

Probabilistic Short-term Wind Forecasting Based on Pinball Loss Optimization

Mucun Sun, Cong Feng, Jie Zhang
The University of Texas at Dallas
Richardson, TX 75080, USA

Erol Kevin Chartan, Bri-Mathias Hodge
National Renewable Energy Laboratory
Golden, CO 80401, USA

Abstract—Probabilistic wind power forecasts that quantify the uncertainty in wind output have the potential to aid in the economic grid integration of wind power at large penetration levels. In this paper, a novel probabilistic wind forecasting approach based on pinball loss optimization is proposed, in conjunction with a multi-model machine learning based ensemble deterministic forecasting framework. By assuming the point-forecasted value as the mean at each point, one unknown parameter (i.e., standard deviation) of a predictive distribution at each forecasting point is determined by minimizing the pinball loss. A surrogate model is developed to represent the unknown distribution parameter as a function of deterministic forecasts. This surrogate model can be used together with deterministic forecasts to predict the unknown distribution parameter and thereby generate probabilistic forecasts. Numerical results of case studies show that the proposed method has improved the pinball loss by up to 35% compared to a baseline quantile regression forecasting model.

Index Terms—probabilistic wind forecasting, optimization, surrogate model, machine learning, pinball loss.

I. INTRODUCTION

The uncertain and variable nature of wind imposes challenges to integrate wind power, particularly at large penetration levels. Improved wind forecasts are needed to assist power system planning and operations.

A number of wind forecasting technologies have been developed in the literature and have also been applied to a variety of power system operation and planning problems. For example, Lee *et al.* [1] used improved wind power forecasts to reduce the cost of system ancillary services (AS) and to conduct a system risk analysis. Botterud *et al.* [2] applied wind power forecasts in unit commitment (UC) and economic dispatch (ED) decision making to provide dynamic operating reserves, which benefits system operators and electricity traders. In electricity markets for energy, conventional deterministic forecasts are not sufficient to describe the inherent unpredictability of wind power, which accommodates through operating reserves. As a result, probabilistic forecasts that provide quantitative uncertainty information associated with wind power are expected to assist power system operations better. The output of a probabilistic wind forecast usually takes the form of probability distribution associated with point forecasts, namely the expectation. Methods of constructing predictive distributions can be mainly classified into parametric and non-parametric approaches in terms of distribution shape assumptions [3]. A prior assumption of the predictive

distribution shape is made via parametric methods. Once an analytical form of the predictive distribution is defined, parameters describing this distribution can be determined from data, which generally requires low computational cost. Distribution parameters can be estimated through different methods, and non-linear time series is one of the most popular used methods. For example, Pinson *et al.* [4] proposed a conditional parametric autoregression model to estimate the parameters of a Generalized Logit-Normal (GL-normal) distribution which is a discrete-continuous mixture of GL-normal distribution and two probability masses.

For distribution-free non-parametric approaches, the predictive distribution is estimated through a finite number of observations. Quantile regression (QR) and kernel density estimation (KDE) are traditional non-parametric probabilistic forecasting methods [5]. Haben *et al.* [6] proposed a non-parametric hybrid method, which combines the KDE and QR together to generate probabilistic load forecasts. Ordiano *et al.* [7] conducted probabilistic solar power forecasting using a Nearest-Neighbor based non-parametric method.

Pinball loss is one of most popular used metrics for evaluating the performance of probabilistic forecasting [8]. In this paper, a novel two-step probabilistic wind forecasting method is developed based on pinball loss optimization. First, deterministic forecasts are generated (with any deterministic forecasting methods). Second, a set of unknown parameters in the predictive distribution are optimized determined by minimizing the pinball loss. The optimal prediction distribution parameters are first determined in the training dataset. A surrogate model is developed to represent the unknown distribution parameter as a function of deterministic forecasts. At the forecasting state, the surrogate model is then used together with deterministic forecasts to predict the unknown distribution parameter and thereby generate probabilistic forecasts. The main contribution of this paper is to develop a two-step probabilistic wind forecasting methodology based on pinball loss optimization, which quantifies the uncertainties of wind speed/power and improves the forecasting accuracy.

The remainder of the paper is organized as follows. Section II describes the proposed probabilistic forecasting method, including a multi-distribution model and pinball loss based optimization process and a deterministic forecasting method. Section III applies the developed pinball loss based probabilistic forecasting method to multiple wind datasets, and compares

the forecasting performance with benchmark models. Concluding remarks and future work are discussed in Section IV.

II. METHODOLOGY

An optimal pinball loss based short-term probabilistic forecasting method is developed in this paper and the overall framework is illustrated in Fig. 2. This is a two-step probabilistic forecasting method, consisting of deterministic forecasts generation and predictive distribution (type and parameters) determination. A machine learning based multi-model forecasting framework (MMFF) is first adopted to generate short-term deterministic wind forecasts (i.e., 1-hour-ahead (1HA) here). To generate probabilistic forecasts, deterministic forecasts are considered as means of predictive distributions as described in Fig. 1, and unknown parameters of the predictive distributions are solved by minimizing pinball loss. The distribution with the minimum pinball loss in conjunction with a surrogate model, are used to generate probabilistic forecasts.

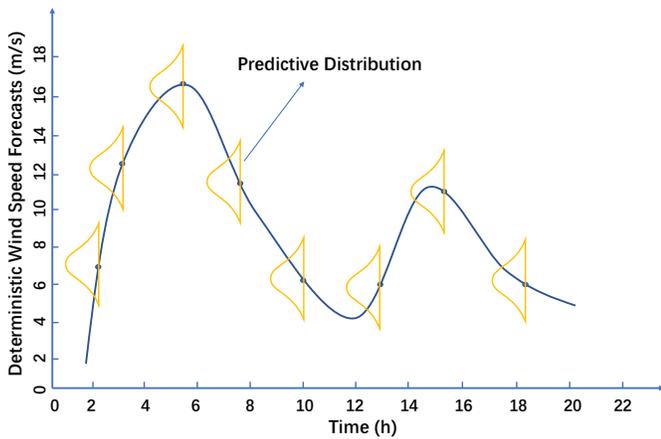


Fig. 1: Deterministic wind speed forecasts and predictive distribution

A. MMFF deterministic forecasting

The proposed pinball loss based probabilistic forecasting methodology is a two-step method, which can be applied with any deterministic wind forecasts. In this paper, a MMFF system consisting of an ensemble of four single machine learning algorithms with various kernels is adopted to generate deterministic forecasts. Details of the MMFF method can be found in [9].

B. Multi-distribution model

A multi-distribution database is formulated to model the possible shapes of the predictive distribution. Four distribution types are considered, which are Gaussian, Gamma, Laplace, and non-central t distributions. Probability density functions (PDFs) of the four aforementioned distributions are listed in [10]. Parameters used to describe the four predictive PDFs are all related to their mean and standard deviation values. Therefore, all of the PDFs can be represented in the form of mean μ and standard deviation σ , namely $f(x|\mu, \sigma)$. A

cumulative distribution function (CDF), $F(x|\mu, \sigma)$, can be deduced through the integration of a PDF.

C. Pinball loss based optimization

Pinball loss is one of the most popular metrics for evaluating probabilistic forecasts [8], and is a function of observations and quantiles of a forecast distribution. A smaller pinball loss value indicates a better probabilistic forecast.

$$L(q_m, x_i) = \begin{cases} (1 - \frac{m}{100}) \times (q_m - x_i), & x_i < q_m \\ \frac{m}{100} \times (x_i - q_m), & x_i \geq q_m \end{cases} \quad (1)$$

Where x_i represents the i th hour observation, m represents a quantile percentage from 1 to 99, and q_m represents the predicted quantile. For a given m percentage, the quantile q_m represents the value of random variable whose accumulated probability density (i.e., CDF) is m percentage. The quantiles of different distributions types are represented by a standard deviation σ . In this paper, the optimal standard deviation σ is determined by minimizing the sum of pinball loss function $L(\cdot)$, by considering appropriate constraints. A genetic algorithm (GA) is used to solve this optimization problem. GA is a widely used heuristic method for solving both constrained and unconstrained optimization problems [11]. In this study, the maximum number of iterations is set to 100 and the iteration stops if the improvement is less than 0.001. At each deterministic forecasting time point, an optimal standard deviation which minimizes the pinball loss of this single point is found accordingly. The optimization problem is formulated as follows.

$$\begin{aligned} \min_{\sigma} \sum_{m=1}^{99} L(q_m(\sigma), x_i) \\ \text{subject to} \\ \sigma_l \leq \sigma \leq \sigma_u \end{aligned} \quad (2)$$

Where σ_l and σ_u represent lower and upper bounds of the unknown standard deviation, respectively. In this paper, the lower and upper bounds of standard deviations are set to be 0 m/s and 10 m/s, respectively [12]. The distribution with the minimum pinball loss is selected as the predictive distribution shape. The optimal σ 's estimated using the training data are used to construct a surrogate model to be used in the forecasting stage.

D. Surrogate model

To generate probabilistic forecasts, an optimal standard deviation value is needed at every forecasting time point. To obtain this optimal standard deviation value, a surrogate model is developed to represent the optimal standard deviation as a function of deterministic forecasting value based on the training data, which is expressed by:

$$\hat{\sigma} = f(x_p) \quad (3)$$

Where x_p is a point forecast and $f(\cdot)$ is a surrogate model of the optimal standard deviation of the predictive distribution. A support vector regression method is used in this paper to

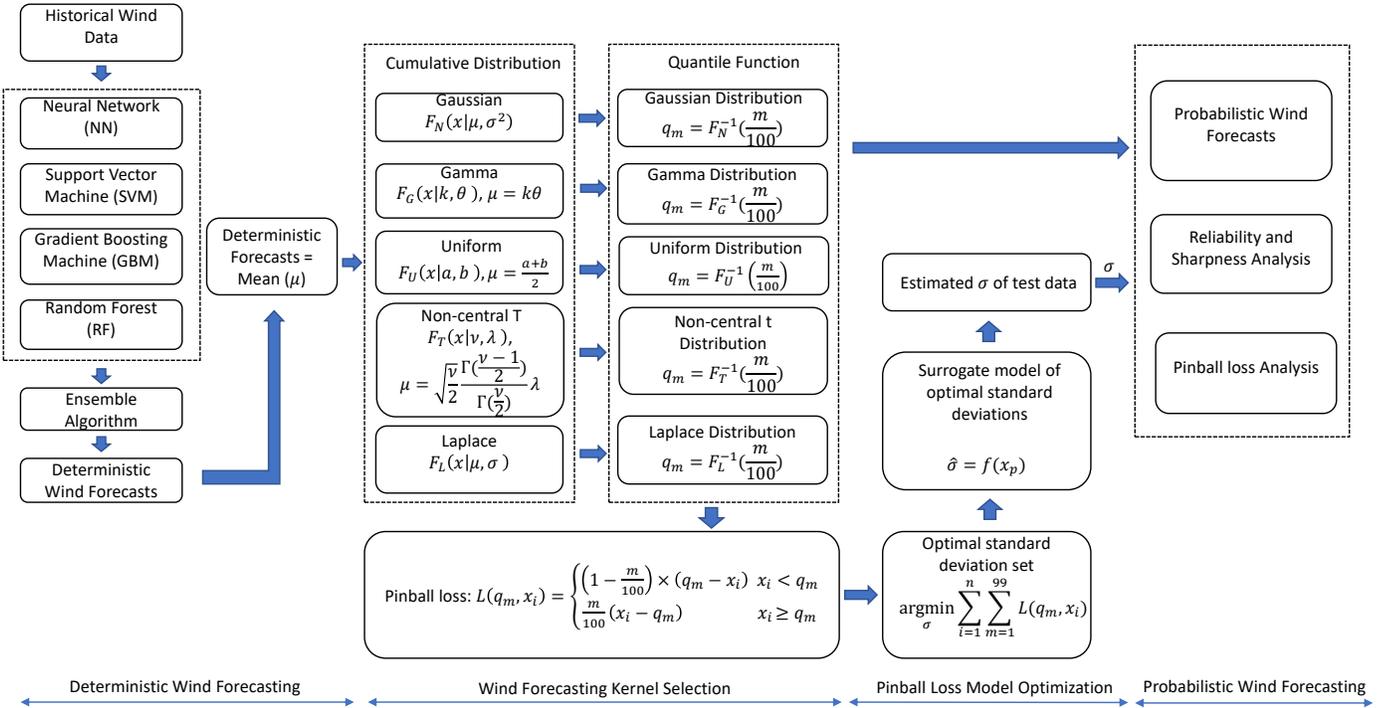


Fig. 2: Overall framework of the pinball loss based probabilistic wind forecasting framework

construct the surrogate model. This surrogate model is used to estimate the standard deviation of the predictive distribution in the forecasting stage, thereby generating the final probabilistic forecasts.

III. CASE STUDIES AND RESULTS

A. Data Summary

The proposed pinball loss based probabilistic forecasting approach is applied to 7 locations for wind speed forecasts. The wind speed data is collected near hub height with a 1-hour resolution [13]. The duration of collected data is summarized in Table I. For all locations, the first 2/3 of data is used as training data. The last 1/12 of the training data is used to build a surrogate model between optimal standard deviation and deterministic forecasts. The effectiveness of forecasts is evaluated by the remaining 1/3 of data. While the proposed method is capable of generating forecasts at multiple forecasting timescales, only 1HA forecasts are generated in this study.

TABLE I: Data duration at selected sites

Site	Data duration
Boulder_NWTC	2009-01-02 to 2012-12-31
Bovina50	2010-10-10 to 2012-10-08
Bovina100	2010-03-03 to 2012-03-01
CapeMay	2007-09-26 to 2009-09-24
CedarCreek_H06	2009-01-02 to 2012-12-31
Goodnoe_Hills	2007-01-01 to 2009-12-31
Megler	2010-11-03 to 2012-11-01

B. Pinball loss optimization results

Pinball loss values with different predictive distributions are listed in Table II. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the maximum wind speed at each site. A lower loss score indicates a better probabilistic forecast. It can be seen that the Laplace distribution with MMFF has the smallest pinball loss value at all locations except CedarCreek_H06. The lower pinball loss in CedarCreek_H06 using Laplace distribution with persistence method is mainly due to that the persistence deterministic forecasts perform better than MMFF forecasts. A quantile regression (QR) method, persistence-laplace (PS_Laplace) method with pinball loss optimization, and a MMFF-Laplace method without pinball loss optimization are used as baselines in case studies. The MMFF-Laplace forecasts have improved the pinball loss by up to 35% compared to the three benchmark models. Therefore, the Laplace distribution is finally chosen to generate probabilistic wind speed forecasts. It is also important to note that the methods of MMFF-Gaussian, MMFF-Gamma, and MMFF-Laplace perform similarly, which indicates that the optimization can help achieve a better accuracy with different predictive distribution types. For the baseline method of MMFF-Laplace without pinball loss optimization, a random standard deviation value is selected from the range between the minimum and maximum values of the optimal σ . We repeat this process 30 times to obtain an average sum of pinball loss without optimization. The training time of the MMFF-laplace method ranges from 1 to 2 hours, and the forecasting time

ranges from 2 to 5 minutes.

C. Deterministic forecasting results

Standard metrics of root mean squared error (RMSE), mean absolute error (MAE), and their corresponding normalized indices, *i.e.*, NMAE and NRMSE, are adopted to evaluate deterministic forecasting performance. For these metrics, a smaller value indicates better performance. Deterministic forecasting errors using MMFF at the selected locations are summarized in Table III. It is shown that the 1HA NMAE and NRMSE are in the range of 3%-5% and 4%-7%, respectively. An example of the forecasts at the Megler site from 2012-02-01 to 2012-02-04 are shown in Fig. 3. The persistence method is used as a baseline and the forecasting errors are also summarized in Table III. Overall, the accuracies of MMFF deterministic forecasts are better than those of persistence forecasts except CedarCreek_H06.

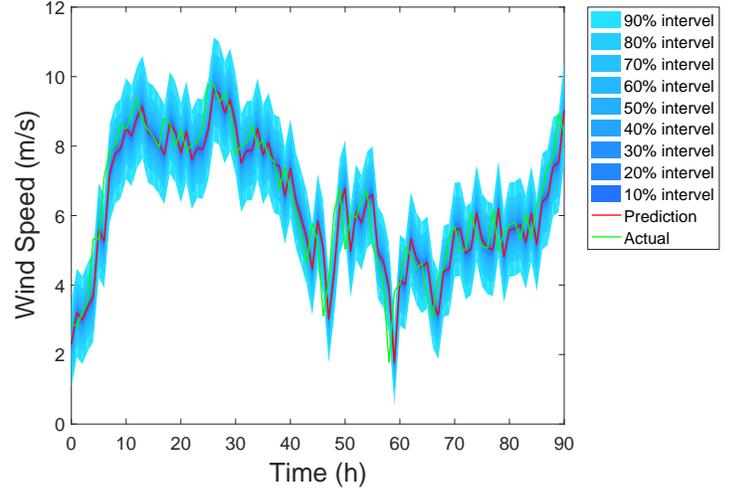
D. Probabilistic forecasting results

With estimated scale parameters through pinball loss minimization and surrogate modeling, predictive wind speed distributions are determined and the quantiles q_1, q_2, \dots, q_{99} can be calculated. To better visualize probabilistic forecasts, the 99 quantiles are converted into nine predictive intervals I_β ($\beta=10, \dots, 90$) in a 10% increment. Fig. 3(a) shows an example of probabilistic wind speed forecasts at the Megler site from 2012-02-01 to 2012-02-04. It is seen that the width of the predictive interval varies with the level of wind speed fluctuation. When the wind speed fluctuates significantly, the predictive interval tends to be wider, *i.e.*, the uncertainty in wind speed forecasts is relatively higher. Fig. 3(b) shows probabilistic forecasts generated from the baseline quantile regression method at the same site and time period. It is seen that the predictive intervals of the proposed MMFF-Laplace with pinball loss optimization method are narrower than those of the QR method. Thus, there is less uncertainty in the proposed probabilistic forecasts.

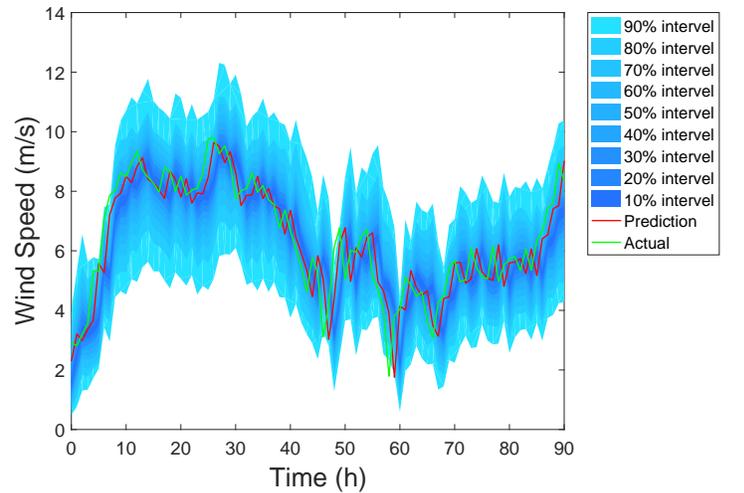
1) *Reliability*: Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies [14].

$$RE = \left[\frac{\xi^{(1-\alpha)}}{N} - (1 - \alpha) \right] \times 100\% \quad (4)$$

where N is the number of test samples, and $\xi^{(1-\alpha)}$ is the number of times that the actual test samples lie within the α th prediction interval. With measured empirical coverage, a reliability diagram can be plotted to describe the quantile forecast series with different nominal proportions. A reliability plot shows whether a given method tends to systematically underestimate or overestimate the uncertainty. In this study, the nominal coverage rates range from 10% to 90% with a 10% increment. Fig. 4 shows the reliability curves of probabilistic forecasts at the CapeMay, Megler, and Bovina50 sites. A forecast presents better reliability when the curve is closer to the diagonal. It is seen that overall the QR has better reliability performance, due to the fact that the confidence band of QR is



(a) MMFF-laplace with pinball loss optimization



(b) Quantile regression

Fig. 3: MMFF-Laplace and quantile regression forecasts at the Megler site

much wider than that of the proposed MMFF-Laplace method. A wider confidence band indicates that the results take more errors into consideration. However, it is important to note that the reliability over the 90th confidence intervals is similar between the proposed method and the baseline QR method, which is generally more important in probabilistic forecast applications in power system operations. Also, it is seen that overall the MMFF-Laplace with pinball loss optimization has much better reliability than that of the MMFF-Laplace without pinball loss optimization, which indicates effectiveness of the pinball loss optimization.

2) *Sharpness*: Sharpness indicates the capacity of a forecasting system to forecast extreme probabilities [15]. This criterion evaluates the predictions independently of the observations, which gives an indication of the level of usefulness of the predictions. For example, a system that only provides uniformly distributed predictions is less useful for decision making under uncertainty. Predictions with perfect

TABLE II: Normalized optimal sum of pinball loss

Site	Boulder_NWTC	Megler	CedarCreek_H06	Goodnoe_Hills	Bovina50	Bovina100	CapeMay
QR	2.22	1.76	2.03	1.96	2.56	2.44	1.95
MMFF_Gaussian	1.74	1.26	1.44	1.35	1.86	1.69	1.27
MMFF_Gamma	1.74	1.26	1.43	1.35	1.87	1.69	1.27
MMFF_Laplace	1.72	1.25	1.43	1.35	1.85	1.68	1.26
MMFF_noncentral_t	1.74	1.81	2.20	2.21	2.68	3.41	2.56
MMFF_Laplace (without opt)	2.94	2.93	2.40	2.39	3.53	3.08	2.45
PS_Laplace	1.81	1.29	1.34	1.38	1.92	1.69	1.32

Note: The smallest normalized optimal sum of pinball loss at each location is in boldface.

TABLE III: Deterministic forecasting results using MMFF and PS

Method	Site	Boulder_NWTC	Megler	CedarCreek_H06	Goodnoe_Hills	Bovina50	Bovina100	CapeMay
MMFF	NMAE(%)	4.71	3.36	3.86	3.72	5.00	4.56	3.39
	NRMSE(%)	6.87	4.65	5.43	5.09	6.64	6.28	4.74
PS	NMAE(%)	4.97	3.51	3.69	3.85	5.27	4.68	3.59
	NRMSE(%)	7.23	4.87	5.08	5.27	7.04	6.46	4.98

sharpness are discrete predictions with a probability of one (i.e., deterministic predictions). The sharpness is measured by the average size of the predictive intervals. The sharpness of the proposed pinball loss based MMFF-Laplace forecasts, QR, pinball loss based MMFF with other distribution types, and MMFF-Laplace without pinball loss optimization at the CapeMay, Megler, and Bovina50 site are compared in Fig. 5. It is seen that the sharpness of pinball loss based forecasts are better than that of the baseline QR method. It is also observed that the expected intervals size increases with increasing nominal coverage rate. Also, the MMFF-Laplace with pinball loss optimization has much better sharpness than that of MMFF-Laplace without pinball loss optimization. The intervals size of pinball loss based MMFF-Laplace forecasts ranges from 2% up to 18%, which indicates low sharpness.

IV. CONCLUSION

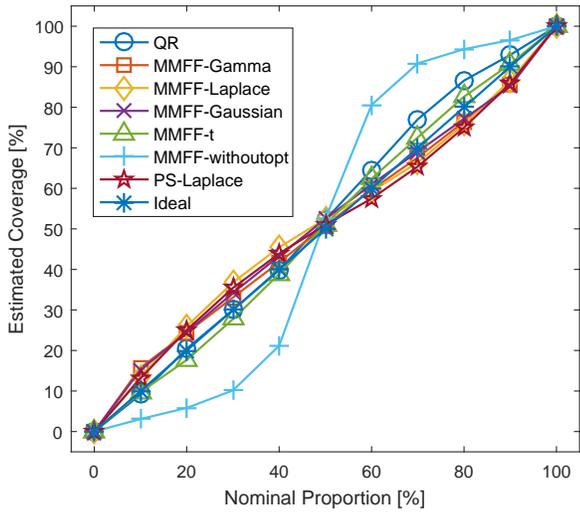
In this paper, an optimal pinball loss based probabilistic wind forecasting method was developed, in conjunction with a multi-model deterministic forecasting framework. Different shapes of predictive distributions are tested and compared, including Gaussian, Gamma, Laplace, and non-central t distributions. The optimal shape parameter of the predictive distribution is determined by minimizing the sum of pinball loss using training data. This optimal shape parameter is used in the forecasting stage through surrogate modeling. We found that the laplace distribution presents the best pinball loss. Results showed that the proposed probabilistic forecasting method could reduce the pinball loss by up to 35% compared to the baseline methods. The relationship between the accuracy of deterministic and probabilistic forecasts will be explored in future work.

V. ACKNOWLEDGEMENT

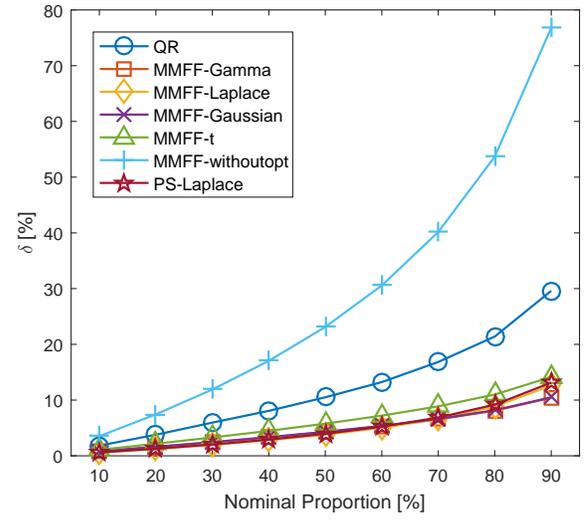
This work was supported by the National Renewable Energy Laboratory under Subcontract No. XGJ-6-62183-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308).

REFERENCES

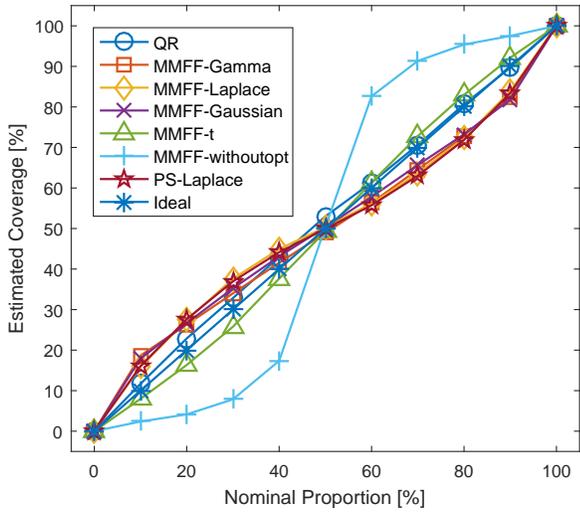
- [1] D. Lee *et al.*, “Wind power forecasting and its applications to the power system,” Ph.D. dissertation, 2015.
- [2] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, “Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 250–261, 2013.
- [3] S. Lefèvre, C. Sun, R. Bajcsy, and C. Laugier, “Comparison of parametric and non-parametric approaches for vehicle speed prediction,” in *American Control Conference (ACC), 2014.* IEEE, 2014, pp. 3494–3499.
- [4] 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, no. 4, pp. 555–576, 2012.
- [5] Y. Zhang, J. Wang, and X. Wang, “Review on probabilistic forecasting of wind power generation,” *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.
- [6] S. Haben and G. Giasemidis, “A hybrid model of kernel density estimation and quantile regression for gefcom2014 probabilistic load forecasting,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 1017–1022, 2016.
- [7] J. Á. G. Ordiano, W. Doneit, S. Waczowicz, L. Gröll, R. Mikut, and V. Hagenmeyer, “Nearest-neighbor based non-parametric probabilistic forecasting with applications in photovoltaic systems,” *arXiv preprint arXiv:1701.06463*, 2017.
- [8] I. Steinwart, A. Christmann *et al.*, “Estimating conditional quantiles with the help of the pinball loss,” *Bernoulli*, vol. 17, no. 1, pp. 211–225, 2011.
- [9] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, “A data-driven multi-model methodology with deep feature selection for short-term wind forecasting,” *Applied Energy*, vol. 190, pp. 1245–1257, 2017.
- [10] M. Evans, N. Hastings, and B. Peacock, “Statistical distributions,” 2000.
- [11] L. Scrucca *et al.*, “Ga: a package for genetic algorithms in r,” *Journal of Statistical Software*, vol. 53, no. 4, pp. 1–37, 2013.
- [12] C. H. Hansen, C. J. Doolan, and K. L. Hansen, *Wind Farm Noise: Measurement, Assessment, and Control.* John Wiley & Sons, 2017.
- [13] J. Zhang, C. Draxl, T. Hopson, L. Delle Monache, E. Vanvyve, and B.-M. Hodge, “Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods,” *Applied Energy*, vol. 156, pp. 528–541, 2015.
- [14] J. Juban, N. Siebert, and G. N. Kariniotakis, “Probabilistic short-term wind power forecasting for the optimal management of wind generation,” in *Power Tech, 2007 IEEE Lausanne.* IEEE, 2007, pp. 683–688.
- [15] C. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, “Online quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power,” *Energy*, vol. 113, pp. 355–365, 2016.



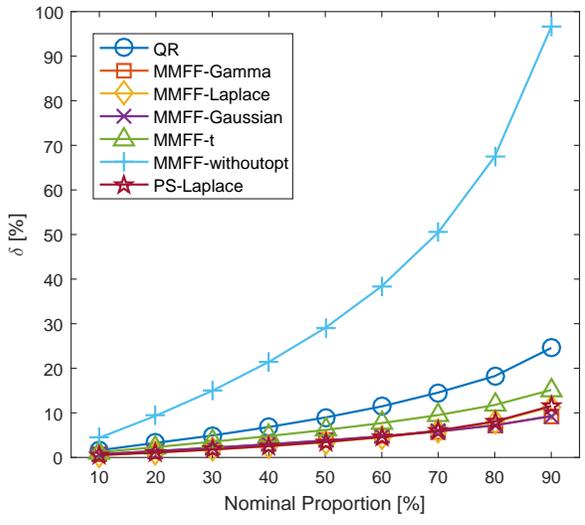
(a) CapeMay



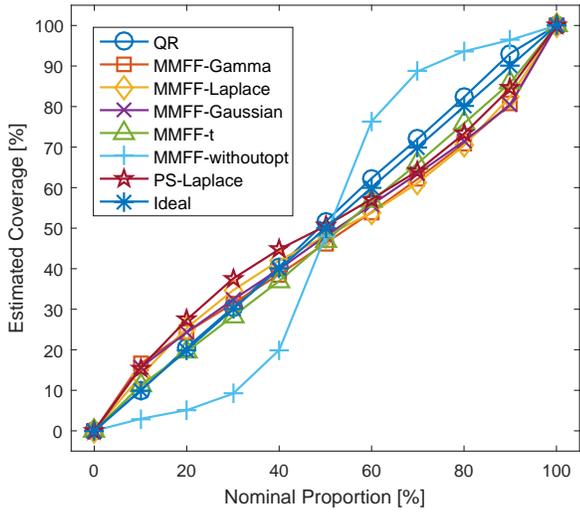
(a) CapeMay



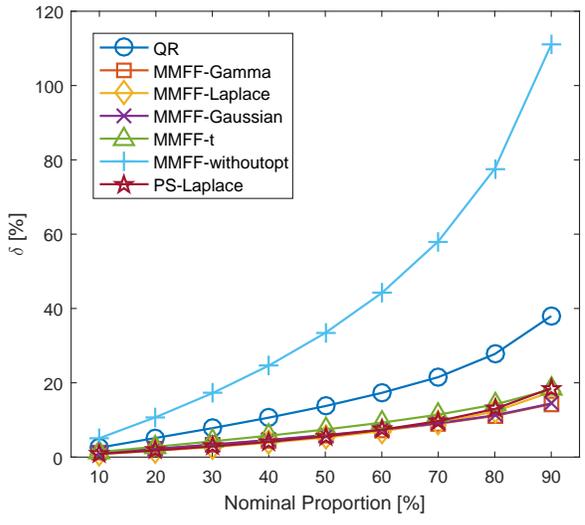
(b) Megler



(b) Megler



(c) Bovina50



(c) Bovina50

Fig. 4: Reliability of probabilistic forecasts on selected sites

Fig. 5: Sharpness of probabilistic forecasts on selected sites