Aggregated Probabilistic Wind Power Forecasting Based on Spatio-Temporal Correlation

Mucun Sun, Cong Feng, Jie Zhang
The University of Texas at Dallas
Richardson, TX 75080, USA
Email: jiezhang@utdallas.edu

***Abstract*—In this paper, an aggregated probabilistic wind power forecasting method by considering spatio-temporal effects is proposed, which consists of three major components: Q-learning enhanced deterministic forecasting, marginal distribution fitting, and aggregated probabilistic wind power forecasting. A high-dimensional joint distribution is used to model the spatio-temporal correlation among the member wind farms through Copula, where the marginal distributions are built from historical aggregated wind power observations and forecasts from member wind farms. Then, a conditional distribution of the aggregated wind power is deduced through the Bayesian theory, which is used for aggregated probabilistic forecasts. Numerical results of a 3 wind farms case study show that the performance of the aggregated probabilistic wind power forecasting is enhanced by considering the spatio-temporal correlation among individual wind farms.**

***Index Terms*—probabilistic wind forecasting, aggregated probabilistic forecasting, copula, pinball loss, spatio-temporal correlation.**

I. Introduction

The uncertain and variable nature of wind makes it challenging to be integrated into power systems, particularly at ever-increasing level of wind penetration. Studies have shown that the integration of geographically dispersed wind farms could reduce extreme power output, which is referred to as smoothing effect [1]. In addition, power produced from one wind farm at different times is typically temporally correlated [2]. It would be interesting to explore the impacts of spatio-temporal correlation on the performance of wind forecasting. The benefits of spatio-temporal modeling for wind power forecasting at aggregated levels have been briefly discussed in [3].

A number of wind power forecasting technologies that considers spatio-temporal effects have been developed in the literature and also applied to a variety of power system operation and planning problems. For example, He *et al.* [4] used finite-state Markov chains to account for the non-stationary and periodicity of wind power generation across multiple years between different wind farms. The developed wind power forecasts was used in stochastic unit commitment and economic dispatch models. Xie *et al.* [5] leveraged the spatio-temporal correlation in wind speed and direction among geographically dispersed wind farms to generate wind power forecasts. Probabilistic wind forecasting technologies based on spatio-temporal effects have also been developed in the literature. For example, Zhang *et al.* used off-site information of geographically dispersed wind farms to capture spatio-temporal correlation and generated quantile forecasts. Then an Alternating Direction Method of Multipliers (ADMM)-based method was used to generate distributed probabilistic forecasts. Dowell *et al.* [6] proposed a probabilistic wind power forecasting method and the spatio-temporal correlation was captured through Sparse Vector Autoregression. In addition to wind power, spatio-temporal correlation modeling has also been applied to wind speed forecasting [7] and solar power forecasting [8].

One of the most intuitive ways of modeling spatio-temporal correlation is to use a joint distribution. Based on the Sklar's theorem, the joint distribution can be modeled through univariate marginal-distribution functions and a Copula that describes the dependence structure between the variables [9]. By modeling and considering the spatio-temporal correlation between different wind farms, this paper develops a conditional aggregated probabilistic wind power forecasting model based on the Copula theory. First, deterministic wind power forecasts are generated for member wind farms (with selected deterministic forecasting methods). Second, a Copula method is used to build a spatio-temporal correlated joint model between the aggregated wind power and forecasted wind power of each wind farm. The conditional distribution of aggregated wind power is deduced through Bayesian theory, which is then used (in conjunction with deterministic forecasts) to generate probabilistic wind power forecasts at the forecasting stage.

The rest of the paper is organized as follows. Section II describes the proposed aggregated probabilistic forecasting method, which consists of a deterministic forecasting method, marginal probability distribution selection, and a spatio-temporal correlation based Copula model. Section III applies the developed spatio-temporal correlation based aggregated probabilistic wind power forecasting method to three wind farms. Concluding remarks and future work are discussed in Section IV.

II. Methodology

The overall framework of the aggregated probabilistic wind power forecasting methodology by considering spatio-temporal correlation is illustrated in Fig. 1, which consists of three major steps: Q-learning enhanced deterministic forecasting, marginal distribution fitting, and aggregated probabilistic

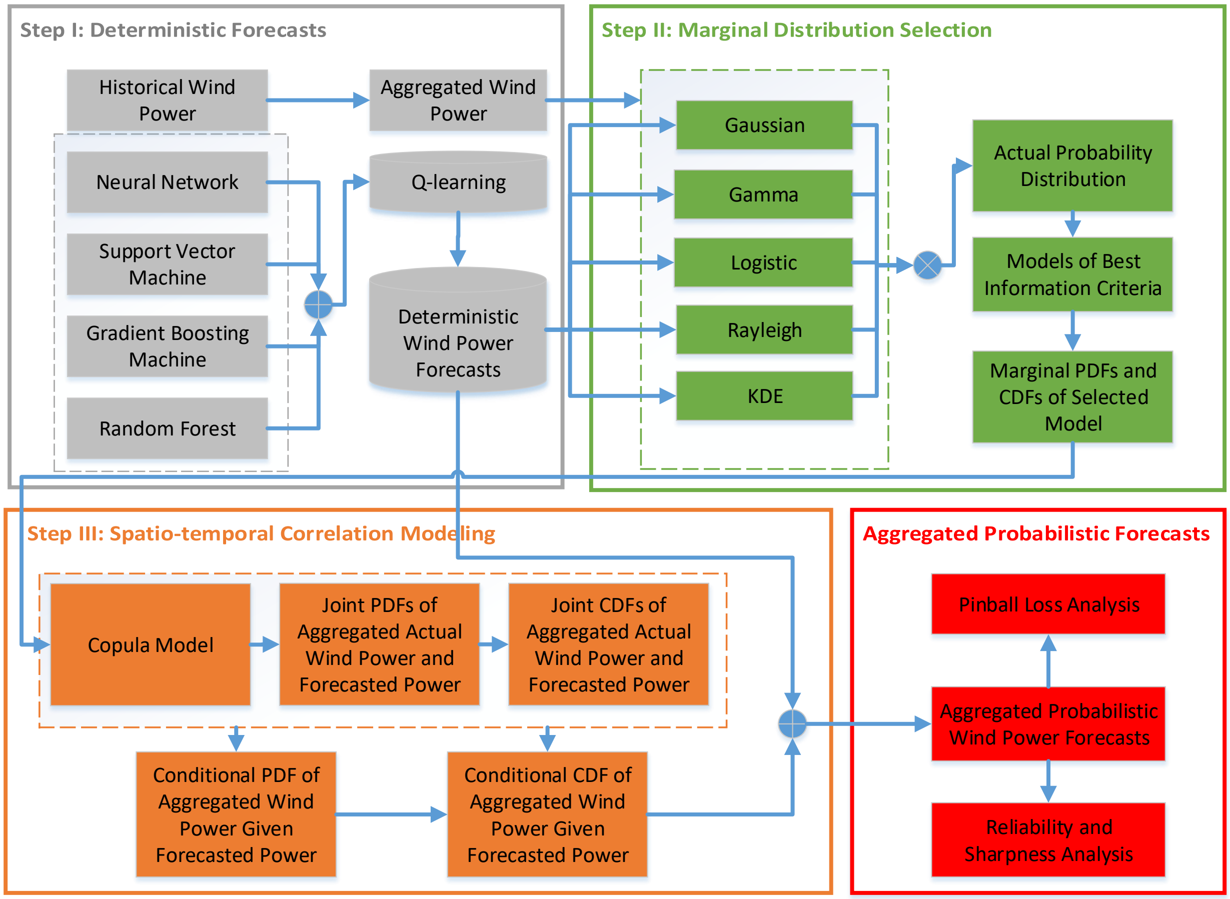

Fig. 1: The Overall framework of the developed aggregated probabilistic wind power forecasts

wind power forecasting. The three major steps are briefly described as follows:

1) Step 1: A Q-learning based ensemble deterministic forecasting method is used to select the best forecasting model from a pool of state-of-the-art machine learning based forecasting models at each time step, thus generating deterministic wind power forecasts for all member wind farms.
2) Step 2: A number of selected probability distribution types are used to fit the probability distribution functions (PDFs) of historical aggregated actual wind power and the wind power forecasts at each wind farm.
3) Step 3: Based on the Copula theory, the joint distribution of historical aggregated actual wind power and wind power forecasts at each member wind farm is constructed.

A. *Q-learning Enhanced Deterministic Forecasting*

The developed spatio-temporal correlation based aggregated probabilistic forecasting methodology is constructed based on deterministic forecasts. A large collection of methods have been done to effectively perform deterministic wind forecasting. However, most of existing deterministic methods are either selected based on the overall performance or ensembled by multiple models. Selecting a model based on the overall forecasting performance generally neglects the local performance of the selected model.

In this paper, a Q-learning enhanced deterministic forecasting method is adopted, which can choose the best forecasting model from a pool of state-of-the-art machine learning based forecasting models (i.e., artificial neural network, support vector machine, gradient boosting machine, and random forest) at each time step. To be more specific, the developed method trains Q-learning agents based on the rewards of transferring from the current model to the next model. For example, a Q-learning agent will receive a reward by transferring from the current forecasting model M_i to the next forecasting model M_j in each training step, from which the Q-learning agent will learn the optimal policy of the model selection. Then, this optimal policy will be applied to select the best model for forecasting in the next step based on the current model in the forecasting stage. The dynamic model selection process is expressed as:

$$\boldsymbol{S} = \{\boldsymbol{s}\} = \{s_1, s_2, ..., s_I\} \tag{1}$$

$$\boldsymbol{A} = \{\boldsymbol{a}\} = \{a_1, a_2, ..., a_I\} \tag{2}$$

$$R^t(s_i, a_j) = ranking(M_i) - ranking(M_j) \tag{3}$$

where $\boldsymbol{S}$, $\boldsymbol{A}$, and R are state space, action space, and reward function in the dynamic model selection Markov Decision Process, respectively. s and a are possible state and action, respectively. I is the number of models (M) in the model pool. The reward function is defined as the model performance improvement, which ensures the effective and efficient convergence of Q-learning. More details about the Q-learning enhanced deterministic forecasting can be found in Ref. [10].

B. Spatio-temporal Correlation Modeling Among Wind Farms

Copula theory is one of the most widely used methods for modeling dependency between different random variables, which is adopted in this paper for spatio-temporal correlation modeling. A multi-distribution database is formulated to model the possible shapes of the predicted wind power distribution, which consists of 5 distribution types: Gaussian, Gamma, Logistic, Rayleigh, and kernel density estimation (KDE). The aggregated actual wind power p^{Σ} can be expressed as follows:

$$p^{\Sigma} = \sum_{i=1}^{N} p_i \tag{4}$$

where p_i is the actual wind power of the ith wind farm, and N is the total number of wind farms to be aggregated. The parameter $\hat{p}_i$ denotes the corresponding forecasted wind power, and $f(\hat{p}_i)$ denotes the marginal PDF of the forecasted wind power at the ith wind farm. Similarly, $F(\hat{p}_i)$ denotes the marginal cumulative distribution function (CDF) of the forecasted wind power at the ith wind farm. In the Copula theory, the joint CDF of the predicted wind power at different member wind farms and the aggregated actual wind power (of all farms), $F(p^{\Sigma}, \hat{p}_1, ..., \hat{p}_N)$, can be modeled through their marginal CDFs and a Copula function, which is expressed by:

$$F(p^{\Sigma}, \hat{p}_1, ..., \hat{p}_N) = C(F(p^{\Sigma}), F(\hat{p}_1), ...F(\hat{p}_N)) \tag{5}$$

where $C(\cdot)$ is the Copula function. Similarly, the joint PDF of the forecasted wind power at different member farms and the aggregated actual wind power (of all farms) is expressed as:

$$f(p^{\Sigma}, \hat{p}_1, ..., \hat{p}_N) = c(F(p^{\Sigma}), F(\hat{p}_1), ...F(\hat{p}_N)) \cdot f(p^{\Sigma}) \cdot \prod_{i=1}^{N} f(\hat{p}_i) \tag{6}$$

where the marginal PDF is modeled through the aforementioned 5 different distribution types based on historical actual and forecasting data. Therefore, the conditional joint PDF of the aggregated wind power given all the member farm power forecasts could be deduced from Bayesian theory, given by

$$f(p^{\Sigma} | \hat{p}_1, ..., \hat{p}_N) = \frac{c(F(p^{\Sigma}), F(\hat{p}_1), ...F(\hat{p}_N))}{c(F(\hat{p}_1), ...F(\hat{p}_N))} \cdot f(p^{\Sigma}) \tag{7}$$

The conditional distribution of aggregated wind power given all the member forecasts can be trained through historical actual and forecasting data. Based on the copula model and trained conditional PDF in Eq. 7, scenarios of aggregated wind power can be generated.

Aggregated probabilistic wind power forecasts are generated by sampling from the conditional distribution given deterministic forecasts of each wind farm. An inverse transform method is used to sample from the conditional distribution to generate a large number of aggregated wind power forecasting scenarios.

III. Case Study and Results

A. Data Summary

The developed spatio-temporal correlation based probabilistic forecasting approach is applied to three wind farms for aggregated wind power forecasting. The wind power data is collected from the Wind Integration National Dataset (WIND) Toolkit with a 1-hour resolution [11]. The duration of the collected data is summarized in Table I. For all the 3 wind farms, the first 3/4 of data is used as training data. The correlation matrix based on the training data is visualized in Fig. 2. It is seen that the forecasted wind power of each wind farm and the aggregated wind power (of all the 3 farms) is highly correlated. The number of scenarios generated from the conditional distribution is set as N_s=5,000. The accuracy of the forecasts is evaluated by the remaining 1/4 of data. Although the developed method is capable of generating forecasts at multiple forecasting horizons, only 1-hour-ahead (1HA) forecasts are generated in this study.

TABLE I: Data summary of the selected sites in WIND Toolkit

Case No.	Site ID	Data duration	Capacity (MW)
C1	10069	2010-01-01 to 2012-12-31	16
C2	10526	2010-01-01 to 2012-12-31	16
C3	10527	2010-01-01 to 2012-12-31	2

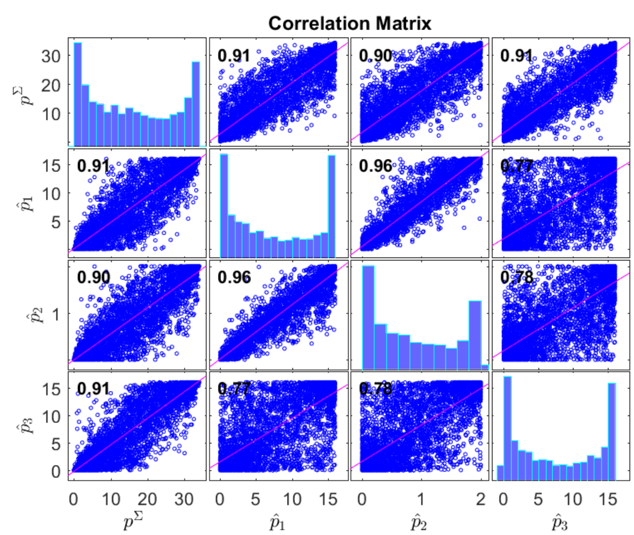

Fig. 2: Correlation matrix between aggregated wind power and forecasted wind power of each wind farm

B. Comparison of Different Marginal Distribution Models

Fig. 3 shows the marginal probability distributions of wind power forecasts from five distribution types at the C3 site. The

Akaike information criterion (AIC) and the Log-Likelihood (LL) are used to evaluate the estimated wind power distribution accuracy. The preferred model is the one that has the lowest AIC and the biggest LL [12]. The AIC and LL values of the three wind farms with different distribution types are summarized in Table II. Results show that the KDE distribution outperforms other single distributions for modeling wind power forecasts. Therefore, the estimated distribution through KDE is used as marginal distribution of the Copula model. The CDF $\hat{F}$ corresponding to the estimated PDF is expressed as:

$$\hat{F}(p) = \frac{1}{n}\sum_{m=1}^{M}\phi(\frac{p-p_m}{h}) \tag{8}$$

where h is the bandwidth, ϕ is the CDF of the kernel, and M is the number of kernels.

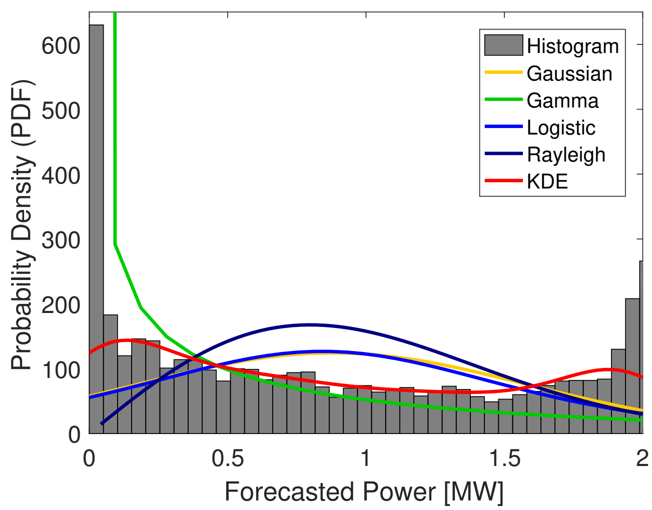

Fig. 3: Probability distributions of wind power forecasts at the C3 site

TABLE II: Information criteria of the estimated distribution

Model	C1 AIC	C1 LL	C2 AIC	C2 LL	C3 AIC	C3 LL
Gaussian	27450	-13720	27140	-13570	8859	-4428
Gamma	26010	-13010	25940	-12970	7523	-3759
Logistic	27960	-13980	27620	-13810	9298	-4647
Rayleigh	29910	-14950	29130	-14570	11250	-5624
KDE	**21900**	**-12100**	**22180**	**-11980**	**6864**	**-3310**

Note: The best information criterion at each location is in boldface.

C. Deterministic Forecasting Results

Standard metrics of root mean squared error (RMSE), mean absolute error (MAE), and their corresponding normalized indices, *i.e.*, NMAE and NRMSE, are adopted to evaluate the deterministic forecasting performance. For these metrics, a smaller value indicates better performance. Deterministic forecasting errors using Q-learning at the selected locations are summarized in Table III. It is shown that the 1HA forecasting NMAE and NRMSE are in the ranges of 6%-7% and 10%-11%, respectively. The persistence method (PS) is used as a baseline. The percentage improvement of NMAE and NRMSE by the Q-learning model are in the ranges of 3%-4% and 5%-7%, respectively, over the PS benchmark method.

TABLE III: 1HA deterministic forecasting performance using Q-learning and PS methods

Method	Site	C1	C2	C3
Q-learning	NMAE(%)	6.70	6.74	6.85
	NRMSE(%)	11.00	10.66	10.72
PS	NMAE(%)	6.93	7.03	7.11
	NRMSE(%)	11.76	11.43	11.39

D. Probabilistic Forecasting Results

With the KDE-based wind power marginal distributions, a joint distribution between the aggregated wind power (of all the farms) and the forecasted power of each member is determined. The conditional distribution of the aggregated wind power given forecasts of each member can be calculated. This conditional distribution is used to generate aggregated wind power forecasting scenarios. To better visualize probabilistic forecasts, a large number of wind power forecasting scenarios (i.e., 5000) is shown in Fig. 4 from 2012-07-05 to 2012-07-09. It is observed that for all the representative periods, the aggregated wind power reasonably lies in the scenario intervals.

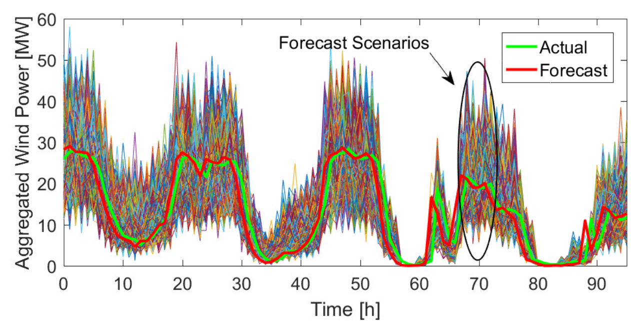

Fig. 4: Aggregated probabilistic wind power forecasts

1) Pinball Loss: Pinball loss is a widely used metric to evaluate the overall performance of probabilistic forecasts, which is defined by:

$$L_{m,t}(q_{m,t},p_t) = \begin{cases} (1-\dfrac{m}{100}) \times (q_{m,t}-p_t), & p_t < q_{m,t} \\ \dfrac{m}{100} \times (p_t - q_{m,t}), & p_t \geq q_{m,t} \end{cases} \tag{9}$$

where $q_{m,t}$ represents the mth quantile at time t. A smaller pinball loss value indicates better forecasting performance. Numerical summation (NS) of the quantiles from individual wind farms is used as the baseline method for comparison. The quantile forecasts of individual wind farms are generated by quantile regression. Results show that the normalized pinball loss of the developed aggregated method by considering spatio-temporal correlation (i.e., **4.04**) is smaller than the NS from quantile regression (i.e., **5.42**).

2) Reliability: Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies [13]:

$$RE = \left[\frac{\xi^{(1-\alpha)}}{N} - (1-\alpha) \right] \times 100\% \tag{10}$$

where N is the number of test samples, and $\xi^{(1-\alpha)}$ is the number of times that the actual test samples lie within the αth prediction interval. A reliability plot shows whether a given method tends to systematically underestimate or overestimate the uncertainty. In this study, the nominal coverage rates range from 10% to 90% with a 10% increment. The blue curve in Fig. 5 shows the reliability of the aggregated probabilistic forecasts. A forecast presents better reliability when the curve is closer to the diagonal. Overall the reliability of the developed method is better than the baseline NS method, especially over the 90th confidence interval, which is important in power system operations.

3) Sharpness: Sharpness indicates the capacity of a forecasting system to forecast extreme probabilities [13]. This criterion evaluates the predictions independently of the observations, which gives an indication of the level of usefulness of the predictions. For example, a system that provides only uniformly distributed predictions is less useful for decision-making under uncertainty. Predictions with perfect sharpness are discrete predictions with a probability of one (i.e., deterministic predictions). The sharpness is measured by the average size of the predictive intervals. The sharpness of the aggregated probabilistic forecasting is shown by the orange curve in Fig. 5. Overall, the sharpness of the developed method ranges from 2% to 45%, which is smaller than the sharpness of the the baseline NS method.

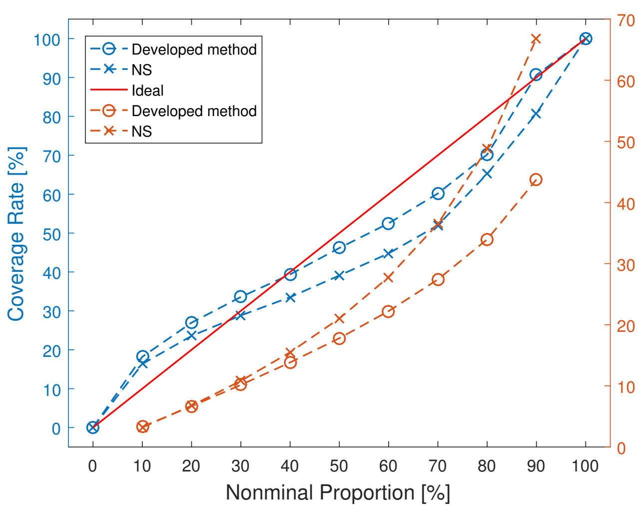

Fig. 5: Reliability and sharpness of the aggregated probabilistic wind power forecasts and baseline NS forecasts

IV. Conclusion

In this paper, a spatio-temporal correlation based aggregated probabilistic wind power forecasting method was developed, in conjunction with a Q-learning based deterministic forecasting method. The Copula method was used to build the spatio-temporal related joint distribution of the aggregated wind power and the forecasted power of each member wind farm. Different shapes of marginal distributions of wind power were tested and compared, including Gaussian, Gamma, Logistic, Rayleigh, and KDE distributions. The conditional distribution was deduced and used for generating aggregated probabilistic wind power forecasts. We found that the KDE distribution present the best marginal distribution shape. Results of the case study showed that the developed aggregated probabilistic forecasting method could provide reliable probabilistic forecasts by utilizing spatio-temporal correlations among individual wind farms. Different mixed models will be explored in future work to simulate the marginal wind power distribution.

V. Acknowledgement

This work was supported by the National Renewable Energy Laboratory under Subcontract No. XAT-8-82151-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308).

References

[1] H. Louie, "Correlation and statistical characteristics of aggregate wind power in large transcontinental systems," *Wind Energy*, vol. 17, no. 6, pp. 793–810, 2014.

[2] A. Malvaldi, S. Weiss, D. Infield, J. Browell, P. Leahy, and A. M. Foley, "A spatial and temporal correlation analysis of aggregate wind power in an ideally interconnected europe," *Wind Energy*, vol. 20, no. 8, pp. 1315–1329, 2017.

[3] A. Lenzi, I. Steinsland, and P. Pinson, "Benefits of spatiotemporal modeling for short-term wind power forecasting at both individual and aggregated levels," *Environmetrics*, vol. 29, no. 3, p. e2493, 2018.

[4] M. He, L. Yang, J. Zhang, and V. Vittal, "A spatio-temporal analysis approach for short-term forecast of wind farm generation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1611–1622, 2014.

[5] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, "Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 511–520, 2014.

[6] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 763–770, 2016.

[7] Q. Zhu, J. Chen, L. Zhu, X. Duan, and Y. Liu, "Wind speed prediction with spatio–temporal correlation: A deep learning approach," *Energies*, vol. 11, no. 4, p. 705, 2018.

[8] X. G. Agoua, R. Girard, and G. Kariniotakis, "Probabilistic model for spatio-temporal photovoltaic power forecasting," *IEEE Transactions on Sustainable Energy*, 2018.

[9] L. Rüschendorf, "On the distributional transform, sklar's theorem, and the empirical copula process," *Journal of Statistical Planning and Inference*, vol. 139, no. 11, pp. 3921–3927, 2009.

[10] C. Feng and J. Zhang, "Reinforcement learning based dynamic model selection for short-term load forecasting," *arXiv preprint arXiv:1811.01846*, 2018.

[11] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.

[12] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[13] M. Sun, C. Feng, J. Zhang, E. K. Chartan, and B.-M. Hodge, "Probabilistic short-term wind forecasting based on pinball loss optimization," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 2018, pp. 1–6.