

Multi-dimensional scenario forecast for generation of multiple wind farms

Ming YANG (✉), You LIN, Simeng ZHU,
Xueshan HAN, Hongtao WANG



Abstract A novel multi-dimensional scenario forecast approach which can capture the dynamic temporal-spatial interdependence relation among the outputs of multiple wind farms is proposed. In the proposed approach, support vector machine (SVM) is applied for the spot forecast of wind power generation. The probability density function (PDF) of the SVM forecast error is predicted by sparse Bayesian learning (SBL), and the spot forecast result is corrected according to the error expectation obtained. The copula function is estimated using a Gaussian copula-based dynamic conditional correlation matrix regression (DCCMR) model to describe the correlation among the errors. And the multi-dimensional scenario is generated with respect to the estimated marginal distributions and the copula function. Test results on three adjacent wind farms illustrate the effectiveness of the proposed approach.

Keywords Wind power generation forecast, Multi-dimensional scenario forecast, Support vector machine (SVM), Sparse Bayesian learning (SBL), Gaussian copula, Dynamic conditional correlation matrix

1 Introduction

The wind power has been developing very fast in China since 2006. Within the mainland, the total installed capacity of wind power had reached up to 91424 MW by the end of 2013 [1]. Since many wind farms are centrally constructed in the wind-rich zones, some regional power grids in China have already had a relatively high wind power penetration [2].

To improve the operating security and economics of the power grid which is integrated with large-scale wind power, a project named as collaborative optimization of the thermal power, hydro power and wind power in extremely cold areas was carried out by Shandong University and Heilongjiang Electric Power Company. Developing a short-term wind power forecast program is the main and fundamental research objective of this project.

Heilongjiang power grid has 45 integrated wind farms, and the installed capacity reaches up to 3153 MW which accounts for 14.8% of the total installed capacity in that region. Because of the rapid growth of wind power and relatively slow expansion of the transmission networks, transmission congestion happens from time to time in the grid. To consider the transmission constraints during the scheduling process, forecast is required to be performed for each single wind farm as well as the whole region. Moreover, the cross-correlation among the outputs of multiple wind farms is expected to be estimated to make full use of the adjustable capacity of the power grid.

Although great efforts have been made to improve the forecast accuracy, it is still hard to predict the wind power generation precisely. As a result, estimating the uncertainty of the forecast result is believed to be crucial for the operation of power systems [3, 4]. By now, several parametric or non-parametric approaches, e.g., the quantile regression approaches [5], the interval estimation approaches [6, 7],

CrossCheck date: 3 December 2014

Received: 4 July 2014 / Accepted: 4 March 2015 / Published online: 7 May 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

M. YANG, Y. LIN, X. HAN, H. WANG, Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Laboratory of Electric Vehicles Engineering of Shandong Province, Shandong University, Jinan 250061, China
(✉) e-mail: myang@sdu.edu.cn

S. ZHU, Zaozhuang Power Supply Company, Zaozhuang 277800, China



and the probability density forecast approaches [8, 9], have been proposed to achieve this aim. These approaches can provide end-users with forecast uncertainty information in various ways.

Temporal-spatial dependence relation among the outputs of wind farms is the valuable information for the power system operation [10, 11]. In [12], a short-term joint probability density function (JPDF) forecast approach was proposed to include the temporal correlation of forecast errors into the distribution forecast results. The errors were assumed to follow a joint Gaussian distribution and the correlation matrix was estimated by the recursive statistic estimation. Reference [13] introduced an approach to consider the temporal interdependence structure in the quantile regression based probabilistic forecast approach. In the approach, the interdependence structure was summarized by a unique covariance matrix through the conversion of the prediction errors to a multivariate Gaussian random vector. The approaches mentioned in [12, 13] are instructive. However, the spatial dependence structure is ignored in the approaches.

In this paper, a novel multi-dimensional scenario forecast approach which can capture the dynamic temporal-spatial interdependence relation among the outputs of multiple wind farms is proposed. The advantages of the proposed approach are as follows.

- 1) The temporal-spatial dependence relation of the forecast errors is included into the probabilistic forecast result, and the approach can provide more useful information for the system operation.
- 2) By using the kernel based sparse learning approaches and the error correction strategy, the accuracy of the spot and probabilistic forecast results is guaranteed.
- 3) The dependence structure of the forecast errors is well represented by the Gaussian copula, and it is not necessary to make any assumption on the distributions of the errors.
- 4) The multi-dimensional scenario is generated with respect to the error distributions and the copula function.

2 Overviews of the proposed approach

The work is carried out on three wind farms located in the same region. The relative location of the wind farms is shown in Fig. 1.

Wind speed and direction of the forecast target period are provided by the commercial numerical weather prediction (NWP) service, and the wind power generation data are collected from the supervisory control and data acquisition (SCADA) system. The time span of the data is from

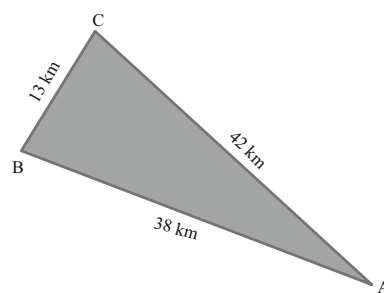


Fig. 1 Relative location of wind farm A, B and C

July 1st 2009 to December 31th 2010. In the forecast models, the wind speed and wind power generation are normalized according to the maximum wind speed and the installed capacity respectively. And the wind direction is presented by its sinusoidal and cosinusoidal values.

The samples are divided into a training set, a test set and a validation set. The training data set is used to train the support vector machine (SVM) models for the spot forecast. The test data set is used to produce SVM forecast error samples which are applied to train the sparse Bayesian learning (SBL) models and estimate the parameters of the copula-based dynamic conditional correlation matrix regression (DCCMR) model. The validation data set is used to evaluate the performance of the proposed approach.

In the proposed approach, the wind power generation of each wind farm is treated as a random variable. The forecast result is the possible trajectories of the outputs of the wind farms which are referred to as the multi-dimensional scenarios. Assuming L wind farms and T look-ahead hours are considered, the dimension of one multi-dimensional scenario is $K = LT$. The framework of the proposed approach is shown in Fig. 2.

The approach includes two main parts, i.e., the training part and the forecast part. In the training part, K SVMs are trained using the data of the training set. Then, a virtual spot forecast is performed on the test data set and the forecast error samples are collected. After that, K SBL models are trained using the SVM forecast results and the corresponding NWP data. In parallel, the parameters of the copula-based DCCMR model are estimated. The outputs of the training part are K SVM models for the spot forecast, K SBL models for the error distribution forecast, and one copula-based DCCMR model describing the temporal-spatial interdependence structure of the errors.

The forecast part of the approach has three main modules, i.e., the spot forecast module using the SVMs, the probability density function (PDF) forecast module using the SBLs, and the scenario sampling module. The outputs of this part include the expected wind power generation trajectory of each wind farm, the joint cumulative

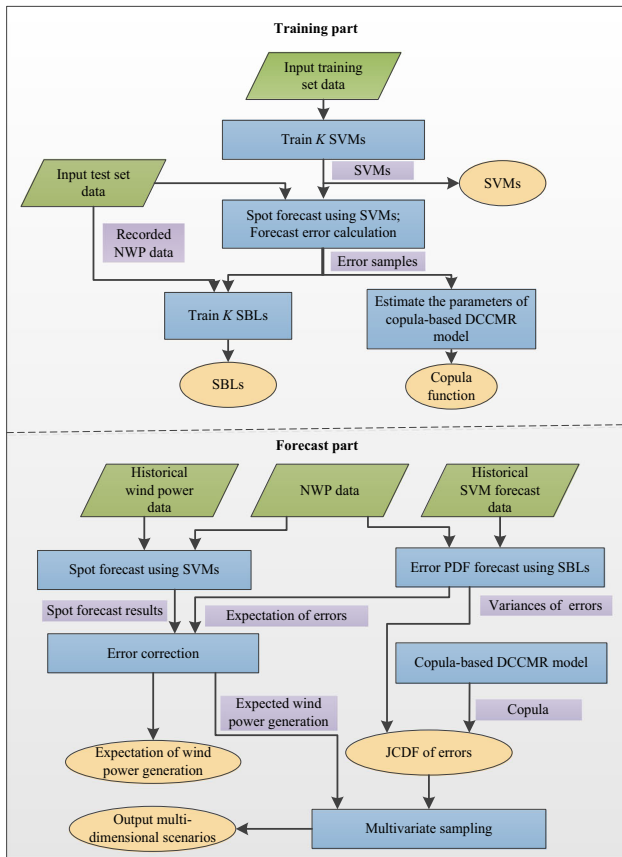


Fig. 2 Framework of the proposed approach

distribution function (JCDF) of the forecast errors, and the corresponding multi-dimensional scenarios.

3 Spot forecast and error distribution

3.1 Spot forecast based on SVM

SVM is an effective statistical machine learning approach which is suitable for the high-order non-linear regression problem [14, 15]. SVM regression model can be expressed by:

$$y_{\text{output}} = \sum_{i=1}^M w_i K(\mathbf{x}_{\text{input}}, \mathbf{x}_i) + w_0 + \varepsilon \quad (1)$$

where y_{output} is the random variable to be predicted; $\mathbf{x}_{\text{input}}$ is the input vector; \mathbf{x}_i is the input vector corresponding to the i^{th} training sample; $K(\cdot)$ is the kernel function (Gaussian kernel function); w_i is the i^{th} weight coefficient; and ε is the residual term.

SVM is used to predict the output of each wind farm in the forecast target period. The detailed training and forecast procedures have been explained in [15]. Based on the

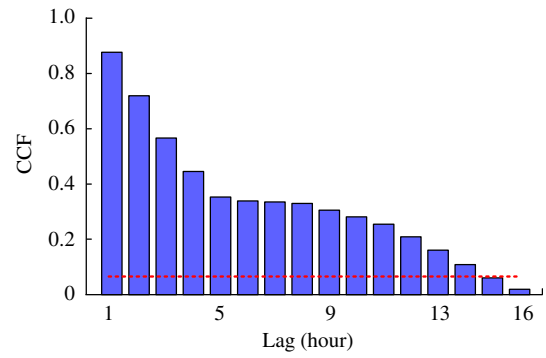


Fig. 3 ACF values of SVM forecast error series

correlation analysis result, wind speed, wind direction and historical generation data are selected as the input data of the model.

3.2 Statistical analyses of forecast errors

Statistical analysis is essential for choosing a reasonable forecast strategy. The statistical properties of the SVM forecast error will be explained in this subsection.

Auto-correlation function (ACF) indicator is employed to test the auto-correlation property of the SVM forecast error, which can be expressed by [16]:

$$\rho(a) = \frac{E((e_t - \mu)(e_{t-a} - \mu))}{h} \quad (2)$$

where $\rho(a)$ is the value of ACF at lag a ; e_t is the t^{th} sample of the series; μ is the mean value of the series; and h is the variance of the series.

Figure 3 shows the ACF values of the 1-h-ahead forecast error series at lag a , $a = 1, 2, \dots, 16$. In this figure, the red dotted line represents the upper confidence limit of the ACF values. It is observed from the figure that the SVM forecast error has significant auto-correlation at the first several lags, which suggests that the historical forecast errors can be applied as the explanatory variables when predicting the error distribution.

Cross-correlation function (CCF) [16] is applied to explore the temporal and spatial dependence relation among the forecast errors. CCF between two forecast error series is defined by:

$$\rho_{i,j}(a) = \frac{E((e_{i,t} - \mu_i)(e_{j,t-a} - \mu_j))}{\sqrt{h_i} \sqrt{h_j}} \quad (3)$$

where i and j are the indices of the error series.

Figure 4 describes the CCF values between the 1-h-ahead forecast error series and the forecast error series from 2-h-ahead to 48-h-ahead. In this figure, the red dotted lines represent the upper and lower confidence limits of the CCF values. It can be seen from the figure that the dependence

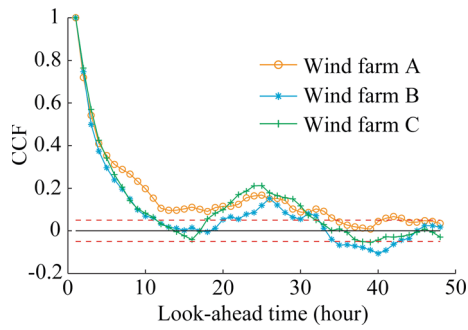


Fig. 4 CCF values between 1-h-ahead forecast error series and error series from 2-h-ahead to 48-h-ahead

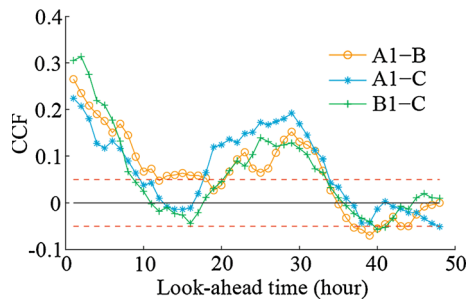


Fig. 5 CCF values between 1-h-ahead forecast error series and all the forecast error series of another wind farm

relation is strong between the 1-h-ahead and 2-h-ahead forecast error series, and the CCF value decreases rapidly when the lag increases.

The CCF values between the error series corresponding to different wind farms are depicted in Fig. 5 to test the temporal-spatial dependence relation of the forecast errors. In the figure, A1-B stands for the CCF values between the 1-h-ahead forecast error series of wind farm A and the error series from 1-h-ahead to 48-h-ahead of wind farm B. So do A1-C and B1-C. The relatively large CCF values verify the existence of the temporal and spatial dependence relation among the forecast errors.

The PDF of the spot forecast error is predicted by SBL in this paper. SBL is a parametric forecast approach which assumes that the wind power generation forecast error at each moment follows a Gaussian distribution. Sometimes this assumption is criticized because the usual statistical distribution of the forecast error is non-Gaussian [17]. Taking a recorded 1-h-ahead forecast error series shown in Fig. 6a as an example, the sharp peak of its statistical distribution distinguishes the error variable from a Gaussian random variable, as shown in Fig. 6b.

The criticism seems reasonable. However, the statistical distribution should not be identified as the distribution at each moment considering the non-stationary feature of the error series [8]. To confirm the validity of this argument, Fig. 6c shows the forecasted PDFs corresponding to the

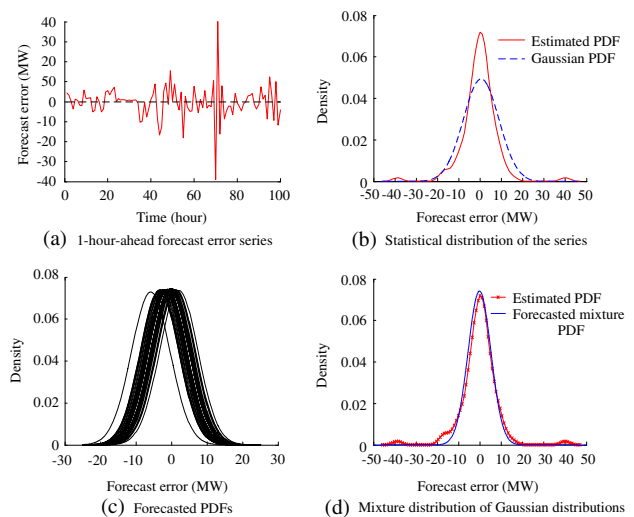


Fig. 6 Comparison between estimated statistical PDF and forecasted PDF

error samples described in Fig. 6a using SBL. In the figure the parameters of the forecasted Gaussian distributions are time-varying, which reflects the non-stationary nature of the error series. The mixture distribution [18] of the Gaussian distributions, which represents the realizations of all the Gaussian variables as one random variable, is calculated according to (4) and is shown in Fig. 6d. By comparison, the mixture distribution is very similar to the statistical distribution, which illustrates that SBL is able to capture the non-Gaussian statistical feature of the SVM forecast error even it assumes that the error at each moment follows a Gaussian distribution.

$$\tilde{f}(x) = \sum_{i=1}^N \frac{1}{N} f_i(x) \tag{4}$$

where $\tilde{f}(x)$ is the PDF of the mixture distribution; $f_i(x)$ is the PDF of the i^{th} Gaussian distribution; N is the number of the Gaussian PDFs. In Fig. 6, $N = 100$.

Additionally, the cross-correlation between the spot forecast error and the corresponding wind speed is tested. According to the test result, the cross-correlation is significant. Therefore, the wind speed data provided by NWP should be incorporated into the input data of the SBL model.

3.3 PDF forecast based on SBL

The distribution of the spot forecast error is estimated by SBL in this paper. SBL is a kernel-based sparse learning model which has significant generalization capability. The parameters of the SBL model are estimated by the maximal posteriori probability estimation according to the Bayesian inference [19, 20]. SBL can provide reliable PDF forecast result, which has been fully verified in [8] and [12].

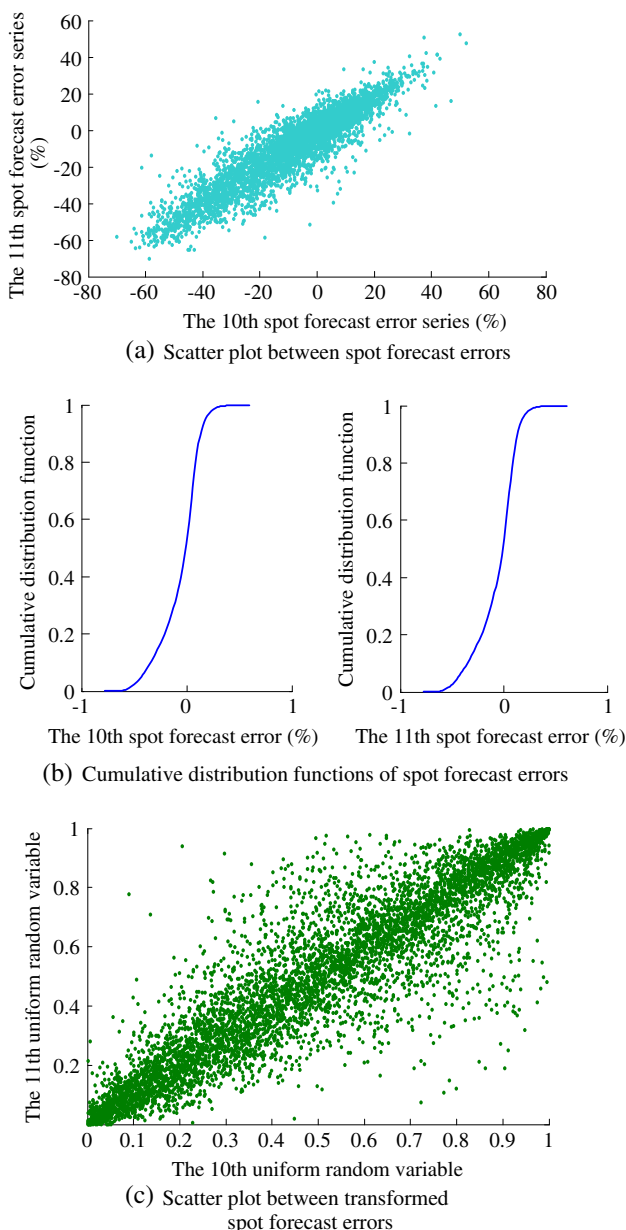


Fig. 7 Analysis of the dependence structure between the random variables using the scatter plot

Thorough descriptions of the SBL model and the corresponding forecast procedure have been given in [8]. The historical forecast data, wind speed and wind direction are selected as the input data of the SBL model in this paper. The outputs of the model are composed of the expectation and variance of the spot forecast error.

4 Multi-dimensional scenario forecast

Multi-dimensional scenarios can be generated from the JCDF of the random variables. To avoid making

assumptions on the type of the joint distribution, copula-based DCCMR is applied here to estimate the JCDF of the spot forecast errors.

4.1 Copula-based DCCMR for modeling the time-varying temporal-spatial dependence structure

1) Basic concepts of the copula function

Copula function is a bridge connecting the marginal and joint distributions of the random variables [21]. Multi-dimensional copula function can be expressed as:

$$F(e_1, e_2, \dots, e_K) = C(F_1(e_1), F_2(e_2), \dots, F_K(e_K)) \quad (5)$$

where e_k is the k^{th} random variable; $F_k(\cdot)$ is the cumulative distribution function (CDF) of the k^{th} random variable; $F(\cdot)$ is the JCDF of the random variables; K is the number of the random variables; and $C(\cdot)$ is the copula function.

According to Sklar's theorem [22], if all the CDFs are continuous, the copula function $C(\cdot)$ is unique. Therefore, the dependence structure of the random variables can be uniquely represented by the corresponding copula function.

2) Selection of the copula function

Many categories of copulas, e.g., Gaussian copulas, Archimedean copulas and extreme-value copulas can be used to model the dependence structure of random variables according to the statistical properties of the variables. Scatter plot, which is able to reveal the relationship between two random variables, is applied here for the copula selection.

Figure 7 provides the scatter plot analysis result corresponding to the wind power generation spot forecast errors. Figure 7a shows the scatter plot between the 10-h-ahead and 11-h-ahead spot forecast error series of wind farm A. The CDFs of the two series are shown in Fig. 7b. And the scatter plot between the transformed error series which are obtained according to the probability integral transform rules [23] is shown in Fig. 7c. In the figure, all the scatter plots have an obvious symmetrical dependence structure, which implies that the dependence relation among the errors can be modeled by the Gaussian copula appropriately [24].

Therefore, a K -dimensional Gaussian copula is selected to model the dependence structure of the spot forecast errors. Gaussian copula has an explicit formula and its computational complexity is moderate. The K -dimensional Gaussian copula can be defined by [24]:

$$\begin{aligned} F(e_1, e_2, \dots, e_K; \mathbf{R}) &= C(F_1(e_1), F_2(e_2), \dots, F_K(e_K); \mathbf{R}) \\ &= \Phi_{\mathbf{R}}(\Phi^{-1}(F_1(e_1)), \Phi^{-1}(F_2(e_2)), \\ &\quad \dots, \Phi^{-1}(F_K(e_K))) \end{aligned} \quad (6)$$

where Φ^{-1} is the inverse of the one-dimensional standard Gaussian CDF; $\Phi^{-1}(F_k(e_k))$ is a random variable following the standard Gaussian distribution; and Φ_R stands for a K -dimensional Gaussian JCDF with zero means, unit marginal variances and the covariance matrix/correlation matrix \mathbf{R} .

It can be seen from (6) that the JCDF $F(e_1, e_2, \dots, e_K; \mathbf{R})$ of the forecast errors can be obtained by estimating the CDFs of the forecast errors and the corresponding correlation matrix \mathbf{R} .

In this paper, the CDFs are forecasted according to the process mentioned in Section 3, and the matrix \mathbf{R} is estimated using the following copula-based DCCMR model.

3) Copula-based DCCMR model

In (6) the random vector $[\Phi^{-1}(F_1(e_1)), \Phi^{-1}(F_2(e_2)), \dots, \Phi^{-1}(F_K(e_K))]$ follows the K -dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$ where the correlation matrix \mathbf{R} can be estimated dynamically as follows [25]:

$$\mathbf{R}_t = \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \tag{7}$$

where \mathbf{R}_t is the time-varying correlation matrix; t is the index indicating the time to estimate the correlation matrix; \mathbf{Q}_t can be expressed by

$$\mathbf{Q}_t = \left(1 - \sum_{i=1}^I \alpha_i - \sum_{j=1}^J \beta_j \right) \bar{\mathbf{Q}} + \sum_{i=1}^I \alpha_i \mathbf{v}_{t-i} \mathbf{v}_{t-i}^T + \sum_{j=1}^J \beta_j \mathbf{Q}_{t-j} \tag{8}$$

where α_i and β_j are the parameters need to be identified; I and J are the orders of the model; \mathbf{v}_{t-i} is the realization of the random vector $[\Phi^{-1}(F_1(e_1)), \Phi^{-1}(F_2(e_2)), \dots, \Phi^{-1}(F_K(e_K))]$ at time period $t - i$; \mathbf{Q}_{t-j} is the covariance matrix estimated at time period $t - j$; and $\bar{\mathbf{Q}}$ can be expressed as

$$\bar{\mathbf{Q}} = \frac{1}{\Omega} \sum_{k=1}^{\Omega} \mathbf{v}_k \mathbf{v}_k^T \tag{9}$$

where Ω is the number of the samples of the random vector $[\Phi^{-1}(F_1(e_1)), \Phi^{-1}(F_2(e_2)), \dots, \Phi^{-1}(F_K(e_K))]$.

The parameters α_i and β_j in the above equations are estimated using the composite maximum likelihood approach, and the detailed estimation process is explained in [25].

4.2 Generate multi-dimensional scenarios

According to the spot forecast result, PDF of the spot forecast error, and the correlation matrix of the errors, V groups of K -dimensional wind power generation scenarios can be generated by taking the following steps.

Step 1: Generate V groups of K -dimensional random samples according to the Gaussian copula described in (6). And the i^{th} group of the samples is represented by $[u_{i,1,1}, u_{i,1,2}, \dots, u_{i,1,T}, \dots, u_{i,L,1}, u_{i,L,2}, \dots, u_{i,L,T}]$, in which L is the number of the wind farms and T is the number of the look-ahead periods.

Step 2: Generate V groups of error samples through the inverse transform process [24]. With respect to the forecasted marginal distribution functions, the i^{th} error sample vector can be transformed from the samples obtained in Step 1 as:

$$\mathbf{e}_i = \left[F_{1,1}^{-1}(u_{i,1,1}), F_{1,2}^{-1}(u_{i,1,2}), \dots, F_{1,T}^{-1}(u_{i,1,T}), \dots, F_{L,1}^{-1}(u_{i,L,1}), F_{L,2}^{-1}(u_{i,L,2}), \dots, F_{L,T}^{-1}(u_{i,L,T}) \right] \tag{10}$$

Step 3: Generate V groups of multi-dimensional scenarios of wind power generation. And the i^{th} scenario can be generated by:

$$\mathbf{s}_i = \boldsymbol{\eta} + \mathbf{e}_i \tag{11}$$

where $\boldsymbol{\eta} = [\hat{p}_{1,1}, \hat{p}_{1,2}, \dots, \hat{p}_{1,T}, \dots, \hat{p}_{L,1}, \hat{p}_{L,2}, \dots, \hat{p}_{L,T}]$ is the generation predicted by the spot forecast module.

5 Test results and discussions

5.1 Performance evaluation indicators

The following indicators are applied to evaluate the performance of the proposed approach.

1) Indicator for the expectation forecast result

Normalized mean absolute error (NMAE) [8] is employed here to evaluate the accuracy of the forecasted expectation of wind power generation. The indicator can be expressed by:

$$N_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |e_i| \tag{12}$$

where e_i is the i^{th} forecast error sample and N is the number of the samples.

2) Indicators for the distribution forecast result

Indicators including the distortion rate (DR), marginal calibration, sharpness, and the continuous ranked probability score (CRPS) are applied to evaluate the distribution forecast performance of the proposed approach.

DR [8] of a forecasted PDF is defined by:

$$D_R = \frac{1}{2N} \sum_{i=1}^H |N_{i,a} - N_{i,f}| \times 100\% \tag{13}$$

where $N_{i,a}$ is the actual times that the wind power generation sample falls into the i^{th} probability interval; $N_{i,f}$ is the expected times that the wind power generation sample

should fall into the i^{th} probability interval according to the forecasted PDF; and H is the number of the probability intervals.

Marginal calibration is an indicator concerning about the equality of the observed CDF and the forecasted CDF [26]. To calculate the indicator, the observed CDF is represented by the average value of the indicator functions:

$$\bar{G}_N(p) = \frac{1}{N} \sum_{i=1}^N 1\{p_i \leq p\} \tag{14}$$

where $1\{\cdot\}$ is a $\{0, 1\}$ indicator function which takes value 1 when the condition is satisfied; p is the normalized wind power generation; and p_i is the i^{th} sample of the wind power generation.

The corresponding forecasted CDF is represented by the average forecasted CDF:

$$\bar{F}_N(p) = \frac{1}{N} \sum_{i=1}^N F_i(p) \tag{15}$$

where $F_i(\cdot)$ is the forecasted CDF corresponding to the i^{th} wind power generation sample.

The marginal calibration, which is a function of p , measures the difference between the forecasted CDF and the observed CDF:

$$M_C = \bar{F}_N(p) - \bar{G}_N(p) \tag{16}$$

Sharpness is another important performance indicator for evaluating the forecasted PDF. Obviously, the sharper the forecasted distribution is, the better the probabilistic forecast approach will be, since a sharper distribution means less volatility of the forecast result. In this paper, the sharpness of the forecasted PDF is assessed by the coverage of the central probability intervals.

Moreover, in order to measure the overall performance of the probabilistic forecast approach, the CRPS, which can address the calibration and sharpness simultaneously [26], is calculated by:

$$C_{RPS} = \frac{1}{N} \sum_{i=1}^N \left[\int_0^1 (F_i(p) - 1\{p_i \leq p\})^2 dp \right] \tag{17}$$

3) Quality evaluation of multi-dimensional scenarios

Energy score (ES), which is a multivariate verification tool for the forecasted scenarios, is applied to evaluate the quality of the generated multi-dimensional scenarios [27]. The indicator is a negatively-oriented score. The lower the indicator is, the better the forecast result will be. ES is defined by:

$$E_s = \frac{1}{V} \sum_{i=1}^V \|p - s_i\|_2 - \frac{1}{2V^2} \sum_{i=1}^V \sum_{j=1}^V \|s_i - s_j\|_2 \tag{18}$$

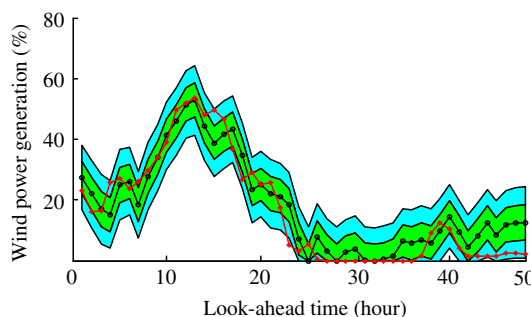


Fig. 8 An example of PDF forecast result

where $\|\cdot\|_2$ is the Euclidean norm; s_i and s_j are the predicted scenarios; V is the number of the scenarios; and p is the measured wind power generation series.

5.2 Performance evaluation of the proposed approach

Data collected from the three adjacent wind farms are used to illustrate the effectiveness of the proposed approach. 5000 times forecast tests are implemented on the wind farms and each test will forecast the PDFs of wind power generation for the further 48 hours. An example of the PDF forecast result is shown in Fig. 8, where the black solid line with circles and the asterisked red line stand for the forecasted wind power generation curve and the actual wind power generation curve respectively. The central 0.65 and 0.95 confidence intervals are represented in the figure by two different colors.

In Fig. 8, most of the actual wind power generation samples fall into the 0.65 confidence interval and very few samples fall outside the 0.95 confidence interval. The result indicates that the forecasted wind power generation distribution can reflect the real distribution appropriately.

The persistence (PER) model [12], the common SVM model and the linear quantile regression model are selected as the competitive models to evaluate the expectation forecast accuracy of the proposed approach.

Table 1 shows the average NMAE of 48 look-ahead time periods. It can be seen from the table that the SVM model has a remarkable superiority to the PER model, and the accuracy of the SVM forecast result is improved significantly by using the error correction strategy proposed in this paper. Also, the proposed approach has better performance than the linear quantile regression model.

Figure 9 depicts the NMAE values corresponding to the forecast results of wind farm A for the 48 look-ahead time periods. It can be seen from the figure that the proposed approach is much better than the other three benchmark approaches on the expectation forecast accuracy. Similar conclusions can be found from the test results of the other two wind farms.



Table 1 Average NMAE of 48 look-ahead time periods

Wind farms	Average NMAE (%)			
	PER	Quantile regression	SVM	Proposed approach
Wind farm A	23.18	12.16	11.36	9.20
Wind farm B	29.37	15.27	11.18	8.85
Wind farm C	21.99	16.39	10.70	9.06

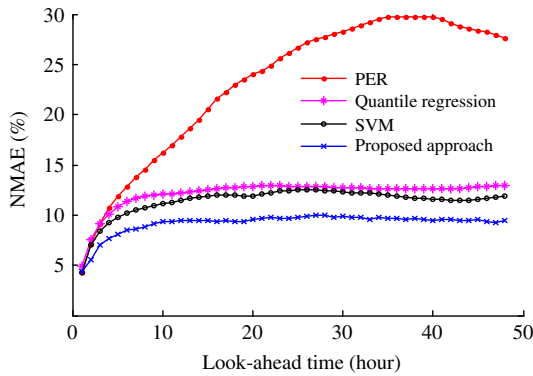


Fig. 9 NMAE curves of wind farm A

Table 2 Information of the probability intervals

Intervals	Corresponding probability	Theoretical falling points
$(-\infty, -2\sigma]$	0.0228	114
$(-2\sigma, -\sigma]$	0.0228	680
$(-\sigma, 0]$	0.3413	1706
$(0, \sigma]$	0.3413	1706
$(\sigma, 2\sigma]$	0.1359	680
$(2\sigma, \infty]$	0.0228	114

Table 3 Average DR of 48 look-ahead time periods

Approaches	Average DR (%)		
	Wind farm A	Wind farm B	Wind farm C
Empirical approach	19.67	13.50	13.09
Quantile approach	15.84	19.35	15.23
Proposed approach	13.57	10.64	11.25

Empirical distribution estimation is a popular non-parametric distribution estimation approach, which estimates the probability distribution of a random variable by analyzing its historical realizations [28]. The empirical approach and the linear quantile regression approach are applied here as benchmarks to evaluate the probabilistic forecast performance of the proposed approach.

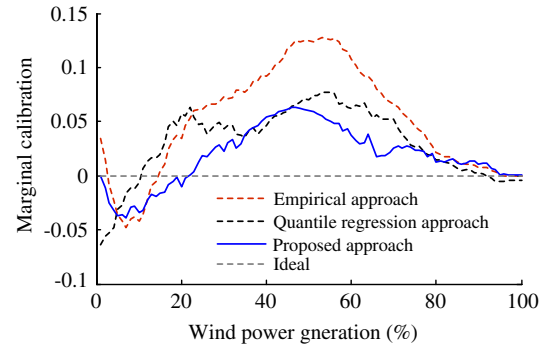


Fig. 10 Marginal calibration curves of 6-h-ahead forecast results

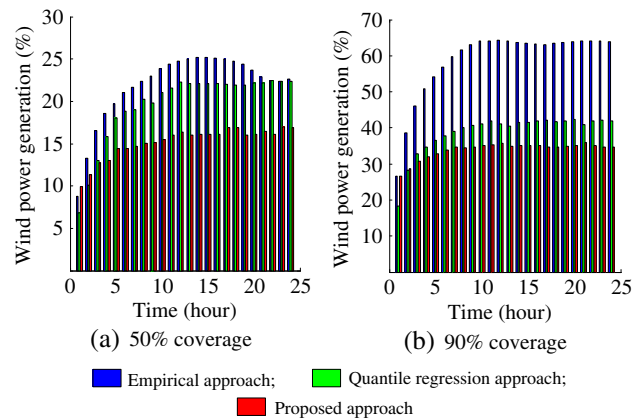


Fig. 11 Coverage of 50% and 90% central probability intervals

To calculate the DR indicator of the PDF forecast result, six probability intervals have been specified according to the variance of the Gaussian distribution σ . The probabilities corresponding to the intervals are given in Table 2. And the number of the theoretical falling points for each interval in the 5000 times forecast tests is also presented in the table.

Table 3 summaries the average DR values of 48 look-ahead time periods according to the forecast results. It can be seen in the table that the proposed approach has lower average DR values for all the three wind farms than the other two approaches.

The marginal calibration curves corresponding to the 6-h-ahead distribution forecast results are shown in Fig. 10. In the figure the proposed approach has relatively lower marginal calibration values, which means the CDF forecasted by the proposed approach is much closer to the real CDF.

The coverage of the 50% and 90% central probability intervals corresponding to the results of the two approaches is described in Fig. 11. It can be seen from the figure that the coverage of the proposed approach is almost always

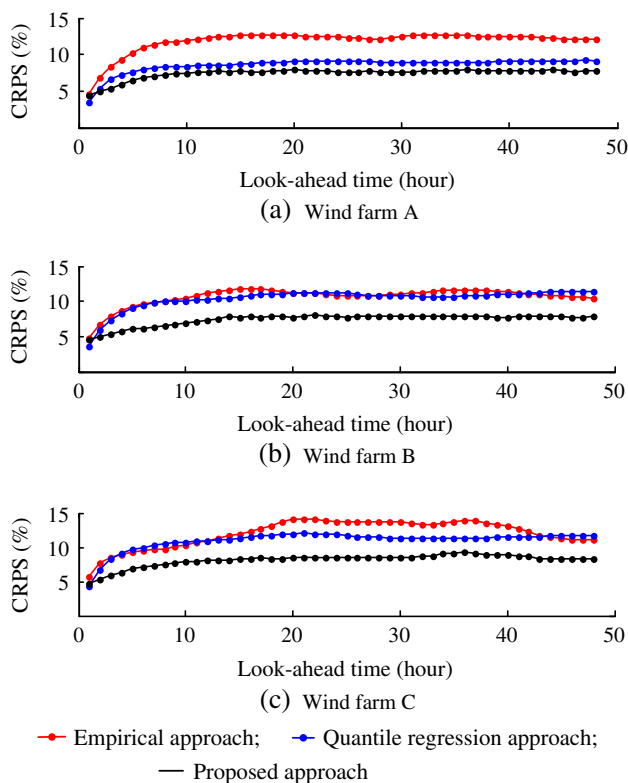


Fig. 12 CRPS values corresponding to two probabilistic forecast approaches for three wind farms

smaller than that of the other two approaches, suggesting that the proposed approach can provide less volatile forecast results.

The CRPS values of the three approaches are depicted in Fig. 12. This figure shows that the proposed approach has lower CRPS values in almost all the periods, meaning that the proposed approach has better overall distribution forecast performance.

A set of 50 multi-dimensional scenarios corresponding to the predicted PDF and the estimated Gaussian copula is depicted in Fig. 13. In the figure, it can be seen that the actual wind power generation curve is well covered by the scenarios, which means that the scenarios can reflect the real wind power generation properly.

At the same time, a set of 2000 scenarios is generated for the quantitative evaluation. Table 4 summarizes the ES values corresponding to the generated 2000 scenarios. According to the table, the ES indicator is lower when the correlation information is included in the forecast result. The test result indicates that the temporal-spatial correlation information has positive effects on improving the quality of the forecasted scenarios.

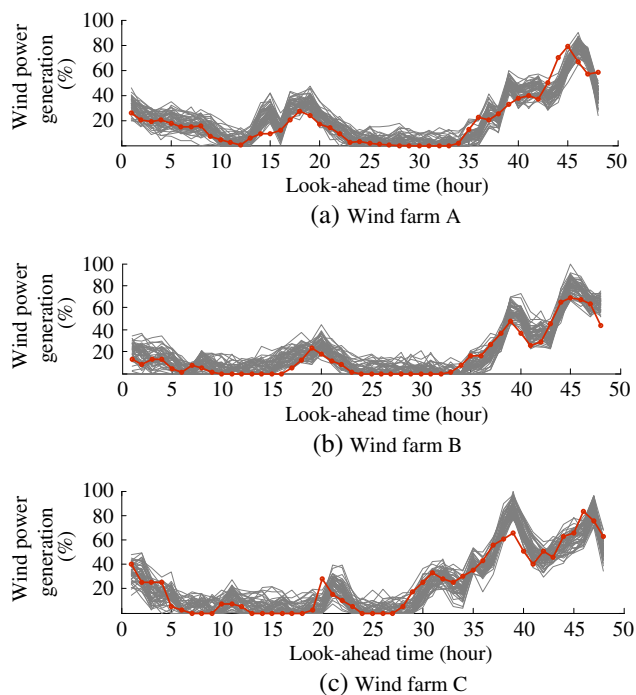


Fig. 13 Generated multi-dimensional scenarios

Table 4 ES values corresponding to the forecasted scenarios

Approaches	ES values (%)		
	Wind farm A	Wind farm B	Wind farm C
Without considering correlation information	0.5634	0.6722	0.8879
Considering correlation information	0.5572	0.6654	0.8798

6 Conclusions

A multi-dimensional scenario forecast approach is proposed in this paper. In the proposed approach, SVM is used to perform the spot forecast of wind power generation. The expectation and variance of the spot forecast error are estimated by SBL. Then the SVM forecast result is corrected using the estimated error expectation. The dependence structure of the forecast errors is reflected by the Gaussian copula, which is estimated using the copula-based DCCMR model. Therefore, the multi-dimensional scenarios of wind power generation are produced with respect to the spot forecast result, PDF of the spot forecast error, and the Gaussian copula. The proposed approach is tested on three adjacent wind farms, and the test results illustrate the effectiveness of the approach.

Acknowledgements This work is supported by National Natural Science Foundation of China (No. 51007047, No. 51077087), Shandong Provincial Natural Science Foundation of China (No. 20100131120039), and National High Technology Research and Development Program of China (863 Program) (No. 2011AA05A101).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- [1] Global wind statistics in 2013 (2014) Global wind energy council (GWEC), Brussels
- [2] Cui J, Chi YN, Li Y et al (2012) Study on grid capability to accommodate wind power based on DSM. In: Proceedings of the 2012 Asia-Pacific power and energy engineering conference (APPEEC'12), Shanghai, China, 27–29 Mar 2012, 5 pp
- [3] Methaprayoon K, Yingvivatanapong C, Lee WJ et al (2007) An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty. *IEEE Trans Ind Appl* 43(6):1441–1448
- [4] Bessa RJ, Matos MA, Costa IC et al (2012) Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study. *IEEE Trans Sustain Energy* 3(4):827–836
- [5] Bremnes JB (2004) Probabilistic wind power forecasts using local quantile regression. *Wind Eng* 7(1):47–54
- [6] Kou P, Gao F, Guan XH (2013) Sparse online warped Gaussian process for wind power probabilistic forecasting. *Appl Energy* 108(C):410–428
- [7] Pinson P, Kariniotakis G (2010) Conditional prediction intervals of wind power generation. *IEEE Trans Power Syst* 25(4):1845–1856
- [8] Yang M, Fan S, Lee WJ (2013) Probabilistic short-term wind power forecast using componential sparse Bayesian learning. *IEEE Trans Ind Appl* 49(6):2783–2792
- [9] Juban J, Siebert N, Kariniotakis GN (2007) Probabilistic short-term wind power forecasting for the optimal management of wind generation. In: Proceedings of the 2007 IEEE Lausanne power tech conference, Lausanne, Switzerland, 1–5 July 2007, pp 683–688
- [10] Kanga AFK, Voller S, Verstege JF (2009) Congestion management in transmission systems with large scale integration of wind energy. In: Proceedings of the 2009 CIGRE/IEEE PES joint symposium on integration of wide-scale renewable resources into the power delivery system, Calgary, Canada, 29–31 July 2009, 11 pp
- [11] Xie L, Gu YZ, Zhu XX et al (2011) Power system economic dispatch with spatio-temporal wind forecasts. In: Proceedings of the 2011 IEEE Energytech conference, Cleveland, OH, USA, 25–26 May 2011, 6 pp
- [12] Zhu SM, Yang M, Li M et al (2013) One parametric approach for short-term JPDF forecast of wind generation. In: Proceedings of the 2013 IEEE industry applications society annual meeting, Lake Buena Vista, FL, USA, 6–11 Oct 2013, 7 pp
- [13] Pinson P, Madsen H, Nielsen HA et al (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Eng* 12(1):51–62
- [14] Cortes C, Vapnik V (1995) Support vector machine. *Mach Learn* 20(3):273–297
- [15] Zhu SM, Yang M, Han XS (2012) Short-term generation forecast of wind farm using SVM-GARCH approach. In: Proceedings of the 2012 IEEE international conference on power system technology (POWERCON'12), Auckland, New Zealand, 30 Oct–2 Nov 2012, 6 pp
- [16] Chatfield C (2004) The analysis of time series: an introduction, 6th edn. Chapman & Hall/CRC, Boca Raton
- [17] Bludszweit H, Dominguez-Navarro JA, Llobert A (2008) Statistical analysis of wind power forecast error. *IEEE Trans Power Syst* 23(3):983–991
- [18] Sylvia FS (2006) Finite mixture and Markov switching models. Springer, New York
- [19] Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res* 1:211–244
- [20] Tipping ME, Faul AC (2003) Fast marginal likelihood maximization for sparse Bayesian models. In: Proceedings of the 9th international workshop on artificial intelligence and statistics, Key West, FL, USA, 3–6 Jan 2003, 13 pp
- [21] Bessa RJ, Miranda V, Botterud A et al (2012) Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renew Energy* 40(1):29–39
- [22] Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges, vol 8. de L'Institut de Statistique de L'Université de Paris, Paris, pp 229–231
- [23] Angus JE (1994) The probability integral transform and related results. *SIAM Rev* 36(4):652–654
- [24] Nelsen RB (1999) An introduction to copulas. Springer, New York
- [25] Bodnar T, Hautsch N (2013) Copula-based dynamic conditional correlation multiplicative error processes. Center for Financial Studies (CFS) Working Paper Series 2013/19
- [26] Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *J R Stat Soc B* 69(2):243–268
- [27] Gneiting T, Stanberry LI, Gritmit EP et al (2008) Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test* 17(2):211–235
- [28] Williams WH, Goodman ML (1971) A simple method for the construction of empirical confidence limits for economic forecasts. *J Am Stat Assoc* 66:752–754

Ming YANG received the Ph.D. degree in electrical engineering from Shandong University, Jinan, China, in 2009. Currently, he is an associate professor at Shandong University. His research interests include energy system forecasting, power system analysis, and power system operation optimization.

You LIN received the B.S. degree in electrical engineering from Shandong University, Jinan, China, in 2012. Currently, she is pursuing the M.S. degree in electrical engineering at Shandong University. Her research focuses on energy system forecasting.

Simeng ZHU received the M.S. degree in electrical engineering from Shandong University, Jinan, China, in 2014. Currently, he is an Engineer at Zaozhuang Power Supply Company, Zaozhuang, China. His research focuses on energy system forecasting.

Xueshan HAN received the Ph.D. degree in electrical engineering from Harbin Institute of Technology University, Harbin, China, in 1994. Currently, he is a Professor at Shandong University, Jinan, China. His research interests include energy system forecasting, power system analysis, and power system operation optimization.

Hongtao WANG received the Ph.D. degree in electrical engineering from Shandong University, Jinan, China, in 2005. Currently, he is a professor at Shandong University. His research interests include power system dispatching with intermittency and renewable energy, power system stability analysis, and power system restoration.