A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization

Mucun Sun[a], Cong Feng[a], Erol Kevin Chartan[b], Bri-Mathias Hodge[b,c], Jie Zhang[a,*]

[a] University of Texas at Dallas, Richardson, TX 75080, USA
[b] National Renewable Energy Laboratory, Golden, CO 80401, USA
[c] University of Colorado Boulder, Boulder, CO 80309, USA

HIGHLIGHTS

- A pinball loss optimization based probabilistic forecasting method is developed.
- The best shape of a predictive distribution is explored and optimized.
- The proposed method reduces pinball loss by up to 35% compared to baselines.

ABSTRACT

With increasing wind penetrations into electric power systems, probabilistic wind forecasting becomes more critical to power system operations because of its capability of quantifying wind uncertainties. In this paper, a two-step probabilistic wind forecasting approach based on pinball loss optimization is developed. First, a multimodel machine learning-based ensemble deterministic forecasting framework is adopted to generate deterministic forecasts. The deterministic forecast is assumed to be the mean value of the predictive distribution at each forecasting time stamp. Then, the optimal unknown parameter (i.e., standard deviation) of the predictive distribution is estimated by a support vector regression surrogate model based on the deterministic forecasts. Finally, probabilistic forecasts are generated from the predictive distribution. Numerical results of case studies at eight locations show that the developed two-step probabilistic forecasting methodology has improved the pinball loss metric score by up to 35% compared to a baseline quantile regression forecasting model.

1. Introduction

The uncertain and variable nature of wind imposes challenges on the grid integration of wind power, particularly at high penetration levels. Wind forecasting plays an important role in reducing the uncertainty of wind power output in operations. This can be useful at different time horizons, from day-ahead for unit commitment to minutes- and hours-ahead for economic dispatch. Probabilistic wind power forecasts provide even more information about the possible wind generation output, thus their inclusion directly in system operations is an active research area.

1.1. Literature review

A number of wind forecasting technologies have been developed in the literature to assist power system operation and planning. For

example, Lee et al. [1] used improved wind power forecasts to reduce the cost of system ancillary services and conduct a system risk analysis. Botterud et al. [2] applied wind power forecasts in unit commitment and economic dispatch decision-making to provide dynamic operating reserves, which provided benefits to system operators and electricity traders. In electricity markets, it has been found that conventional deterministic forecasts might not be sufficient to characterize the inherent uncertainty of wind power. Probabilistic forecasts that provide quantitative uncertainty information associated with wind power are therefore expected to better assist power system operations. Probabilistic wind forecasts usually take the form of probability distributions associated with point forecasts, namely, the expectation. Existing methods of constructing predictive distributions can be mainly classified into parametric and nonparametric approaches in terms of distribution shape assumptions [3]. A prior assumption of the predictive distribution shape is made in parametric methods, and unknown

distribution parameters are estimated based on historical data. Parametric approaches generally require low computational cost. Gaussian [4] and beta [5] distributions are two commonly used predictive distributions in probabilistic wind forecasting. However, Gaussian and beta distributions could not capture the fat tails and double bounded properties of wind power distribution. To better account for the nonlinear and double-bounded properties of wind power generation in short-term probabilistic forecasting, Pinson et al. [6] proposed a generalized logit-normal (GL-normal) distribution.

Once an analytical form of the predictive distribution is defined, distribution parameters can be estimated by using different methods. For estimators of local parameters, non-linear time series is one of the most popular categories of methods. For example, Pinson et al. [6] developed a conditional parametric auto-regression model to estimate the parameters of a GL-normal distribution, which is a discrete-continuous mixture of the GL-normal distribution and two probability masses. For estimators of scale parameter, autoregression-generalized autoregressive with conditional heteroscedasticity models are one of the most popular used methods [7]. Other methods to estimate distribution parameters in probabilistic wind forecasts include maximum likelihood method [8], least squares method [9], method of moments [10], and the fast Bayesian approach [11]. Overall, none of the distribution parameter estimation methods mentioned above aims at optimizing the probabilistic forecasting metrics (e.g., pinball loss). While it is challenging to select a universal predictive distribution under different wind conditions, this paper seeks to explore the hypothesis that adaptively optimizing the prediction distribution (e.g., standard deviation) could further improve the performance of probabilistic forecasting.

In addition to parametric approaches, nonparametric approach is another way to provide probabilistic forecasts. Instead of assuming a predictive distribution, the quantiles are estimated through a finite number of observations. Quantile regression (QR) is one of the traditional nonparametric probabilistic forecasting methods [12]. However, the widely used QR is a direct function of the point forecast and predictors, and it can only provide the range of the given percentage [13]. Haben et al. [14] developed a nonparametric hybrid method that combines KDE and QR to generate probabilistic load forecasts. Ordiano et al. [15] conducted probabilistic solar power forecasting using a nearest-neighbor-based nonparametric method. Most existing parametric and nonparametric probabilistic wind forecasting approaches focus on statistical methods. In addition to traditional statistical approaches, the performance of probabilistic wind forecasting can be further improved by machine learning techniques. For example, Wan et al. [16] used an extreme learning machine to predict the optimal prediction interval without using statistical inferences and distribution assumptions. In the Global Energy Forecasting Competition 2014 (GEFCom2014), Landry et al. [17] used gradient-boosted machines (GBM) for multiple quantile regression to fit each quantile and zone independently and generate probabilistic forecasts. Zhang et al. [18] developed a probabilistic forecasting method based on k-nearest neighbor point forecasts through KDE. Wang et al. [19] used deep convolutional neural network and wavelet transform to quantify the wind power uncertainties with respect to model misspecification and data noise.

1.2. Research objective

To adaptively optimize the predictive distribution shape and combine the advantages of statistical and machine learning approaches, this paper develops a two-step probabilistic forecasting method based on pinball loss optimization. Pinball loss is one of the most popular metrics for evaluating the performance of probabilistic forecasting [20]. First, deterministic forecasts are generated by a machine learning-based multi-model (M3) forecasting framework. Second, a set of unknown parameters in the predictive distribution are determined by minimizing

the pinball loss using the generic algorithm. *Note that the optimal distribution parameter is adaptive and dynamically updated based on the point forecast value at each time stamp.* The optimal adaptive predictive distribution parameters are first determined offline with the historical training data. Then a surrogate model is developed to represent the optimizeddistribution parameter as a function of the deterministic forecast. At the online forecasting stage, the surrogate model is used together with deterministic forecasts to adaptively predict the unknown distribution parameters and thereby generate probabilistic forecasts. The main contributions of this paper include:

- Develop a two-step probabilistic wind forecasting methodology based on pinball loss optimization.
- Select the best predictive distribution and adaptively optimize the predictive distribution shape simultaneously.
- Explore the relationship between deterministic and probabilistic forecasting accuracies.

The remainder of the paper is organized as follows. Section 2 describes the proposed pinball loss-based probabilistic forecasting method, including a multi-distribution database, a pinball loss-based optimization process, and a deterministic forecasting method. Section 3 validates the effectiveness of the proposed method by means of a comparison with multiple benchmark models at eight locations. The relationship between the deterministic and probabilistic forecasting accuracies is also explored in this section. Concluding remarks and future work are discussed in Section 4.

2. Pinball loss-based short-term probabilistic forecasting

The overall framework of the proposed optimal pinball loss-based short-term probabilistic forecasting method is illustrated in Fig. 1. This is a two-step probabilistic forecasting method, consisting of deterministic forecast generation and predictive distribution (type and parameters) determination. In the first step, the machine learning-based multimodel (M3) forecasting framework is adopted to generate short-term deterministic wind forecasts (i.e., 1-h-ahead), which are considered as a means of predictive distributions at each forecasting time stamp. In the second step, a set of optimal standard deviation values are determined through pinball loss optimization at the training stage. The relationship between the deterministic forecasts and the corresponding optimal standard deviations is quantified through a SVR surrogate model. At the forecasting stage, we generate deterministic forecasts first. Then, we feed the deterministic forecasts to the surrogate model built at the training stage to estimate a new set of standard deviation values (pseudo-optimal standard deviation). Finally, these estimated pseudo-optimal standard deviation values and the deterministic forecasts are used together to generate probabilistic forecasts. The pseudodocode of the probabilistic forecasting model is illustrated in Algorithm 1.

Algorithm 1. M3 probabilistic forecasting method based on pinball loss optimization

---

Data: Deterministic wind power forecasts
Result: Probabilistic forecasts
1 Initialization: Obtain PDF of a single model and represent it in the form of mean μ and standard deviation σ as $f(x|\mu, \sigma)$;
2 Calculate CDF $F(x|\mu, \sigma)$ of the predictive distribution;
3 Calculate quantile function through $q_i(F^{-1}(x|\mu, \sigma, \frac{i}{100}))$;
4 Calculate optimal σ's through pinball loss optimization;
5 Build a surrogate model between deterministic forecasts and optimal σ's;
6 Estimate pseudo-optimal σ based on deterministic forecasts;
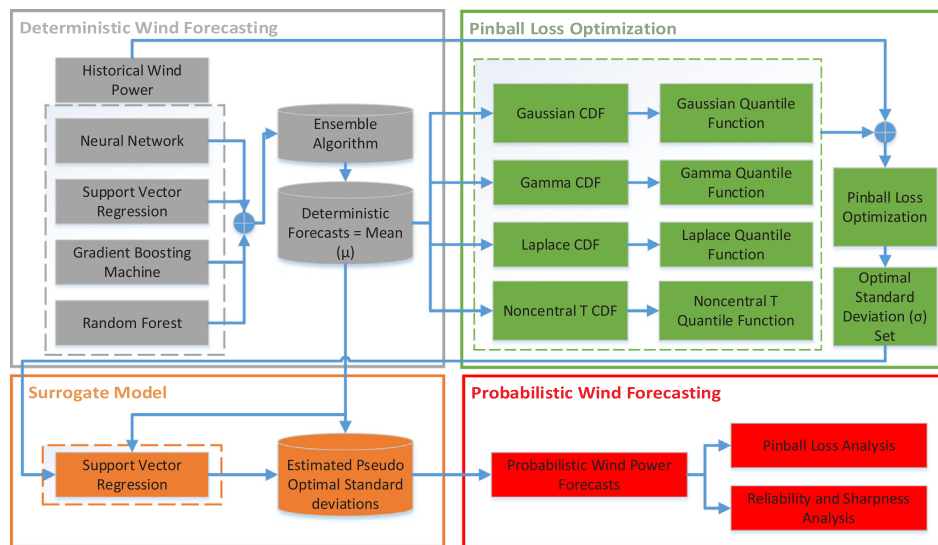7 Generate probabilistic forecasts.

Fig. 1. Overall framework of the pinball loss optimization-based probabilistic wind forecasting framework.

2.1. Machine learning-based multimodel (M3) deterministic forecasting

M3, a two-layer, short-term forecasting method, is adopted to generate deterministic forecasts. Multiple sets of deterministic forecasts are generated by using different machine learning algorithms with different kernels in the first layer. Then, the forecasts are blended by another machine learning algorithm in the second layer to generate the final forecasts. Machine learning algorithms used in the M3 method include artificial neural networks (ANNs), support vector regression (SVR), GBMs, and random forests. Details about the M3 method can be found in [21].

2.2. Multi-distribution database

A multi-distribution database is formulated to model the possible shapes of the predictive distribution. Four widely used predictive distribution types are considered: Gaussian, Gamma, Laplace, and noncentral t distributions. Probability density functions (PDFs) of the four distribution types are summarized in [22]. PDFs of the four distributions can be represented by the mean μ and standard deviation σ as $f(x|\mu, \sigma)$, and the corresponding cumulative distribution functions (CDFs) can be deduced and denoted as $F(x|\mu, \sigma)$.

2.3. Pinball loss-based optimization

Pinball loss is one of the most popular metrics for evaluating probabilistic forecasts [20]; it is a function of observations and quantiles of a forecast distribution. A smaller pinball loss value indicates better probabilistic forecasting. The pinball loss value of a certain quantile L_m is expressed as:

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

where x_i represents the ith 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 a random variable whose CDF is m percentage. The quantiles of different distribution types are represented by a standard deviation σ, denoted as $q_m(\sigma)$. At the offline training stage where x's are available, the optimal standard deviation σ is determined by minimizing the pinball loss summation of the 1st to 99th quantiles at each point, which is formulated as follows:

$$\min_\sigma \sum_{m=1}^{99} L(q_m(\sigma), x_i)$$

subject to   (2)

$$\sigma_l \leqslant \sigma \leqslant \sigma_u$$

where σ_l and σ_u represent the lower and upper bounds of the unknown standard deviation, which are selected based on the forecasting target [23]. The genetic algorithm [24] is adopted in this paper to solve this optimization problem. In this study, the maximum number of iterations is set to be 100, and the iteration stops if the improvement is less than 0.001. The distribution with the minimum pinball loss is selected as the predictive distribution shape. The optimal standard deviation's estimated using the training data are used to construct a surrogate model to be used at the forecasting stage.

2.4. Surrogate model

To generate probabilistic forecasts, a pseudo-optimal standard deviation value is needed at every forecasting time point, which is estimated by a surrogate model. Several possible surrogate model types can be used, such as SVR, radial basis function, kriging, and ANN. SVR is adopted in this paper because it is more accurate than other surrogate models in the case studies. The surrogate model is constructed by fitting the optimal standard deviation as a function of the deterministic forecasting value, which is expressed by:

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

where x_p is a point forecast, and $f(\cdot)$ is the SVR surrogate model of the optimal standard deviation of the predictive distribution. This surrogate model is used to estimate the standard deviation of the predictive distribution at the online forecasting stage.

3. Case studies and results

3.1. Data summary

The proposed pinball loss optimization-based probabilistic forecasting approach is applied to eight locations to generate wind speed forecasts. The wind speed data were collected near hub height with 1-h resolution [25]. The duration and measurement height of the collected data at all locations are summarized in Table 1. For all locations, the first 2/3 of data are used as training data, in which the first 11/12 is

Table 1
Data duration at selected sites.

Case No.	Site	Data duration	Height (m)
C1	Boulder_NWTC	2009-01-02 to 2012-12-31	80
C2	Megler	2010-11-03 to 2012-11-01	53.3
C3	CedarCreek_H06	2009-01-02 to 2012-12-31	69
C4	Goodnoe_Hills	2007-01-01 to 2009-12-31	59.4
C5	Bovina50	2010-10-10 to 2012-10-08	50
C6	Bovina100	2010-03-03 to 2012-03-01	100
C7	CapeMay	2007-09-26 to 2009-09-24	100
C8	Cochran	2008-06-30 to 2011-06-29	70

used to train M3 and the remaining 1/12 of the training data is used to build the SVR surrogate model of the optimal standard deviation. The accuracy of the forecasts is evaluated by the remaining 1/3 of data. Although the proposed method is capable of generating forecasts at multiple forecasting horizons, only 1-h-ahead forecasts are explored in this study.

3.2. Deterministic forecasting results

Standard metrics root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), normalized RMSE (NRMSE), and normalized MAE (NMAE), are adopted to evaluate the performance of deterministic forecasts. They are defined by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}} \tag{4}$$

$$NRMSE = \frac{1}{x_{max}}\sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - x_i| \tag{6}$$

$$NMAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{x}_i - x_i}{x_{max}}\right| \tag{7}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right| \tag{8}$$

where $\hat{x}_i$ is the forecasted wind speed, x_i is the actual wind speed, x_{max} is the maximum actual wind speed, and n is the sample size.

For these metrics, a smaller value indicates better performance. Deterministic forecasting errors using M3 at the selected locations are summarized in Table 2. It is seen that the 1-h-ahead NMAE, NRMSE, and MAPE are in the range of 3–5%, 4–7%, and 18–40%, respectively. Two examples of the forecasts at the C2 and C5 sites from 2012-02-01 to 2012-02-04 are illustrated in Figs. 2 and 3, respectively. The

Table 2
Deterministic forecasting results using M3 and PS.

Method	Metric	Site							
		C1	C2	C3	C4	C5	C6	C7	C8
M3	MAE (m/s)	1.32	0.69	1.26	0.99	1.03	1.10	0.94	0.78
	NMAE (%)	4.77	2.89	3.86	3.72	4.89	4.47	3.45	3.98
	RMSE (m/s)	1.93	0.96	1.78	1.35	1.37	1.53	1.31	1.06
	NRMSE (%)	6.95	3.99	5.48	5.10	6.54	6.22	4.81	5.41
	MAPE (%)	0.40	0.21	0.18	0.19	0.24	0.18	0.27	0.19
PS	MAE (m/s)	1.38	0.84	1.20	1.02	1.10	1.15	0.98	0.85
	NMAE (%)	4.97	3.51	3.69	3.85	5.27	4.69	3.59	4.37
	RMSE (m/s)	2.00	1.17	1.65	1.40	1.48	1.59	1.36	1.16
	NRMSE (%)	7.23	4.87	5.08	5.27	7.04	6.46	4.98	5.92
	MAPE (%)	0.41	0.24	0.17	0.20	0.25	0.20	0.29	0.23

persistence (PS) method is used as a baseline, and the forecasting errors are also summarized in Table 2. Overall, the accuracies of the M3 deterministic forecasts are better than those of persistence forecasts, except C3. The smallest NMAE, NRMSE, and MAPE at C3 are produced by the model using the persistence method. This is mainly because C3 has less wind speed variance.

3.3. Surrogate model accuracy

SVR is adopted in this paper to build the surrogate model of the optimal predictive distribution parameter (i.e., standard deviation). The NMAE and NRMSE of surrogate modeling are summarized in Table 3. A smaller NMAE/NRMSE value indicates better performance. It is observed that the NMAE and NRMSE are in the range of 5–9% and 8–14%, respectively. Overall, the accuracy of the SVR surrogate model is satisfactory.

3.4. Pinball loss optimization results

Pinball loss values with different predictive distributions are summarized in Table 4. 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. Table 4 shows that the M3-Laplace with pinball loss optimization (M3-Laplace) has the smallest pinball loss value at all locations except C3. The smallest pinball loss at C3 is produced by the model using Laplace distribution with the persistence method (PS-Laplace). This is mainly because that the persistence deterministic forecasts are more accurate than the M3 forecasts at this location. *The models of quantile regression, PS-Laplace with pinball loss optimization, and M3-Laplace without pinball loss optimization (M3-Laplace-w) are used as benchmark models in case studies. The reasons for choosing these three baselines are: (i) quantile regression is a widely used method in probabilistic forecasts; (ii) the PS-Laplace method allows us to explore the impacts of point forecasts on this two-step probabilistic forecasting framework; (iii) the M3-Laplace-w method allows us to explore the effectiveness of pinball loss optimization.* Results show that the M3-Laplace model has improved the pinball loss by up to 35% compared to the three benchmark models and M3 forecasts with other predictive distributions (i.e., M3-Gaussian, M3-Gamma, and M3-Noncentral T (M3-ncT)). Therefore, the Laplace distribution is finally chosen to generate probabilistic wind speed forecasts. Note that the models of M3-Gaussian, M3-Gamma, and M3-Laplace perform similarly, which indicates that the optimization can help achieve better accuracies with different predictive distribution types. For the baseline model of M3-Laplace-w, a random standard deviation value is selected within 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.

3.5. Probabilistic forecasting results

With estimated scale parameters through pinball loss minimization and surrogate modeling, predictive wind speed distributions are determined and the quantiles $q_1, q_2, ..., q_{99}$ can be calculated. To better visualize probabilistic forecasts, the 99 quantiles are converted into nine prediction intervals I_β ($\beta = 10, ..., 90$) in a 10% increment. Figs. 2(a) and 3(a) show two examples of probabilistic wind speed forecasts at the C2 and C5 site from 2012-02-01 to 2012-02-04. The width of the prediction interval varies with the wind speed variability. When the wind speed fluctuates frequently, the prediction interval tends to be wider, and thereby the uncertainty in wind speed forecasts is relatively higher. Figs. 2(b) and 3(b) show probabilistic forecasts generated from the baseline quantile regression method at the same sites and time periods. The prediction intervals of the proposed M3-Laplace method are **narrower** than those of the quantile regression method. Thus, there is less uncertainty in the M3-Laplace probabilistic
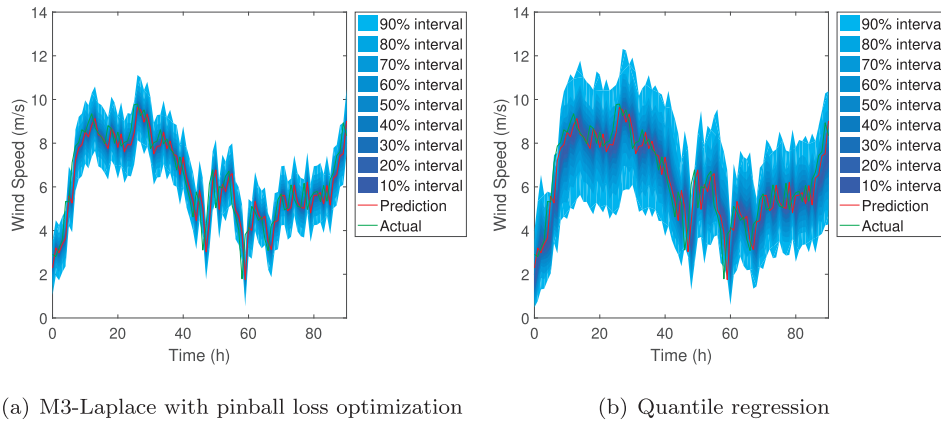
(a) M3-Laplace with pinball loss optimization (b) Quantile regression

Fig. 2. M3-Laplace and quantile regression forecasts at the C2 site.

wind forecasts.

In addition to pinball loss, two more standard metrics, i.e., sharpness and reliability, are used to assess the probabilistic forecasting accuracy.

3.5.1. Sharpness

Sharpness indicates the capacity of a forecasting system to forecast wind power with extreme probability [26]. The sharpness is measured by the average size of the prediction intervals. The sharpness of the proposed M3-Laplace model, quantile regression, M3 with other distribution types, and M3-Laplace-w at the eight sites are compared in Fig. 4. The sharpness of the pinball loss-based forecasts is better than that of the baseline quantile regression model. Also, the expected interval size increases with increasing nominal coverage rate, and the M3-Laplace has much better sharpness than that of the M3-Laplace-w. The interval size of the M3-Laplace forecasts ranges from 2% to 18%, which indicates low sharpness.

3.5.2. Reliability

Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies [27]. A reliability plot shows whether a given method tends to systematically underestimate or overestimate the uncertainty. In this study, the nominal coverage rate ranges from 10% to 90% with a 10% increment. Fig. 5 shows the reliability plots of the probabilistic forecasts at the eight test sites. A forecast presents better reliability when the curve is closer to the diagonal. Fig. 5 shows that overall quantile regression has better reliability performance, because the confidence band of QR is much wider than that of the proposed M3-Laplace method. A wider confidence band indicates that the result takes more errors into consideration; however,

Table 3
NMAE and NRMSE of the SVR surrogate model.

Metrics	Site							
	C1	C2	C3	C4	C5	C6	C7	C8
NMAE (%)	8.57	5.26	8.64	7.82	6.89	7.50	6.87	7.84
NRMSE (%)	13.48	8.16	13.27	11.51	10.13	11.15	10.64	11.96

Table 4
Normalized optimal averaged sum of pinball loss.

Model	Site							
	C1	C2	C3	C4	C5	C6	C7	C8
QR	2.22	1.76	2.03	1.96	2.56	2.44	1.95	1.68
M3-Gaussian	1.74	1.26	1.44	1.36	1.86	1.69	1.27	1.59
M3-Gamma	1.74	1.26	1.43	1.36	1.87	1.69	1.27	1.58
M3-Laplace	**1.72**	**1.25**	1.43	**1.35**	**1.85**	**1.63**	**1.26**	**1.57**
M3-ncT	1.74	1.81	2.20	2.21	2.68	3.41	2.56	2.86
M3-Laplace-w	2.94	2.93	2.40	2.39	3.53	3.08	2.46	2.72
PS-Laplace	1.81	1.29	**1.34**	1.38	1.92	1.70	1.32	1.64

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

note that the reliability over the 90th confidence interval is similar between M3-Laplace and the quantile regression, which is generally more important in probabilistic forecasting applications in power system operations. Also, the M3-Laplace has much better reliability than M3-Laplace-w at all selected locations, which indicates the effectiveness of the pinball loss optimization.
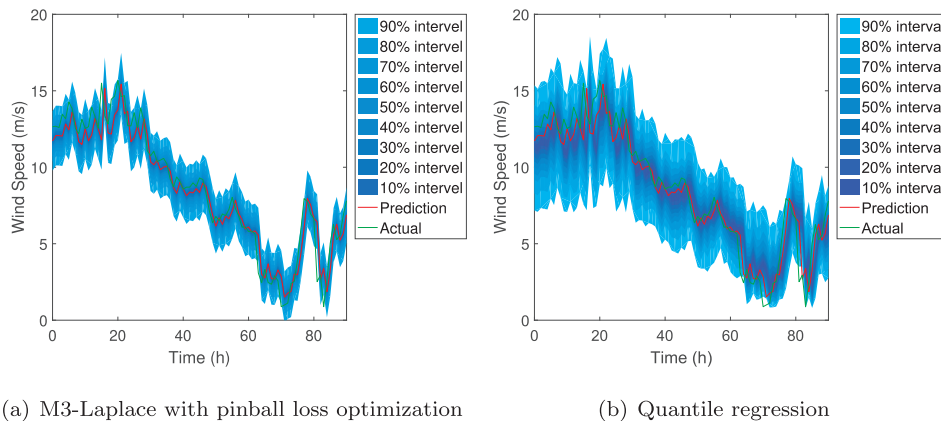

(a) M3-Laplace with pinball loss optimization (b) Quantile regression

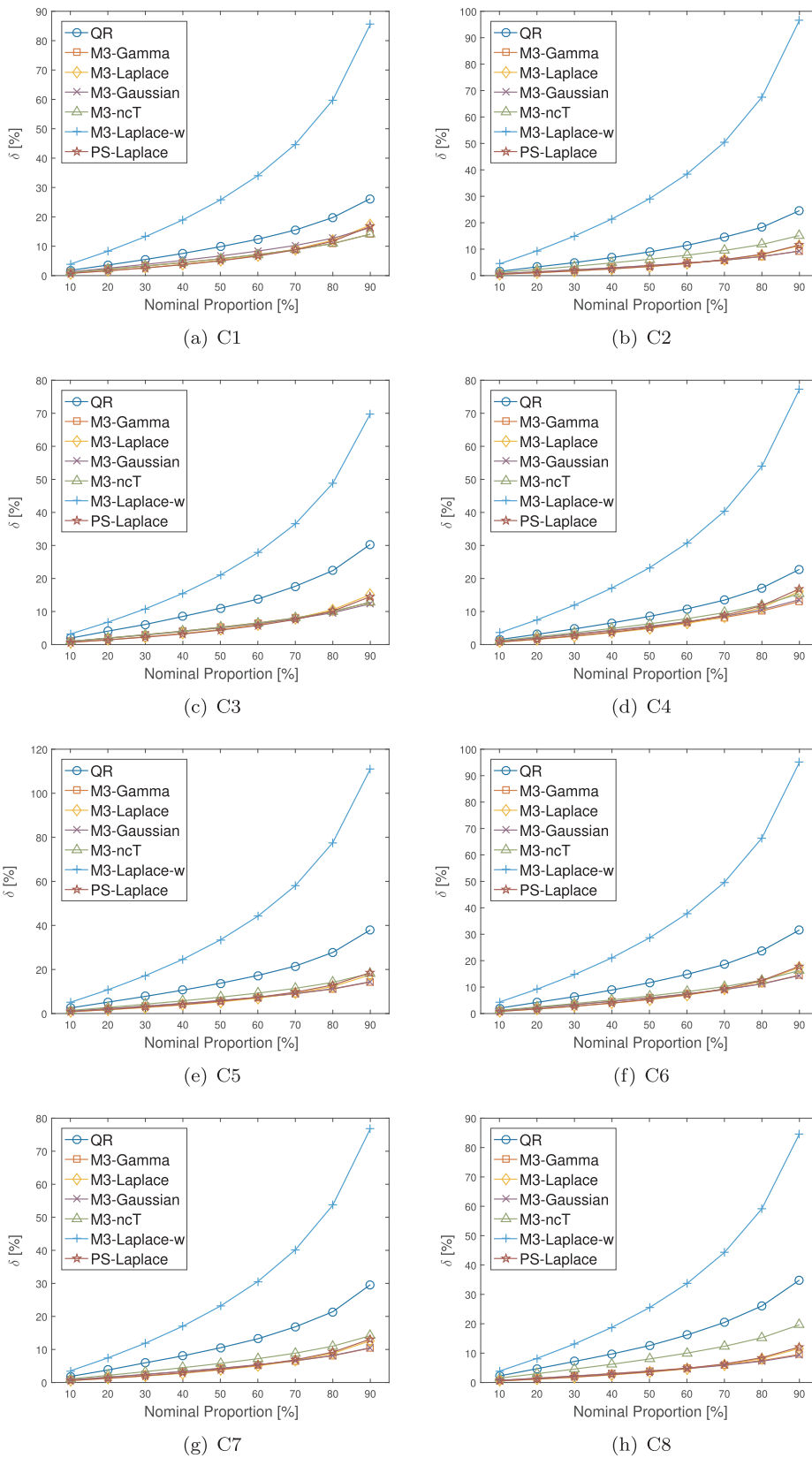**Fig. 3.** M3-Laplace and quantile regression forecasts at the C5 site.

Fig. 4. Sharpness of probabilistic forecasts at selected sites.

3.6. Relationship between deterministic and probabilistic forecasts

Because the proposed method is a two-step probabilistic forecasting approach, it is interesting to explore the inherent relationship between the first deterministic forecasting step and the second probabilistic forecasting step. To this end, the relationship between a deterministic forecasting metric and a probabilistic forecasting metric is quantified. NMAE is used to represent the deterministic forecasting accuracy,
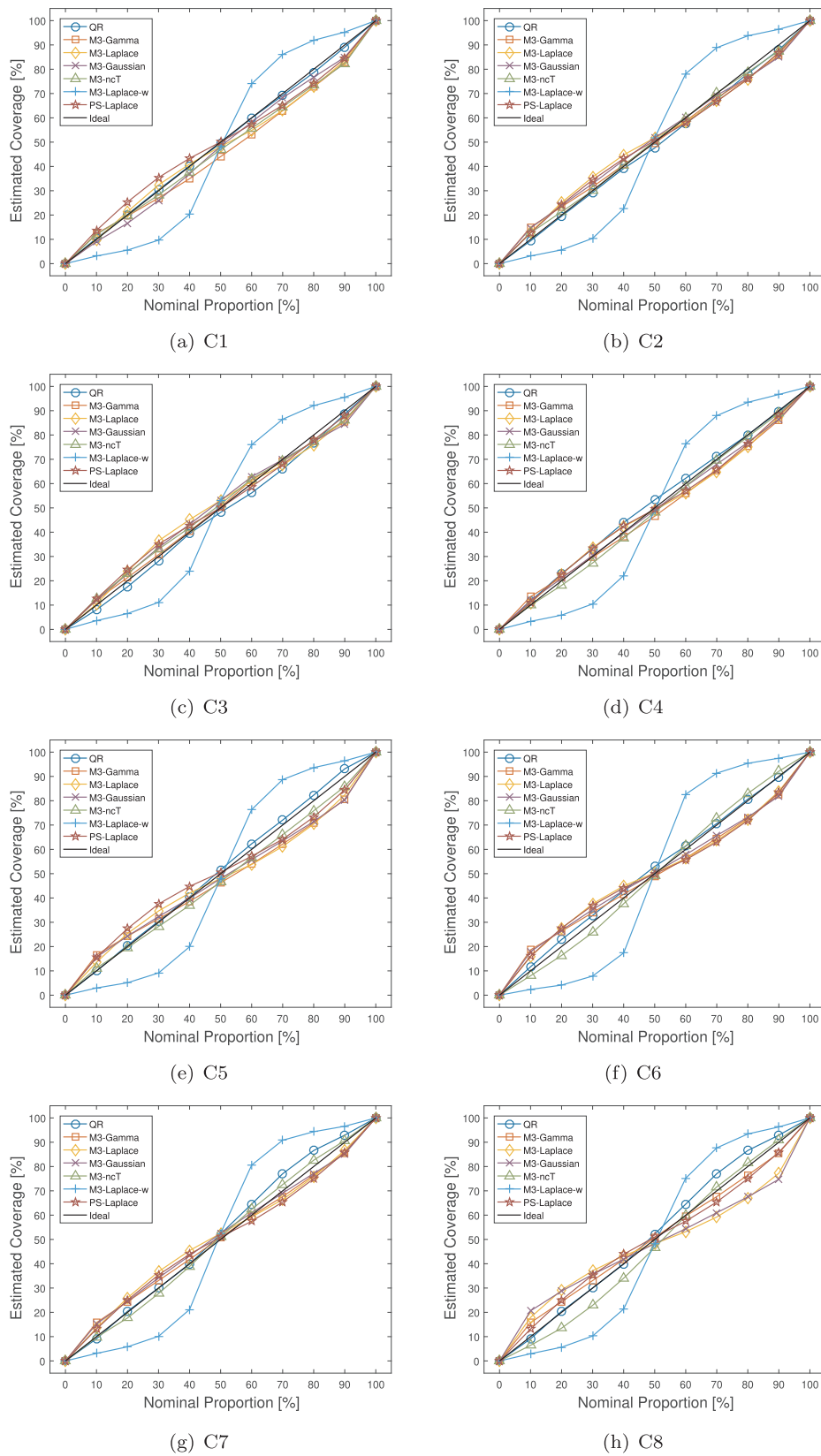
(a) C1

(b) C2

(c) C3

(d) C4

(e) C5

(f) C6

(g) C7

(h) C8

Fig. 5. Reliability of probabilistic forecasts at selected sites.

and normalized pinball loss (NPL) is used to represent the performance of probabilistic forecasts. To generate different NMAE and NPL scenarios, four single machine learning algorithms—i.e., ANN, SVR, GBM, and RF with different kernels are used to produce 14 deterministic

forecasts, including:

- Three SVR models with linear (SVR_l), polynomial (SVR_p), and radial base (SVR_r) kernels;

Table 5
One-hour-ahead forecasting NMAE and NPL of single-algorithm models with different kernels.

Method	Metrics	Site							
		C1	C2	C3	C4	C5	C6	C7	C8
SVR_r	NMAE (%)	5.101	3.114	6.229	3.799	5.145	4.554	3.662	4.014
	NPL	1.845	1.381	2.606	1.372	1.974	1.655	1.360	1.974
SVR_l	NMAE (%)	**4.765**	**2.886**	3.927	**3.718**	**4.891**	**4.466**	**3.449**	**3.984**
	NPL	1.723	*1.253*	1.468	1.350	*1.853*	*1.627*	*1.261*	*1.853*
SVR_p	NMAE (%)	4.772	2.919	4.267	3.734	4.913	4.572	3.553	4.009
	NPL	1.727	1.270	1.592	1.352	1.859	1.662	1.296	1.859
ANN1	NMAE (%)	4.793	2.921	4.155	3.738	4.936	4.671	3.717	4.007
	NPL	1.721	1.267	1.580	1.353	1.874	1.690	1.356	1.874
ANN2	NMAE (%)	4.789	2.938	4.042	3.735	4.939	4.536	3.560	4.012
	NPL	1.722	1.275	1.524	1.352	1.874	1.650	1.293	1.873
ANN3	NMAE (%)	4.817	2.927	4.096	3.738	4.932	4.494	3.502	4.017
	NPL	1.729	1.270	1.549	1.354	1.875	1.636	1.281	1.875
ANN4	NMAE (%)	4.792	2.906	4.022	3.735	4.924	4.481	3.500	4.005
	NPL	1.722	1.264	1.504	1.351	1.873	1.629	1.281	1.873
ANN5	NMAE (%)	4.793	2.902	**3.859**	3.727	4.899	4.487	3.480	4.006
	NPL	*1.719*	1.260	*1.431*	*1.349*	1.861	1.628	1.274	1.861
GBM1	NMAE (%)	4.822	2.945	4.468	3.739	4.961	4.479	3.562	4.022
	NPL	1.733	1.284	1.731	1.351	1.872	1.629	1.300	1.873
GBM2	NMAE (%)	4.808	2.941	4.474	3.736	4.963	4.478	3.550	4.020
	NPL	1.735	1.283	1.732	1.350	1.872	1.631	1.295	1.872
GBM3	NMAE (%)	4.806	2.936	4.730	3.768	4.969	4.491	3.504	4.002
	NPL	1.743	1.285	1.860	1.360	1.882	1.628	1.291	1.882
GBM4	NMAE (%)	4.845	2.946	4.348	3.754	4.974	4.544	3.554	4.023
	NPL	1.742	1.282	1.671	1.357	1.878	1.650	1.302	1.878
RF1	NMAE (%)	4.965	3.060	4.207	3.883	5.115	4.703	3.715	4.159
	NPL	1.796	1.338	1.577	1.407	1.941	1.701	1.357	1.941
RF2	NMAE (%)	4.920	3.012	4.221	3.852	5.057	4.637	3.659	4.110
	NPL	1.777	1.312	1.596	1.393	1.913	1.681	1.336	1.913

Note: The smallest NMAE (%) at each location is in boldface. The smallest NPL at each location is in italic.

- Five ANN models with different numbers of hidden layers (n_l), neurons in each layer (n_o), and weight decay parameter (n_d) values. Our selected models employ the feed-forward back propagation learning function and sigmoid activation function;
- Four GBM models based on different loss functions (Gaussian and Laplacian) and parameters, i.e., number of trees, learning rate (λ), maximum depth of variable interactions, and minimum number of observations in the terminal nodes;
- Two random forest models with different numbers of variables that are randomly sampled as candidates at each split.

The proposed pinball loss optimization method with Laplace distribution is used to generate probabilistic forecasts. The NMAE values of the 14 deterministic forecasts and their corresponding NPL of probabilistic forecasts are summarized in Table 5. A linear regression method [28] is used to fit the relationship between NMAE and NPL. Fig. 6 shows the relationship between NMAE and NPL at the eight sites. A linear relationship is observed between NMAE and NPL at all test locations, which indicates that a better deterministic forecast model will very likely result in a more accurate probabilistic model with the proposed two-step method. The R^2 values of the eight locations of the linear least squares fit are listed in Table 6. The R^2 values are close to 1, which also indicates the strong correlation between the deterministic and probabilistic forecasting steps.

4. Conclusion

This paper developed a two-step probabilistic wind forecasting method based on pinball loss optimization, in conjunction with a multi-model deterministic forecasting framework. Different types of predictive distributions were compared, and the Laplace distribution was found to be the most suitable predictive distribution type. The optimal shape parameter (i.e., standard deviation) of the predictive distribution was determined by minimizing the sum of pinball loss in the training

stage. A surrogate model of the optimal shape parameter was used to estimate a pseudo-optimal shape parameter in the forecasting stage. Results showed that the M3-Laplace model could reduce the pinball loss score metric by up to 35% compared to benchmark models. Also, M3-Laplace showed better reliability than that of the M3-Laplace-w, which indicates the effectiveness of the pinball loss optimization. The sharpness intervals size of the M3-Laplace forecasts ranges from 2% to 18%, which indicates low sharpness.

Results also showed a linear relationship between the deterministic forecasting metric (i.e., NMAE) and the probabilistic forecasting metric (i.e., NPL). This indicates that a better deterministic model will very likely result in a more accurate probabilistic model with the developed framework.

The potential future work will (i) quantitatively evaluate the forecasting performance by considering spatial-temporal effects, and (ii) study how to aggregate probabilistic forecasts of multiple wind locations.

Acknowledgment

Fig. 6. The relationship between NMAE and NPL.

Table 6
R-square of the least square fit.

Site	C1	C2	C3	C4	C5	C6	C7	C8
R^2	0.97	0.98	0.99	0.98	0.95	0.99	0.95	0.99

U.S. Government purposes.

References

[1] Lee D, et al. Wind power forecasting and its applications to the power system [Ph.D. thesis]; 2015.
[2] Botterud A, Zhou Z, Wang J, Sumaili J, Keko H, Mendes J, et al. Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: a case study of illinois. IEEE Trans Sustain Energy 2013;4(1):250–61.
[3] Lefèvre S, Sun C, Bajcsy R, Laugier C. Comparison of parametric and non-parametric approaches for vehicle speed prediction. In: American Control Conference (ACC), 2014. IEEE; 2014. p. 3494–9.
[4] Lange M. On the uncertainty of wind power predictionsanalysis of the forecast accuracy and statistical distribution of errors. J Sol Energy Eng 2005;127(2):177–84.
[5] Bludszuweit H, Domínguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. IEEE Trans Power Syst 2008;23(3):983–91.
[6] Pinson P. Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. J Roy Stat Soc: Ser C (Appl Stat) 2012;61(4):555–76.
[7] Meitz M, Saikkonen P. Parameter estimation in nonlinear ar–garch models. Econometr Theory 2011;27(6):1236–78.
[8] Delignette-Muller ML, Dutang C, et al. fitdistrplus: an r package for fitting distributions. J Stat Softw 2015;64(4):1–34.
[9] Kantar YM. Generalized least squares and weighted least squares estimation methods for distributional parameters. REVSTATStat J 2015;13:263–82.
[10] Hall AR. Generalized method of moments. Oxford University Press; 2005.
[11] Jin B. Fast bayesian approach for parameter estimation. Int J Numer Methods Eng 2008;76(2):230–52.
[12] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70.
[13] Lee D, Baldick R. Probabilistic wind power forecasting based on the laplace distribution and golden search. Transmission and distribution conference and exposition (T&D), 2016 IEEE/PES. IEEE; 2016. p. 1–5.
[14] Haben S, Giasemidis G. A hybrid model of kernel density estimation and quantile regression for gefcom2014 probabilistic load forecasting. Int J Forecast 2016;32(3):1017–22.
[15] Ordiano JÁG, Doneit W, Waczowicz S, Gröll L, Mikut R, Hagenmeyer V. Nearest-neighbor based non-parametric probabilistic forecasting with applications in photovoltaic systems. arXiv preprint arXiv:1701.06463.
[16] Wan C, Xu Z, Pinson P, Dong ZY, Wong KP. Optimal prediction intervals of wind power generation. IEEE Trans Power Syst 2014;29(3):1166–74.
[17] Landry M, Erlinger TP, Patschke D, Varrichio C. Probabilistic gradient boosting machines for gefcom2014 wind forecasting. Int J Forecast 2016;32(3):1061–6.
[18] Zhang Y, Wang J. Gefcom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator. Power & energy society general meeting, 2015 IEEE. IEEE; 2015. p. 1–5.
[19] Wang H-z, Li G-q, Wang G-b, Peng J-c, Jiang H, Liu Y-t. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[20] Steinwart I, Christmann A, et al. Estimating conditional quantiles with the help of the pinball loss. Bernoulli 2011;17(1):211–25.
[21] Feng C, Cui M, Hodge B-M, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[22] Forbes C, Evans M, Hastings N, Peacock B. Statistical distributions. John Wiley & Sons; 2011.
[23] Hansen CH, Doolan CJ, Hansen KL. Wind farm noise: measurement, assessment, and control. John Wiley & Sons; 2017.
[24] Scrucca L, et al. Ga: a package for genetic algorithms in r. J Stat Softw 2013;53(4):1–37.
[25] Zhang J, Draxl C, Hopson T, Delle Monache L, Vanvye E, Hodge B-M. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. Appl Energy 2015;156:528–41.
[26] Gallego-Castillo C, Bessa R, Cavalcante L, Lopez-Garcia O. On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power. Energy 2016;113:355–65.
[27] Juban J, Siebert N, Kariniotakis GN. Probabilistic short-term wind power forecasting for the optimal management of wind generation. Power tech, 2007 IEEE Lausanne. IEEE; 2007. p. 683–8.
[28] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 1963;11(2):431–41.