Probability Density Forecasting of Wind Speed Based on Quantile Regression and Kernel Density Estimation

Lei Zhang, Lun Xie, Qinkai Han, Zhiliang Wang and Chen Huang

1 School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; thesun19821025@163.com (L.Z.); xielun@ustb.edu.cn (L.X.); wzl@ustb.edu.cn (Z.W.)
2 The State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
3 TAIJI Computer Corporations Limited, Beijing 100083, China; zdswzl@163.com

* Correspondence: hanqinkai@hotmail.com

Abstract: Based on quantile regression (QR) and kernel density estimation (KDE), a framework for probability density forecasting of short-term wind speed is proposed in this study. The empirical mode decomposition (EMD) technique is implemented to reduce the noise of raw wind speed series. Both linear QR (LQR) and nonlinear QR (NQR, including quantile regression neural network (QRNN), quantile regression random forest (QRRF), and quantile regression support vector machine (QRSVM)) models are, respectively, utilized to study the de-noised wind speed series. An ensemble of conditional quantiles is obtained and then used for point and interval predictions of wind speed accordingly. After various experiments and comparisons on the real wind speed data at four wind observation stations of China, it is found that the EMD-LQR-KDE and EMD-QRNN-KDE generally have the best performance and robustness in both point and interval predictions. By taking conditional quantiles obtained by the EMD-QRNN-KDE model as the input, probability density functions (PDFs) of wind speed at different times are obtained by the KDE method, whose bandwidth is optimally determined according to the normal reference criterion. It is found that most actual wind speeds lie near the peak of predicted PDF curves, indicating that the probabilistic density prediction by EMD-QRNN-KDE is believable. Compared with the PDF curves of the 90% confidence level, the PDF curves of the 80% confidence level usually have narrower wind speed ranges and higher peak values. The PDF curves also vary with time. At some times, they might be biased, bimodal, or even multi-modal distributions. Based on the EMD-QRNN-KDE model, one can not only obtain the specific PDF curves of future wind speeds, but also understand the dynamic variation of density distributions with time. Compared with the traditional point and interval prediction models, the proposed QR-KDE models could acquire more information about the randomness and uncertainty of the actual wind speed, and thus provide more powerful support for the decision-making work.

Keywords: probability density forecasting; wind speed series; quantile regression; kernel density estimation; signal decomposing algorithm

1. Introduction

Because of the independence from fossil energy and low environmental costs, wind energy has become an important part of sustainable development strategies in many countries [1]. Wind power generation is the conversion of air kinetic energy into electrical energy, and its characteristics will be directly affected by the characteristics of wind speed, which is a stochastic variable and intermittent. With the increase of the proportion of wind energy capacity in the system, the impact of fluctuating...
wind power on the grid system becomes more and more obvious. High wind speed disturbance will cause great changes in the voltage and frequency of the system and may cause the system to lose stability in serious cases. Therefore, accurate prediction of wind speed is meaningful for the optimal control of wind turbine operation in wind farms, the reasonable formulation of power system dispatch plan, and adverse effect reduction of the wind power on the whole grid [2–7].

In the literature, much attention has been paid to developing accurate wind speed forecasting models, which are mainly designed for point predictions. According to the implementation mechanism, current point forecasting models mainly include two categories: one is called the physical model, which is based on numerical weather prediction. The other is based on historical data to construct statistical models to predict future wind speed. Traditional statistical models are represented by time series models, i.e., autoregressive moving average (ARMA) models [8]. In recent years, artificial intelligence and machine learning (AI/ML) models, such as artificial neural networks (ANN) [9], support vector machine (SVM) [10], extreme learning machines (ELM) [11], and deep learning networks (DLN) [12], have been widely used in point prediction of short-term wind speed. In order to improve the prediction accuracy and robustness, hybrid or combined models [13] integrating the advantages of single models are attracting more and more attention.

As nondeterminacy exists in actual wind speed samples, the point prediction model could not always give satisfactory results, which makes the decision-making work face certain risks. Interval prediction is an effective tool to describe and quantify the uncertainty of wind speed. Some scholars have carried out research on wind speed interval prediction in recent years. Song, Jiang, and Zhang [14] examined a Markov-switching model in wind speed forecasting. Unlike the traditional point forecast of wind speeds, such a model could offer both the point and interval forecasts of wind speed series. Iverson et al. [15] used stochastic differential equations (SDEs) to model short-term wind speed. They showed that SDEs could effectively capture the time dependence structure of wind speed prediction errors naturally and, most importantly, derive point and interval forecasts using one SDE model. Recently, wind speed interval forecast using AI/ML models under the multi-objective optimization framework has received more and more attention [16–21]. The multi-objective optimization aims at concurrently minimizing the width and maximizing the coverage probability of the constructed intervals. Several AI/ML models, such as SVMs [16–18], back propagation neural networks (BPNNs) [19], radial basis function neural networks (RBFNNs) [20], and deep belief networks (DBNs) [21], have been successfully applied in wind speed interval forecasting. Some signal processing algorithms, including variational mode decomposition (VMD) [17], complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [18], and wavelet transform (WT) [19,21], were also utilized to reduce the noise and complexity of the raw wind speed data.

Although interval prediction can give upper and lower boundaries of future wind speed, it cannot describe the probability distribution of wind speed. It is well known that the probability density function (PDF) could completely model the probabilistic characteristics of random variables. Thus, probability density prediction can describe the uncertainty of actual wind speed data more accurately and can provide fully predictive information for decision-making work. Some scholars have already made some fruitful attempts. Gneiting and his collaborators [22,23] first used multiple estimates of the current state of wind speed to generate an ensemble of deterministic predictions. They then adopted Bayesian model averaging (BMA) as a statistical post-processing method to predict the PDFs of wind speed. Their results showed that the BMA method could provide calibrated and sharp probabilistic forecasts of wind speed. A similar idea was also adopted by Baran [24,25], but using two different distributions (log normal and truncated normal distributions) to calibrate the probabilistic wind speed forecasts. Both Gneiting’s [22,23] and Baran’s [24,25] studies need multiple runs of numerical weather forecasting models with various initial conditions to obtain the ensembles of forecasts. Unlike these studies, Hu and Wang [26] proposed a hybrid probability prediction model based on wind speed historical data. The hybrid model is composed of empirical WT and Gaussian process regression (GPR). Experimental results showed that the hybrid GPR model can provide the
most likely value and the probability information corresponding to the wind speed forecast based on the predictive PDF [26]. Compared with the large number of studies on point and interval predictions (especially for the point prediction), the studies on wind speed probability density prediction are relatively few, and more effective models should be explored to meet engineering needs.

Quantile regression (QR) is one of regression analysis methods. It was first proposed by Koenker and Bassett in 1978 [27]. QR estimates conditional quantiles of response variables from explained variables and further deduces the conditional probability distribution of response variables without assuming the distribution type of random variables [28]. Therefore, QR could be a good choice for probability density forecasting of wind speed. In order to improve the performance of linear QR (LQR) in complex nonlinear problems, some scholars integrated the AI/ML algorithms with QR and proposed quantile regression neural networks (QRNN) [29], quantile regression random forests (QRRF) [30], quantile regression support vector machine (QRSVM) [31], etc. Very recently, He and his collaborators [32–35], respectively, applied QRSVM and QRNN for short-term probability prediction of wind power. Their results showed that QRNN has excellent performance in wind power probability forecasting. However, in the current literature, it is rare to see the QR for wind speed probability prediction. Zheng et al. [36] performed pioneering work in this area. They put forward a theoretical framework for wind speed probability density prediction based on composite QR and outlier-robust ELM (CQR-ORELM) with feature selection and parameter optimization. A detailed analysis of actual wind speed data showed that the CQR-ORELM model can well describe the conditional distribution and provide satisfactory wind speed forecasts. Research in this area will be continued in this study, using both linear QR and nonlinear QR (NQR, i.e. QRNN, QRRF, QRSVM, etc.) to predict the probability density of short-term wind speed, in order to promote the further study of wind speed probability prediction.

Density estimation methods should be used to assist QR to obtain the PDF of predictive wind speed. Unlike the traditional parametric estimation methods, the kernel density estimation (KDE) method does not require prior knowledge of the data distribution and does not add any assumptions to the data distribution. Because of its flexibleness and robustness, KDE has been widely used in wind speed probability distribution estimation and wind energy assessment [6,37]. The core of KDE is the selection of the kernel function and the determination of bandwidth. In this study, the Gaussian kernel function will be chosen for its generality. The normal reference criterion (NRC) [38] will be used for calculating the bandwidth. According to the NRC, the optimal bandwidth is achieved by minimizing the value of the mean integrated squared error (MISE). By choosing the Gaussian kernel function and optimal bandwidth, the KDE method is utilized to handle future wind speed data predicted by QR models to acquire the overall PDFs at any moment. In addition, empirical mode decomposition (EMD) [39,40], which is a famous adaptive de-noising algorithm based on local characteristics of the signal, will be implemented to make the raw wind speed series less noisy and more stable.

The above literature review indicates that only a few studies have focused on the wind speed probability prediction; this is particularly true for the QR-KDE based probability density forecasting of short-term wind speed. In addition, few efforts have been made to comprehensively compare the performance and robustness of both the LQR and NQR models in point, interval, and probabilistic density predictions of short-term wind speed. Therefore, the novelty and contributions of this study can be summarized as follows:

- A framework for probability density forecasting of wind speed based on QR and KDE is proposed. EMD is implemented to reduce the noise of raw wind speed series.
- Both LQR and NQR (QRNN, QRRF, and QRSVM) are, respectively, utilized to study the de-noised wind speed signal. By choosing the Gaussian kernel function and optimal bandwidth, the KDE method is utilized to handle future wind speed data predicted by QR models to acquire the overall PDFs at any moment.
- Various experiments are conducted on the real wind speed data at four wind sites in China. The performance and robustness of various QR-KDE models in point, interval, and density
predictions of short-term wind speed are compared comprehensively. The best QR-KDE based probabilistic density forecast model is then recommended for real applications.

The content of this paper is organized as follows. Section 2 explains the structures and procedures of QR-KDE based density forecast models. Section 3 introduces the measurement of wind speed data at four sites of China. Section 4 presents the evaluation of model parameters and compares the performances of various models, and Section 5 summarizes the conclusions of the study.

2. Methods

2.1. Linear Quantile Regression

Different from the classical regression analysis, QR aims at estimating the conditional quantiles of the response variable under given independent variables and then the conditional density distribution of the response variable. Similar to linear regression, LQR is also based on the method of least squares. If the independent variables are \( x_i(t) \) \((i = 1, 2, ..., I)\), regression coefficients are \( m_i \), and the intercept term is \( b \), then we can get the \( \tau \)th \((0 \leq \tau \leq 1)\) quantile of the response variable \( \hat{y}_\tau(t) \) as:

\[
\hat{y}_\tau(t) = \sum_{i=1}^{I} m_i x_i(t) + b. \quad (1)
\]

From LQR, \( \hat{y}_\tau(t) \) could be estimated by minimizing the quantile regression error function:

\[
E_\tau = \frac{1}{N} \sum_{t=1}^{N} \rho_\tau(y(t) - \hat{y}_\tau(t)), \quad (2)
\]

where \( y(t) \) is the observed value of response variable at time \( t \) \((t = 1, 2, ..., N)\). \( \rho_\tau(u) \) is known as the pinball loss function, and its expression is given by:

\[
\rho_\tau(u) = \begin{cases} 
\tau u & \text{if } u \geq 0; \\
(\tau - 1)u & \text{if } u < 0.
\end{cases} \quad (3)
\]

The detailed optimization algorithm was outlined by Koenker [28]. The whole conditional density distribution can be obtained by continuously taking the value of \( \tau \) in the range of \((0, 1)\). When dealing with complex nonlinear problems, the performance of LQR might be poor. Thus, NQR models integrating with AI/ML algorithms have also been proposed in the literature. In this study, these models will also be utilized for probability density forecasting of wind speed data.

2.2. Nonlinear Quantile Regression

2.2.1. Quantile Regression Neural Network

The ANN family are widely used AI/ML algorithms. In the forecasting field, the single hidden-layer feedforward network is the most commonly used algorithm. Cannon [29] combined LQR and this type of ANN to propose the QRNN model. By applying the hyperbolic tangent to the inner product between independent variables \( x_i(t) \) and weights \( w^{(h)}_{ij} \) and the bias \( b^{(h)}_j \) of the hidden-layer, we can obtain the output of the \( j \)th hidden-layer node as follows:

\[
g_j(t) = \tanh \left( \sum_{i=1}^{I} x_i(t) w^{(h)}_{ij} + b^{(h)}_j \right). \quad (4)
\]

An estimate of the conditional \( \tau \)-quantile \( \hat{y}_\tau(t) \) is then given by:

\[
\hat{y}_\tau(t) = f \left( \sum_{j=1}^{J} g_j(t) w^{(o)}_j + b^{(o)} \right), \quad (5)
\]
in which \( w^{(b)}, b^{(b)} \) are, respectively, the weights and bias of output-layer nodes. The output-layer transfer function \( f(\cdot) \) is usually considered as the linear function. The number of hidden-layer nodes \( f \), which controls the model complexity, should be set carefully to avoid overfitting. Moreover, we can also use the weight decay regularization [41] to help prevent overfitting. A positive constant \( \lambda \), which is called the weight penalty, should also be determined to control the relative contribution of the weight decay term.

### 2.2.2. Quantile Regression Random Forest

Both LQR and QRNN obtain the optimal parameters by minimizing the quantile regression error function (see Equation (2)). Meinshausen [30] proposed a different approach (QRRF), which is based on random forests (RFs) instead of directly optimizing the minimum error function. For QRRF, the ensemble of trees is grown as in the standard RF algorithm by employing random node and split point selection. Then, the conditional distribution of the response variable is estimated by the weighted distribution of observed response variables, where the weights attached to observations are identical to the original RF algorithm [30].

Following the notation of Breiman [42], the random parameter vector \( \theta_k \) describes how a tree is grown. The observed values of one leaf are \( f(x, \theta_k) \). Let the weight vector \( \omega_i(x, \theta_k) \) be given by a positive constant if observation \( X_i \) is part of leaf \( f(x, \theta_k) \) and zero if it is not. The conditional distribution \( \hat{F}(y|X = x) \) could be approximated by the weighted mean over the observations of \( 1_{\{y \leq y_i\}} \) as:

\[
\hat{F}(y|X = x) = \frac{1}{N_{\text{tree}}} \sum_{i=1}^n \sum_{k=1}^{N_{\text{tree}}} \omega_i(x, \theta_k) 1_{\{y \leq y_i\}}.
\]  

Estimations of the \( \tau \)th conditional quantiles \( \hat{y}_\tau(t) \) could then be obtained according to the definition:

\[
\hat{y}_\tau(t) = \inf\{y : \hat{F}(y|X = x) \geq \tau\}.
\]

The number of trees \( N_{\text{tree}} \) and the number of leaves grown in every tree \( m_{\text{try}} \) are two tuning parameters of the QRRF. The values of these two parameters will be fine-tuned on the out-of-bag samples.

### 2.2.3. Quantile Regression Support Vector Machine

Takeuchi et al. [31] first proposed the QRSVM model, which integrates SVM into the QR to construct an NQR model. By minimizing the error function plus a regularizer, QRSVM could provide the conditional quantiles of response variable. Similar to the regular SVM for regression [43], the independent variable vector could be projected into a higher dimensional feature space by using the kernel function defined nonlinear mapping relations, i.e., \( f(x) = \langle \phi(x), w \rangle + b \), in which \( w \) is the weight vector, \( \phi(x) \) is the mapping function of independent variable \( x \), and \( b \) is the dual variable to the constraint. Using the connection between reproducing the kernel Hilbert space and feature space, we obtain the dual optimization problem, which is equivalent to the minimization of error function plus the regularizer:

\[
\begin{align*}
\min_{w, b, \xi^*, \xi^+} \quad C \sum_{i=1}^m \tau \xi^*_i + (1 - \tau) \xi^+_i + \frac{1}{2}||w||^2, \\
\text{s.t.} \quad y_i - \langle \phi(x), w \rangle - b \leq \xi^*_i, \quad \langle \phi(x), w \rangle + b - y_i \leq \xi^+_i, \quad \xi^*, \xi^+ \geq 0.
\end{align*}
\]  

(8)

Here, we use \( C = 1/(\lambda_s m) \), in which \( \lambda_s \) is the regularization parameter. Equation (8) could be solved straightforwardly using the Lagrange multipliers. \( f(x) \) is also represented by the kernel expansion form as \( f(x) = \sum \alpha_i k(x, x_i) + b \), where \( \alpha_i \) is the Lagrange multiplier and \( k(x, x_i) \) is the kernel function. In our study, we chose the radial basis function as the kernel function. The bandwidth of kernel function and regularization parameter \( \lambda_s \) determine the performance of QRSVM and should be tuned to make the QRSVM in the optimal state.
2.3. Kernel Density Estimation

KDE is used to estimate the PDF from conditional quantiles obtained by QR models. By centering a smooth kernel function at each data point, KDE then sums to get a density estimate. The expression of the basic kernel estimator is given as:

$$\hat{f}_{\text{kde}}(x) = \frac{1}{n} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right),$$

(9)
in which $K(\cdot)$ is the kernel function and $h$ is the bandwidth. The Gaussian kernel function is chosen in this study for its generality. The selection of the bandwidth would significantly affect the estimation results. NRC [38] is used by minimizing the MISE to select the optimal bandwidth. The optimization problem is given by:

$$\min (MISE) = E \left( \int (\hat{f}(x) - f(x))^2 \, dx \right).$$

(10)

As the form of $f(x)$ is unknown, so the solution of Equation (10) is not easy. Usually, we can assume that the kernel density function obeys the normal distribution, so that the optimal bandwidth can be calculated directly from the following equation:

$$h_{\text{opt}} = \left( \frac{4}{(d+2)n} \right)^{1/(d+4)} \hat{\sigma},$$

(11)
in which $d$ is the lag number, $n$ is the sample number, and $\hat{\sigma}$ is the standard error of the sample. The $\hat{\sigma}$ could be replaced by the inter-quartile range (IQR) to make it less sensitive to outliers.

2.4. Empirical Mode Decomposition

In order to reduce the noisy of original wind speed series, EMD is implemented before applying the QR-KDE models for probability density forecasting. After the process of EMD, the original wind speed data are decomposed into a finite number of intrinsic mode functions (IMFs) and residuals. Compared with the original data, each IMF behaves as more stable and regular. In order to determine whether the decomposed signal is an IMF or not, two conditions should be satisfied [40]: (a) the number of extrema and the number of zero crossings are equal or differ at most by one; (b) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The detailed computation of the EMD can be found in [40].

2.5. Probabilistic Density Forecast Models Based on QR-KDE

Based on the principles of LQR, NQR, KDE, and EMD introduced in previous sections, the solution process of QR-KDE based probabilistic density forecast models is proposed and shown in Figure 1. It can be summarized as the following steps:

Step 1: For the original wind speed data, EMD is implemented to obtain a finite number of IMFs and residuals. These IMFs and residuals are then studied by the LQR or NQR (QRNN, QRRF, and QRSVM) models, respectively. An ensemble of conditional quantiles of wind speed is obtained accordingly.

Step 2: Based on the optimal bandwidth by NRC, KDE estimates fully the PDFs of predictive wind speed from the ensemble of conditional quantiles. If the prediction error does not meet the requirement, then Steps 1 and 2 are repeated by tuning and selecting the best parameters of each LQR and NQR models.

Step 3: Under the 90% and 80% confidence levels, the forecasting PDF curves are constructed, respectively. Both point and interval predictions under the two confidence levels are calculated in for model performance comparisons.
According to different QR models used, the wind speed density prediction models proposed in this study are divided into four types, namely EMD-LQR-KDE, EMD-QRNN-KDE, EMD-QRRF-KDE, and EMD-QRSVM-KDE. We can also directly apply the QR models to the original wind speed data without the implementation of EMD. Here, these four models, including LQR-KDE, QRNN-KDE, QRRF-KDE, and QRSVM-KDE, are also investigated for comparisons with EMD based models.

2.6. Forecasting Performance Evaluation

In order to compare the performance of various models in point and interval predictions, several evaluation metrics should be defined. For the point prediction, the probability mean and median are of the most interest. Here, the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) are utilized to evaluate the point forecasting performance. Their expressions are given by:

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}},
\]

\[
MAE = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t|,
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_t - \hat{y}_t|}{y_{t,\text{max}}} \times 100%,
\]

where \(y_t, \hat{y}_t\) denote the actual and predicted values and \(n\) is the sample number. Due to the randomness and volatility, wind speed may be close to zero or even equal to zero at some times. The maximal value of wind speed \(y_{t,\text{max}}\) is used as the denominator in the analysis of the MAPE to avoid the instance that the error value tends to infinity.

There are usually two metrics for the performance evaluation of interval predictions. The first one is called prediction interval coverage probability (PICP) [44], which is expressed as follows:

\[
\text{PICP} = \frac{1}{n} \sum_{i=1}^{n} c_i,
\]

where \(n\) is the sample number. If the actual observations of the \(i\)th wind speed sample fall into the predicted interval, then \(c_i = 1\); otherwise, \(c_i = 0\). The predicted interval is unreliable if the PICP is significantly lower than the predetermined confidence level.

The normalized mean prediction interval width (NMPIW) [44] is the other metric to evaluate the accuracy of interval prediction. Its mathematical expression is written as:

\[
\text{NMPIW} = \frac{1}{nR} \sum_{i=1}^{n} (U(x_i) - L(x_i)),
\]

where \(U(x_i), L(x_i)\) are the upper and lower interval boundaries of the \(i\)th sample, respectively. \(R\) is the variation in range of actual wind speed. If the NMPIW is large enough, PICP can always reach 100% to satisfy the predetermined confidence level. However, such an interval width is too large and useless for engineering applications. The goal of constructing the prediction interval is to make NMPIW as small as possible under the premise that PICP is greater than the predetermined confidence level.
From a practical point of view, we expect higher PICP and lower NMPIW. However, these two objectives always conflict with each other. Under certain conditions, higher PICP would lead to higher NMPIW, and lower NMPIW would cause PICP become lower. Therefore, a coverage width based criterion (CWC) [45] is defined as a comprehensive evaluation metric:

Figure 1. Solution process of quantile regression (QR)-kernel density estimation (KDE) based probabilistic density forecast models. IMF, intrinsic mode function; LQR, linear QR; NQR, nonlinear QR.
\[ CWC = NMPIW + \gamma(PICP)e^{-\eta|PICP-\mu|}, \]  
\( \mu \) is the predetermined confidence level, \( \eta \) is the penalty parameter (usually a large number), and:

\[ \gamma(PICP) = \begin{cases} 0 & \text{if } PICP \geq \mu; \\ 1 & \text{if } PICP < \mu. \end{cases} \]

When PICP is greater than \( \mu \), the CWC is equal to NMPIW. If PICP is lower than \( \mu \), then both PICP and NMPIW determine the value of the CWC, and PICP would have a greater impact on the CWC. In short, the lower the CWC, the better the accuracy of interval prediction is.

For the probabilistic density prediction, the continuous ranked probability score (CRPS), defined by Matheson and Winkler [46] and Gneiting and Raftery [47], could be utilized as the evaluation metric. Its expression is presented as follows:

\[ CRPS = \int_{-\infty}^{\infty} (F(x) - 1\{x \geq y\})^2 dx. \]

where \( F(x) \) is the predictive cumulative distribution function and \( y \) is the verifying observation. Generally, the lower the CRPS metric, the higher the accuracy of probability density forecasting.

3. Wind Data Description

With the aid of an anemometer, the wind speed information can be recorded continuously. Usually, the time interval of recorded wind speed data is 10 min. For the convenience of following the analysis, the data are averaged hourly. Four wind speed observation stations in China, whose geographic information is listed in Table 1, are concerned in this study. For each observation station, we select a wind speed dataset with a sample size of 1000, of which the first 800 is the training set and the latter 200 is the testing set. Time plots for each wind speed dataset are given in Figure 2.

Table 1. Locations of four observation stations in China.

| Observation Station      | Longitude | Latitude   | Altitude (m) | Data Length            |
|--------------------------|-----------|------------|--------------|------------------------|
| AnHui (AH)               | 117°17’E  | 31°52’N   | 20           | 7/8/2011–8/19/2011     |
| GuangDong (GD)           | 113°17’E  | 23°8’N    | 11           | 12/2/2011–1/13/2012    |
| GanSu (GS)               | 103°44’E  | 36°2’N    | 1500         | 1/17/2014–2/28/2014    |
| HeiLongJiang (HLJ)       | 126°38’E  | 45°45’N   | 128          | 11/1/2015–12/13/2015   |

Table 2 shows the statistical characteristics of wind speed data at four sites. It can be seen that the maximum wind speeds of the four sites are all greater than 10 m/s. The average wind speed of the HeiLongJiang (HLJ) site is the highest with a large deviation, while that of the AnHui (AH) site is the lowest with a small deviation. Although the kurtosis indexes of the four sites have little difference, the difference in the skewness index is evident. There are not only sites with high skewness, such as the GanSu (GS) site (0.9460), but also sites with low skewness, such as the GuangDong (GD) site (−0.0621). Therefore, the wind speed data of the four selected points are representative, which can provide data support for the research of the prediction method in this paper.

Table 2. Statistical features of wind speed data at four wind sites.

| Observation Station | Max-Min Values (m/s) | Mean (m/s) | Standard Deviation (m/s) | Skewness | Kurtosis |
|---------------------|----------------------|------------|--------------------------|----------|----------|
| AH                  | 10.41-0              | 4.6009     | 1.6494                   | 0.1648   | 2.9031    |
| GD                  | 13.05-0              | 6.7162     | 2.3287                   | -0.0621  | 2.5962    |
| GS                  | 18.54-0              | 5.8787     | 3.3801                   | 0.9460   | 3.8657    |
| HLJ                 | 16.56-0              | 8.5214     | 2.8915                   | -0.3391  | 3.0986    |
4. Results and Discussions

Before applying the QR-KDE models, EMD is used for signal de-noising. In this study, two IMFs and residuals are used to represent the original series and are respectively studied by the QR-KDE models to obtain the ensemble of conditional quantiles of predictive wind speed.

The input variables of the model are wind speed data with different lag periods. The lag order of different wind speed data is determined by the partial autocorrelation function (PACF). Taking the original data of the AH site as an example, the PACF is drawn and shown in Figure 3. When the lag order is greater than four, the PACF value is small and can be ignored. Therefore, the lag order is four. For the other data, the same method can be used to determine the lag order.

Quantiles from 0.05 to 0.95 with an interval of 0.01 are used to construct the 90% confidence level, and quantiles in the range of 0.1–0.9 with the interval of 0.01 are adopted to construct the 80% confidence level. For the LQR model, the least squares method is used for model estimation. However, the parameters of various NQR models are different, and there is no uniform parameter determination method. Therefore, aiming at minimizing the MAE results of point prediction, various NQR models (QRNN, QRRF, and QRSVM) with different parameter combinations are tested for each wind speed dataset of four observation stations, and the optimal parameters are obtained, as shown in Tables 3–5.
Table 3. Parameters used in the QRNN for different wind speed datasets.

|      | Weight Penalty | Number of Hidden-Layer Nodes |
|------|----------------|-----------------------------|
| AH   |                |                             |
| Original data | 7.0         | 4                           |
| IMF1  | 0.3           | 6                           |
| IMF2  | 2.8           | 7                           |
| Residuals | 0.5         | 7                           |
| GD   |                |                             |
| Original data | 1.0         | 2                           |
| IMF1  | 8.2           | 6                           |
| IMF2  | 7.0           | 6                           |
| Residuals | 3.2         | 1                           |
| GS   |                |                             |
| Original data | 9.5         | 1                           |
| IMF1  | 0.3           | 4                           |
| IMF2  | 4.4           | 9                           |
| Residuals | 3.8         | 3                           |
| HLJ  |                |                             |
| Original data | 7.6         | 5                           |
| IMF1  | 8.0           | 2                           |
| IMF2  | 1.9           | 7                           |
| Residuals | 4.9         | 2                           |

Figure 3. Partial autocorrelation function (PACF) plot of the original wind data at the AH site.

Using these models, we compare the accuracy of both the point and interval predictions of various QR-KDE models. Then, taking the conditional quantiles obtained by the QR-KDE model with the best performance as the input, PDFs of wind speed at different times are obtained by the KDE method and verified by comparing with the actual wind speed values.
Table 4. Parameters used in the QRRF for different wind speed datasets.

|       | Number of Trees | Number of Leaves |
|-------|----------------|-----------------|
| AH    |                |                 |
| Original data | 600            | 5               |
| IMF1  | 400            | 6               |
| IMF2  | 300            | 8               |
| Residuals | 800            | 9               |
| GD    |                |                 |
| Original data | 500            | 5               |
| IMF1  | 500            | 1               |
| IMF2  | 900            | 1               |
| Residuals | 200            | 2               |
| GS    |                |                 |
| Original data | 700            | 8               |
| IMF1  | 700            | 2               |
| IMF2  | 300            | 8               |
| Residuals | 500            | 2               |
| HLJ   |                |                 |
| Original data | 100            | 9               |
| IMF1  | 100            | 3               |
| IMF2  | 500            | 1               |
| Residuals | 700            | 2               |

Table 5. Parameters used in the QRSVM for different wind speed datasets.

|       | Bandwidth of Kernel Function | Regularization Parameter |
|-------|-------------------------------|--------------------------|
| AH    |                               |                          |
| Original data | 100             | 4                       |
| IMF1  | 0.5                           | 0.25                     |
| IMF2  | 2                             | 0.25                     |
| Residuals | 2                            | 64                      |
| GD    |                               |                          |
| Original data | 0.5             | 0.25                    |
| IMF1  | 1                             | 64                      |
| IMF2  | 0.5                           | 4                       |
| Residuals | 10                           | 256                     |
| GS    |                               |                          |
| Original data | 1              | 0.25                    |
| IMF1  | 2                             | 4                       |
| IMF2  | 0.5                           | 0.25                    |
| Residuals | 4                            | 256                     |
| HLJ   |                               |                          |
| Original data | 2               | 0.25                    |
| IMF1  | 4                             | 64                      |
| IMF2  | 4                             | 64                      |
| Residuals | 4                            | 64                      |

4.1. Point Prediction

Point prediction includes the probabilistic mean and median prediction. Various QR-KDE models with/without EMD are applied for point predictions. In addition, the ARMA model is also a common model in the point prediction analysis. Bayesian information criteria (BICs) are used to determine the optimal order of the ARMA model. Taking the original data of the AH site as an example, the BIC values of the ARMA models with different order combinations are calculated and shown in Table 6. Obviously, the BIC value corresponding to ARMA(1,1) is the lowest. Therefore, ARMA(1,1) is determined as the prediction model of AH wind data, and the model parameters are calculated by the maximum likelihood estimation. The order and parameter estimation of ARMA models for other wind data can be determined according to the above process.
Table 6. BIC values of candidate ARMA models for the AH site.

| Order | MA(0)       | MA(1)       | MA(2)       | MA(3)       | MA(4)       | MA(5)       |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| AR(0) | 3045.58     | 2336.09     | 2017.82     | 1862.14     | 1787.76     | 1752.68     |
| AR(1) | 1620.40     | 1618.84     | 1623.77     | 1629.48     | 1635.54     | 1641.64     |
| AR(2) | 1619.64     | 1624.22     | 1629.52     | 1635.64     | 1633.74     | 1639.74     |
| AR(3) | 1623.47     | 1629.35     | 1635.47     | 1633.29     | 1634.36     | 1639.52     |
| AR(4) | 1629.36     | 1634.10     | 1633.04     | 1639.19     | 1641.12     | 1642.45     |
| AR(5) | 1635.40     | 1641.57     | 1639.21     | 1644.46     | 1642.76     | 1660.15     |

Tables 7–10, respectively, show the RMSE, MAE, and MAPE results of the wind speed dataset at the four wind stations. It is shown that:

The introduction of EMD greatly improves the point prediction accuracy of the QR-KDE models. The results of the AH station (see Table 7) are taken as an example. Without EMD, the lowest RMSE, MAE, and MAPE of wind speed mean are all obtained by ARMA, i.e., 0.8692 m/s, 0.6833 m/s, and 19.68%. With EMD, the lowest RMSE, MAE, and MAPE of wind speed mean are gained by EMD-QRNN-KDE and EMD-LQR-KDE, i.e., 0.4635 m/s, 0.3498 m/s, and 11.08%. The prediction accuracy is increased by nearly 50%. For the wind speed median, the prediction accuracy is also increased 50% after considering the EMD. Similar accuracy improvement could also be found in the other three stations.

Table 7. Point prediction results at the AH station. EMD, empirical mode decomposition.

|               | Mean     | Median   |               | Mean     | Median   |
|---------------|----------|----------|---------------|----------|----------|
|               | RMSE (m/s) | MAE (m/s) | MAPE (%)  | RMSE (m/s) | MAE (m/s) | MAPE (%)  |
| ARMA          | 0.8692   | 0.6833   | 19.68        | -         | -         | -         |
| LQR-KDE       | 0.8806   | 0.6938   | 19.79        | 0.8799    | 0.6949    | 19.79     |
| QRNN-KDE      | 0.8817   | 0.6985   | 20.63        | 0.8776    | 0.6959    | 20.52     |
| QRRF-KDE      | 0.9862   | 0.7765   | 23.60        | 0.9793    | 0.7725    | 23.60     |
| QRSVM-KDE     | 1.6811   | 1.3021   | 46.57        | 1.6446    | 1.2611    | 44.42     |
| EMD-ARMA      | 0.4741   | 0.3639   | 11.84        | -         | -         | -         |
| EMD-LQR-KDE   | 0.4726   | 0.3550   | 11.08        | 0.4868    | 0.3451    | 10.89     |
| EMD-QRNN-KDE  | 0.4635   | 0.3498   | 11.09        | 0.4849    | 0.3475    | 10.63     |
| EMD-QRNN-KDE  | 0.5099   | 0.4509   | 13.73        | 0.5810    | 0.4479    | 12.94     |
| EMD-QRSVM-KDE | 0.5470   | 0.4089   | 13.29        | 0.4939    | 0.3704    | 11.06     |

Table 8. Point prediction results at the GD station.

|               | Mean     | Median   |               | Mean     | Median   |
|---------------|----------|----------|---------------|----------|----------|
|               | RMSE (m/s) | MAE (m/s) | MAPE (%)  | RMSE (m/s) | MAE (m/s) | MAPE (%)  |
| ARMA          | 1.1173   | 0.8128   | 13.75        | -         | -         | -         |
| LQR-KDE       | 1.1252   | 0.8132   | 13.69        | 1.1298    | 0.8166    | 13.69     |
| QRNN-KDE      | 1.1220   | 0.8119   | 13.87        | 1.1240    | 0.8127    | 13.77     |
| QRRF-KDE      | 1.1252   | 0.8419   | 15.01        | 1.1409    | 0.8481    | 15.05     |
| QRSVM-KDE     | 2.2610   | 1.8092   | 31.67        | 2.2059    | 1.7501    | 30.64     |
| EMD-ARMA      | 0.5149   | 0.4006   | 6.66         | -         | -         | -         |
| EMD-LQR-KDE   | 0.4950   | 0.3890   | 6.50         | 0.4693    | 0.3670    | 6.16      |
| EMD-QRNN-KDE  | 0.5131   | 0.4045   | 6.98         | 0.4689    | 0.3668    | 6.16      |
| EMD-QRRF-KDE  | 0.7764   | 0.5510   | 9.37         | 0.7938    | 0.5573    | 9.27      |
| EMD-QRSVM-KDE | 2.1864   | 1.6820   | 24.21        | 1.8578    | 1.3318    | 21.63     |
Generally, ARMA, LQR, and QRNN have better point prediction performance than QRRF and QRSVM. At the AH station (see Table 7), except for the MAPE of the mean and the MAE of the median, the lowest values of other metrics of both the mean and median prediction are obtained by EMD-QRNN-KDE. At the GD station (see Table 8), EMD-LQR-KDE has the lowest RMSE, MAE, and MAPE of the wind speed mean, while EMD-QRNN-KDE gets the lowest metrics of wind speed median. Except for the RMSE of the wind speed median at the GS station (see Table 9) and the RMSE and MSE of the wind speed median at the HLJ station (see Table 10), EMD-LQR-KDE gains the lowest values of other metrics at the two stations. Thus, we can conclude that for point prediction, EMD-LQR-KDE and EMD-QRNN-KDE have better performance than the other seven models.

### 4.2. Interval Prediction

Interval predictions under two confidence levels (90% and 80%) are carried out by various QR-KDE models. Both the PICP and NMPIW of the predicted intervals for the four observation stations are then calculated and given in Tables 11–14, respectively. It is shown that the PICPs of predicted intervals by QR-KDE models without EMD (LQR-KDE, QRNN-KDE, QRRF-KDE, and QRSVM-KDE) are always lower than the predetermined confidence levels. After the introduction of EMD, the PICPs of various QR-KDE models are significantly increased. The most obvious increase is found in the EMD-QRRF-KDE model. Taking the results at the GD station for example (see Table 12), we can see that the PICPs of EMD-QRRF-KDE are 99% and 95%, greatly exceeding the confidence levels (90% and 80%). However, if we pay attention to the NMPIW, we can find that the NMPIW of EMD-QRRF-KDE (36.3% and 28%) becomes wider than that of QRRF-KDE (30.3% and 22.7%). This indicates that for the QRRF model, the cost of introducing EMD to increase PICP is the widening of NMPIW, which is
contrary to the goal of interval prediction. For the QRNN model, after the introduction of EMD, we can find that PICPs increase to 90.0% and 81%, which are in the vicinity of the predetermined confidence levels. Most importantly, the NMPIW is reduced greatly, from 22.3% and 15.2% of QRNN-KDE to 16.4% and 11.8% of EMD-QRNN-KDE. The positive effect of EMD on interval prediction accuracy could also be found in the LQR model. Although EMD could reduce the NMPIW of QRSVM-KDE, its increased effect on NICP is not as significant as the QRNN and LQR models. Similar findings could also be gained from the results of the other three stations.

Table 11. Interval prediction results at the AH station. NMPIW, normalized mean prediction interval width; CWC, coverage width based criterion.

|          | PICP (%) | NMPIW (%) | CWC (%) |
|----------|----------|-----------|---------|
|          | 90%      | 80%       | 90%     | 80%     | 90%     | 80%     |
| LQR-KDE  | 81.0     | 65.5      | 21.5    | 16.1    | 21.5    | 16.1    |
| QRNN-KDE | 78.0     | 66.0      | 21.6    | 16.7    | 21.6    | 16.7    |
| QRFF-KDE | 80.0     | 69.0      | 27.9    | 20.9    | 27.9    | 20.9    |
| QRSVM-KDE| 79.0     | 63.5      | 42.6    | 31.8    | 42.6    | 31.8    |
| EMD-LQR-KDE | 90.0   | 80.5      | 15.5    | 11.6    | 15.5    | 11.6    |
| EMD-QRNN-KDE | 94.0  | 86.5      | 17.1    | 13.6    | 17.1    | 13.6    |
| EMD-QRFF-KDE | **98.5** | **96.0** | 30.7    | 23.6    | 30.7    | 23.6    |
| EMD-QRSVM-KDE | 95.5  | 92.0      | 25.5    | 17.7    | 25.5    | 17.7    |

Table 12. Interval prediction results at the GD station.

|          | PICP (%) | NMPIW (%) | CWC (%) |
|----------|----------|-----------|---------|
|          | 90%      | 80%       | 90%     | 80%     | 90%     | 80%     |
| LQR-KDE  | 78.0     | 67.0      | 21.9    | 15.3    | 21.9    | 15.3    |
| QRNN-KDE | 79.0     | 65.5      | 22.3    | 15.2    | 22.3    | 15.2    |
| QRFF-KDE | 85.0     | 75.5      | 30.3    | 22.7    | 30.3    | 22.7    |
| QRSVM-KDE| 77.0     | 69.5      | 55.6    | 42.2    | 55.6    | 42.2    |
| EMD-LQR-KDE | 90.0   | 80.5      | 16.5    | 11.8    | 16.5    | 11.8    |
| EMD-QRNN-KDE | 90.0  | 81.0      | 16.4    | 11.8    | 16.4    | 11.8    |
| EMD-QRFF-KDE | **99.0** | **95.0** | 36.3    | 28.0    | 36.3    | 28.0    |
| EMD-QRSVM-KDE | 80.0  | 80.0      | 45.0    | 27.5    | 45.0    | 27.5    |

Table 13. Interval prediction results at the GS station.

|          | PICP (%) | NMPIW (%) | CWC (%) |
|----------|----------|-----------|---------|
|          | 90%      | 80%       | 90%     | 80%     | 90%     | 80%     |
| LQR-KDE  | 89.5     | 75.0      | 29.3    | 21.5    | 29.3    | 21.5    |
| QRNN-KDE | 91.0     | 74.5      | 29.0    | 21.4    | 29.0    | 21.4    |
| QRFF-KDE | 89.5     | 79.5      | 36.3    | 26.9    | 36.3    | 26.9    |
| QRSVM-KDE| 82.0     | 70.5      | 49.6    | 42.1    | 49.6    | 42.1    |
| EMD-LQR-KDE | 96.0   | 89.5      | 23.5    | 16.3    | 23.5    | 16.3    |
| EMD-QRNN-KDE | 96.0  | 89.5      | 23.4    | 16.3    | 23.4    | 16.3    |
| EMD-QRFF-KDE | **100** | **97.0** | 40.6    | 31.1    | 40.6    | 31.1    |
| EMD-QRSVM-KDE | 90.0  | 90.0      | 48.4    | 44.0    | 48.4    | 44.0    |

Under the two confidence levels, Tables 11–14 also present the results of the CWC metric at the four stations, to directly show the performance of various QR-KDE models without and with EMD. According to the definition of Equation (17), the QR-KDE model with the lower CWC has higher accuracy and better performance. It is shown that except for the case of the AH station, both EMD-LQR-KDE and EMD-QRNN-KDE have the lowest CWC. At the AH station (see Table 11), the CWC of EMD-QRNN-KDE is slightly higher than that of EMD-LQR-KDE, while it is still much lower than the CWCs of the other QR-KDE models. Among the considered QR-KDE models, EMD-LQR-KDE and EMD-QRNN-KDE have the best performance in interval prediction. However, the introduction of EMD is not always useful to improve the accuracy of interval prediction, especially
for the QRRF models. We can see that the performance of QRRF-KDE becomes worse after considering the EMD. For QRSVM-KDE, it is found that EMD is helpful for accuracy improvement in most cases. Only at the GS station, the introduction of EMD seems to have little effect (90% confidence level; see Table 13) and even a negative effect (80% confidence level; see Table 13) on accuracy improvement. Figures 4–7 give the upper and lower bounds of wind speed predicted by EMD-QRNN-KDE at the four stations, respectively. The interval widths of the 90% confidence level are basically larger than those of the 80% confidence level. At some times, such as the 110th hour (AH site in Figure 4b), the 150th hour (GD site, Figure 5b), the 30th hour (GS site, Figure 6a), and the 50th hour (HLJ site, Figure 7a), the interval predicted by the model can not well contain the actual wind speed values. However, these moments are in the minority. As most real wind speeds were covered by the upper and lower bounds, we can find that the performance of EMD-QRNN-KDE is good.

### Table 14. Interval prediction results at the HLJ station.

| Method        | PICP (%)  | NMPIW (%) | CWC (%) |
|---------------|-----------|-----------|---------|
|               | 90%       | 80%       | 90%     | 80%     | 90%     | 80%     |
| LQR-KDE       | 86.5      | 80.5      | 26.0    | 19.9    | 26.0    | 19.9    |
| QRNN-KDE      | 88.0      | 80.0      | 26.3    | 20.0    | 26.3    | 20.0    |
| QRRF-KDE      | 80.5      | 80.5      | 33.9    | 25.8    | 33.9    | 25.8    |
| QRSVM-KDE     | 85.0      | 80.5      | 18.3    | 13.2    | 18.3    | 13.2    |
| EMD-LQR-KDE   | 93.0      | 85.0      | 18.3    | 13.2    | 18.3    | 13.2    |
| EMD-QRNN-KDE  | 90.5      | 85.0      | 18.3    | 13.2    | 18.3    | 13.2    |
| EMD-QRRF-KDE  | 99.5      | 98.0      | 18.3    | 13.2    | 18.3    | 13.2    |
| EMD-QRSVM-KDE | 92.0      | 86.0      | 23.9    | 22.4    | 23.9    | 22.4    |

**Figure 4.** Interval predictions by EMD-QRNN-KDE at the AH station under two confidence levels: (a) 0–100 h and (b) 101–200 h.
Figure 5. Interval predictions by EMD-QRNN-KDE at the GD station under two confidence levels: (a) 0–100 h and (b) 101–200 h.

Figure 6. Interval predictions by EMD-QRNN-KDE at the GS station under two confidence levels: (a) 0–100 h and (b) 101–200 h.
Figure 7. Interval predictions by EMD-QRNN-KDE at the HLJ station under two confidence levels: (a) 0–100 h and (b) 101–200 h.

4.3. Probabilistic Density Prediction

From the analysis of the previous sections, one can see that the performance of the LQR and QRNN models on point and interval predictions could be greatly improved by EMD. In this section, by taking conditional quantiles obtained by various QR-KDE models as the input, the PDFs of wind speed at 200 h are obtained by the KDE method. The optimal bandwidth of the KDE method is determined by the NRC. Tables 15–18 present the CRPS values of both LQR and nonlinear QR models under the 80% and 90% confidence levels. Obviously, through the pre-processing of wind speed by EMD, the CRPS metric of probability density prediction is significantly reduced, except for the QRRF-EMD model. It is indicated that the introduction of EMD can improve the accuracy of probability density prediction for most QR-KDE models. In addition, the CRPS value of the EMD-QRNN-KDE model is the lowest, indicating that the EMD-QRNN-KDE model performs best in wind speed probability density prediction.
Here, the specific PDF curves predicted by EMD-QRNN-KDE at nine hours (the 1st, 25th, 50th, 75th, 100th, 125th, 150th, 175th, and 200th) under the 80% and 90% confidence levels for four wind observation stations are, respectively, given by Figures 8–11. In order to make the comparisons, the actual wind speed values at these hours are also presented in the figures. Except for the 150th hour of the AH station in Figure 8g, the 25th hour of the GD station in Figure 9b, the 125th hour of the GS station in Figure 10f, and the 175th hour of the HLJ station in Figure 11h, most actual wind speeds are always near the peaks of the predicted PDF curves, indicating that the probabilistic density prediction by EMD-QRNN-KDE is believable. Compared with the curves of the 90% confidence level, the density curves of the 80% confidence level usually have narrower wind speed ranges and higher peak values. At some times, the density curves might be biased distributions (see the 25th hour of the AH station in Figure 8b, the 125th hour of the GD station in Figure 9f, the 1st hour of the GS station in Figure 10a, the 1st hour of the HLJ station in Figure 11a, etc.), bimodal distributions (see the 175th hour of the AH station in Figure 8i, the 75th hour of the GD station in Figure 9c, etc.), or even multi-modal distributions (see the 75th hour of the AH station in Figure 8c, the 50th hour of the GD station in Figure 9c, the 175th hour of the HLJ station in Figure 11h, etc.).

| Table 15. Probabilistic density prediction results at the AH station. CRPS, continuous ranked probability score. |
| CRPS  | 90%      | 80%      |
|-------|----------|----------|
| LQR-KDE | 2.3533   | 2.0208   |
| QRNN-KDE | 2.3489   | 2.0183   |
| QRRF-KDE | 3.2095   | 2.6987   |
| QRSVM-KDE | 3.3704   | 2.8403   |
| EMD-LQR-KDE | 2.0415   | 1.7377   |
| EMD-QRNN-KDE | 2.0327   | 1.7262   |
| EMD-QRRF-KDE | 3.5994   | 3.0331   |
| EMD-QRSVM-KDE | 3.0755   | 2.6502   |

| Table 16. Probabilistic density prediction results at the GD station. |
| CRPS  | 90%      | 80%      |
|-------|----------|----------|
| LQR-KDE | 2.6622   | 2.0624   |
| QRNN-KDE | 2.6531   | 2.0534   |
| QRRF-KDE | 3.6923   | 3.0677   |
| QRSVM-KDE | 3.8345   | 3.2408   |
| EMD-LQR-KDE | 2.3575   | 1.9552   |
| EMD-QRNN-KDE | 2.3477   | 1.9318   |
| EMD-QRRF-KDE | 5.0087   | 4.2460   |
| EMD-QRSVM-KDE | 3.6470   | 3.1872   |

| Table 17. Probabilistic density prediction results at the GS station. |
| CRPS  | 90%      | 80%      |
|-------|----------|----------|
| LQR-KDE | 4.4190   | 3.4790   |
| QRNN-KDE | 4.4161   | 3.4782   |
| QRRF-KDE | 5.8681   | 4.7660   |
| QRSVM-KDE | 5.6816   | 4.6067   |
| EMD-LQR-KDE | 3.4384   | 2.8105   |
| EMD-QRNN-KDE | 3.4375   | 2.7995   |
| EMD-QRRF-KDE | 6.5087   | 5.4665   |
| EMD-QRSVM-KDE | 5.4326   | 4.3725   |
Table 18. Probabilistic density prediction results at the HLJ station.

| Model          | CRPS   | 90%   | 80%   |
|----------------|--------|-------|-------|
| LQR-KDE        | 4.4642 | 3.6488|
| QRNN-KDE       | 4.4660 | 3.6234|
| QRRF-KDE       | 5.4351 | 4.5577|
| QRSVM-KDE      | 5.5742 | 4.7362|
| EMD-LQR-KDE    | 3.1572 | 2.6379|
| EMD-QRNN-KDE   | 3.1565 | 2.6354|
| EMD-QRRF-KDE   | 6.0577 | 5.0770|
| EMD-QRSVM-KDE  | 5.3849 | 4.4406|

Based on the EMD-QRNN-KDE model, we can not only get the specific PDF curves of wind speeds, but also obtain the dynamic change of density distributions with time. This means that compared with traditional point and interval prediction models, the proposed QR-KDE models could acquire more information about the randomness and uncertainty of the actual wind speed.

Figure 8. Probabilistic density curves based on EMD-QRNN-KDE at the AH station: (a) 1st hour, (b) 25th hour, (c) 50th hour, (d) 75th hour, (e) 100th hour, (f) 125th hour, (g) 150th hour, (h) 175th hour, and (i) 200th hour. Dashed line: 80% confidence level; solid line: 90% confidence level; star: real value of wind speed.
Figure 9. Probabilistic density curves based on EMD-QRNN-KDE at the GD station: (a) 1st hour, (b) 25th hour, (c) 50th hour, (d) 75th hour, (e) 100th hour, (f) 125th hour, (g) 150th hour, (h) 175th hour, and (i) 200th hour. Dashed line: 80% confidence level; solid line: 90% confidence level; star: real value of wind speed.

Figure 10. Probabilistic density curves based on EMD-QRNN-KDE at the GS station: (a) 1st hour, (b) 25th hour, (c) 50th hour, (d) 75th hour, (e) 100th hour, (f) 125th hour, (g) 150th hour, (h) 175th hour, and (i) 200th hour. Dashed line: 80% confidence level; solid line: 90% confidence level; star: real value of wind speed.
5. Conclusions

A framework for probability density forecasting of wind speed based on QR and KDE is proposed. EMD is implemented to reduce the noise of raw wind speed series. Both LQR and NQR (QRNN, QRRF, and QRSVM) are, respectively, utilized to study the de-noised wind speed series. By taking the predicted conditional quantiles as the input, PDFs of wind speed at different times are obtained by the KDE method. Various experiments and comparisons are conducted on the real wind speed data at four wind observation stations in China. The conclusions are summarized as follows:

1. EMD-LQR-KDE and EMD-QRNN-KDE have the best performance and robustness in both point and interval predictions.

2. Most actual wind speeds lay near the peak of the predicted PDF curves, indicating that the probabilistic density prediction by EMD-QRNN-KDE is believable.

3. With the change of times, the predicted density curves might be biased distributions, bimodal distributions, or even multi-modal distributions.

The results show that the QR-KDE model can not only provide point prediction and interval prediction results, but also provide the probability density distribution of wind speed at any moment. Therefore, the research results will help to deepen the understanding of the randomness and uncertainty of actual wind speeds.

Author Contributions: Conceptualization, L.X. and Q.H.; methodology, software, validation, L.Z. and Q.H.; writing, original draft preparation, L.Z.; writing, review and editing, L.X. and Q.H.; supervision, L.X.; project administration, Z.W. and C.H. All authors read and agreed to the published version of the manuscript.

Funding: The research work described in the paper was supported by National Key Research and Development Program (No. 2017YFB1302200), Intelligent robot and system high level innovation center open fund (No.2018IRS01), and the National Science Foundation of China under Grant No. 61672093/11872222.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Global Wind Reports Published by GWEC. Available online: http://gwec.net/cost-competitiveness-puts-wind-in-front/ (accessed on 25 April 2018).
2. Han, Q.; Meng, F.; Hu, T.; Chu, F. Non-parametric hybrid models for wind speed forecasting. *Energy Convers. Manag.* 2017, 148, 554–568. [CrossRef]
3. He, Q.; Wang, J.; Lu, H. A hybrid system for short-term wind speed forecasting. *Appl. Energy* 2018, 226, 756–771. [CrossRef]
4. Liu, H.; Duan, Z.; Li, Y.; Lu, H. A novel ensemble model of different mother wavelets for wind speed multi-step forecasting. *Appl. Energy* 2018, 228, 1783–1800. [CrossRef]
5. Tian, C.; Hao, Y.; Hu, J. A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization. *Appl. Energy* 2018, 231, 301–319. [CrossRef]
6. Han, Q.; Hao, Z.; Hu, T.; Chu, F. Non-parametric models for joint probabilistic distributions of wind speed and direction data. *Renew. Energy* 2018, 126, 1032–1042. [CrossRef]
7. Han, Q.; Ma, S.; Wang, T.; Chu, F. Kernel density estimation model for wind speed probability distribution with applicability to wind energy assessment in China. *Renew. Sustain. Energy Rev.* 2019, 115, 109387. [CrossRef]
8. Lydia, M.; Kumar, S.; Selvakumar, A.; Kumar, G. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* 2016, 112, 115–124. [CrossRef]
9. Noorollahi, Y.; Jokar, M.; Kalhor, A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. *Energy Convers. Manag.* 2016, 115, 17–25. [CrossRef]
10. Zhou, J.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* 2011, 52, 1990–1998. [CrossRef]
11. Salcedo-Sanz, S.; Pastor-Sánchez, A.; Prieto, L.; Blanco-Aguilera, A.; García-Herrera, R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—Extreme learning machine approach. *Energy Convers. Manag.* 2014, 87, 10–18. [CrossRef]
12. Chen, J.; Zeng, G.; Zhou, W.; Du, W.; Lu, K. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Convers. Manag.* 2018, 165, 681–695. [CrossRef]
13. Tascikaraoglu, A.; Uzunoglu, M. A review of combined approaches for prediction of short-term wind speed and power. *Renew. Sustain. Energy Rev.* 2014, 34, 243–54. [CrossRef]
14. Song, Z.; Jiang, Y.; Zhang, Z. Short-term wind speed forecasting with Markov-switching model. *Appl. Energy* 2014, 130, 103–112. [CrossRef]
15. Iverson, E.; Morales, J.; Moller, J.; Madsen, H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *Int. J. Forecast.* 2016, 32, 981–990. [CrossRef]
16. Shrivastava, N.; Lohia, K.; Panigrahi, B. A multiobjective framework for wind speed prediction interval forecasts. *Renew. Energy* 2016, 87, 903–910. [CrossRef]
17. Li, R.; Jin, Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. *Appl. Energy* 2018, 228, 2207–2220. [CrossRef]
18. Wang, J.; Niu, T.; Lu, H.; Guo, Z.; Yang, W.; Du, P. An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms. *Appl. Energy* 2018, 211, 492–512. [CrossRef]
19. Qin, S.; Liu, F.; Wang, J.; Song, Y. Interval forecasts of a novelty hybrid model for wind speeds. *Energy Rep.* 2015, 1, 8–16. [CrossRef]
20. Zhang, C.; Wei, H.; Xie, L.; Shen, Y.; Zhang, K. Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework. *Neurocomputing* 2016, 205, 53–63. [CrossRef]
21. Wang, H.; Wang, G.; Li, G.; Peng, J.; Liu, Y. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl. Energy* 2016, 182, 80–93. [CrossRef]
22. Thorarinsdottir, T.; Gneiting, T. Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *J. R. Stat. Soc. Ser. A* 2010, 173, 371–388. [CrossRef]
23. Slaughter, J.; Gneiting, T.; Raftery, A. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* 2010, 105, 25–35. [CrossRef]
24. Baran, S. Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components. *Comput. Stat. Data Anal.* 2014, 75, 227–238. [CrossRef]

25. Baran, S.; Lerch, S. Log-normal distribution based ensemble model output statistics models for probabilistic wind-speed forecasting. *Q. J. R. Meteorol. Soc.* 2015, 141, 2289–2299. [CrossRef]

26. Hu, J.; Wang, J. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy* 2015, 83, 1456–1466. [CrossRef]

27. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* 1978, 46, 33–50. [CrossRef]

28. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.

29. Cannon, A. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Comput. Geosci.* 2011, 37, 1277–1284. [CrossRef]

30. Meinshausen, N. Quantile regression forests. *J. Mach. Learn. Res.* 2006, 7, 983–999.

31. Takeuchi, I.; Le Q.; Sears, T.; Smola, A. Nonparametric quantile regression. *J. Mach. Learn. Res.* 2005, 7, 1001–1032.

32. He, Y.; Liu, R.; Sa, A. Short-term power load probability density forecasting method based on real time price and support vector quantile regression. *Proc. CSEE* 2017, 37, 758–776.

33. He, Y.; Liu, R.; Li, H.; Wang, S.; Lu, X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory. *Appl. Energy* 2017, 185, 254–266. [CrossRef]

34. He, Y.; Xu, Q.; Wan, J.; Yang, S. Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. *Energy* 2016, 114, 498–512. [CrossRef]

35. He, Y.; Li, H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. *Energy Convers. Manag.* 2018, 164, 374–384. [CrossRef]

36. Zheng, W.; Peng, X.; Lu, D.; Zhang, D.; Liu, Y.; Lin, Z.; Lin, L. Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: A new approach. *Energy Convers. Manag.* 2017, 151, 737–752. [CrossRef]

37. Wang, J.; Hu, J.; Ma, K. Wind speed probability distribution estimation and wind energy assessment. *Renew. Sustain. Energy Rev.* 2016, 60, 881–899. [CrossRef]

38. Zamborn, A.; Dias, R. A review of kernel density estimation with applications to econometrics. *Int. Econom. Rev.* 2013, 5, 20–42.

39. Huang, N.E.; Shen, Z.; Long, S. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. A* 1998, 454, 903–995. [CrossRef]

40. Boudraa, A.; Cexus, J. EMD-Based Signal Filtering. *IEEE Trans. Instrum. Meas.* 2007, 56, 2196–2202. [CrossRef]

41. Bishop, C. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, UK; New York, NY, USA, 1995.

42. Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32. [CrossRef]

43. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.

44. Khosravi, A.; Nahav, S.; Creighton, D. A prediction interval based approach to determine optimal structures of neural network metamodels. *Expert Syst. Appl.* 2010, 37, 2377–2387. [CrossRef]

45. Khosravi, A.; Nahav, S.; Creighton, D. Lower upper bound estimation method for construction of neural network based prediction intervals. *IEEE Trans. Neural Netw.* 2011, 22, 337–346. [CrossRef] [PubMed]

46. Matheson, J.; Winkler, R. Scoring rules for continuous probability distributions. *Manag. Sci.* 1976, 22, 1087–1096. [CrossRef]

47. Gneiting T, Raftery A. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* 2007, 102, 359–378. [CrossRef]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).