

Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression

He, Yaoyao; Li, Haiyan ; Wang, Shuo; Yao, Xin

DOI:

[10.1016/j.neucom.2020.10.093](https://doi.org/10.1016/j.neucom.2020.10.093)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

He, Y, Li, H, Wang, S & Yao, X 2021, 'Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression', *Neurocomputing*, vol. 430, pp. 121-137.
<https://doi.org/10.1016/j.neucom.2020.10.093>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression

Yaoyao He ^{a,b,*}, Haiyan Li^{a,b}, Shuo Wang^c, Xin Yao^{d,c}

^a School of Management, Hefei University of Technology, Hefei 230009, China

^b Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230009, China

^c CERCIA, the School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

^d Shenzhen Key Laboratory of Computational Intelligence, School of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Abstract

Accurate forecasting of wind power plays an important role in an effective and reliable power system. However, the fact of non-schedulability and fluctuation of wind power significantly increases the uncertainty of power systems. The output power of a wind farm is usually mixed with uncertainties, which reduce the effectiveness and accuracy of wind power forecasting. In order to handle the uncertainty of wind power, this paper first proposes to conduct outlier detection and reconstruct data before the prediction. Then, a wind power probability density forecasting method is proposed, based on cubic spline interpolation and support vector quantile regression (CSI-SVQR), which can better estimate the whole wind power probability density curve. However, the probability density prediction method can not acquire the optimal point prediction and interval prediction results at the same time. In order to analyze the uncertainty of wind power, the present study considers the prediction results from the perspective of probabilistic point prediction and interval prediction respectively. Three sets of real-world wind power data from Canada and China are used to validate the CSI-SVQR method. The results show that the proposed method not only efficiently eliminates the outliers of wind power but also provides the probability density function, offering a complete description of wind power generation fluctuation. Furthermore, more accurate point prediction and prediction interval (PI) can be obtained compared to existing methods. **Wilcoxon signed rank test is used to verify that CSI can improve the performance of forecasting methods**

Keywords: Wind power forecasting; Support vector quantile regression (SVQR); Cubic spline interpolation (CSI) function; Probability density forecasting.

*Corresponding author. School of Management, Hefei University of Technology, Hefei 230009, China

Email addresses: hy-342501y@163.com (Yaoyao He ^{a,b,*}), 2465460550@qq.com (Haiyan Li^{a,b}), s.wang.2@bham.ac.uk (Shuo Wang^c), xiny@sustc.edu.cn, X.Yao@cs.bham.ac.uk (Xin Yao^{d,c})

1. Introduction

Wind power forecasting is one of the most important ways to reduce negative impact from wind power uncertainty [1]. Wind power is an inherently uncertain and highly random energy resource, which poses huge challenges to wind power forecasting. Accurate wind power prediction can help to regulate and schedule the power system and cut down the hazard of multiple grids integration [2–4]. The forecast of short-term wind power is important to the system safety and stability, which help to achieve better performance for the power system [5–7].

A lot of research has been carried out in the field of wind power forecasting in recent years. Existing wind power prediction methods fall into four categories [8]: 1) statistical methods, such as autoregressive integrated moving average (ARIMA) and Markov chain [9, 10]; 2) artificial intelligence methods such as artificial neural network (ANN) [11, 12], fuzzy logic method [13, 14]; 3) physical methods that employ numerical weather prediction (NWP) and the surrounding physical information of the wind farm, for instance, surface roughness, pressure and obstacles, to build forecasting models [15–17]; and 4) hybrid methods that combine the above categories.

The statistical and artificial intelligence methods do not use relevant physical information for forecasting and usually suffer from higher prediction errors in short-term wind power forecasting. Hybrid methods become more popular because they can combine the merits of the other categories. For example, in [18], vector autoregressive moving average-generalized autoregressive conditional heteroscedastic (VARMA- GARCH) was applied to wind direction and speed data, the forecasting results are converted to a wind power density function by conditional kernel density (CKD). In [19], a support vector machine (SVM) combined with empirical mode decomposition (EMD) was implemented. The wind power time series were classified into several sequences by EMD, and the SVM was used to optimize the training parameters for the accurate prediction result.

In terms of the output, traditional wind power forecasting methods only produce a point value, or the conditional expectation of the output at a time point. It is difficult to get the comprehensive knowledge of future events, leading to increased uncertainties for wind power operation [20]. To quantify the uncertainty of wind power systems, probabilistic forecasting takes the form of prediction intervals (PIs) or predictive density functions. In [21], a bootstrap-based extreme learning machine (BELM) was proposed to construct the PIs of wind power time series. In [22], a neural network (NN)-based method for the construction of PIs with a prescribed probability called the confidence level was proposed. Particle swarm optimization (PSO) based on the lower upper bound estimation (LUBE) was used to solve this problem. As to probabilistic forecasting methods, the ideal forecasting results should contain information on the probability distribution of expected wind power. However, due to the intermittent and volatile characteristics of the wind power data distribution, the complete probability distribution of wind power time series is hard to be estimated.

Koenker and Bassett [23] proposed a quantile regression (QR) method, which has the capability to solve point prediction of various quantiles. Any information on the shape of the distributions can easily be determined by means of the estimation throughout the range of quantiles. Hence, complete information on the prediction can be obtained without any distributional assumptions. In [24], a local quantile regression (LQR) was used to define 90% and 50% PIs. A large number of quantiles of the probability distribution was estimated using wind power data. Bessa et al. [25] proposed a time-adaptive quantile-couple estimator to select the appropriate kernel function for wind power probabilistic forecasting, which was validated on two real wind farms. In [26], a probabilistic wind power forecasting model was put forward based on a parametric additive quantile regression (PAQR) method, which was implemented in a practical wind power system using R software. PAQR is used to fit the training data by means of the sum of spline functions for a given degree of freedom. It is evaluated on the testing set, when parameters of spline basis functions are successfully estimated. If model parameters and input data are certain, the prediction of PAQR is invariable. As an open wind power probabilistic forecasting method based on QR, the R-script code of PAQR can be downloaded in [26]. Hence, it is well suited as the comparative model.

The SVM was proposed by Vapnik [27], which has been widely used in regression problems with small sample sizes and high dimensions. Considering the complexity and potential nonlinearity of wind power, support vector regression (SVR) has been used to solve this problem because of its outstanding performance in real-world applications [28–31]. It is a kernel-based method, which can translate nonlinear regression to linear regression problems. Due to the strong generalization capability of Gaussian kernels, the SVR model constructed by Gaussian kernel functions (i.e. radial basis functions) has been widely applied in the forecasting field [32, 33]. However, the traditional SVR model does not provide the probability density prediction. In [34], in order to quantify the forecasting uncertainty, SVM was used to improve QR. The support vector quantile regression (SVQR) model was proposed, which can deal with nonlinear structural problems in economic systems. Finally, Xu et.al [35, 36] investigated the relationship between linear and nonlinear value at risk (VaR) through the SVQR model. The prediction results show that the method is better than traditional methods.

The kernel functions in SVQR have a great effect on the prediction performance. Gaussian kernel is shown to be the excellent kernel function in SVQR [37]. However, a unitary SVQR method is insufficient to describe the probability density function of wind power. As explained in Parzen and Rosenblatt [38, 39], Kernel density estimation (KDE) is a non-parametric method used to estimate the probability density curve of a dataset with an unknown distribution [18]. It does not assume any data distribution, which is an advantage over parametric methods. For KDE, Epanechnikov kernel function is optimal in a mean square error sense [40]. This paper proposed a method based on the SVQR method and Epanechnikov KDE for predicting wind power. The complete probability density curve at arbitrary moment in the future can be achieved.

Another difficult in wind power forecasting is outliers. Due to the inherent variability in wind power, not all wind farm data are of good quality. Outliers pollute wind farm data, which can significantly affect the forecasting accuracy [17]. Such outliers may come from wind speed, fan blades, wind farm maintenance operation and shut-down, and meter measuring errors, etc [41]. In [42], outliers were removed during the wind generation time periods, using consistency examination detection. It deletes outliers, and reduces the sample size, which may lead to inaccurate prediction. It needs a large amount of samples and its calculation process is relatively complex.

To simplify the computation and obtain accurate results, this article adopts the quartile method to detect the anomaly of time series in the initial data processing, then a cubic spline interpolation function (CSI) is applied to replace the abnormal points before forecasting the wind power. A novel method of wind power forecasting is presented, combining the advantages of CSI function and SVQR model (CSI-SVQR). This method considers the wind power abnormal data for the stable operation of electricity markets. Three real-world wind power data sets are used to test its feasibility and effectiveness. The power curve outliers are analyzed by the quartile method. The CSI function is used to reconstruct the time series. The optimal interval prediction and mode point forecasting criteria are adopted to analyze the uncertainty of forecasting results. The proposed method is compared with existing ϵ -SVR, ν -SVR [28], Gaussian process regression (GPR) [43], extreme learning machine (ELM) [43–45], **random vector functional link network (RVFL)** [46–48] point forecasting methods and PAQR [26] probabilistic forecasting method, to evaluate the quality of probability density forecasting through the reliability criterion. In addition to producing the probability density function, the comparative results indicate that the method can produce more accurate prediction.

The main contributions of this paper are: 1) The present study proposes a new CSI-SVQR method for wind power probability density forecasting. The combination of the outlier detection and SVQR model such that the outliers are removed from the final prediction by means of the quartile method. CSI functions are adopted to modify the original data to reduce the noisy of original wind power times series. 2) By means of the Epanechnikov KDE, the complete probability density curve of wind power is obtained, and more comprehensive information is obtained. 3) Different from existing trial and error methods, the grid search method is adopted to explore the optimal parameters of SVQR and CSI-SVQR. 4) **Probability density forecasting methods need to consider several metrics. It is difficult to optimize these metrics at the same time. Here, the optimal interval prediction and mode point prediction metrics are adopted to analyze the uncertainty of wind power probability density prediction results, respectively.** 5) The experimental results indicate that the accuracy of PI and point prediction values have been remarkably improved compared with ϵ -SVR, ν -SVR, GRP, ELM, **RVFL**, PAQR, and SVQR. In addition, **Wilcoxon signed rank test verifies that the methods combined with the CSI function are better than the forecasting technique without using CSI.**

The remainder of this paper is organized as follows. Section 2 introduces the method of outlier detection and interpolation. The forecasting method based on CSI-SVQR and the evaluation metrics of forecasting

results are presented in Section 3. Section 4 discusses wind power probability density forecasting through three case studies. Finally, Section 5 draws conclusion and points out future research directions for wind power probability density forecasting.

2. Data preprocessing

The collected wind power data often contain accidental errors and system errors. These errors may be caused by the sensor malfunctions and faults and the instability of wind turbines. Previously proposed methods incline to predict the output of a wind farm based on the collected data sets without data preprocessing, which can greatly reduce the prediction accuracy. The present study proposes to use the quartile data preprocessing method to identify exceptional values and clear up the original data before prediction.

2.1. Outlier detection

The quartile method [41] can be used to identify abnormal values in data without the assumption of the data distribution. The quartile method can be illustrated by Fig. 1.

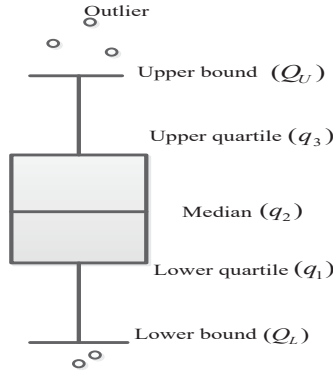


Figure 1: Quartile method [41]

If a test data point lies outside the interquartile range $[Q_L, Q_U]$, it can be considered as an outlier [41]. The fences formulation is calculated as follows:

$$[Q_L, Q_U] = [q_1 - 1.5 * Q_R, q_3 + 1.5 * Q_R] \quad (1)$$

In Eq (1), lower quartile q_1 and upper quartile q_3 are the first and third quartiles, respectively. The median (q_2) is the 50th percentile of the data. The first quartile is the 25th percentile, and the third quartile is the 75th percentile. Q_R is the interquartile range, namely, $Q_R = q_3 - q_1$, which is a comparatively steady statistic in terms of the standard deviation. The lower bound is the Q_L and the upper bound is the Q_U

of data, and any data lying outside these defined bounds can be considered an outlier. This method has certain advantage that it can be employed even when data spots are not normally distributed, because the quartile method relies on the median and not the mean of the samples [41]. Considering the high uncertainty of the wind power data, the quartile method is used to identify outliers in this paper.

2.2. Cubic spline interpolation function

The above quartile method finds out the outliers in the collected wind power data sets. These outliers can actually be used to improve the prediction precision by means of spline interpolation. Spline interpolation not only obtains a higher degree of polynomial interpolation, but also keeps the stability because of Runge's phenomenon [49]. In this paper, the outliers are replaced by a cubic spline interpolation (CSI) function, which is the most commonly used interpolating method [49].

The basic concept of the cubic spline is to acquire a smooth curve by a series of points. Supposing that there are $N + 1$ data samples $\{(x_i, y_i), i = 0, 1, 2, \dots, N\}$ and $x_0 \leq x_1 \leq x_2 \leq \dots \leq x_N$, the present study constructs a spline function $S(x)$ which conforms the requirements $S(x_i) = y_i$. A spline is a cubic polynomial function in the interval $x \in [x_i, x_{i+1}]$. $S(x)$ is equal to $a_i + b_i x + c_i x^2 + d_i x^3$. The coefficients of the cubic polynomial a_i, b_i, c_i and d_i are used to guarantee the smoothness of generated data. Function $S(x)$ must ensure that $S(x), S'(x), S''(x)$ are continuous everywhere.

Through the above outlier detection and data transformation steps, the outliers found by the quartile method are transformed by the CSI function. A continuous and smooth curve is obtained. The points on this curve will be used to build the model for wind power forecasting.

3. Probability density forecasting based on CSI-SVQR method

3.1. SVQR model

The least squares estimation is a traditional method to obtain the approximate results of the explained variable, which gives specific values of the explanatory variable. Koenker and Bassett [23] proposed a quantile regression (QR) method in order to estimate the conditional median and other quantiles of the explanatory variable. Compared with the least squares regression, the major advantage of QR is the robustness against outliers. Suppose that there is a set of stochastic vectors $Y = [Y_1, Y_2, \dots, Y_n]$ and independent vectors $X = [X_1, X_2, \dots, X_n]$, where n is the sample size, and $X_i = [x_{i1}, x_{i2}, \dots, x_{ir}]'$. The response variable Y_i is characterized by the cumulative distribution function $F_{Y_i}(y) = P(Y_i \leq y)$, and the τ -th quantile of Y_i is described as follows:

$$Q_{Y_i}(\tau) = F_{Y_i}^{-1} = \inf(y : F_{Y_i}(y) \geq \tau) \quad (2)$$

where quantile fractile $0 < \tau < 1$. In [38], the linear quantile regression model is defined as:

$$Q_{Y_i}(\tau|X_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \beta_2(\tau)x_{i2} + \dots + \beta_r(\tau)x_{ir} \quad (3)$$

where r is the number of quantile, and $Q_{Y_i}(\tau|X_i)$ is the τ conditional quantile of dependent variable Y_i under independent variable X_i . The regression coefficient vector is $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), \dots, \beta_r(\tau)]'$. The optimum vector $\beta(\tau)$ can be estimated by the optimization function $h(\beta)$:

$$h(\beta) = \min_{\beta} \sum_{i=1}^n \rho_{\tau}(Y_i - X_i' \beta) \quad (4)$$

where $\rho_{\tau}(Y_i - X_i' \beta)$ is the loss function:

$$\rho_{\tau}(Y_i - X_i' \beta) = \begin{cases} (Y_i - X_i' \beta)\tau, & Y_i \geq X_i' \beta \\ (Y_i - X_i' \beta)(\tau - 1), & Y_i < X_i' \beta \end{cases} \quad (5)$$

The response variable Y_i of conditional quantile at any quantiles is obtained by the estimated parameter vector $\beta(\tau)$, and then the conditional density forecasting is able to be achieved. According to formulas (3) and (4), the linear structure is adopted by the QR model. It means that the model of the explained variables and explanatory variables is linear. However, the relationship of wind power time series is nonlinear. The simple QR is hard to solve complex wind power forecasting problems.

SVM is a supervised learning model in machine learning. It has been successfully used to solve nonlinear and regression problems. SVR [50] modifies SVM to solve linear regression problems. The main idea is that the vector can be mapped into a multidimensional feature space by nonlinear mapping χ . The constructed SVR model is as follows:

$$f(x) = \omega^T \chi(x) + b \quad (6)$$

The parameter vector ω and threshold value b can be obtained by minimizing the following

$$\min_{\omega, b} \left(\frac{1}{2} \omega^T \omega + C \sum_{i=1}^{n_t} |y_i - f(x_i)| \right) \quad (7)$$

where C is the penalty parameter, n_t is the sample size of the training set, and x_i, y_i are the input and output vector of the training data sets.

In order to solve the nonlinear and heterogeneous problems, the SVQR model is proposed by Takeuchi [34]. Shim [51] introduced a simple SVQR model on the basis of the semiparametric method. The model is obtained by using QR to substitute the penalty function in the SVR model, which is defined as :

$$\min_{\omega_{\tau}, b_{\tau}} \left(\frac{1}{2} \omega_{\tau}^T \omega_{\tau} + C \sum_{i=1}^{n_t} \rho_{\tau}(y_i - b_{\tau} - \gamma_{\tau}^T z_i - \omega_{\tau}^T \chi(x_i)) \right) \quad (8)$$

where ω_{τ} is the parameter vector at τ quantile, and $\rho_{\tau}(\cdot)$ is the loss function as defined in Eq (4). z_i, x_i represent, the linear and nonlinear vector. b_{τ} is threshold value at τ quantile, γ_{τ} is regression vector at τ quantile, and y_i is the output vector of the training data sets.

Slack variable and the Lagrange multiplier [52] are used to estimate the SVQR model. Eq.(8) can be converted into an unconstrained problem, then the optimal values of $\omega_\tau, b_\tau, \gamma_\tau$ can be represented as:

$$\begin{cases} \omega_\tau = \sum_{i=1}^{n_t} (\varepsilon_i - \varepsilon_i^*) \chi(x_i) \\ (b_\tau, \gamma_\tau) = (A_S^T A_S)^{-1} A_S^T (y_S - \kappa_S(\varepsilon - \varepsilon^*)) \end{cases} \quad (9)$$

where $\varepsilon_i, \varepsilon_i^*$ are the optimized Lagrange multipliers, the design matrix is $A_S = (1, z_i^T)$ and the condition of $i \in I_S$ must be met. The index set of support vectors $I_S = \{i = 1, 2, \dots, n_t \mid 0 < \varepsilon_i < \tau C, 0 < \varepsilon_i^* < (1 - \tau)C\}$ is obtained by Karush-Kuhn-Tucher conditions. In the above formula, κ_S is the kernel matrix and its elements are $\kappa(x_j, x_i) = \varphi(x_j)^T \varphi(x_i)$ with $j = 1, 2, \dots, n_t$ and $i \in I_S$. The Gaussian function is selected to calculate the kernel matrix, the formula is

$$\kappa(x_j, x_i) = e^{-\frac{\|x_j - x_i\|^2}{2\sigma^2}} \quad (10)$$

σ^2 is the free parameters of the kernel matrix.

3.2. Evaluation metrics for predicted results

3.2.1. Point prediction evaluation metrics

In order to evaluate the proposed CSI-SVQR method, the paper selects the mean absolute percentage error (MAPE) and relative mean square error (RMSE) as the evaluation criteria of point prediction. Considering that wind power is highly volatile and the output value of wind farm may be 0 or very close to 0, which is not favorable for statistical analysis, the present study uses the maximum history value Y_{max} as the denominator of relative errors [53]. The following point forecasting evaluation metrics are defined as:

$$MAPE = \frac{1}{n_f} \sum_{t=1}^{n_f} \left| \frac{Y_t - p_t}{Y_{max}} \right| \quad (11)$$

$$RMSE = \frac{\sum_{t=1}^{n_f} (Y_t - p_t)^2}{\sum_{t=1}^{n_f} Y_t^2} \quad (12)$$

where n_f is the number of forecasting samples, Y_t and p_t are the actual value and predicted result at the t -th moment, respectively.

This paper chooses mode and median of wind power probability density curve as the point prediction values. Mode is the highest probability value of the probability density curve and median is the middle value of a dataset. Supposing that there is a set of wind power predicted values at the i -th moment from small to large $f_{i1}, f_{i2}, \dots, f_{ir}$, the mode and median are calculated as follows respectively:

$$Median = \begin{cases} f_{i\frac{r+1}{2}} & , r \text{ is odd} \\ f_{i\frac{r}{2}} + f_{i\frac{r+2}{2}} & , r \text{ is even} \end{cases} \quad (13)$$

$$Mode = \arg \max(\hat{f}_h(x)) \quad (14)$$

where $\arg \max(\cdot)$ denotes the value of x when the function is maximal.

In order to evaluate the point forecasting performance of probability density methods, the mode forecasting error is selected as a evaluation metric

$$SSE_{mode} = \sum_{t=1}^{n_f} (Y_t - Mode_t)^2 \quad (15)$$

where SSE_{mode} is the sum of squared errors of mode and $Mode_t$ is forecasting value of mode at the t -th moment.

3.2.2. Evaluation metrics for prediction intervals

Different from point forecasting that simply provides the forecasting errors, PIs not only offer a prediction interval, in which practical values are most likely to be included, but also produce the coverage probability. Prediction interval normalized average width (PINAW) and prediction interval coverage probability (PICP) are commonly used. Their definitions are [22, 54]:

$$PINAW = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{U_i - L_i}{R} \quad (16)$$

$$PICP = \frac{1}{n_f} \sum_{i=1}^{n_f} \lambda_i \quad (17)$$

where n_f is the number of predictions, L_i, U_i are the lower bound and upper bound at the i -th moment. R represents the range between maximum upper bound minus minimum lower upper of the target values, which is used to normalize the PIs average width; λ_i is a Boolean function. It is defined as follows:

$$\lambda_i = \begin{cases} 1, & Y_i \in [L_i, U_i] \\ 0, & Y_i \notin [L_i, U_i] \end{cases} \quad (18)$$

PINAW is an objective measurement that is used for computing the average width of PIs. A small PINAW means narrow width of PIs, which is more advantageous for decision making. It is insufficient to only consider PINAW but ignore the coverage probability of PIs. As the cardinal feature of PIs, PICP expresses the percentage value of actual values covered by the PI, and it is expected to be greater than the level of nominal confidence [54, 55]. In the practical application of wind power prediction, a small PINAW and a high PICP are preferable.

To comprehensively assess the quality of PIs, Ref [22] designed an entertaining metrics, named coverage width-based criterion (CWC) for balancing the effect both PICP and PINAW, which is shown as follows:

$$CWC = PINAW + \rho(PICP)e^{-\gamma(PICP-\beta)} \quad (19)$$

where β is predefined as the nominal confidence level that should be satisfied. γ is a scaling factor that amplifies the diversity between PICP and β . $\rho(PICP) = 1$ in the training process. For testing set, $\rho(PICP)$ is a step function. During the evaluation of the test set PIs, if PICP satisfies the assigned nominal confidence level β , $\rho(PICP) = 0$, and CWC is equal to PINAW. Otherwise, $\rho(PICP) = 1$, and the exponential penalty term will be evaluated by CWC.

3.3. Wind power probability density forecasting based on CSI-SVQR

SSE_{mode} and CWC are two different types of evaluation indicators. It is hard to obtain simultaneous optimal results. High quality interval prediction may partially sacrifice point prediction performance. In this subsection, the article describes the CSI-SVQR from the optimal SSE_{mode} and CWC, respectively. The flowchart of the proposed method is shown in Fig. 2.

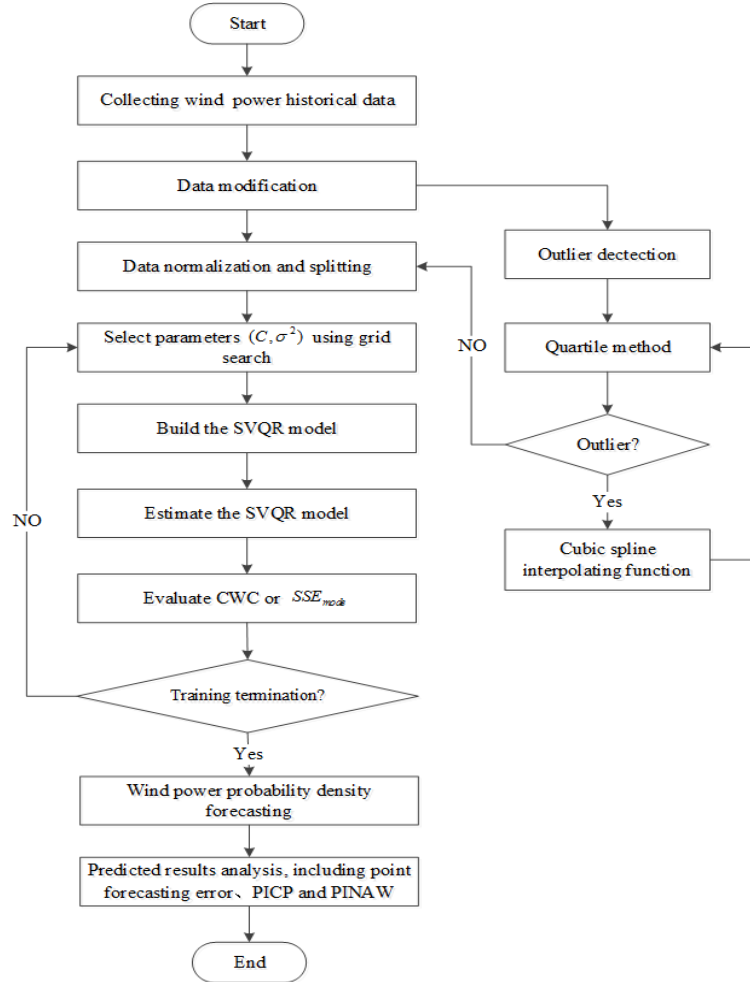


Figure 2: The flowchart of CSI-SVQR probability density forecasting method

3.3.1. Data modification

The original data need to be pre-processed because the outliers may seriously affect the forecasting accuracy. In this article, the quartile method is applied to detect the abnormal values which are embedded in the data sets. After the outliers are determined, they are replaced by the CSI function. Then, modified wind power time series are obtained. The detailed description of this part has been explained in Section 2.

3.3.2. Data normalization and splitting

The above preprocessed data are normalized, with a weakened uniform dimension from the time series, which is calculated by:

$$x_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad i = 1, 2, \dots, n \quad (20)$$

where x_i is the processed data, X_i is the unprocessed data, and X_{\min} , X_{\max} are the minimum and maximum of original wind power series, respectively; n is the total number of samples. The obtained new data sets are further divided into two sets: a training set and a test set. The training data set is applied to train the CSI-SVQR model. The test data set is used to test the performance of the proposed method.

3.3.3. Select parameters using grid search

The selection of parameters C , σ^2 may affect the performance of SVQR. Similar to SVR [27], the present study selects the parameters of SVQR using the grid search method, which adopts a grid that is constructed by exponentially growth sequences of C and σ^2 . Namely, in all the candidate parameters, each possibility is tried through a loop, and the best performing parameter is the final result.

3.3.4. Build the SVQR model

The modified wind power time series x_i are used as the input to SVQR according to Eq (8).

3.3.5. Estimate the SVQR model

The estimator of SVQR is obtained by Eq (9). Then, the conditional quantile at different quantiles can be obtained based on the optimized parameter vector ω_τ and threshold value b_τ , and the conditional quantile of wind power forecasting is defined as follows:

$$Q_{y_i}(\tau|z_i, x_i) = b_\tau + \gamma_\tau^T z_i + K_i(\varepsilon - \varepsilon^*) \quad (21)$$

where K_i is the i -th quantile of kernel matrix.

3.3.6. Evaluate CWC or SSE_{mode}

In order to measure the interval prediction quality of SVQR, we try to minimize CWC

$$\min(CWC(C, \sigma^2)) \quad (22)$$

Furthermore, the point forecasting results of mode is assessed by means of minimizing SSE_{mode}

$$\min(SSE_{mode}(C, \sigma^2)) \quad (23)$$

Then, the ability of uncertainty analysis of SVQR is shown from the perspective of probabilistic point prediction and interval prediction respectively.

3.3.7. Training termination

The training process is terminated if the optimal $\omega_\tau, b_\tau, \gamma_\tau$ are obtained. Otherwise, it continues and returns to Step 3.

3.3.8. Wind power probability density forecasting

In this paper, kernel density estimation (KDE) is adopted to forecast the probability density function of wind power. The conditional quantile $Q_{y_i}(\tau|z_i, x_i)$ is taken as the input of KDE function to achieve the predicted probability density curve. Suppose there is a set of random variables X_1, X_2, \dots, X_r that are taken from a identically and independent distributed sample set, the KDE is:

$$\hat{f}_h(x) = \frac{1}{rh} \sum_{i=1}^r K_E\left(\frac{X_i - x}{h}\right) \quad (24)$$

where r is the number of quantiles, $K_E(\cdot)$ is the kernel function. h is the bandwidth of kernel function, which seriously affects the accuracy of forecasting results.

There are many commonly used kernel functions: Triangular, Uniform, Epanechnikov, Gaussian and so on [56]. Among those functions, the loss of efficiency and mean square error of Epanechnikov kernel are both less than the other kernel functions, and this kernel is convenient to calculate [40, 57]. Based on these reasons, the Epanechnikov kernel is selected as a kernel function for probability density forecasting and its mathematical expression is shown:

$$K_E(\eta) = \begin{cases} \frac{3}{4}(1 - \eta^2) & , \quad \eta \in [-1, 1] \\ 0, & \eta \notin [-1, 1] \end{cases} \quad (25)$$

and $\eta = \frac{X_i - x}{h}$ in Eq (25).

3.3.9. Prediction results analysis

Once training CSI-SVQR terminates, the parameters of C, σ^2 are chosen to test the proposed method. The CSI-SVQR method is adopted to obtain the conditional quantile $Q_{y_i}(\tau|z_i, x_i)$. The conditional quantile of wind power forecasting under all quantiles are inputted into the kernel function to build probability density functions. Then the minimum and maximum values of probability density functions is defined as the upper and lower bounds of the prediction intervals. The evaluation criteria in Section 3.2 are applied to

evaluate the performance of wind power probability density forecasting results. Compared with point and interval prediction, the proposed probability density forecast method can provide more useful informations for the decision-making department.

4. Case studies

4.1. Data sources

The real wind power data from a wind farm in Canada [58] and a wind turbine in China are adopted to verify the effectiveness of the proposed algorithm. These two different places represent different types of wind power profiles. The wind power data from China is more complex than those from Canada because wind power data in a turbine may include many 0 power values. In the first case study, the data from Canada are relatively clean without outlier. Data in the second case study are also from Canada, which is more noisy. Data in the third case study from China contain more outliers. The wind power data of the first case were collected from January 20, 2015 to January 24, 2015, from Ontario in Canada, with 24 points in every day (120 samples). In the second case study, data were collected hourly from June 1, 2014 to June 28, 2014 in Ontario, Canada (672 samples). The third dataset was collected every fifteen minutes from May 30, 2006 to June 3, 2006 in Mathematical modeling competition of electrical cup of China of 2011 [59] (480 samples). The boxplots of outlier detection for three cases are shown in Fig. 3, including wind farm data without outliers in case 1, wind farm data with outliers in case 2, and wind turbine data with outliers in case 3. It is easy to discover that several outliers exist in cases 2 and 3, which may influence the accuracy of forecasting methods.

Each data set is divided into two subsets: In the first case, the data from January 20, 2015 to January 23, 2015 are used for training, and the data of January 24, 2015 are used for testing (24 test samples); In the second case, the wind power data of Ontario from June 1, 2014 to June 21, 2014 are considered as the training data, and the data from June 22, 2014 to June 28, 2014 are used as the testing data (168 test samples); In the third case, the data of China from May 30, 2006 to June 2, 2006 are selected for training, while the data of June 3, 2006 are selected for testing (96 test samples). By trial and error, we select previous 2-10h wind power data as the input data, showing similar forecasting results. For the Canada data sets, the input of previous 7 hours is slightly better; for the China data set, the input of previous 4 quarterly-hours is slightly better.

In all cases, 20 quantiles with the interval of 0.05 are chosen to perform the probability density forecasting, and the quantile is from 0.01 to 0.96 (i.e., the confidence level is 95%). To evaluate the performance of CSI-SVQR model, the point prediction methods, including ϵ -SVR, ν -SVR, GPR [43], ELM [43–45], RVFL [46–48], CSI- ϵ -SVR (the ϵ -SVR model combined with CSI), CSI- ν -SVR (the ν -SVR model combined with CSI), CSI-GPR (the GPR model combined with CSI), CSI-ELM (the ELM model combined with CSI), and CSI-

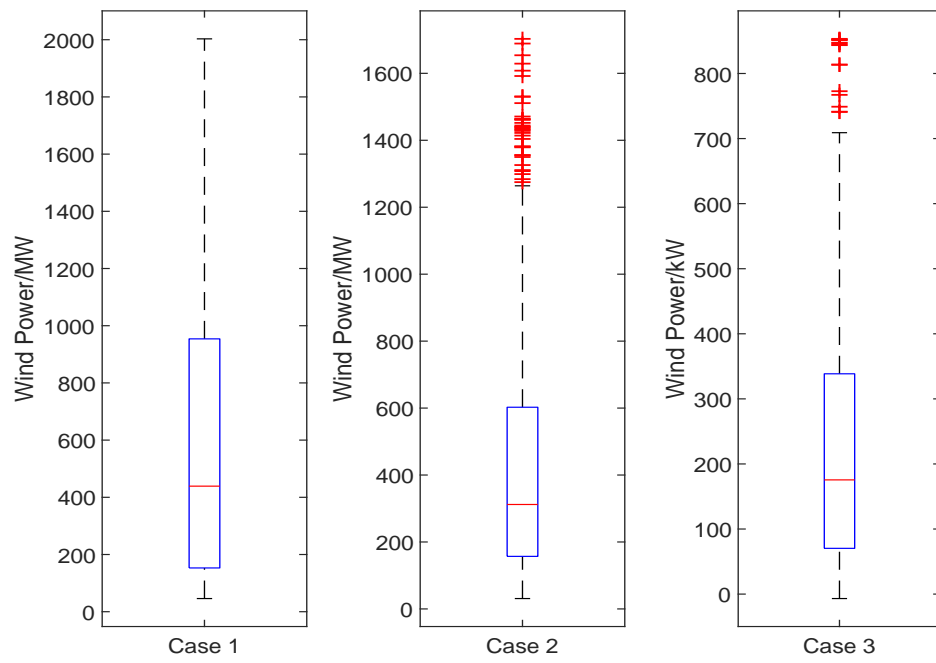


Figure 3: The boxplots of outlier detection for three cases

RVFL (the RVFL model combined with CSI) are compared in this paper. Their point prediction results are compared with the mode and median of SVQR and CSI-SVQR. Meantime, the present study modified the PAQR probability wind power forecasting method according to R code presented in [26]. PAQR is combined with CSI and KDE to construct PAQR and CSI-PAQR probability density density forecasting method in comparison with CSI-SVQR. Based on the optimal CWC and SSE_{mode} , two point prediction assessment metrics and two PIs assessment metrics, including MAPE, RMSE, PICP and PINAW are calculated to evaluate the presented method. The mode of probability density curve is used to verify the performance of probability density forecasting method. The parameters of the ϵ -SVR, ν -SVR, CSI- ϵ -SVR, and CSI- ν -SVR are determined based on LIBSVM 3.23 toolbox in Matlab [28]. The likelihood function of GPR and CSI-GPR are the Gaussian and isotropic rational quadratic covariance functions. The node number of the hidden layer of ELM and CSI-ELM is set to 30 based on the Sigmoidal function. For the optimal results, the number of hidden neurons of RVFL and CSI-RVFL is set to 1000. The "radbas" activation function and the scaling range of the randomization for uniform diatribution (Scalemode=3) are adopted in cases of Canada; the "sigmoid" activation function and the scaling features for all neurons (Scalemode=1) are adopted in case of China [48]. The optimal results of PAQR and CSI-PAQR depend on the degree of freedom of basis functions [26]. The optimal results of SVQR and CSI-SVQR depend on C, σ^2 . According to the training data, the optimal parameters of SVQR and CSI-SVQR using the grid-search approach are determined by exponentially growing sequences (e.g., $C = 2^{-1}, 2^0, 2^1, \dots, 2^{12}$ and $\sigma^2 = 2^{-1}, 2^0, 2^1, \dots, 2^{12}$, then a 14*14 grid is constructed). Four SVR methods, two SVQR methods, two GPR methods, two ELM methods, and two RVFL methods were implemented on Matlab 2016b. Two PAQR methods were implemented on R 3.43 software. It is worth noting that the obtained results of SVQR, CSI-SVQR, PAQR, CSI-PAQR, GPR, CSI-GPR, RVFL and CSI-RVFL are unaffected by the software version because random choice policy is not adopted in the process of training and testing. The prediction results of four SVR methods, and two ELM methods may be influenced by the running environment. For a fair comparison, ELM and CSI-ELM are run repeatedly for 30 times, and the optimal values are adopted as the final results.

4.2. Case study 1

Table 1: Forecasting errors of two traditional SVR, GRP, ELM and RVFL methods for case 1.

Methods	MAPE(%)	RMSE(%)
ϵ -SVR	7.39	3.31
ν -SVR	6.00	2.38
GPR	5.65	1.82
ELM	6.61	2.81
RVFL	7.03	3.35

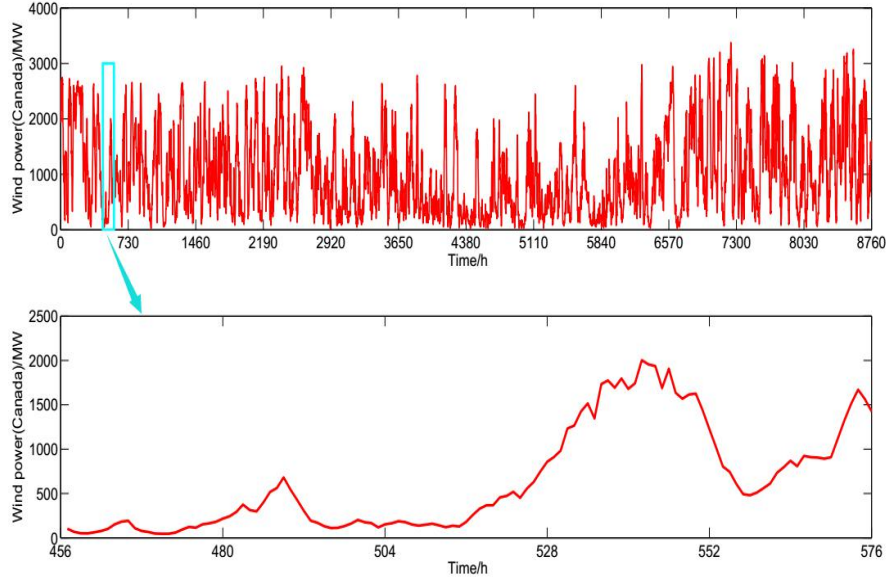


Figure 4: Hourly wind power data of Canada in 2015 (top) and the time series of case 1 (bottom)

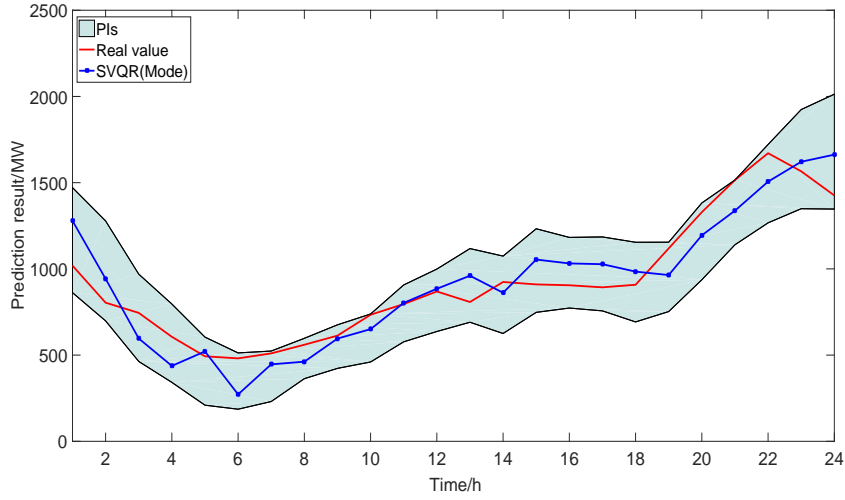


Figure 5: PIs and mode of the SVQR based on the optimal CWC for the first case

Table 2: The prediction results of SVQR and PAQR based on the optimal CWC for the first case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	100	23.27	5.94	5.35	1.93	1.54
PAQR	95.83	28.61	6.64	7.07	1.87	2.15

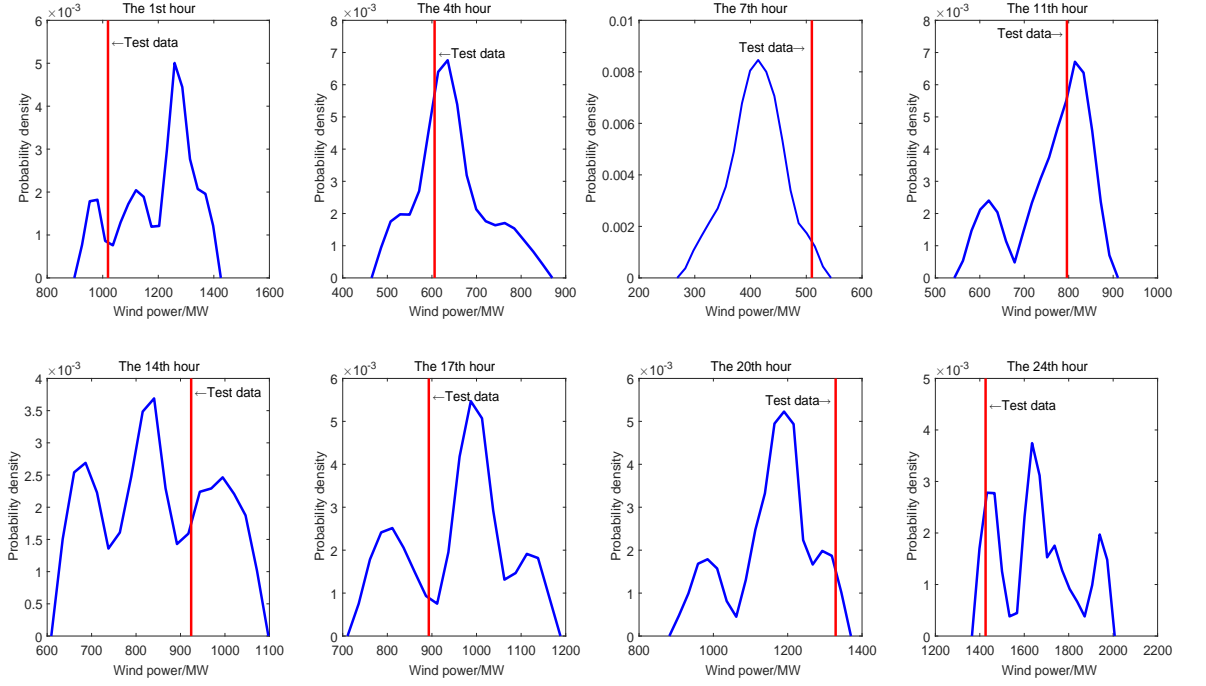


Figure 6: The diagram of probability density curves of the SVQR method based on the optimal SSE_{mode} for the first case

Table 3: The prediction results of SVQR and PAQR based on the optimal SSE_{mode} for the first case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	100	25.02	4.62	5.34	1.29	1.61
PAQR	95.83	29.08	5.30	5.64	1.18	1.22

In the first case, the hourly wind power curves of Ontario, Canada in 2015 are plotted in Fig. 4. The hourly data from January 20, 2015 to January 23, 2015 are selected to predict the wind power on January 24, 2015. After outlier identification, we found that no outliers are detected in this case. Therefore, the original data are used in predicting the wind power. The first 7-hour data are used as the input data. The selected optimal parameters of SVQR model based on the optimal CWC are $C = 2048, \sigma^2 = 8$. Based on the optimal SSE_{mode} , the parameters are $C = 256, \sigma^2 = 8$. To fully analyze the results of the algorithm proposed in this paper, PICP and PINAW are used in comparing algorithms. MAPE and RMSE are selected in comparing methods of point forecasting.

Table 1 shows the errors of ϵ -SVR, ν -SVR, GPR, ELM, and RVFL. The results of the PIs and point predictions of SVQR and PAQR are shown in Tables 2 and 3. By comparing the forecasting errors in Tables 1 - 3, it is easy to find that the error values of SVQR are smaller than PAQR and five traditional point forecasting methods, and the constructed PI obtained by SVQR is superior to that evaluated by PAQR. The accuracy of GPR is superior to that of ϵ -SVR, ν -SVR, ELM, and RVFL. The PICP value of SVQR is 100%, which means all actual values are covered by the obtained upper and lower bounds. Furthermore, when CWC is adopted as the fitness function, the PINAW is 23.27%, which is the optimal PI. But the MAPE is close to traditional ν -SVR. The mode point forecasting is even less than that of median. In comparison, when the minimal SSE_{mode} is selected as the objective function, the MAPE of SVQR evaluated by mode is only 4.62%, which is optimal point forecasting results. But, the PINAW achieves 25.02%. It is obvious that CWC and SSE_{mode} cannot be optimal at the same time in case 1.

The optimal PI and mode point forecasting results are shown in Figs 5 and 6. Based on the optimal CWC, Fig 5 indicates that all test samples (red line) are located in the narrow predicted intervals, and the mode curve (blue line) is near to the real targets curve. Fig. 6 provides the complete wind power probability density curves in the 1st, 4th, 7th, 11th, 14th, 17th, 20th and 24th hours on January 31, 2015, obtained by SVQR based on the optimal SSE_{mode} . In Fig. 6, all real values of testing samples appear in the probability density curve.

4.3. Case study 2

In this case, the wind power data of Ontario in Canada are selected for predicting wind power from June 22, 2014 to June 28, 2014 (one week in summer). Through the quartile method, 19 outliers are found, namely 2.83% of the total data (19/672). The hourly wind power curve of Ontario in 2014 and the outliers in June 2014 are plotted in Fig. 7. The first 7-hourly data are used as the input samples. Based on the optimal CWC, the selected parameters of SVQR and CSI-SVQR models are both $C = 4096, \sigma^2 = 1$. Based on the optimal SSE_{mode} , parameters C of SVQR and CSI-SVQR are set to 256 and 128, respectively. The parameters σ^2 of both models are 16.

The forecasting errors of traditional point forecasting methods are shown in Table 4, and obtained results

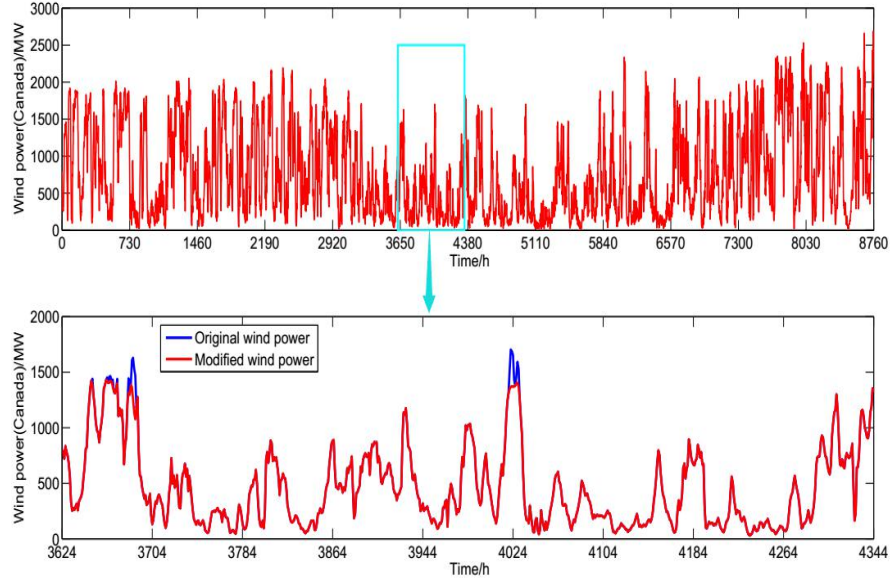


Figure 7: Hourly wind power data of Canada in 2014 (top), and the comparison between the original wind power series and modified series of case 2 (bottom)

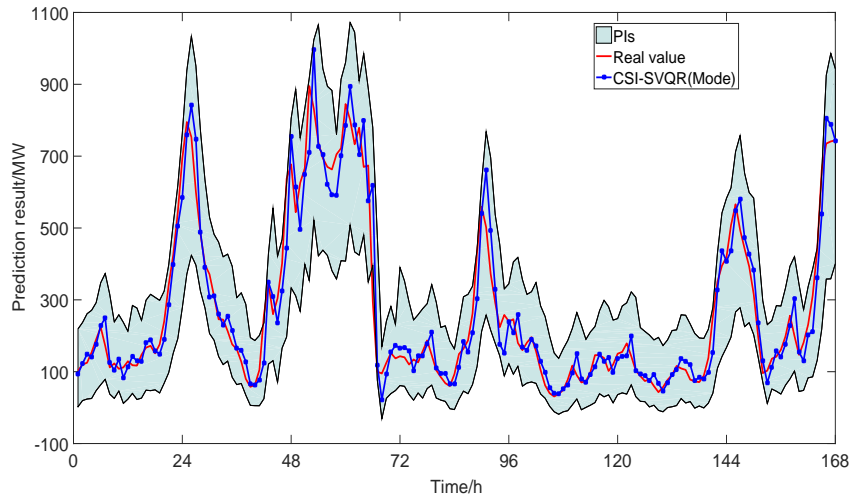


Figure 8: PIs and mode of the CSI-SVQR based on the optimal CWC for the second case

Table 4: Forecasting errors of **ten** point forecasting methods for the second case.

Methods	MAPE(%)	RMSE(%)
ϵ -SVR	4.12	5.63
CSI- ϵ -SVR	3.59	4.46
ν -SVR	2.59	3.24
CSI- ν -SVR	2.53	3.15
GPR	2.61	3.29
CSI-GPR	2.57	3.21
ELM	2.57	3.18
CSI-ELM	2.57	3.14
RVFL	3.89	2.54
CSI-RVFL	2.76	3.49

of four probability density forecasting methods based on the optimal CWC and the SSE_{mode} are shown in Tables 5 and 6. For the point prediction results in Table 4, the predicting errors of the CSI- ϵ -SVR, CSI- ν -SVR, and CSI-GPR wind power forecasting models are smaller than that of ϵ -SVR, ν -SVR, and GRP. The MAPE of CSI-ELM is the same as that of ELM, but its RMSE is smaller than that of ELM. **CSI-RVFL reduces the MAPE of RVFL, but fails to obtain better RSME.** In **ten** point forecasting methods, CSI- ν -SVR obtains the best MAPE and the third best RMSE (3.15%), which is less than that of **RVFL (2.54%)** and CSI-ELM (3.14%). From Tables 5 and 6, the forecasting errors (MAPE, RMSE) of CSI-SVQR are smaller than SVQR and PICPs all satisfy predefined nominal confidence level (namely, 95%). In terms of the forecasting results of the optimal CWC in Table 5, the PINAW of CSI-SVQR is narrower than that of other methods though its PICP is less than that of PAQR and CSI-PAQR. The mode of CSI-SVQR obtains the optimal point forecasting results. The point forecasting results of PAQR are better than that of CSI-PAQR. This suggests that CSI-PAQR need to sacrifice the point forecasting results in order to obtain better interval prediction results. Furthermore, in terms of the forecasting results of the optimal SSE_{mode} in Table 6, the values of mode calculated by the above four methods are better than those of median, and the mode obtained by CSI-SVQR is better than other methods mentioned above. The point forecasting results and PINAW of two SVQR methods are better than that of two PAQR methods. But, the PINAWs of CSI-SVQR and CSI-PAQR are slightly less than the results of SVQR and PAQR, indicating the CSI may affect interval prediction performance for acquiring better mode point forecasting results. Meantime, the PICPs of four methods are all satisfied the preassigned confidence level (95%). The PINAW values obtained by PAQR and CSI-PAQR are more than 30% though the results of PICP are 100%. On the whole, CSI-SVQR achieves the most satisfying result.

The prediction results of one week obtained by CSI-SVQR based on the optimal CWC are presented in Fig. 8. Fig. 8 shows that the real values lie within the predicted upper and lower bounds with a high percentage and narrow PINAW. The actual test samples are close to the mode. Fig. 9 shows the complete wind power probability density curves in the 1st, 25th, 49th, 73rd, 97th, 121st, 145th and 168th hours of the week from June 22, 2014 to June 28, 2014, which are evaluated by the CSI-SVQR based on the optimal SSE_{mode} and the kernel density estimation function. It is clearly seen from the figure that all real values of the wind power appear in the probability density curves; they are also near the peak of the probability density curves. Experimental results show that the proposed CSI-SVQR method can better handle the uncertainty of wind power.

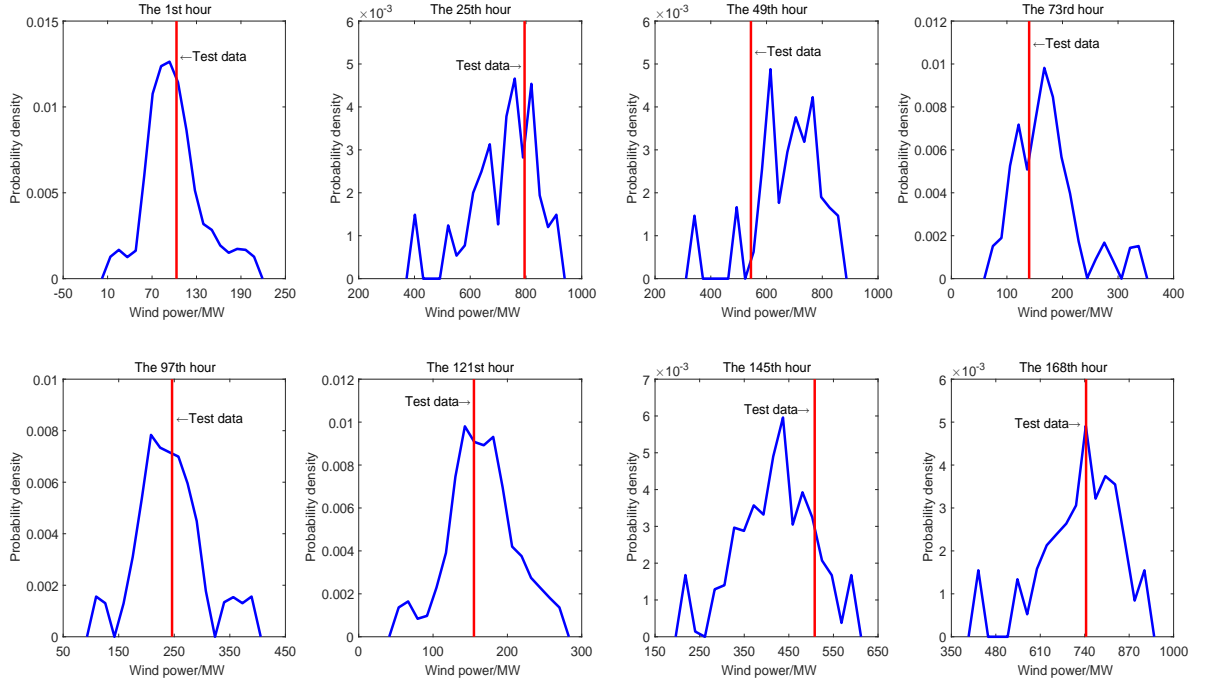


Figure 9: The diagram of probability density curves of the CSI-SVQR method based on the optimal SSE_{mode} for the second case

The data set in case 2 is collected from historical wind power in the summer. Apparently, the distribution characteristics of the wind power data in the winter are different from those in the summer. In order to consider the impact of meteorological (seasonal) factors on wind power, the quartile method is used to detect outliers in the winter wind power data for Ontario, Canada from 2013 to 2020. Fig. 10 presents the boxplots of outlier detection produced by the quartile method for ten months in the winter of 2018 to 2020. It shows that there is no outlier existing in the Canada winter data set. Thus, it is unnecessary to apply CSI on this

Table 5: Comparison of four probability density forecasting methods based on the optimal CWC for the second case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	95.24	23.16	3.04	3.13	4.33	5.05
CSI-SVQR	95.24	22.62	3.0	3.21	4.44	5.27
PAQR	97.61	27.24	5.02	5.70	3.28	4.03
CSI-PAQR	98.21	25.79	5.19	6.52	3.78	4.62

Table 6: Comparison of four probability density forecasting methods based on the optimal SSE_{mode} for the second case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	96.43	26.57	2.47	2.95	3.29	3.98
CSI-SVQR	97.62	27.9	2.41	2.96	2.93	4.0
PAQR	100	34.73	4.78	5.39	3.23	3.60
CSI-PAQR	100	35.62	4.71	5.76	3.18	4.03

data set.

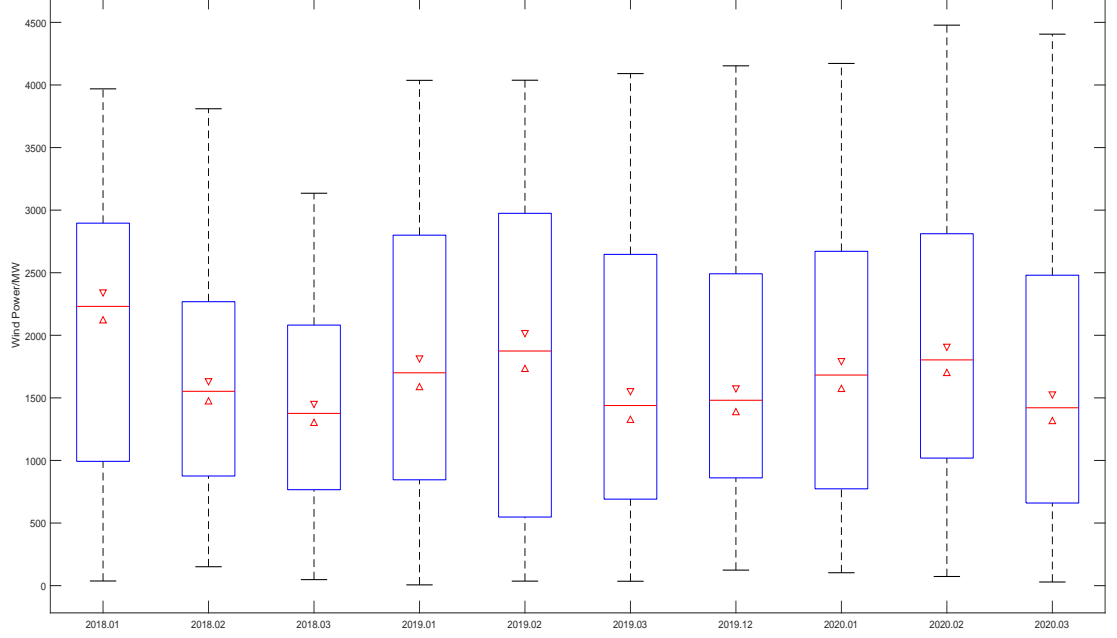


Figure 10: The boxplots of outlier detection for ten months in the winter of 2018 to 2020

4.4. Case study 3

The third case collects data at every 15 minutes from May 30, 2006 to June 3, 2006 in Mathematical modeling competition of electrical cup of China of 2011 [59]. In this case, the wind power of previous 4 quarterly-hours are used as the input to forecast 15-minutes interval wind power data of June 3, 2006. Because a large amount of outliers exist, the preprocessing method was applied first. The quartile method found 15 outliers in the wind power data, namely 3.125% of the total data (15/480), which were then transformed by the CSI function. The 15-minute wind power data and the modified data set are plotted in Fig. 11. Based on the optimal CWC, the optimal parameters of CSI-SVQR are $C = 2048, \sigma^2 = 8$, and the parameters of SVQR models are $C = 1024, \sigma^2 = 2$. Based on the optimal SSE_{mode} , parameters C of SVQR and CSI-SVQR are set to 2048 and 4096, respectively. The parameters σ^2 of both models are 16.

The prediction errors of the compared models are drawn in Table 7. According to Table 7, the number of the prediction errors of CSI- ϵ -SVR, CSI- ν -SVR and CSI-ELM is smaller than that of traditional ϵ -SVR, ν -SVR and ELM methods. CSI-ELM is the best method out of the ten methods of point forecasting. CSI-GPR is awarded the second best MAPE, though its RMSE is slightly less than GPR and CSI-ELM. CSI-RVFL also

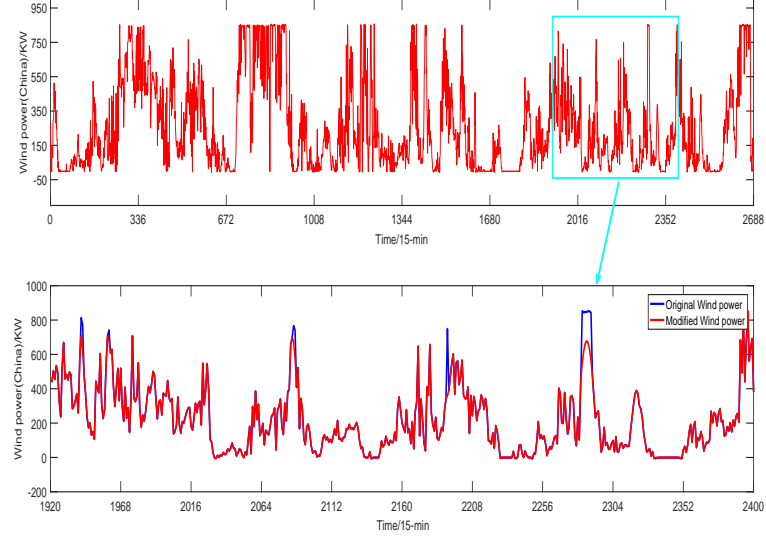


Figure 11: Wind power data from a wind turbine of China (top) and the comparison between the raw and revised wind power data set of case 3 (bottom)

Table 7: Forecasting errors of **ten** point forecasting methods for the third case.

Methods	MAPE(%)	RMSE(%)
ϵ -SVR	7.63	14.27
CSI- ϵ -SVR	7.41	13.83
ν -SVR	7.02	13.68
CSI- ν -SVR	6.99	13.87
GPR	7.00	13.54
CSI-GPR	6.77	13.62
ELM	6.86	15.51
CSI-ELM	6.65	13.49
RVFL	6.95	13.58
CSI-RVFL	6.82	13.81

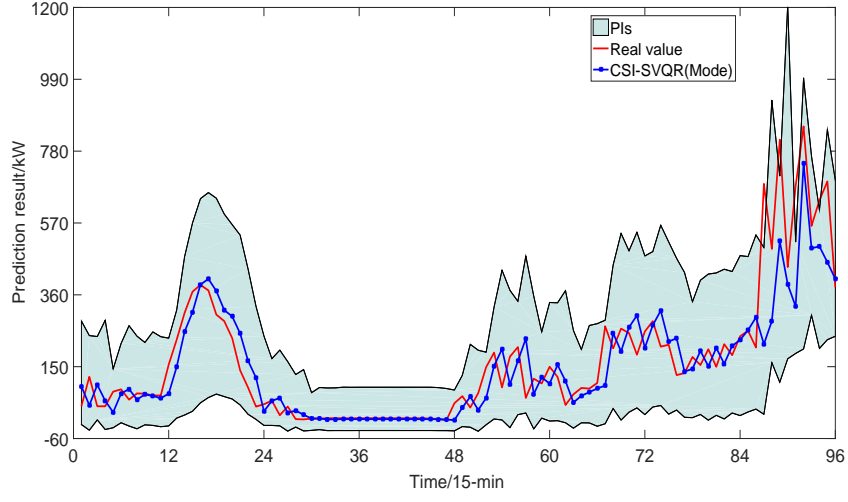


Figure 12: PIs and mode of the CSI-SVQR based on the optimal CWC for China data set

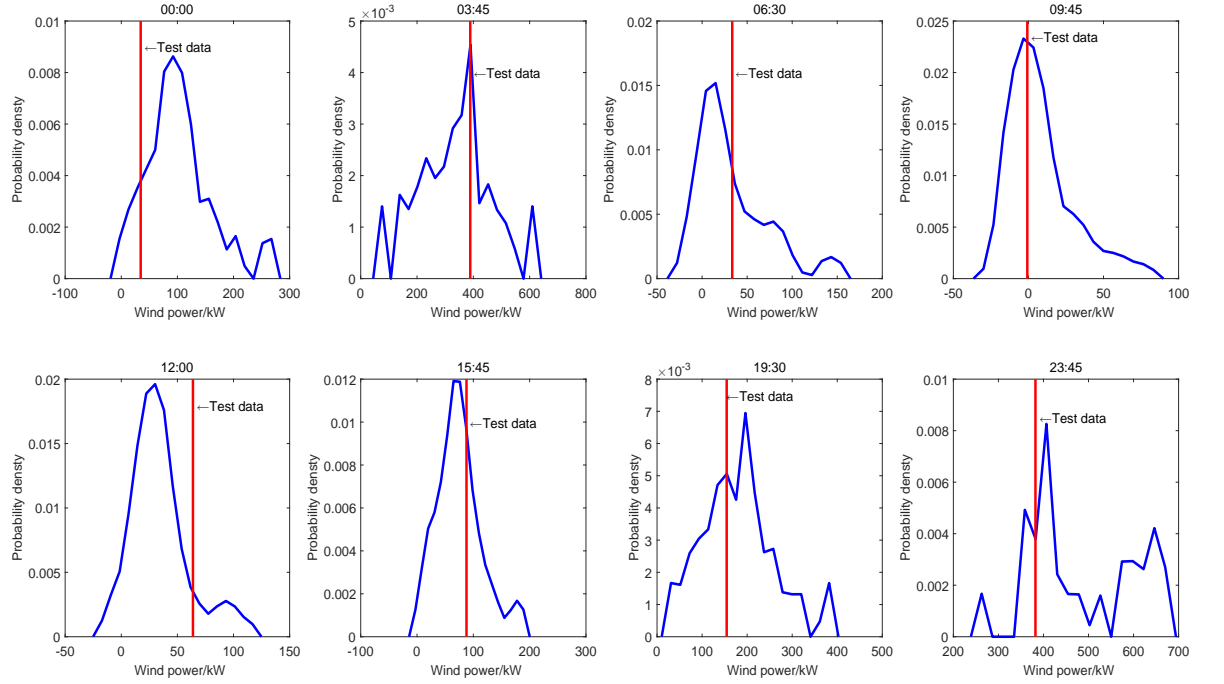


Figure 13: The diagram of probability density curves of the CSI-SVQR method based on the optimal SSE_{mode} for China data set

fails to acquire better RMSE compared to RVFL though its MAPE is smaller than that of RVFL. Tables 8 and 9 show the number of prediction errors, PICP and PINAW of the four probability density forecasting methods. Similar to cases 1 and 2, the PICPs of all methods are also satisfied the preassigned confidence level (95%). By comparing with the values of Table 8 based on the optimal CWC, CSI-SVQR obtains the optimal PINAW and MAPE. The PINAW of CSI-PAQR is less than that of PAQR though the point forecasting results of CSI-PAQR are better than PAQR. The RMSE of SVQR and CSI-SVQR evaluated by Mode is less than that of Median. This should be a normal phenomenon for it is hard to control the point forecasting results when the optimal CWC is adopted as the objective function. In Table 9, CSI-SVQR also acquires the optimal PINAW and MAPE, when the optimal SSE_{mode} is adopted as the objective function. The point forecasting results of CSI-PAQR are superior to that of PAQR, but the quality of PI obtained by CSI-PAQR is less than that of PAQR. Though PAQR achieves the highest PICP, its PINAW is close to 40%. It is obvious to discover that PAQR and CSI-PAQR have difficulties in achieving comprehensive perfect results because it needs to drastically sacrifice PINAW in order to gain higher PICP and lower forecasting error. CSI-SVQR can obtain more reasonable PINAW and point forecasting results based on a satisfied confidence level (over 95%). Hence, the performance of CSI-SVQR is optimal.

Table 8: Comparison of four probability density forecasting methods based on the optimal CWC for the third case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	95.83	29.25	7.9	8.28	19.87	18.46
CSI-SVQR	95.83	26.53	7.32	7.99	20.24	17.21
PAQR	100	30.04	7.81	10.34	18.53	23.18
CSI-PAQR	98.96	30.99	7.53	10.25	17.55	20.83

Table 9: Comparison of four probability density forecasting methods based on the optimal SSE_{mode} for the third case.

Methods	PICP(%)	PINAW(%)	MAPE(%)		RMSE(%)	
			Mode	Median	Mode	Median
SVQR	95.83	30.54	6.69	7.53	13.21	15.49
CSI-SVQR	95.83	27.24	6.44	7.72	13.37	16.11
PAQR	98.96	39.4	6.68	11.37	15.32	22.65
CSI-PAQR	96.88	40.56	6.58	8.29	14.65	17.79

As shown in Fig. 12 based on the optimal CWC, the actual values fall between the upper and lower bounds with a high probability, and the proposed CSI-SVQR model achieves a higher point prediction accuracy than the existing wind power forecasting methods. Based on the optimal SSE_{mode} , Fig. 13 shows the curves of wind power probability density distribution at the selected 8 quarterly-hours (00:00, 03:45, 06:30, 09:45, 12:00, 15:45, 19:30 and 23:45) on 3 June 2006. It is clear that all test samples follow the probability density curves, and are near the peak of curves.

4.5. Wilcoxon signed rank test for performance evaluation of CSI

To validate the effectiveness of CSI, two-sample Wilcoxon signed rank test (also called Mann-Whitney test) [60] is used to contrast the performance of different forecasting methods with and without the use of CSI. One-sided hypothesis test with the significance level of 5% is employed. The obtained MAPE values are divided into two pairs of samples. One set of samples records the MAPE of ϵ -SVR, ν -SVR, GPR, ELM, RVFL, SVQR (mode), and PAQR (mode); another one records the MAPE of these methods with the use of CSI. The null hypothesis is that the average MAPE of compared methods without the use of CSI is equal to the paired methods using CSI. The alternative is that the average MAPE of compared methods without the use of CSI is greater than the paired methods using CSI, namely CSI can improve the forecasting accuracy of different methods. Cases 2 and 3 are conducted to evaluate CSI. The calculated p-values are 0.01776 and 0.007813, respectively. Because the p-values are less than the significance level, the null hypothesis is rejected. It means that CSI can significantly improve the prediction accuracy of the methods adopted.

4.6. Test based on out-of-sample data

In order to test the performance of selected hyper-parameters in SVQR and CSI-SVQR, we extend the experiment to out-of-sample data. For case 2, SVQR and CSI-SVQR are used to forecast wind power of next week (from June 29, 2014 to July 5, 2015), when the optimal C and σ^2 are certain. It's worth noting that the maximal power (1821MW) in this week is more than the one (1703MW) of original data. Hence, it can be used to detect the impact of the extreme value on the proposed method. From the results based on the optimal CWC in Table 10, the PINAW of CSI-SVQR is narrower than that of SVQR (namely, the basic target can be met.), but their PICPs are all lower than predefined confidence level, and the mode point forecasting results are not better than that of median. From Table 11, when the optimal SSE_{mode} is adopted, all metrics of CSI-SVQR except RMSE are less than that of SVQR though the target where the mode is greater than the median can be satisfied. To comprehensively measure the generalization performance of the proposed probability density prediction method, the MAPE and RMSE evaluated by probability mean of the probability density function (referred to as "Mean") introduced in [12] are shown in Tables 10 and 11. Then, the optimal MAPE can be obtained from CSI-SVQR.

For case 3, SVQR and CSI-SVQR are used to forecast wind power of 96 points of next day (June 4, 2006). Tables 10 and 11 show that the interval prediction and mode point forecasting results of CSI-SVQR are significantly better than that of SVQR, although the PICP is hard to satisfy the confidence level. The point forecasting performance of probability mean is better than mode and median. As for as the metric of the optimal CWC, the MAPE and RMSE of CSI-SVQR based on probability mean are better than that of SVQR. But, if the optimal SSE_{mode} is adopted, two methods have same MAPE in probability mean, and the RMSE of CSI-SVQR based on probability mean is slightly less than that of SVQR.

On the whole, for the out-of-sample data, the PICP of SVQR and CSI-SVQR can not keep the predefined confidence level because of the impact of the cumulative error. The main metrics of CSI-SVQR are better than that of SVQR.

Table 10: Forecasting results of SVQR and CSI-SVQR based on the optimal CWC for out-of-sample data

Data sets	Methods	PICP	PINAW	MAPE (%)			RMSE (%)		
				Mode	Median	Mean	Mode	Median	Mean
Case 2	SVQR	86.90	16.33	6.60	5.38	5.55	3.94	2.73	3.05
	CSI-SVQR	86.31	14.71	5.71	5.60	5.29	3.34	3.72	3.22
Case 3	SVQR	88.54	35.63	11.44	12.62	10.84	31.59	33.50	27.37
	CSI-SVQR	91.67	34.47	10.76	11.51	10.13	27.04	27.73	24.30

Table 11: Forecasting results of SVQR and CSI-SVQR based on the optimal SSE_{mode} for out-of-sample data

Data sets	Methods	PICP	PINAW	MAPE (%)			RMSE (%)		
				Mode	Median	Mean	Mode	Median	Mean
Case 2	SVQR	93.45	22.23	4.32	4.67	4.15	1.86	2.14	1.75
	CSI-SVQR	92.86	23.14	4.41	5.36	4.11	1.83	2.84	1.70
Case 3	SVQR	91.67	36.93	10.92	10.69	10.10	27.75	25.07	23.65
	CSI-SVQR	93.75	36.04	10.60	11.23	10.10	27.62	26.42	24.16

4.7. Discussion on calculation time

The running time of the algorithm is also a problem worthy of attention. Probability density forecasting methods construct probability density functions by means of the forecasting results under different quantiles, which need to spend more training time. Generally speaking, it is meaningless to compare the running time of point prediction method and probability density forecasting method. In this paper, all the algorithms were repeated 10 times on a 3.20-GHz-based Intel six-core processor (i7-8700) with 32 GB of random access

memory. The running time of CSI is only 0.06s. The speed of PAQR and CSI-PAQR are not obviously affected by the data size, because they use spline function to directly fit the prediction model. In all three cases, the average computing time of two PAQR method is about 1.5 seconds (ranges from 0.65 to 2.46 seconds), when the degree of freedom (it must be a positive integer) of the spline function is given. The presented SVQR and CSI-SVQR search parameters C and σ^2 on a 14*14 grid, consuming a lot of computation time. The parameters of SVQR is predefined by trial and error in the previous work [37], which is compared with PAQR given the degree of freedom. The average calculation time of two SVQR methods for three cases is shown in Table 12. It is obvious that the computing time of SVQR-typed methods increases significantly with the increase of the amount of data in the training sample. The speed of CSI-SVQR is slightly faster than that of SVQR, because CSI improves the quality of data with lower computing cost. In case 1, the speed of SVQR predefined parameters is close to that of PAQR. But in cases 2 and 3, it significantly slower than PAQR.

Table 12: Calculation time of SVQR and CSI-SVQR for three cases

Data sets	Methods	Calculation time(s)	
		Grid search	Predefined
Case 1	SVQR	80.33	0.67
Case 2	SVQR	1119.34	14.51
	CSI-SVQR	1107.34	13.46
Case 3	SVQR	566.44	4.28
	CSI-SVQR	563.83	3.63

5. Conclusion and future work

In this paper, a wind power probability density forecasting method based on CSI-SVQR is presented. It can handle uncertainty in training data better than other methods from the perspective of probabilistic interval prediction and point prediction, respectively. To detect and transform the outliers in the data, this paper proposes to use the quartile method and CSI function to preprocess the data first, and SVQR is then employed to build predictive models, in which the Epanechnikov kernel function is used to produce probability density functions.

Simulation **experiments show** that the proposed method has the following advantages. In case study 1 without outliers, SVQR is superior than PAQR and **five** point forecasting methods. In cases 2 and 3 with outliers, CSI can improve the accuracy of all compared algorithms. Classical point prediction methods can also obtain very low of RMSE and MAPE values for the comparative experiments are all based on the optimal parameters. When the optimal SSE_{mode} is selected as the objective function, all probabilistic point

forecasting results are superior in terms of Mode values. However, the results in terms of Median values may be better than that of Mode, if the optimal CWC is considered as the objective function. That is probably because the probability density forecasting method mainly focuses on quantitatively analysing the uncertainty of PIs, which is difficult to justify the best point forecasting metrics unless a coercive strategy is adopted. Compared with the traditional methods, CSI-SVQR can produce more accurate point prediction results and PIs. In addition, CSI-SVQR presents higher PICP and narrower PINAW than SVQR. Meanwhile, the overall performance of CSI-SVQR is superior to that of PAQR and CSI-PAQR whether using the optimal CWC criterion or the optimal SSE_{mode} criterion. When SVQR and CSI-SVQR are used to forecast out-of-sample wind power data, the performance of CSI-SVQR is slightly better than SVQR, though the PICP of them can not satisfy the predefined confidence level. In addition, CSI-SVQR reduces the computation time of SVQR, which is also a obvious advantage of CSI-SVQR.

In the future studies, we will look into the following problems that have not been solved: 1) There are plentiful advanced QR methods. Hence, it is worth exploring probability density forecasting based on other QR methods to find better results. 2) Some other factors have an impact on wind power forecasting, such as temperature, wind speed, wind direction and humidity, which should be considered. 3) There is a trade-off between PICP and PINAW. A better objective function will be proposed and studied to balance the trade-off. 4) Last but not the least, intelligent optimization algorithms will be considered to find the optimal parameters in CSI-SVQR.

ACKNOWLEDGEMENT

The authors would like to thank the National Natural Science Foundation (Nos 71771073 and 61329302.), the Fundamental Research Funds for the Central Universities (PA2020GDKC0006), the Science and Technology Innovation Committee Foundation of Shenzhen (Grant No. ZDSYS201703031748284) and EPSRC (Grant No. EP/K001523/1). Xin Yao was also supported by a Royal Society Wolfson Research Merit Award.

References

- [1] Yan J, Liu Y, Han S, Wang Y, Feng S. Reviews on uncertainty analysis of wind power forecasting. *Renewable and Sustainable Energy Reviews* 2015;52:1322–30.
- [2] Abdoos AA. A new intelligent method based on combination of vmd and elm for short term wind power forecasting. *Neurocomputing* 2016;203:111–20.
- [3] Li X, Cao J, Du D. Probabilistic optimal power flow for power systems considering wind uncertainty and load correlation. *Neurocomputing* 2015;148:240–7.
- [4] Sorknæs P, Djørup SR, Lund H, Thellufsen JZ. Quantifying the influence of wind power and photovoltaic on future electricity market prices. *Energy conversion and management* 2019;180:312–24.

- [5] Zameer A, Arshad J, Khan A, Raja MAZ. Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. *Energy conversion and management* 2017;134:361–72.
- [6] Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama R. A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. *Neurocomputing* 2019;.
- [7] Zhang Z, Qin H, Liu Y, Wang Y, Yao L, Li Q, et al. Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction. *Energy Conversion and Management* 2019;192:37–51.
- [8] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews* 2014;31:762–77.
- [9] Chen P, Pedersen T, Bak-Jensen B, Chen Z. Arima-based time series model of stochastic wind power generation. *IEEE Transactions on Power Systems* 2010;25(2):667–76.
- [10] Papaefthymiou G, Klockl B. Mcmc for wind power simulation. *IEEE Transactions on Energy Conversion* 2008;23(1):234–40.
- [11] De Felice M, Yao X. Short-term load forecasting with neural network ensembles: A comparative study [application notes]. *IEEE Computational Intelligence Magazine* 2011;6(3):47–56.
- [12] He Y, Li H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. *Energy conversion and management* 2018;164:374–84.
- [13] Amjady N, Keynia F, Zareipour H. Short-term wind power forecasting using ridgelet neural network. *Electric power systems Research* 2011;81(12):2099–107.
- [14] Kukolj D, Levi E. Identification of complex systems based on neural and takagi-sugeno fuzzy model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 2004;34(1):272–82.
- [15] Khalid M, Savkin AV. A method for short-term wind power prediction with multiple observation points. *IEEE Transactions on Power Systems* 2012;27(2):579–86.
- [16] Brown BG, Katz RW, Murphy AH. Time Series Models to Simulate and Forecast Wind Speed and Wind Power . *Journal of Applied Meteorology* 1984;23:1184–95. doi:10.1175/1520-0450(1984)023<1184:TSMTSA>2.0.CO;2.
- [17] Shen Z, Ritter M. Forecasting volatility of wind power production. *Applied Energy* 2016;176:295–308.
- [18] Jeon J, Taylor JW. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association* 2012;107(497):66–79.
- [19] Zhang W, Liu F, Zheng X, Li Y. A hybrid EMD-SVM based short-term wind power forecasting model. In: *Power and Energy Engineering Conference (APPEEC), 2015 IEEE PES Asia-Pacific*. IEEE; 2015, p. 1–5.
- [20] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews* 2014;32:255–70.
- [21] Wan C, Xu Z, Pinson P, Dong ZY, Wong KP. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems* 2014;29(3):1033–44.
- [22] Quan H, Srinivasan D, Khosravi A. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Transactions on Neural Networks and Learning Systems* 2014;25(2):303–15.
- [23] Koenker R, Bassett Jr G. Regression quantiles. *Econometrica: journal of the Econometric Society* 1978;33–50.
- [24] Bremnes JB. Probabilistic wind power forecasts using local quantile regression. *Wind Energy* 2004;7(1):47–54.
- [25] Bessa RJ, Miranda V, Botterud A, Zhou Z, Wang J. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy* 2012;40(1):29–39.
- [26] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy* 2006;9(1-2):95–108.
- [27] Vapnik V. *The nature of statistical learning theory*. Springer Science & Business Media; 2013.
- [28] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and*

- technology (TIST) 2011;2(3):27.
- [29] Gu B, Sheng VS. A robust regularization path algorithm for ν -support vector classification. *IEEE Transactions on Neural Networks and Learning Systems* 2016;28(5):1241–8.
 - [30] Díaz S, Carta JA, Matías JM. Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. a case study. *Energy Conversion and Management* 2017;140:334–54.
 - [31] Hong X, Mitchell R, Di Fatta G. Simplex basis function based sparse least squares support vector regression. *Neurocomputing* 2019;330:394–402.
 - [32] Che J, Wang J, Wang G. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. *Energy* 2012;37(1):657–64.
 - [33] Kong X, Liu X, Shi R, Lee KY. Wind speed prediction using reduced support vector machines with feature selection. *Neurocomputing* 2015;169(1):449–56.
 - [34] Takeuchi I, Furuhashi T. Non-crossing quantile regressions by SVM. In: *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*; vol. 1. IEEE; 2004,.
 - [35] Xu Q, Zhang J, Jiang C, Huang X, He Y. Weighted quantile regression via support vector machine. *Expert Systems with Applications* 2015;42(13):5441–51.
 - [36] Xu Q, Jiang C, He Y. An exponentially weighted quantile regression via SVM with application to estimating multiperiod var. *Statistical Methods & Applications* 2016;25(2):285–320.
 - [37] 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. *Applied Energy* 2017;185:254–66.
 - [38] Rosenblatt M, et al. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 1956;27(3):832–7.
 - [39] Pollard D. *Convergence of stochastic processes*. Springer Science & Business Media; 2012.
 - [40] Epanechnikov VA. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* 1969;14(1):153–8.
 - [41] Walfish S. A review of statistical outlier methods. *Pharmaceutical technology* 2006;30(11):82.
 - [42] Kusiak A, Zheng H, Song Z. Models for monitoring wind farm power. *Renewable Energy* 2009;34(3):583–90.
 - [43] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis - combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square svm) forecasts using a gpr (gaussian process regression). *Energy* 2015;93(2):41–56.
 - [44] Mohammadi K, Shamshirband S, Yee PL, Petkovic D, Zamani M, Ch S. Predicting the wind power density based upon extreme learning machine. *Energy* 2015;86(1):232–9.
 - [45] Wu S, Wang Y, Cheng S. Extreme learning machine based wind speed estimation and sensorless control for wind turbine power generation system. *Neurocomputing* 2013;102(1):163–75.
 - [46] Ren Y, Suganthan PN, Srikanth N, Amaratunga GAJ. Random vector functional link network for short-term electricity load demand forecasting. *Information Sciences* 2016;367:1078–93.
 - [47] Katuwal R, Suganthan PN, Zhang L. An ensemble of decision trees with random vector functional link networks for multi-class classification. *Applied Soft Computing* 2018;70:1146–53.
 - [48] Qiu X, Suganthan PN, Amaratunga GAJ. Ensemble incremental learning random vector functional link network for short-term electric load forecasting. *Knowledge Based Systems* 2018;145:182–96.
 - [49] McKinley S, Levine M. Cubic spline interpolation. *College of the Redwoods* 1998;45(1):1049–60.
 - [50] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. *Neural processing letters* 1999;9(3):293–300.
 - [51] Shim J, Kim Y, Lee J, Hwang C. Estimating value at risk with semiparametric support vector quantile regression. *Computational Statistics* 2012;27(4):685–700.

- [52] Bertsekas DP. Nonlinear programming. Athena scientific Belmont; 1999.
- [53] Zhou S, Mao M, Su J. Short-term forecasting of wind power and non-parametric confidence interval estimation. In: Zhongguo Dianji Gongcheng Xuebao(Proceedings of the Chinese Society of Electrical Engineering); vol. 31. Chinese Society for Electrical Engineering; 2011, p. 10–6.
- [54] Khosravi A, Nahavandi S, Creighton D, Atiya AF. Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE Transactions on Neural Networks 2011;22(3):337–46.
- [55] Quan H, Srinivasan D, Khosravi A. Construction of neural network-based prediction intervals using particle swarm optimization. In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE; 2012, p. 1–7.
- [56] Che J, Wang J. Short-term load forecasting using a kernel-based support vector regression combination model. Applied Energy 2014;132:602–9.
- [57] Wand MP, Jones MC. Kernel smoothing. Crc Press; 1994.
- [58] Ieso power data. <http://www.ieso.ca/Pages/Power-Data>; 2014-2015.
- [59] Mathematical modeling competition of "electrical cup" in 2011. <http://www.shumo.com/home/html/1323.html>; 2011.
- [60] Neurath P, Hollander M, Wolfe DA. Nonparametric statistical methods. Contemporary Sociology 1976;5(4):469.



Yaoyao He received the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 2010. He was an academic visitor at the Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA) in the University of Birmingham, UK, from Aug 2016 to Aug 2017. He is currently a Professor in School of Management, Hefei University of Technology, Hefei, China. His current research interests include time series analysis advanced algorithms for the training and construction of

neural networks, quantile regression for probability density forecasting, fuzzy systems and SVMs, as well as advanced swarm intelligence (PSO, DE) algorithms, with applications to power systems and water resources systems.



Haiyan Li received the M.S. degree from school of Management, Hefei University of Technology, Hefei, China, in 2018, supervised by Prof. Yaoyao He. Her research area is time series analysis of power systems.



Shuo Wang is a lecturer at School of Computer Science, the University of Birmingham, UK. She was a research fellow at the Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA) in the University of Birmingham between 2011 and 2018. She received the Ph.D. degree in Computer Science from the University of Birmingham in

2011, sponsored by the Overseas Research Students Award (ORSAS) from the British Government. Dr. Wang's research interests include data stream classification, class imbalance learning and ensemble learning approaches in machine learning, and their applications in social media analysis, software engineering and fault detection. Her work has been published in internationally renowned journals and conferences, such as IEEE Transactions on Knowledge and Data Engineering and International Joint Conference on Artificial Intelligence (IJCAI). One of her fundamental works on multi-class learning, published in IEEE Transactions on Cybernetics, has been cited over 330 times since its publication in 2012. The multi-disciplinary research on class imbalance learning and software engineering, published in IEEE Transactions on Reliability, has been cited by other world-leading software engineering groups. In addition, she has been a guest editor of Neurocomputing and Connection Science and the workshop organizer of IJCAI'17 and ICDM'19. URL: <https://phd-shuowang.weebly.com>



Xin Yao (M91-SM96-F03) received the B.Sc. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 1982 and 1990, respectively. He is a Chair Professor of Computer Science with the Southern University of Science and Technology, Shenzhen, China, and a part-time Professor of Computer Science with the University of Birmingham, U.K. He has been researching multi-objective optimization since 2003, when he published a well-cited EMO03 paper on many-objective optimization.

His current research interests include evolutionary computation, ensemble learning, and their applications in software engineering. Dr. Yao was a recipient of the 2001 IEEE Donald G. Fink Prize Paper Award, the 2010, 2016, and 2017 IEEE Transactions on Evolutionary Computation Outstanding Paper Awards, the 2010 BT Gordon Radley Award for Best Author of Innovation (Finalist), the 2011 IEEE Transactions on Neural Networks Outstanding Paper Award, the Prestigious Royal Society Wolfson Research Merit Award in 2012, the IEEE Computational Intelligence Society (CIS) Evolutionary Computation Pioneer Award in 2013, and many other best paper awards. He was the President of IEEE CIS from 2014 to 2015, and from 2003 to 2008 he was Editor-in-Chief of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION. He is a Distinguished Lecturer of IEEE CIS.