




Examining Information Entropy Approaches as Wind Power Forecasting Performance Metrics

Preprint

Bri-Mathias Hodge, Kirsten Orwig,
and Michael Milligan

Presented at the 12th International Conference on Probabilistic Methods Applied to Power Systems
Istanbul, Turkey
June 10-14, 2012

Conference Paper
NREL/CP-5500-53515
June 2012

Contract No. DE-AC36-08GO28308

NOTICE

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.

Examining Information Entropy Approaches as Wind Power Forecasting Performance Metrics

Bri-Mathias Hodge, Kirsten Orwig, and Michael Milligan
Transmission and Grid Integration Group
National Renewable Energy Laboratory
Golden, CO, USA
bri-mathias.hodge@nrel.gov, kirsten.orwig@nrel.gov, michael.milligan@nrel.gov

Abstract—**Wind power forecasting is expected to play an increasing role in power system operations as the amount of wind capacity on-line continues to increase. Traditional forecasting metrics, such as MAE and RMSE, neglect some of the information inherent in forecasting error distributions. Information entropy approaches, based on the Rényi entropy, have been proposed as an alternative metric to assess different forecasting methods. In this work, we examine the parameters associated with the calculation of the Rényi entropy in order to further the understanding of its application to assessing wind power forecasting errors.**

Keywords- Power generation; stochastic systems; power engineeringand energy; wind power generation

I. INTRODUCTION

Wind power installed capacity has been increasing significantly in the United States over the past few years. Since wind is a variable and uncertain source of power, many questions have been asked about how these large additions of wind power can be integrated into the existing electricity system, designed for more predictable and less variable generation. One way that the uncertainty associated with wind power production can be reduced is through forecasting. As the amount of wind power in the system increases, the importance of wind power forecasting in power system operations is also expected to increase. Wind forecasting can be undertaken at a number of different geographical and organizational scales. In this work, we are primarily concerned with wind power forecasting for bulk power system operations, thus there is a focus on centralized forecasting systems employed by Independent System Operators (ISOs), though consideration is also given to forecasting individual wind plant output.

Porter and Rogers [1] provide an overview of the centralized wind power forecasting systems in use by various system operators across North America. An update to that study by the same authors provides the performance of various wind forecasting systems, though only in terms of the two most common metrics, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) [2]. These two measures correspond to measures of the first and second statistical moments of the error, respectively. However, previous work has established that the forecast error distributions seen in wind power forecasting are non-normal [3, 4], and the impacts of wind power forecasting errors do not scale linearly. Small deviations from the forecast output can be relatively easily compensated for in power system operations. Since load is also variable and uncertain, power systems have mechanisms to account for load forecasting errors; at small time frames this is the function of regulation. Nevertheless, large forecast errors can necessitate unit start-up or out-of-merit dispatch, expensive corrections for over-forecasting events. Under-forecasting events can be more easily compensated for through wind curtailment, though this also includes a lost opportunity cost. For plant owners that participate in day-ahead markets, large forecast errors can have large financial consequences because they will require deviations from their production schedule. By considering the third and fourth standardized moments, skewness and kurtosis, respectively, forecasting error distributions from operational wind forecasting systems may be more accurately represented [4]. Since the MAE and RMSE metrics cannot distinguish between two distributions with the same mean and variance, but different skewness and kurtosis values, they ignore additional information about the forecast errors that could potentially have a significant impact on system operations. In fact, the use of statistical measures beyond the variance has been shown to decrease the total system costs with stochastic unit commitment [5]. A brief discussion of the different metrics that may be applied to wind power forecasting assessment is provided in Giebel et al. [6].

The realization that standard forecasting metrics are only optimal if the error distribution is Gaussian has led to the use of concepts from information theory to identify new metrics that can utilize all of the information present in the forecast error distributions [7]. Bessa et al. [8] explore the use of minimum error entropy and correntropy as training metrics in a neural network that produces wind power forecasts from inputs of wind speed and direction forecasts of a numerical weather prediction model. This approach assumes that the less information contained in the forecast error distribution, the better the forecast. As an illustrative example: a Dirac delta function would contain no information because there is only one possible value realization. However, one limitation of this approach is that it does not matter where along the range of possible forecast values the Dirac delta function lies. It cannot distinguish between the case of all errors occurring at the point of zero forecast error and the point of a full wind plant's capacity. In this work we attempt to determine the suitability of information entropy approaches, based on the Rényi entropy [9], as general metrics for assessing wind power forecasting errors. For this assessment we have identified the main parameters associated with the calculation of the Rényi entropy and performed a sensitivity analysis on their values over realistic ranges. Our goal is to examine the potential benefits and limitations of using metrics based on information theory as wind power forecasting performance metrics.

II. Methods and Data

To assess the suitability of information entropy approaches to serve as wind power forecasting performance metrics, we first define and examine the Rényi entropy concept that serves as a basis for the approach in Section II-A. Section II-B describes the wind power forecasting data taken from operational systems that is used to help examine the metric.

A. *Rényi entropy*

To move beyond the limitations of second-order moment metrics for wind power forecasting, and make use of the full range of information available in the error distributions, concepts developed in the field of information theory have been suggested. The Shannon entropy [10] is a measure of the information content contained in a message, and serves as one of the pillars of modern information theory. The Rényi entropy is a generalization of the classical Shannon entropy and is defined as:

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2 \sum_{i=1}^{n} p_i^\alpha \qquad (1)$$

Here α is a parameter that allows for the creation of a spectrum of Rényi entropies and pi are the probabilities of each discrete section of the distribution {x1, x2, …xn}. Previous work with the Rényi entropy in wind power forecasting [7, 8] has utilized the special case of quadratic entropy, i.e., when $\alpha = 2$. However, the choice of α value used in the calculation of the Rényi entropy can be very important, as it determines the relative weighting of high and low probability events. High values of α will favor higher probability events, while lower values of α weight all of the instances more evenly. This weighting is an important consideration in the application of the metric to wind power forecasting errors. Traditional power system operations without variable generation include the forecasting of demand so that supply can be made available to meet the demand. Therefore, services such as regulation are available to compensate for these relatively small forecast errors, whether they arise due to load or wind power forecasting. It is the low probability, but high magnitude, forecasting events that are particularly troublesome for power system operations. The choice of α value can therefore significantly impact how wind forecasting methods are assessed and improved. For example, it can mean the difference between shifting the focus from preventing large magnitude errors (distribution tail reduction) and increasing the peakedness of the distribution (i.e., shifting shoulder values into the peak while ignoring the tails). The relative weighting of forecast errors, and hence choice of α, is especially important in applying the metric between different power systems. This is because different power systems have different amounts of flexibility inherent in current system operations, often based on the physical characteristics, such as minimum up and down times, ramping rates, etc., of the other power plants in the system. Since the flexibility available in the system may vary depending on what other plants are actually online at a given time, it is easy to imagine that the optimal α value for a particular system will change with time.

In this work, we estimate the probability mass function of the wind power forecasting errors by using a large number of bins, such as would be employed in creating a histogram, to represent the values that the function may take. This method is used both to aggregate the probability mass functions created from wind power forecasting error data, and to estimate the continuous probability density functions used to model the observed distributions. In this way, it is similar to the Parzen window size approach used in kernel density estimation, as it effectively aggregates the observed errors into discrete points from which the probability density function may be constructed. As will be shown in Section III, the method of aggregating the observed errors into discrete groups where a probability of occurrence may be calculated is another critical parameter in the calculation of the Rényi entropy.

B. *Data Utilized*

To assess and compare the various metrics suggested for analyzing wind power forecast errors, we utilize the observed errors from three different operational forecasting systems in the evaluation of the metrics. Each dataset comes from a different geographic area of the United States. The first dataset comes from a single wind plant in the Xcel Energy Colorado territory with a nameplate capacity of approximately 300 MW. The data include three months of wind power forecasts and measured output from the plant. The forecasts are updated every 15 minutes for every hour in the next 72 hours and the actuals provide the plant power output at every hour. The second dataset is one year of day-ahead forecasts, made for every hour of the next day, and actuals for the aggregation of a number of wind plants in the CAISO region. The total capacity of all of the wind plants considered is roughly 940 MW. The final dataset comes from the ERCOT interconnection. The centralized day-ahead forecasts and actuals at hourly intervals for approximately 9,000 MW of wind capacity over a 13-month period are included. Further information on the centralized wind power forecasting systems in use in North America, including those used in creating these datasets, may be found in two reports by Porter and Rogers [1, 2].

While the normal distribution is often used to represent the forecast errors observed in forecasting systems, it tends to under represent the number of large errors in the tails of the distribution [4]. When the forecast error distribution from an operational system is compared to a normal distribution, the observed distribution tends to have a sharper peak, more narrow shoulders, and fatter tails, as shown in Fig. 1. Recognizing the usefulness of having a model distribution with which to represent the distributions obtained from operational forecasting systems, we have chosen the hyperbolic distribution as a more accurate substitute. Distribution parameters were fit to the data using a maximum likelihood method implemented in the *hyperbFit* function of the *HyperbolicDist* package [11] in the *R* statistical computing environment [12].

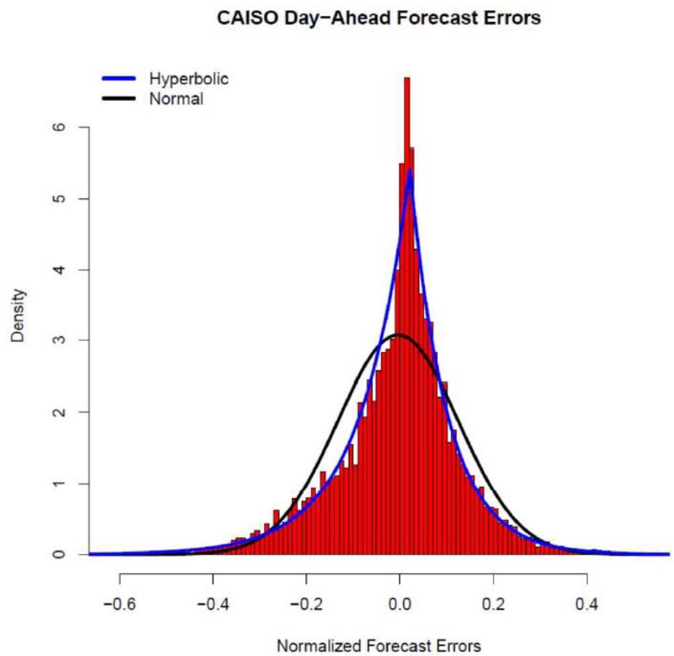

Figure 1. Histogram of the day-ahead forecasts for the CAISO system over a one-year period. This includes forecasts for different timescales, ranging from 18 to 42 hours ahead. $\mu = -0.004$; $\sigma = 0.13$; $\gamma = -0.39$; $\kappa = 1.50$. The black line represents a normal distribution with the same mean and standard deviation. The blue line represents a hyperbolic distribution fit to the data with: $\pi = -0.139$, $\zeta = 9.62\ E^{-5}$, $\delta = 8.79\ E^{-6}$, $\mu = 0.021$.

III. Results

To assess the utility of information entropy techniques as wind forecasting performance metrics, we must first test the parameters associated with their calculation. Most traditional metrics do not involve a choice of parameters, meaning that they may be applied to similar datasets in only one manner. While this simplicity is a benefit for their uniform application, it is also a limitation. For example, the ability to change the α parameter in the calculation of the Rényi entropy could allow for the degree of system flexibility to be considered when deciding which forecasting systems should be employed in a certain region. In this section, we first make a qualitative comparison between using the Rényi entropy as a wind power forecasting performance metric, and two traditional metrics. We then examine some of the implications of being able to choose the size of the divisions used in binning the forecast error distribution, as well as different values of the α parameter.

A. Comparison of Metrics

While multiple metrics have been proposed for assessing wind power forecasting errors, each of these metrics provides only certain information about the distribution of forecasting errors. Due to the limited information provided by any one metric, it is always wise to consider multiple metrics when comparing alternative forecasting performance methods. However, problems arise when the metrics provide differing assessments of the forecasting performance method's capabilities. This is especially true when trying to discern between very similar methods. As an example of this problem, we have compared the forecast errors from the Xcel data with two traditional metrics, as well as the Rényi entropy. To compare the different metrics we have also included two more distributions, a normal distribution and a hyperbolic distribution. These two alternative distributions were created by fitting the model distributions to the observed Xcel errors and then taking a number of samples equal to the size of the Xcel data from each of the model distributions. The results of these calculations are provided in Table I.

TABLE I. Comparison of Metrics for Xcel Data

	Observed	Normal	Hyperbolic
MAE	87.21	125.82	87.69
RMSE	3.64	3.66	2.86
Rényi entropy[a]	6.20	6.93	7.23

a. Calculated with number of bins = 200 and α = 0.05

If we were to choose which of the two model distributions better fits the observed data, the answer would depend on which of the metrics we were considering. Judging solely on the MAE score, the hyperbolic distribution provides a very close fit to the observed data. For the RMSE values, the normal distribution is much closer to the observed values, however, the hyperbolic distribution has a lower value, meaning that it produces a better forecast than the observed distribution, taking only this single metric into consideration. Using the Rényi entropy as a metric, we find that the normal distribution is preferred over the hyperbolic distribution. This result may be true for arbitrarily chosen parameters, such as when the number of bins used in the calculation of the Rényi entropy is 200 and the α value is 0.05. However, with larger α values, the choice of model distribution is switched. This result clearly necessitates a closer examination of the role of the parameters in the calculation of the Rényi entropy.

B. Effect of Number of Bins

A key consideration when examining any distribution is the degree to which different observations should be grouped. This takes the form of the bin size when creating a histogram or the size of the Parzen window if a non-parametric approach, such as kernel density estimation, is applied. This choice determines how smooth the graphical representation of the distribution will be and can be quite important in the recognition of particular characteristics of the distribution. Guidelines for determining the number of bins to use in a histogram do exist, for example Scott's rule [13]. However, in this work we have chosen to perform a parametric sensitivity analysis to examine the impact of the number of bins on the calculation of the Rényi entropy over a range of possible values.

The number of bins used in a histogram function of the ERCOT data was varied between 10 and 9,500, by intervals of one bin, and the Rényi entropy value was computed with a constant α value over all of the bin sizes. The results of these computations are plotted in Fig. 2, for α = 0.5. It is immediately apparent that the number of bins used has an important impact on the calculation of the Rényi entropy value. At lower numbers of bins the entropy value calculated is reduced, because the information is aggregated into a smaller number of categories. Larger numbers of bins allow the information available in the full distribution to be recognized, resulting in higher entropy values. As might be expected from the Rényi entropy formulation given in eq. 1, the entropy value varies logarithmically with the number of bins. It is interesting to see the step function changes that occur in the entropy value with increasing numbers of bins. Since the number of bins is a discrete parameter, and the parameter was varied by a single increment for each calculation, the step point represents a point where further disaggregation of the data provides additional information.

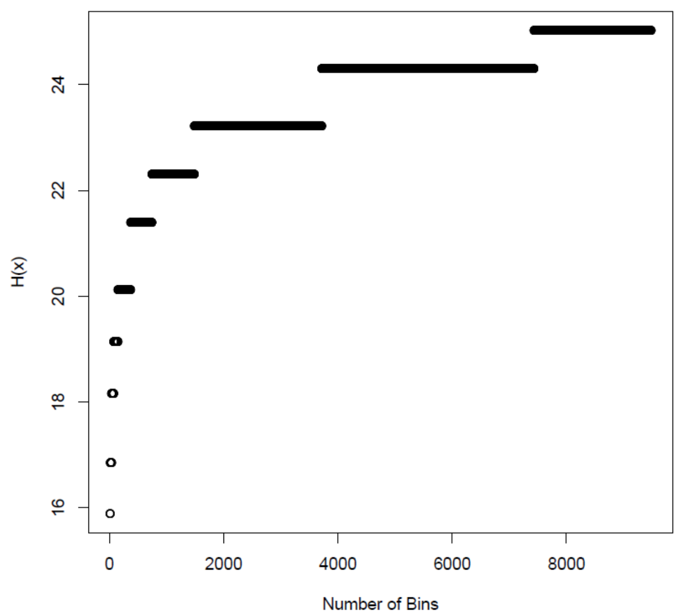

Figure 2. Rényi entropy values for the observed ERCOT day-ahead forecast errors for a number of different histogram bin values. The α value used in this example was held constant at 0.5.

C. *Effect of α Value*

The α parameter value chosen can also be a key consideration in the calculation of the Rényi entropy. As discussed in the Methods and Data section, the parameter value may be selected to adjust the weighting between the peak and tails of the distribution. When dealing with wind power forecasting errors this is a very important consideration, as large forecast errors have a much greater impact on system operations than do small errors. Therefore, a well-designed metric for assessing wind power forecasting errors should weigh larger errors proportionally more than smaller errors. Fig. 3 shows the result of computing the Rényi entropy value for different α values with a constant bin size for the Xcel hour-ahead forecast data. It is important to note the log-scale on the x-axis. Two different ranges of α values were assessed. The first includes values less than one, starting at 0.01 and incrementing by 0.01 until 0.99. These values are meant to examine weightings that emphasize the tails of the distribution. The second ranges from 1.1 to 100 by increments of 0.1. The upper values in this range tend to neglect the tails of the distribution in favor of the peak values. Fig. 3 demonstrates that the Rényi entropy value computed can differ significantly based on the α value utilized.

There is a point in Fig. 3, at approximately α = 0.25, where the information entropy values of the normal distribution and hyperbolic distribution representations cross. This occurrence is significant in that the choice of which model forecast error distribution representation is superior, according to the information entropy metric, is reliant on the α value chosen to assess the models. Presuming a smaller information entropy value is better, the normal distribution would be preferred when using a very low α value, while the hyperbolic distribution would be preferred for a higher α value. This inconsistency is closely related to the relative shapes of the two distributions, which may be seen in Fig. 4. Since higher α values place more emphasis on the smaller magnitude deviations, the hyperbolic distribution is favored due to its more pronounced peak. For the lowest α values, the heavier tails of the hyperbolic distribution are emphasized and the normal distribution has lower Rényi entropy values.

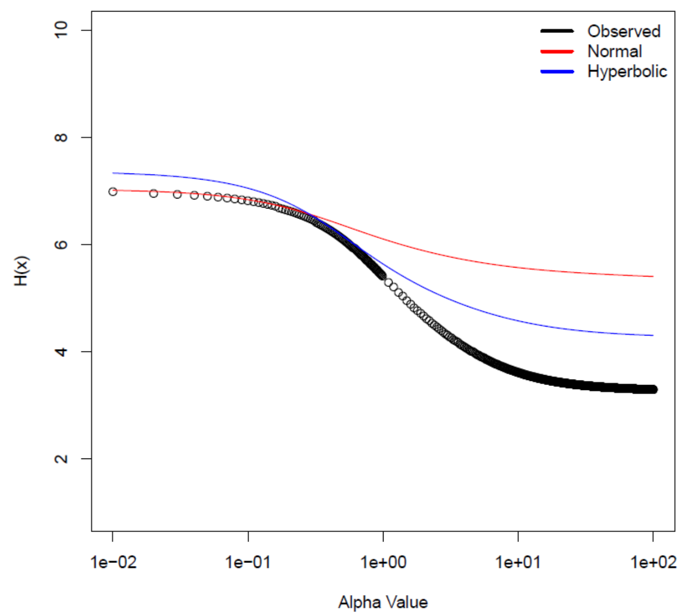

Figure 3. Rényi entropy values for the observed Xcel hour-ahead forecast errors, along with the normal distribution and hyperbolic distribution representations of the observed errors, for a number of different α values. The number of histogram bins used in this example was held constant at 250.

It is also interesting to examine the relative fits of the two model distributions compared to the observed distribution over the entire range of α values in fig. 4. In this particular case, the hyperbolic distribution seems to over-represent the tails of the distribution, leading to a higher Rényi entropy value in the range of the very low values. Conversely, the hyperbolic distribution tends to under-represent the peak of the distribution, as seen in the higher α values. However, the hyperbolic distribution does come closer to matching the observed distribution in this range than does the normal distribution. There is a region of α values, between approximately 0.2 and 0.9, where the hyperbolic distribution matches the observed distribution very closely. It is important to note that these fits are impacted by the step change behavior seen in the choice of the number of bins. Since the location of the step changes may differ for each of the different model distributions, as well as the observed errors, this can play a very large role in assessing which distribution provides a better fit based on the Rényi entropy criterion.

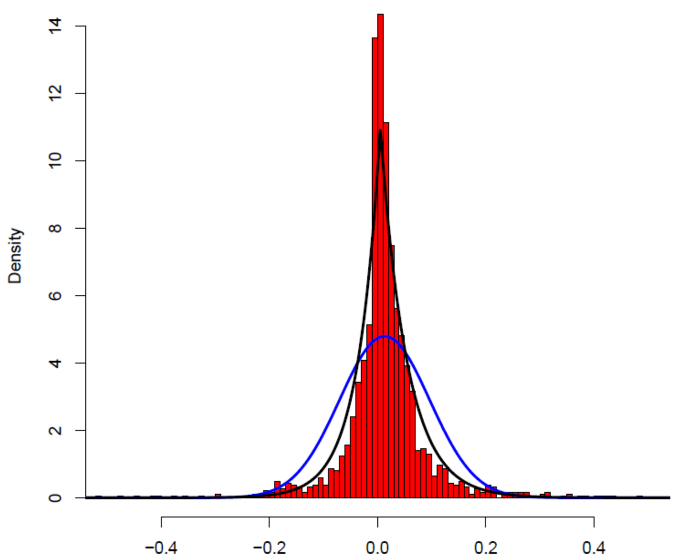

Figure 4. Histogram of the one-hour-ahead forecasts for the Xcel plant over a 3-month period. $\mu = 0.01$; $\sigma = 0.08$; $\gamma = -0.01$; $\kappa = 17.62$. The blue line represents a normal distribution with the same mean and standard deviation. The black line represents a hyperbolic distribution fit to the data with: $\pi = 0.087$, $\zeta = 3.88\ E^{-5}$, $\delta = 1.76\ E^{-6}$, $\mu = 0.005$. The forecast errors have been normalized based on the total wind plant capacity.

D. *The Relationship between the Number of Bins and α*

Having already individually examined the choice of two parameters: number of bins and α value in the computation of the Rényi entropy value for wind power forecasting errors, it is interesting to examine the relationship between the two simultaneously. Fig. 5 shows a three dimensional plot of the Rényi entropy values calculated using the same range ofα values and number of bins as examined in the preceding sections. The step change pattern seen with varying number of bins is clearly apparent on the z-axis, though the degree of change is slightly less at low α values. On the x-axis, the largest changes in Rényi entropy are seen with α values in the zero to one range. The information entropy measure decreases very gradually with increasing α values greater than one.

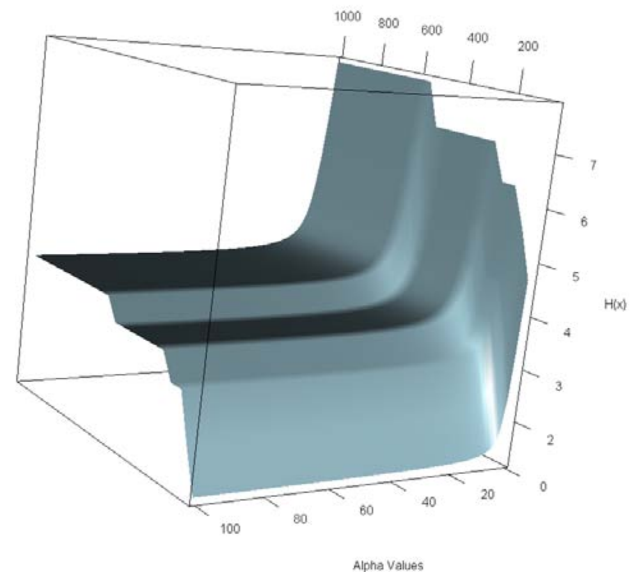

Figure 5. Comparison of the Rényi entropy value of the Xcel hour-ahead forecast errors when calculated using different numbers of bins and α values.

IV. CONCLUSION

Information entropy approaches have been suggested as possible replacements for the traditional performance metrics of wind power forecasting because they allow for the utilization of all of the information in the wind power forecast error distributions, including higher moments. However, these approaches also have limitations based on the choice of parameters utilized. The size of the bin or window used in the estimation of the probability density function is one such parameter that may have a large impact on the resulting information entropy value. In the calculation of the Rényi entropy, the α parameter can also have a significant impact; deciding the weighting between different sections of the distribution in the resulting value. We have highlighted some of the potential pitfalls associated with using the Rényi entropy as the basis for a wind power forecasting performance metric and advocate the use of a systematic approach when applying the metric to real systems. However, despite these limitations, this approach has the potential to tailor the wind forecasting performance metric used based on the state of the electricity system where it is applied. This is an important consideration, because it would allow systems to define the level of wind forecasting error that they can reasonably accommodate, and then focus their forecasting improvement work on eliminating errors outside of their acceptable range.

ACKNOWLEDGMENT

The authors gratefully acknowledge David Maggio at ERCOT, James Blatchford at CAISO, and Keith Parks at Xcel Energy for supplying the wind power forecasting data used in the study.

REFERENCES

[1] K. Porter and J. Rogers, "Central Wind Power Forecasting Programs in North America by Regional Transmission Organizations and Electric Utilities," National Renewable Energy Laboratory, Golden, CO 2009.

[2] K. Porter and J. Rogers, "Status of Centralized Wind Power Forecasting in North America," National Renewable Energy Laboratory, Golden, CO 2010.

[3] B.-M. Hodge and M. Milligan, "Wind Power Forecasting Error Distributions over Multiple Timescales," in *IEEE Power & Energy Society General Meeting*, Detroit, MI, 2011.

[4] B.-M. Hodge, E. Ela, and M. Milligan, "The Distribution of Wind Power Forecast Errors from Operational Systems," in *10th International Workshop on Large-scale Integration of Wind Power*, Aarhus, Denmark, 2011.

[5] C. Lowery and M. O'Malley, "Impact of Wind Forecast Error Statistics upon Unit Commitment," *IEEE Transactions on Sustainable Energy,* 2011.

[6] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The State-Of-The-Art in Short-Term Prediction of Wind Power - A Literature Overview, 2nd Edition," Riso DTU 2011.

[7] R. Bessa, V. Miranda, J. Principe, A. Botterud, and J. Wang, "Information theoretic learning applied to wind power modeling," in *The 2010 International Joint Conference on Neural Networks*, Barcelona, 2010.

[8] R. Bessa, V. Miranda, A. Botterud, and J. Wang, "'Good' or 'bad' wind power forecasts: a relative concept," *Wind Energy,* vol. 14, pp. 625 - 636, 2010.

[9] A. Renyi, "On Measures of Entropy and Intformation," in *4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1960, pp. 547 - 561.

[10] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.

[11] D. Scott, "HyperbolicDist: The hyperbolic distribution," ed, 2009.

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

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