-19189, PNNL, Jan 2010.
[23] P. Frederic, F. Lad, Two moments of the logitnormal distribution, Communications in Statistics - Simulation and Computation , Vol. 37, Iss. 7, (2008) 1263-1269.
[24] U. Focken, M. Lange, K. Mönnich, H.P. Waldl, H.G. Beyer, A. Luig, 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, (2002) 231–246.
[25] G. Giebel, P. Sørensen, H. Holttinen, Forecast error of aggregated wind power, TradeWind Deliverable Report, (2007) Risø-I-2567(EN).

Appendix A

Table A.1 provides details on the ERCOT and MISO data used in this study. Figure A.1 shows the frequency distributions of the estimated uncurtailed wind power in ERCOT, actual wind power in ERCOT and actual wind power in MISO over the respective time periods from Table A.1.

Table A.1: Summary statistics for the ERCOT and MISO wind data used in this study.

	ERCOT	MISO
Time Period	Dec 2008 – Dec 2010	Feb 2011 – May 2012
Average Load	35 GW	61 GW
Maximum Load	66 GW	104 GW
Installed Wind Capacity	8.325 – 9.53 GW	9.125 – 10.79 GW
Actual Wind Capacity Factor	0.28	0.33
Uncurtailed Wind Capacity Factor	0.35	N/A
Ratio of actual wind energy to load	0.069	0.055
Ratio of uncurtailed wind energy to load	0.087	N/A

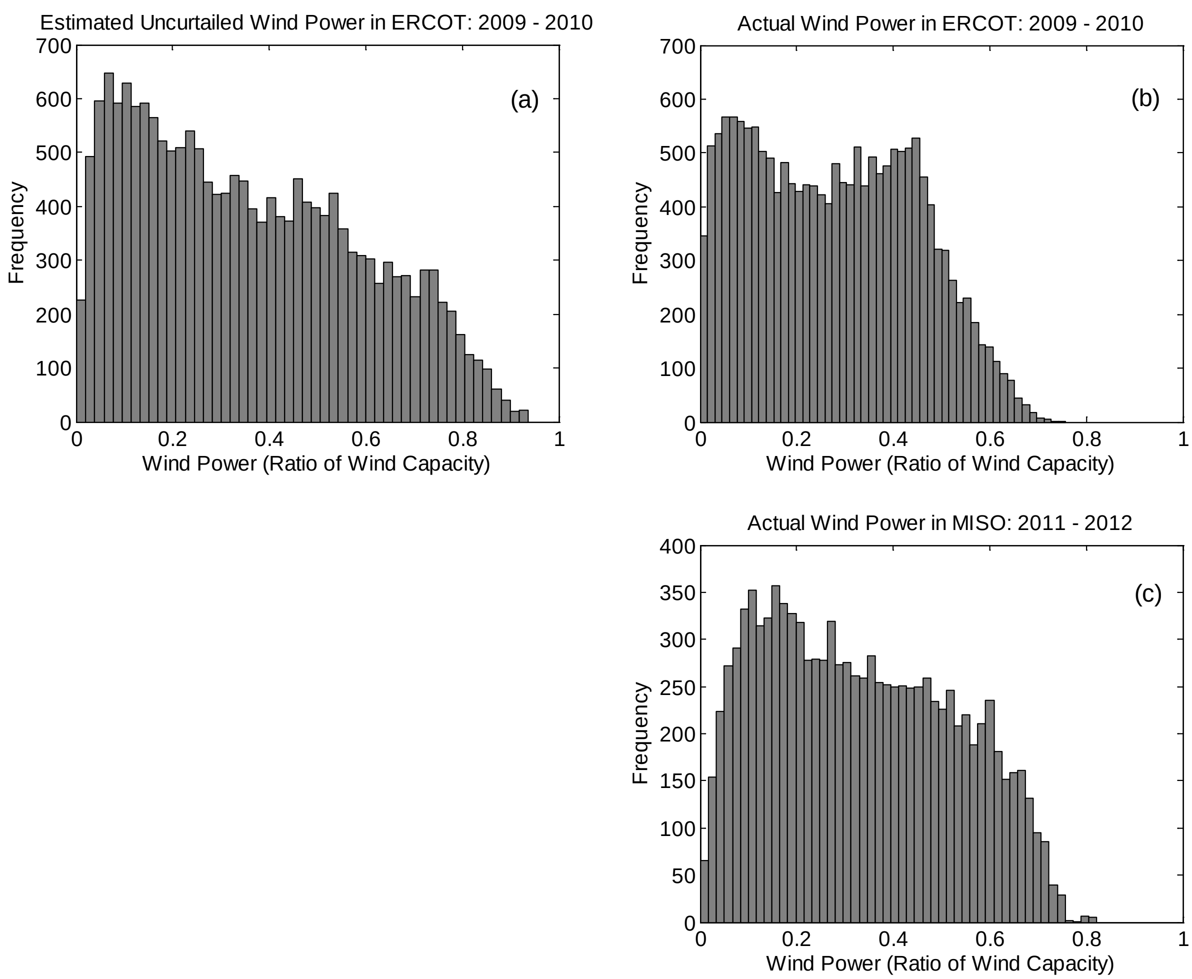

Figure A.1: Frequency distributions of (a) uncurtailed hourly wind power in ERCOT (b) actual hourly wind power in ERCOT and (c) actual hourly wind power in MISO.

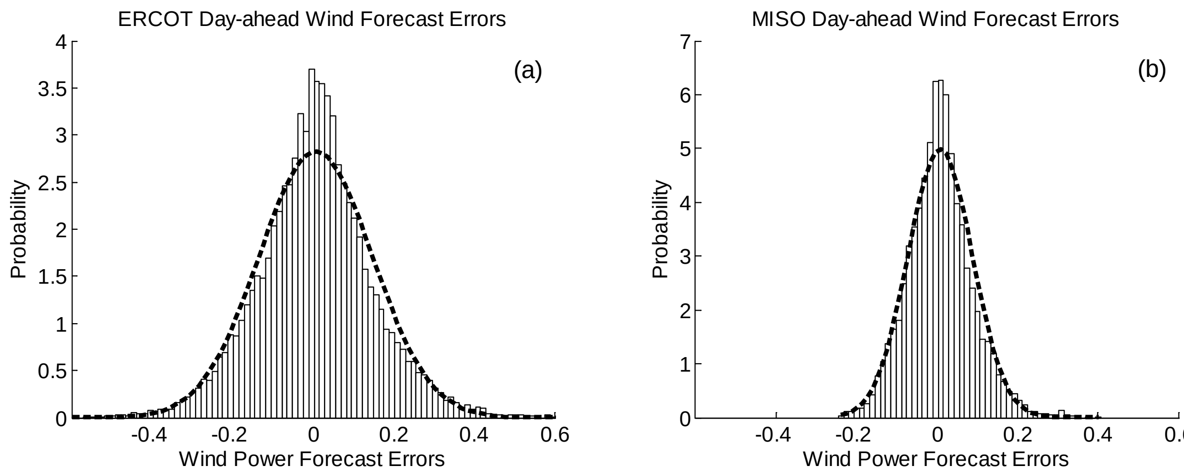

Figure A.2: Frequency distribution (histogram bars) of wind power forecast errors in ERCOT (a) and MISO (b) with a fitted normal distribution (dashed line) overlaid.

Figure A.2 shows the distribution of wind forecast errors in ERCOT and MISO for the respective time periods in our data. The figures above indicate that the forecasts are unbiased on the whole for both regions. As is well-known [6], normal distributions are not well suited for modeling wind forecast errors. MISO day-ahead wind forecasts tend to be more accurate than ERCOT (although the wind capacity is similar in both regions). This may be due to the much larger territory covered by MISO. Large geographic diversity tends to reduce total forecast errors if forecast errors from sub-regions are weakly correlated [24, 25].

Appendix B

Previously, we showed that the logit transformations of the wind power forecast and actual wind power data fit a normal distribution reasonably well. If the logit transformation of a variable is normally distributed, $N(\mu, \sigma^2)$, then the variable itself is distributed logit-normal, $LN(\mu, \sigma^2)$. For the wind forecast (F) and wind power (W), the logit transformations are

$$F^* = \ln\left(\frac{F}{1-F}\right) \qquad W^* = \ln\left(\frac{W}{1-W}\right) \tag{B.1}$$

Each of the transformed variables is normally distributed with a probability density

$$f(X^*) = \frac{1}{\sqrt{2\pi}\sigma_{X^*}} exp\left(-\frac{(X^*-\mu_{X^*})^2}{2\sigma_{X^*}^2}\right) \tag{B.2}$$

The symbols μ_{x^*} and σ_{x^*} represent the mean and standard deviation of the transformed variables. The original variables are logit-normal distributed with a probability density

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma_X X(1-X)} exp\left(-\frac{(X^* - \mu_X)^2}{2\sigma_X^2}\right) \quad \text{(B.3)}$$

The symbols have the same definitions as in Equation A.2. Given that the logit transformed variables are jointly normal, the variables *F* and *W* are jointly logit-normal distributed with probability density defined as

$$f(F,W) = \frac{1}{2\pi\sigma_F\sigma_W\sqrt{1-\rho^2}F(1-F)W(1-W)} exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(F^*-\mu_F)^2}{\sigma_F^2} + \frac{(W^*-\mu_W)^2}{\sigma_W^2} - \frac{2\rho(F^*-\mu_F)(W^*-\mu_W)}{\sigma_F\sigma_W}\right]\right) \quad \text{(B.4)}$$

Figure A.3 shows the distribution of *F* and *W* from the ERCOT data with a fitted logit-normal distribution overlaid.

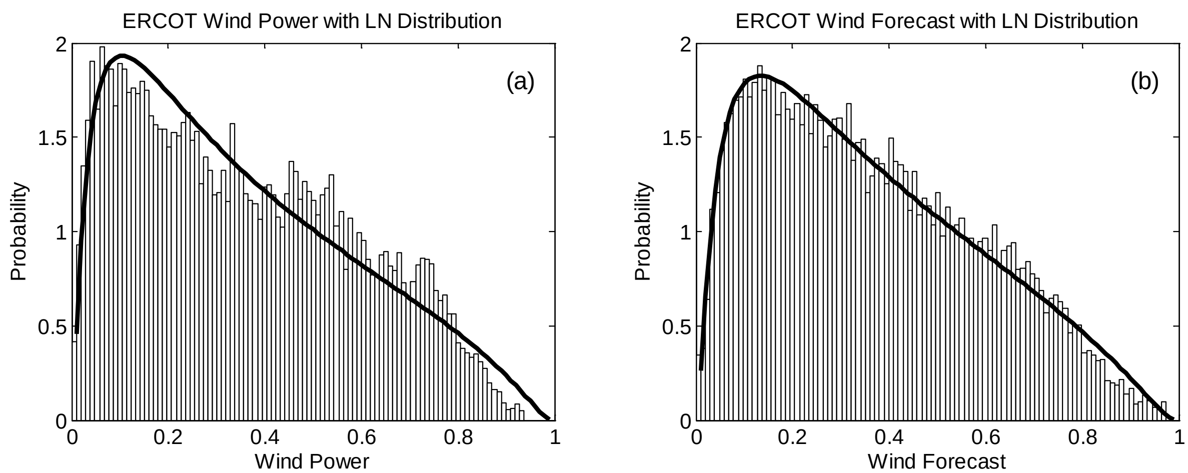

Figure B.1: Distributions of wind power and wind power forecasts from ERCOT with logit-normal distributions overlaid.

Appendix C

As shown in section 3, wind power forecasts exhibit a bias that varies thorough the range of forecast values. At low wind power forecasts the bias is negative while high wind power forecasts have a positive bias. We calculated the mean value of the forecast error conditioned on the forecast value for the ERCOT day-ahead wind power forecasts in the following manner with E[·] denoting the expectation operation.

$$\mathrm{E}[e|F] = \mathrm{E}[(F-W)|F] = F - \mathrm{E}[W|F] \quad \text{(C.1)}$$

For a given forecast, the bias is the forecast minus the expectation of W conditioned on the forecast value. The expectation is determined with the logit-normal distribution as

$$\mathrm{E}[W|F] = \int_0^1 \frac{1}{\sqrt{2\pi}\sigma_{W|F}W(1-W)} exp\left(-\frac{(W^* - \mu_{W|F})^2}{2\sigma_{W|F}^2}\right) W dW \qquad \text{(C.2)}$$

Since there are no closed form solutions for the moments of the logit-normal distribution the expected value must be obtained with numerical integration. We used the quad function in Matlab, which implements Simpson's rule, to evaluate the integral in Equation A.6 over the range of possible values for F. Figure C.1 shows the expected values W conditioned on F and the expected error conditioned on F. In each plot the solid lines were calculated with the logit-normal model, and the dashed lines were calculated using a moving window of width 0.1.

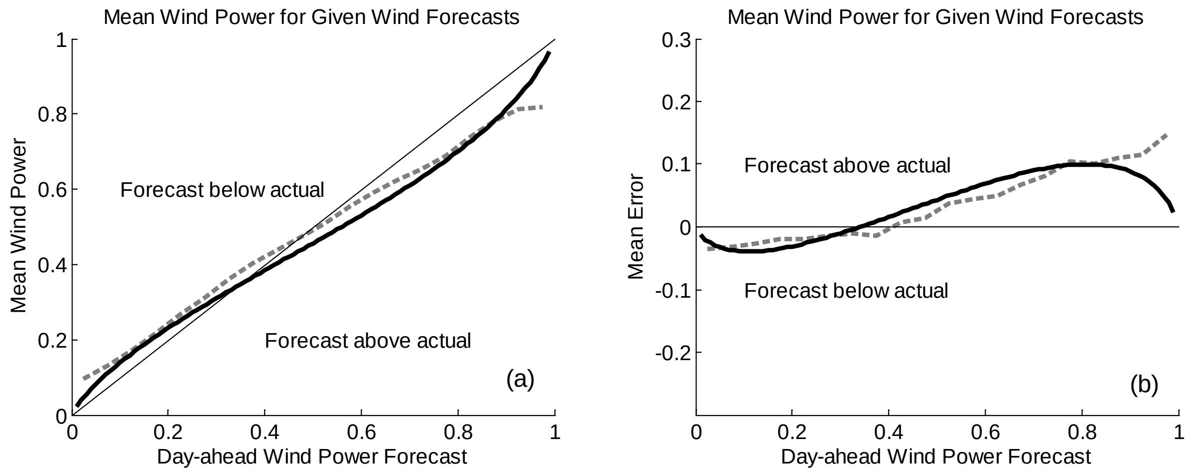

Figure C.1: Mean wind power conditioned on the wind power forecast (a) and mean forecast error conditioned on the wind power forecast (b). Solid lines were calculated directly with binned data, and dashed lines were calculated using the logit-normal distribution.

In order to examine the effect of a forecast algorithm that accounts for the observed forecast bias, we subtracted the bias values calculated with the logit-normal model in Figure A.4 (b) from the ERCOT day-ahead forecasts to remove the conditional bias in the data. Figure A.5 shows the day-ahead wind power forecast distributions in ERCOT with the original data and the unbiased forecasts. The plot of original error distribution is the same as in Figure 2 in the main text. Errors from the original data have a mean value of 0.007 and variance of 0.02. Errors calculated from the unbiased forecasts have a mean value of -0.005 and variance of 0.019.

Removing the conditional bias in the forecasts mean that low forecasts are increased and high forecasts are decreased. As Figure B.1 (b) shows, most forecasts are near 0.2 as a ratio of the wind capacity. Therefore, most errors resulting from the unbiased forecasts are slightly increased compared with the errors from the biased forecasts. The resulting unbiased forecast error distribution in Figure C.2 (b) is skewed, but retains an overall mean near zero and a variance comparable to the biased errors.

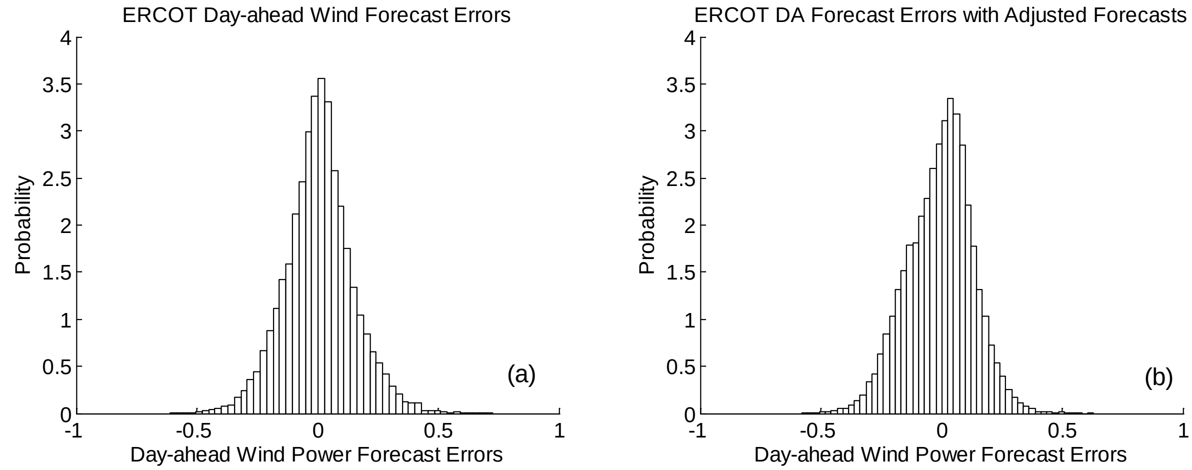

Figure C.2: Wind forecast error distributions from ERCOT with original data (a) and with the conditional bias removed from the forecasts.

Figure C.3 shows the forecast errors plotted against the forecast values with the original ERCOT data and with the data adjusted to remove the bias. In the adjusted case the upward trend in the forecast errors is reduced.

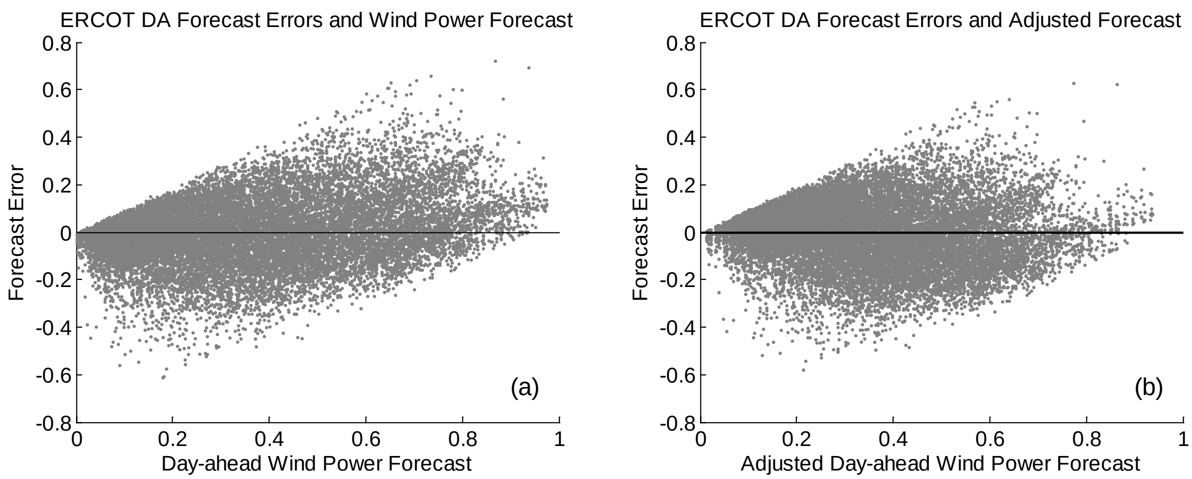

Figure C.3: Wind forecast error distributions from ERCOT with original data (a) and with the conditional bias removed from the forecasts.

As evident in Figure B.1, the data do not fit a logit-normal distribution perfectly. Since the bias was calculated from the fitted logit-normal model, it was also not calculated perfectly. Figure

C.4 shows the conditional bias remaining in the forecast errors after adjusting the forecast data to remove the modeled bias. Residual bias was calculated with a moving window of width 0.1. When compared to the bias calculated for forecast bins in Figure C.1 (b), one can see that the conditional bias is mostly eliminated.

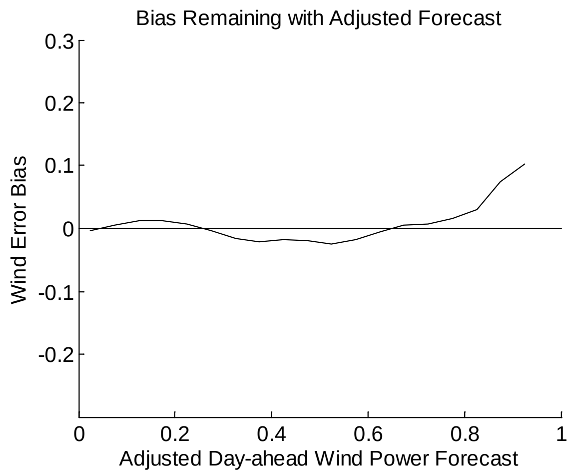

Figure C.4: Bias calculated with binned forecast data after removing the conditional bias calculated with the logit-normal distribution; compare with Figure C.1 (b).