



## Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting

Zhao, Yongning ; Ye, Lin ; Pinson, Pierre; Tang, Yong

*Published in:*  
IEEE Transactions on Power Systems

*Link to article, DOI:*  
[10.1109/TPWRS.2018.2794450](https://doi.org/10.1109/TPWRS.2018.2794450)

*Publication date:*  
2018

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Zhao, Y., Ye, L., Pinson, P., & Tang, Y. (2018). Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. *IEEE Transactions on Power Systems*, 33(5), 5029 - 5040. <https://doi.org/10.1109/TPWRS.2018.2794450>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Correlation-Constrained and Sparsity-Controlled Vector Autoregressive Model for Spatio-Temporal Wind Power Forecasting

Yongning Zhao, *Student Member, IEEE*, Lin Ye, *Senior Member, IEEE*, Pierre Pinson, *Senior Member, IEEE*, and Yong Tang, *Senior Member, IEEE*

**Abstract**—The ever-increasing number of wind farms has brought both challenges and opportunities in the development of wind power forecasting techniques to take advantage of interdependencies between tens or hundreds of spatially distributed wind farms, e.g., over a region. In this paper, a Sparsity-Controlled Vector Autoregressive (SC-VAR) model is introduced to obtain sparse model structures in a spatio-temporal wind power forecasting framework by reformulating the original VAR model into a constrained Mixed Integer Non-Linear Programming (MINLP) problem. It allows controlling the sparsity of the coefficient matrices in direct manner. However this original SC-VAR is difficult to implement due to its complicated constraints and the lack of guidelines for setting its parameters. To reduce the complexity of this MINLP and to make it possible to incorporate prior expert knowledge to benefit model building and forecasting, the original SC-VAR is modified and a Correlation-Constrained SC-VAR (CCSC-VAR) is proposed based on spatial correlation information about wind farms. Our approach is evaluated based on a case study of very-short-term forecasting for 25 wind farms in Denmark. Comparison is performed with a set of traditional local methods and spatio-temporal methods. The results obtained show the proposed CCSC-VAR has better overall performance than both the original SC-VAR and other benchmark methods, taking into account all evaluation indicators, including sparsity-control ability, sparsity, accuracy and efficiency.

**Index Terms**—Wind power, power system operations, forecasting, spatial correlation, sparsity.

## I. INTRODUCTION

WITH the rapid development and large-scale integration of wind power, both stability and reliability of conventional power system operation will be seriously challenged [1], [2]. Wind power forecasting (WPF) has been playing an increasingly important role in helping power system operators and market operators to schedule and trade wind power generation at various spatial and temporal scales [3], [4]. Extensive

This work is supported by the National Natural Science Foundation of China (51477174 and 51677188), the National Key Research & Development Program of China under 2017YFB0902200, the National Natural Science Foundation of China-UK Royal Society International Cooperation & Exchange Program (NSFC-51711530227), the Open Fund of State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, and the China Scholarship Council as well as the Danish Strategic Research Council under the project ‘5s - Future Electricity Markets’ (12-132636/DSF) and CITIES (DSF-1305-00027B).

Y. Zhao and L. Ye are with the Department of Information and Electrical Engineering, China Agricultural University, Beijing 100083, P.R. China (e-mail: zyn@cau.edu.cn; yelin@cau.edu.cn).

P. Pinson is with the Centre for Electric Power and Energy, Technical University of Denmark, Kgs. Lyngby, Denmark (e-mail: ppin@elektro.dtu.dk).

Y. Tang is with China Electric Power Research Institute (CEPRI), Beijing 100192, P.R. China (e-mail: tangyong@epri.sgcc.com.cn).

research work has been conducted over the past few decades to develop efficient and accurate forecasting techniques. A review of state-of-the-art WPF methods can be found in, e.g., [5-7].

Most of the existing WPF methods are optimized for every location individually, using only local on-site information (e.g. wind power measurements, weather forecasts), while overlooking existing spatio-temporal interdependence structures in wind power generation fields [8], [9]. In fact, not only does wind power generation at each site exhibit auto-correlation, but power generation at different wind farms also shows spatial cross-correlation, possibly with a time lag. Therefore, in a region covering tens or even hundreds of wind farms, there might be a lot to gain by better understanding dynamic space-time dependencies among wind farms. Eventually, this may yield a deeper understanding of wind farm dynamics, improved accuracy of wind power forecasts, as well as better models that could be seamlessly used for power system operation and market participation.

Spatial correlation was initially primarily studied and used to simulate wind speed and power time-series by first characterizing the dependencies between two sites or among a group of sites, in both time [10] and frequency domain [11]. The interdependence structure of wind power generation was also investigated in [8] and [12] in terms of WPF errors. The underlying correlation patterns were found in spatio-temporal propagation of forecasting errors, from which probabilistic WPF can be improved [13], [14].

This new insight has motivated researchers to concentrate on WPF that takes more account of spatio-temporal correlations. Different machine learning methods [15-17] were proposed to improve wind speed and power forecasting, where wind speed or power at a target site was predicted based on observations at neighboring sites. Furthermore, regime-switching space-time methods [18-20], multichannel adaptive filters [21] and graph-learning methods [22] are also among approaches proposed for spatio-temporal WPF. In the [Global Energy Forecasting Competition 2014](#), three teams [23-25] ranked in Top 5 of the wind track also used off-site information in probabilistic WPF to improve their forecasting accuracy.

The above spatial methods for WPF generally try to utilize data from all the available neighboring sites. This could be applicable when the number of wind farms is small. However, considering scaling to a large number of wind farms, as is the case in practice where hundreds or even thousands of wind farms are installed over a region, such methods appear unde-

1 **sirable. The excessive use of information from numerous wind**  
 2 **farms may cause over-fitting problems as well as increased**  
 3 **computational burden. Thus, it is necessary to explore low-**  
 4 **dimensional structures to characterize the interactions among**  
 5 **a large number of wind farms [26].**

6 To this end, sparse statistical modeling has attracted much  
 7 attention, where sparsity is to be understood as handling large  
 8 models, though with a large share of coefficients being zeros.  
 9 A sparse model is generally preferable since it can avoid over-  
 10 fitting, while its interpretation is simpler and more intuitive  
 11 than a dense one [27]. Alternative proposals can be found  
 12 for high-dimensional spatio-temporal WPF problems. Among  
 13 others, the compressive sensing and structured-sparse recovery  
 14 algorithms [26], [28], dictionary learning method [29] and  
 15 two-stage sparse vector autoregressive (VAR) model [30] have  
 16 achieved results with reasonably sparse structures. However,  
 17 these methods can only provide overall and fully data-driven  
 18 sparse structures, while forecasters and forecast users may be  
 19 interested in controlling sparsity in a finer way, e.g., by using  
 20 knowledge on space time wind dynamics and layout of wind  
 21 farms as natural constraints on sparsity.

22 In practice, if a user has prior expert knowledge about the  
 23 spatio-temporal relationships between wind farms that can  
 24 improve forecasting, then one may want to finely control  
 25 sparsity in a model structure based on such expert knowledge.  
 26 In addition, there may also be special restrictions on the  
 27 forecasting model. For example, it is difficult to use too many  
 28 wind farms to forecast the target wind farm for some technical  
 29 or financial reason, or historical data from other wind farms is  
 30 not applicable for the target wind farm, or wind farms refuse  
 31 to share their data due to privacy reasons [31].

32 Such expert knowledge or special restrictions can involve  
 33 different aspects, e.g., the number of correlated wind farms to  
 34 explain the target wind farm, the number of past observations  
 35 at each correlated wind farms to explain the target wind farm,  
 36 as well as the overall sparsity of the coefficients. These aspects  
 37 can be used to control the sparse structure of the solution as  
 38 needed. It is not easy to achieve these aspects by using existing  
 39 sparse modeling techniques, as they are generally black-box  
 40 models in terms of sparsity-control. Even though Cavalcante  
 41 et al. [32] explored a set of different sparse structures for  
 42 VAR models based on the Least Absolute Shrinkage and  
 43 Selection Operator (LASSO), this proposal requires setting up  
 44 a new, separate model for each specific sparse structure, and  
 45 this cannot really meet the requirement to freely control the  
 46 sparsity.

47 In this paper, a sparsity-controlled VAR (SC-VAR) model  
 48 [33] is introduced for very-short-term spatio-temporal WPF.  
 49 It reformulates the estimation problem for a VAR model as a  
 50 constrained mixed integer non-linear programming (MINLP)  
 51 problem, which can then be readily solved with a standard  
 52 optimization solver. Sparsity can be explicitly and flexibly  
 53 controlled by setting various parameters in the constraints.  
 54 However the SC-VAR is difficult to implement due to its  
 55 complicated constraints and the lack of guidelines for setting  
 56 its parameters. To reduce the complexity of MINLP, and to  
 57 enable incorporation of prior expert knowledge to benefit the  
 58 forecasting, the original SC-VAR is modified and a correlation-

constrained SC-VAR (CCSC-VAR) is proposed based on the  
 information on spatial correlation between wind farms. The  
 proposed method is compared with different local and spatio-  
 temporal methods using data from 25 wind farms in Denmark.

The paper is structured as following. In Section II, the SC-  
 VAR and CCSC-VAR models for WPF are outlined. Section  
 III introduces a set of benchmark methods and performance  
 metrics. In Section IV, a case study of 25 wind farms in  
 Denmark is carried out to test the proposed methods. Section  
 V concludes the work and results.

## II. THE SPARSITY-CONTROLLED VECTOR AUTOREGRESSIVE MODELS

### A. The VAR Model

A VAR model can capture the dynamic interrelationships  
 between power generation observed at a number of spatially  
 distributed wind farms. Suppose  $y_{i,t}$  is the power generation at  
 wind farm  $i$  and time  $t$ . Power measured at  $N$  wind farms can  
 be integrated into a  $N$ -dimensional vector-valued time series  
 $\{y_t\} \in \mathbb{R}^N$ . To describe  $y_{i,t}$  based on past observations, the  
 VAR process of order  $p$ ,  $\text{VAR}(p)$  is expressed as

$$y_{i,t} = \sum_{j=1}^N \sum_{k=1}^p \alpha_{jk}^i y_{j,t-k} + \varepsilon_{i,t} \quad (1)$$

where  $N$  is the dimension of the vector time series, in this  
 case the number of wind farms that are spatially distributed  
 in a region,  $k = 1, 2, \dots, p$  are time lags,  $\varepsilon_{i,t}$  is zero-mean  
 Gaussian noise of time series  $y_{i,t}$ , and  $\alpha_{jk}^i$  is the autoregressive  
 coefficient that characterizes the contribution of  $y_{j,t-k}$  to  $y_{i,t}$ .  
 The modeling of VAR is to find optimal  $\alpha_{jk}^i$  that can best  
 explain  $y_{i,t}$ .

The VAR in more compact vector form is written as:

$$y_t = \sum_{k=1}^p A_k y_{t-k} + \varepsilon_t \quad (2)$$

where  $A_k = \{\alpha_{jk}^i : i, j = 1, 2, \dots, N\} \in \mathbb{R}^{N \times N}$  is the  
 coefficient matrix for time lag  $k$ , and  $\varepsilon_t \in \mathbb{R}^N$  is the vector of  
 zero-mean Gaussian noise with non-singular covariance matrix  
 for  $N$  wind farms at time  $t$ .

### B. The SC-VAR Model

As it is the most basic multivariate time series model, the  
 VAR is not expected to produce a sparse coefficient structure  
 by itself. Although there is literature devoted to overall sparsity  
 control and different sparse structures for WPF, none of it  
 allows fine-tuning of different aspects of sparsity. In fact, in  
 case of possessing specific information about the data used for  
 VAR modeling or showing special preferences regarding the  
 sparse structure, one may want to control the sparsity of the  
 forecasting model as needed. Some typical aspects of sparsity  
 that can be controlled include:

- 1) The overall number of non-zero coefficients of VAR  
 ( $S_A$ ).
- 2) The number of explanatory wind farms used in VAR to  
 explain the target wind farm  $i$  ( $S_F^i$ ).

- 3) The number of past observations of each explanatory wind farm used to explain the target wind farm  $i$  ( $S_P^i$ ).
- 4) The number of nonzero coefficients used to explain the target wind farm  $i$  ( $S_N^i$ ).

Following [33], by introducing binary variables and adding constraints on the coefficients of VAR, the original VAR can be reformulated as a MINLP problem, which is expressed as

$$\min_{\alpha, \delta, \gamma} \sum_{i=1}^N \sum_{t=p}^T \left( y_{i,t+1} - \sum_{j=1}^N \sum_{k=1}^p \alpha_{jk}^i y_{j,t-k+1} \right)^2 \quad (3a)$$

$$\text{subject to } \delta_{jk}^i \leq \gamma_j^i, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (3b)$$

$$\sum_{j=1}^N \gamma_j^i \leq S_P^i, \forall i \in \mathbf{I} \quad (3c)$$

$$\sum_{k=1}^p \gamma_j^i \delta_{jk}^i \leq S_P^i, \forall i, j \in \mathbf{I} \quad (3d)$$

$$\sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^p \delta_{jk}^i \leq S_A \quad (3e)$$

$$\sum_{j=1}^N \sum_{k=1}^p \delta_{jk}^i \leq S_N^i, \forall i \in \mathbf{I} \quad (3f)$$

$$|\alpha_{jk}^i| \geq \eta_j^i \delta_{jk}^i, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (3g)$$

$$\alpha_{jk}^i (1 - \delta_{jk}^i) = 0, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (3h)$$

$$\delta_{jk}^i, \gamma_j^i \in \{0, 1\}, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (3i)$$

where  $T$  is the length of time series for model training,  $\mathbf{K} = \{1, 2, \dots, p\}$ ,  $\mathbf{I} = \{1, 2, \dots, N\}$ , binary control variable  $\delta_{jk}^i$  indicates whether the coefficient  $\alpha_{jk}^i$  is zero or not, binary control variable  $\gamma_j^i$  indicates whether explanatory wind farm  $j$  can be used to explain target wind farm  $i$ , and threshold  $\eta_j^i$  is the lower bound of  $|\alpha_{jk}^i|$ .

This MINLP optimizes the regression coefficients  $\alpha_{jk}^i$  and binary control variables  $\delta_{jk}^i$  and  $\gamma_j^i$  by minimizing the sum of squared errors. The meaning of each constraint of the SC-VAR is as follows:

- 1) Constraint (3b) forces the  $\gamma_j^i$  to take the value 1 when some  $\delta_{jk}^i$  takes the value 1, i.e. as long as there exists non-zero  $\alpha_{jk}^i$  for at least one lag  $k$ .
- 2) Constraint (3c) limits the number of explanatory wind farms used to explain target wind farm  $i$ .
- 3) Constraint (3d) limits the number of past observations of each explanatory wind farm used to explain target wind farm  $i$ .
- 4) Constraint (3e) limits the overall number of non-zero coefficients of VAR.
- 5) Constraint (3f) limits the number of non-zero coefficients used to explain target wind farm  $i$ .
- 6) Constraint (3g) confines  $|\alpha_{jk}^i|$  to the interval  $[\eta_j^i, +\infty]$  and otherwise assigns zero to any coefficient whose absolute value is less than  $\eta_j^i$ .
- 7) Constraints (3h) links the control variable  $\delta_{jk}^i$  to coefficient  $\alpha_{jk}^i$  so that  $\delta_{jk}^i$  can determine whether  $\alpha_{jk}^i$  is zero or not.

Although the SC-VAR model allows full control of the sparsity, it can be very difficult to implement it in practice.

First, the SC-VAR allows sparsity-control but does not indicate how to control. This requires elaborately setting many parameters, including  $S_P^i$ ,  $S_N^i$  and  $S_A$  for each wind farm  $i$ ,  $\eta_j^i$  for each pair of wind farms and  $S_A$ . This is impracticable when dealing with high dimensional WPF, especially if very limited guidelines are available for setting these parameters. Secondly, in addition to the regression coefficients, the SC-VAR model introduces many binary variables that need to be optimized, i.e.  $\delta_{jk}^i$  for each pair of wind farms at each lag, and  $\gamma_j^i$  for each pair of wind farms, and these are the main causes of computation burden. Finally, as constraint (3e) controls the overall sparsity, all of the coefficients in the objective function (3a) have to be optimized together, which can slow down the computation. If (3e) is removed from the model, the objective function is decomposable so that it can be solved by solving separately the optimization problem for each wind farm  $i$ , considering only the constraints related to wind farm  $i$ . However, removing (3e) will make the SC-VAR lose its global control of sparsity. Even so, the global sparsity can be appropriately controlled via the local control of each wind farm's sparsity using constraints (3f). Therefore, it is worth removing the constraint (3e) as a result of balancing between the computation efficiency and the sparsity-control ability. Also note that, in controlling the total number of non-zero coefficients of a target wind farm, the combination of constraints (3c) and (3d) can serve a similar function to that of the constraint (3f). Thus, based on the above analysis, the constraints (3d) and (3e) can be removed while (3c) and (3f) can be retained to control the sparsity in a computationally efficient way by sacrificing some control flexibility.

Furthermore, as the constraints (3g) and (3h) are non-linear, a reformulation is provided in [33] to linearize them by replacing  $\delta_{jk}^i$  with  $\nu_{jk}^{i+} + \nu_{jk}^{i-}$ . Then (3g) and (3h) can be linearized as  $\alpha_{jk}^i \geq \eta_j^i \nu_{jk}^{i+} - \nu_{jk}^{i-} M$  and  $\alpha_{jk}^i \leq -\eta_j^i \nu_{jk}^{i-} + \nu_{jk}^{i+} M$ , where  $M$  is assumed to be a large positive constant number. But unfortunately many more auxiliary integer variables have to be introduced, which will further make the SC-VAR computationally heavier.

### C. The CCSC-VAR Model

To address the problems of the SC-VAR, in this section the CCSC-VAR model is proposed and formulated as:

$$\min_{\alpha, \delta} \sum_{i=1}^N \sum_{t=p}^T \left( y_{i,t+1} - \sum_{j=1}^N \sum_{k=1}^p \alpha_{jk}^i y_{j,t-k+1} \right)^2 \quad (4a)$$

$$\text{subject to } \delta_{jk}^i \leq \lambda_j^i, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (4b)$$

$$\sum_{k=1}^p \delta_{jk}^i \geq \lambda_j^i, \forall i, j \in \mathbf{I} \quad (4c)$$

$$\sum_{j=1}^N \sum_{k=1}^p \delta_{jk}^i \leq S_N^i, \forall i \in \mathbf{I} \quad (4d)$$

$$|\alpha_{jk}^i| \leq M \cdot \delta_{jk}^i, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (4e)$$

$$\delta_{jk}^i, \lambda_j^i \in \{0, 1\}, \forall k \in \mathbf{K}, i, j \in \mathbf{I} \quad (4f)$$

where  $M$  is any large positive constant number,  $\Lambda_\tau = \{\lambda_j^i : i, j \in \mathbf{I}\} \in \mathbb{R}^{N \times N}$  with regard to parameter  $\tau$

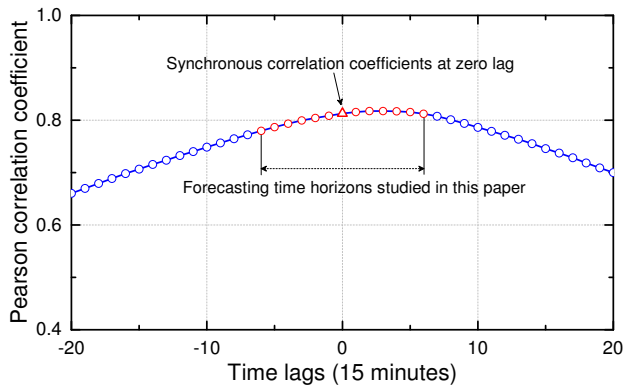


Fig. 1. The Pearson correlation coefficients between the wind power time series of two wind farms at different time lags

is the control matrix derived from cross-correlation matrix  $\Phi = \{\phi_j^i : i, j \in \mathbf{I}\} \in \mathbb{R}^{N \times N}$  of the wind farms. The  $\Phi$  can be obtained in different ways, including Pearson correlation, Kendall rank correlation and graphical modeling. It is found that the results of these correlation analysis approaches are quite close. However, the computation of Kendall rank correlation and graphical modeling is very time-consuming, so the Pearson correlation coefficient is selected in this paper and computed by

$$\phi_j^i = \frac{\sum_{t=1}^T (y_{i,t} - \bar{y}_i)(y_{j,t} - \bar{y}_j)}{\sqrt{\sum_{t=1}^T (y_{i,t} - \bar{y}_i)^2} \sqrt{\sum_{t=1}^T (y_{j,t} - \bar{y}_j)^2}} \quad (5)$$

where  $\bar{y}_i$  and  $\bar{y}_j$  are the average values of  $T$  samples for wind farm  $i$  and  $j$  respectively.

Then, the entries of  $\Lambda_\tau$  are obtained by

$$\lambda_j^i = \begin{cases} 1, & \phi_j^i \geq \tau \\ 0, & \phi_j^i < \tau \end{cases} \quad (6)$$

where  $\tau$  is a predefined threshold. Like the  $\gamma_j^i$  in the SC-VAR,  $\lambda_j^i$  also indicates whether wind farm  $j$  contributes to wind farm  $i$ . However,  $\gamma_j^i$  is a decision variable that needs to be optimized while  $\lambda_j^i$  is an adjustable parameter related to historical data.

Here only the zero-lag correlation coefficient computed by using synchronous time series is considered in generating the control matrix  $\Lambda_\tau$ , because a correlation analysis of the used data shows that the difference among the correlation coefficients at different time lags (each lag is 15 minutes) is not significant in the very-short-term horizons studied in this paper, as shown by an example in Fig. 1. But for longer forecasting time horizons, the correlation coefficients at corresponding lags are suggested to be considered in generating the  $\Lambda_\tau$ .

In the CCSC-VAR model, the implications of the constraints are as follows:

- 1) Constraint (4b): If  $\lambda_j^i = 0$ , the coefficients  $\alpha_{jk}^i$  between  $i$  and  $j$  for all  $p$  time lags are zeros.
- 2) Constraint (4c): If  $\lambda_j^i = 1$ , there should be at least one non-zero coefficient  $\alpha_{jk}^i$  between wind farm  $i$  and  $j$ , i.e., the  $\alpha_{jk}^i$  for at least one lag  $k$  should be non-zero.

- 3) Constraint (4d): It is same as constraint (3f). It is retained from SC-VAR and used together with  $\Lambda_\tau$  to control the sparsity.
- 4) Constraint (4e): The binary control variable  $\delta_{jk}^i$  is linked to  $\alpha_{jk}^i$ , so that  $\delta_{jk}^i$  can determine whether  $\alpha_{jk}^i$  is zero or not, i.e.,

$$|\alpha_{jk}^i| \leq M \cdot \delta_{jk}^i \Leftrightarrow \begin{cases} -M \leq \alpha_{jk}^i \leq M, \delta_{jk}^i = 1 \\ \alpha_{jk}^i = 0, \delta_{jk}^i = 0 \end{cases} \quad (7)$$

Although  $M$  is allowed to take any large positive constant number, it is found that in this work the maximal regression coefficients for all the wind farms are around 1. Therefore, the value of  $M$  is set as 2.

There are two parameters governing the sparsity of the CCSC-VAR, i.e.  $\tau$  and  $S_N^i$ . The sparsity of the CCSC-VAR is directly affected by the sparsity of  $\Lambda_\tau$ , while the sparsity of  $\Lambda_\tau$  is determined by  $\Phi$  and  $\tau$ . So  $\tau$  can serve a similar function as  $S_F^i$  of the SC-VAR in controlling the number of explanatory wind farms. In addition, the CCSC-VAR with both  $S_N^i$  and  $\tau$  can be more flexible in controlling sparsity than the CCSC-VAR without  $S_N^i$ .

Compared with the SC-VAR, the CCSC-VAR has fewer constraints. Also, the compact CCSC-VAR model has many fewer parameters that need to be tuned, but meanwhile it still preserves the ability to control the sparsity. Moreover, when using the transformation in (7), all constraints of the CCSC-VAR are linear without resorting to auxiliary variables like those in the SC-VAR model. Furthermore, as the decision variables  $\gamma_j^i$  are replaced by adjustable parameters  $\lambda_j^i$ , problem solving will be more efficient due to the substantially reduced number of decision variables. Therefore, the CCSC-VAR is much easier to implement than the SC-VAR.

Another significant advantage of the CCSC-VAR over the SC-VAR is that the CCSC-VAR enables incorporation of expert knowledge about the spatio-temporal correlations of the wind farms into the model building. As the  $\Phi$  is deterministic and unique for a specific set of  $N$  wind farms, controlling the sparsity of the CCSC-VAR via tuning  $\tau$  is in fact based on real data information and thus can reduce the uncertainty in sparsity control, which makes the model more capable of characterizing the true relationships between wind farms.

It should be noted that the binary control matrix  $\Lambda_\tau$  used in this paper is just a concrete example of expert knowledge which is derived from basic cross-correlation matrix. However, even such simple expert knowledge can improve the forecasting performance, as will be demonstrated in the case study. Therefore, it is expected that the forecasting accuracy can be further improved by using other specifically designed control matrices. In particular, the control matrix allows element-wise modification when special restrictions are imposed. For example, if historical data of a specific wind farm is not available for some technical reasons or if a wind farm refuses to share its data for privacy reasons, then the entries of the control matrix corresponding to this wind farm can be directly set as zero. However, the special restrictions may contradict the expert knowledge. This means an unavailable wind farm

with regard to the special restriction could be the one preferred by expert knowledge. Therefore, this case should be carefully considered in practice.

#### D. The Decomposition of Sparsity-Controlled Models

Since the objective function and each constraint are decomposable, the CCSC-VAR can be solved separately for each wind farm  $i$ , which is expressed as

$$\min_{\alpha, \delta} \sum_{t=p}^T \left( y_{i,t+1} - \sum_{j=1}^N \sum_{k=1}^p \alpha_{jk}^i y_{j,t-k+1} \right)^2 \quad (8a)$$

$$\text{subject to } \delta_{jk}^i \leq \lambda_j^i, \forall k \in \mathbf{K}, j \in \mathbf{I} \quad (8b)$$

$$\sum_{k=1}^p \delta_{jk}^i \geq \lambda_j^i, \forall j \in \mathbf{I} \quad (8c)$$

$$\sum_{j=1}^N \sum_{k=1}^p \delta_{jk}^i \leq S_N^i \quad (8d)$$

$$|\alpha_{jk}^i| \leq M \cdot \delta_{jk}^i, \forall k \in \mathbf{K}, j \in \mathbf{I} \quad (8e)$$

$$\delta_{jk}^i, \gamma_j^i \in \{0, 1\}, \forall k \in \mathbf{K}, j \in \mathbf{I} \quad (8f)$$

The SC-VAR model can also be decomposed in a similar way if the constraint (3e) is removed. By decomposing the original problem into smaller sub-problems, the dimension of the optimization is significantly reduced, which is very beneficial in high-dimensional spatio-temporal WPF.

In this paper, the decomposed sub-problems are solved in a sequential way on only one computer, i.e. the wind farms are solved one by one. Computational gain will be more significant if they are implemented in the framework of parallel computing or distributed computing, which can concurrently solve the sub-problems. The decomposition enables the forecasting models to be deployed at both wind farm level and the power system operator level and in both centralized and distributed ways. For example, if each wind farm is responsible for solving the forecasting problem for itself, then all the wind farms can solve their own problems respectively and simultaneously in a distributed way.

### III. BENCHMARK METHODS AND PERFORMANCE METRICS

Two classes of forecasting methods are implemented to evaluate the performance of the SC-VAR and the CCSC-VAR: local methods and spatio-temporal methods. As has been discussed, a local method employs only local information to obtain local forecast. Individual local models need to be built for each wind farm. In contrast, the spatio-temporal methods provide an integrated or unified form of model for all wind farms and make use of spatial information to improve forecasting.

#### A. Local Forecasting Methods

1) *Persistence Method*: The persistence method is commonly used as the most basic predictor to benchmark other methods. In this method, the forecast for all times ahead is treated as the present value. The  $k$ -step-ahead persistence method is

$$\hat{y}_{t+k|t} = y_t \quad (9)$$

2) *Autoregressive (AR) Model*: AR is the one-dimensional version of VAR model by setting  $N$  as 1. It models the future wind power generation using only the local wind power time series based on the temporal auto-correlation characteristics.

#### B. Spatio-Temporal Forecasting Methods

1) *LASSO-VAR*: The LASSO-VAR is a multivariate time series forecasting model that is widely used to explore sparse structure and feature selection. The LASSO-VAR in separated form of an individual wind farm  $i$  is expressed by

$$\min_{\mathbf{a}_i} \left[ \sum_{t=p}^T (y_{i,t+1} - \mathbf{a}_i^\top \mathbf{Y}_t)^2 + \mu \|\mathbf{a}_i\|_1 \right] \quad (10)$$

where  $\mathbf{a}_i = (\mathbf{A}_1^i, \mathbf{A}_2^i, \dots, \mathbf{A}_p^i)^\top \in \mathbb{R}^{Np \times 1}$  is coefficients vector of wind farm  $i$  with  $\mathbf{A}_k^i \in \mathbb{R}^{1 \times N}$  being the  $i$ th row of  $\mathbf{A}_k$ ,  $\mathbf{Y}_t = (\mathbf{y}_t^\top, \mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-p+1}^\top)^\top \in \mathbb{R}^{Np \times 1}$ ,  $\|\cdot\|_1$  is the  $\ell_1$ -norm of a vector, and  $\mu$  is a parameter that determines the balance between the estimation error and the degree of sparsity of the solution [34].

According to (10), the LASSO-VAR is a black-box model in terms of sparsity-control. It can only control the overall sparsity of the VAR of wind farm  $i$  by tuning the value of  $\mu$ . The greater the  $\mu$  is, the sparser the coefficients matrix will be. When  $\mu = 0$ , the LASSO-VAR will degenerate to standard VAR. See [32] for details about the LASSO-VAR in WPF.

2) *Simplified SC-VAR model*: As has been mentioned, the complete SC-VAR defined by (3a)-(3i) is very difficult to implement due to its complicated constraints and parameter settings. So a simplified version of the SC-VAR is used for comparison in the cased study. In the simplified SC-VAR, the constraints (3d) and (3e) are removed and all  $\eta_j^i$  are set as 0. As the parameter  $\tau$  of the CCSC-VAR and the parameter  $S_F^i$  of the simplified SC-VAR serve similar functions in controlling the number of wind farms, the CCSC-VAR and the simplified SC-VAR have similar parameter setting complexity and sparsity-control ability, which is fair to the model comparison. **In the following sections, unless otherwise specified, the SC-VAR refers to the simplified SC-VAR.**

#### C. Performance Metrics

For each wind farm  $i$ , the accuracy of the forecasting models are indicated by two metrics following [35]. These are Normalized Root Mean Square Error (NRMSE) and Normalized Mean Absolute Error (NMAE). The average accuracy of a forecasting model for the whole set of studied wind farms is indicated by average NRMSE of all wind farms.

Furthermore, for the LASSO-VAR, SC-VAR and CCSC-VAR, the model sparsity  $\Theta$  is additionally defined as the ratio between the number of zero coefficients and the overall number of coefficients:

$$\Theta = \frac{N_{zero}}{Np^2} \quad (11)$$

where  $N_{zero}$  is the total number of zero coefficients.

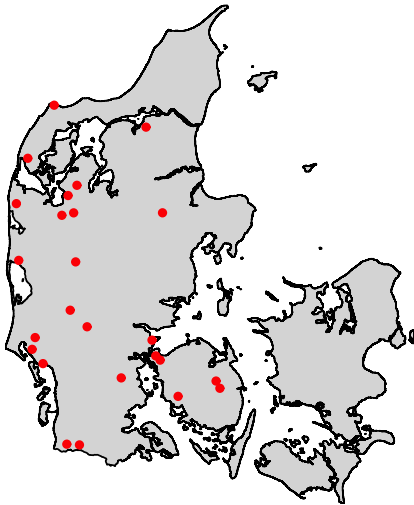


Fig. 2. The geographical distribution of 25 randomly selected wind farms in Denmark

#### IV. APPLICATION AND CASE STUDY

##### A. Data Preparation

The data used in this paper are wind power time series in 2006 from 25 randomly selected wind farms across Denmark. The geographical distribution of these wind farms is shown in Fig. 2. Each time series contains 35 040 consecutive data points with time resolution of 15 min. The dataset of each wind farm is divided into three consecutive parts, including 10 000 data points for training, 10 000 data points for validation (parameter optimization) and the remaining 15 040 data points for out-of-sample predictive performance testing. Each forecasting model is firstly trained using the training data with different parameters. Then the forecasting NRMSE corresponding to different parameters are obtained using validation data to select the optimized parameters for each forecasting model. The performance metrics of the forecasting models with their optimal parameters are finally compared using testing data. In the following sections, the NRMSE computed from the validation data is called validation NRMSE while the NRMSE and MAE computed from the testing data are called testing NRMSE and testing MAE. In this work, the forecasting methods are applied in very-short-term time horizons of up to 6 steps (i.e., 1.5 hours) ahead with each step being 15 minutes.

It is well known that the time series forecasting models are more capable of Gaussian processes. Thus, all the data are normalized to the range of [0,1] and then transformed to Gaussian time series using Logit transformation [36]:

$$y_G = \log\left(\frac{y}{1-y}\right), y \in (0, 1) \quad (12)$$

where  $y$  is the original data and  $y_G$  is the transformed Gaussian data. As the logit transformation requires that  $y \in (0, 1)$ , the data smaller than 0.01 are treated as 0.01 and the data greater than 0.99 are treated as 0.99.

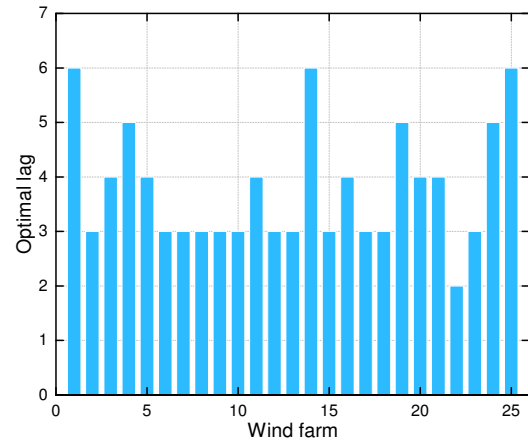


Fig. 3. The optimal order of AR model for each of the 25 wind farms

After finishing the forecasting, all forecasts are transformed back to the normalized range of (0,1) using the inverse transformation

$$y = \frac{\exp(y_G)}{1 + \exp(y_G)} \quad (13)$$

##### B. The Tools for Solving the Forecasting Models

All the forecasting methods are implemented on a 64-bit PC with 2.9GHz Intel Dual Core CPU and 4GB RAM.

Except for the SC-VAR and the CCSC-VAR, all other forecasting methods are implemented on the platform of Matlab R2013b (64-bit). The LASSO-VAR is solved by the Matlab package “Glmnet” using a Coordinate Descent algorithm [37]. Since the VAR is a special case of the LASSO-VAR without the  $\ell_1$ -norm constraint [27], the forecasting of the VAR is achieved by directly setting the shrinkage parameter  $\mu$  of the LASSO-VAR as zero.

The simplified SC-VAR and the CCSC-VAR are solved by the Gurobi solver with mathematical programming language AMPL. Other solvers such as CPLEX and other modeling languages such as Python can also easily deal with these MINLPs.

##### C. Parameter Settings

1) *The Order  $p$  of the Forecasting Models:* Partial autocorrelogram is an effective tool to determine the order of autoregressive models for time series. The partial autocorrelogram calculated from training data is used for each of the 25 wind farms to determine their optimal orders  $p$ , as shown in Fig. 3. The AR model for each wind farm is trained with its corresponding optimal order. However, a spatio-temporal model requires a unified order for all wind farms. It can be seen from Fig. 3 that the optimal  $p$  for most of the wind farms ranges from 3 to 6. Based on the optimal orders of individual wind farms, and considering the fairness of comparison, a moderate  $p = 4$  is configured for all the spatio-temporal methods, i.e., VAR, LASSO-VAR, SC-VAR and CCSC-VAR.

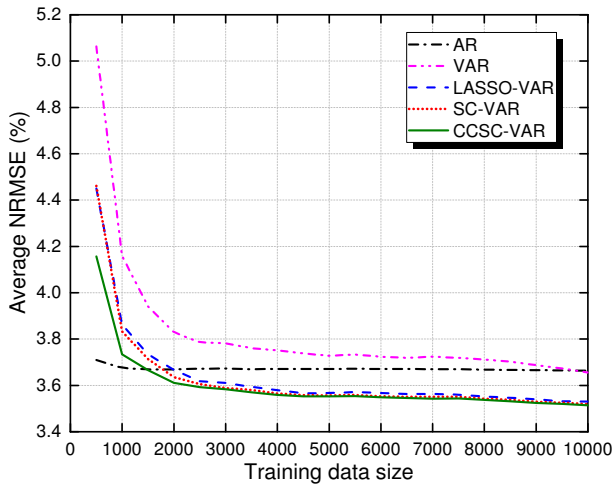


Fig. 4. The average validation NRMSE of 1-step-ahead forecasting by different models with varying training data size

2) *The Impact of Training Data Size on Accuracy:* The size of training data is a very important factor, with significant influence on the accuracy of a forecasting model. The average 1-step-ahead validation NRMSE of different forecasting methods with different training data sizes is given in Fig. 4. Note that the parameters used here are not optimal ones. The  $\mu$  of the LASSO-VAR is 0.001. The  $S_F^i$  and  $S_N^i$  of the SC-VAR are 15 and 20 respectively. The  $\tau$  and  $S_N^i$  of the CCSC-VAR are 0.75 and 30 respectively.

In Fig. 4, the NRMSE values for all the methods tend to decrease with training data size. However, after the data size reaches a specific value (e.g. 5000), the NRMSE decreases very slowly. The AR model is the most insensitive to the training data size. Its performance remains almost unchanged when the data size is greater than 1000. It also has the lowest NRMSE level among all these methods when the data is not sufficient, which indicates AR is a simple but powerful forecasting method. **The VAR model is very data hungry, which is consistent with the conclusion of [30]. It is the worst one in terms of both forecasting accuracy and convergence rate with data size. The NRMSE curves of all the sparse models decrease very quickly to a very low level and then varies closely and slowly when data is more than 4000. The LASSO-VAR and the SC-VAR always keep very close and show almost the same variation tendencies. Among the spatio-temporal methods, the CCSC-VAR significantly outperforms the others when the data size is extremely small (smaller than 1000). This is because the CCSC-VAR that highly dependent on the spatio-temporal correlation information can be effectively characterized even by using very small amounts of data, though sufficient data can be more beneficial. The NRMSE of these models may continue changing if the data size further increases, but it is fair to train all these forecasting models with 10000 data points according to the above analysis.**

3) *Parameters Tuning for Spatio-Temporal Forecasting Model:* This section will discuss in detail the parameters of spatio-temporal forecasting models, including the  $\mu$  of the LASSO-VAR, the  $S_N^i$  for both the SC-VAR and the CCSC-

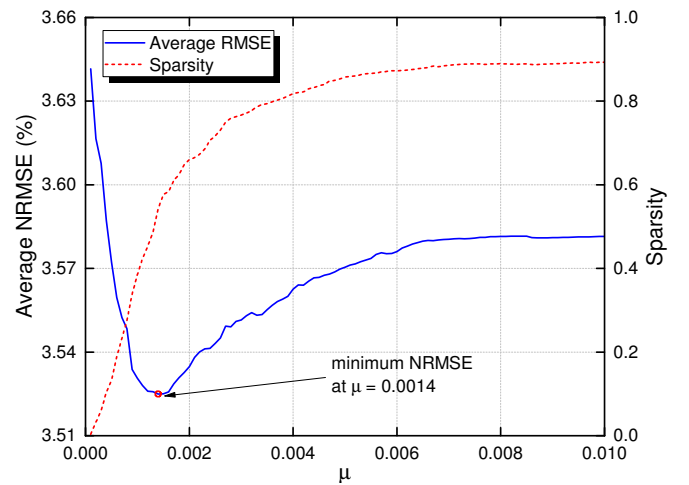


Fig. 5. The average validation NRMSE of 1-step-ahead forecasting and model sparsity with different  $\mu$  of LASSO-VAR

VAR, the  $S_F^i$  for the SC-VAR, and the  $\tau$  for the CCSC-VAR.

Notice that, according to the implications of constraints, the limit of effective  $S_N^i$  for the SC-VAR is  $S_N^i \leq p \cdot S_F^i$ . The impact of constraint (3f) on the SC-VAR will not change with  $S_N^i$  when  $S_F^i$  is fixed and  $S_N^i$  is greater than its upper limit. Similarly, the limit of effective  $S_N^i$  for the CCSC-VAR is  $\sum_{j=1}^N \lambda_j \leq S_N^i \leq p \cdot \sum_{j=1}^N \lambda_j$ , and the impact of constraint (4d) on the CCSC-VAR will not change with  $S_N^i$  when  $\tau$  is fixed and  $S_N^i$  is greater than its upper limit. On the contrary, if  $S_N^i$  is fixed and smaller than its upper limit, then the sparsity will not change with varying  $S_F^i$  or  $\tau$ , but the accuracy of the SC-VAR or the CCSC-VAR could always change with  $S_F^i$  or  $\tau$  respectively, no matter whether  $S_N^i$  is beyond its upper limit or not. In addition, for the CCSC-VAR, the  $S_N^i$  that is smaller than its lower limit will make the optimization problem infeasible due to the contradiction between constraints (4c) and (4d). The  $S_N^i$  and  $S_F^i$  for all wind farms will take the same values in this paper, though different values can be set for different wind farms.

The optimal parameters for the LASSO-VAR, SC-VAR and CCSC-VAR are determined by examining the 1-step-ahead forecasting results. The optimal parameters in other time horizons are set as the same as those in the 1-step horizon.

To investigate the impact of  $\mu$  on the performance of the LASSO-VAR, the average validation NRMSE of 25 wind farms for 1-step-ahead forecasting and the model sparsity in terms of different values of  $\mu$  are obtained and given in Fig. 5. As expected, the model sparsity increases with  $\mu$ , whereas the average NRMSE firstly decreases and then increases with  $\mu$ . The average NRMSE reaches its minimum value 3.5249% at  $\mu = 0.0014$ , with model sparsity  $\Theta = 0.5428$ .

The performance of the SC-VAR is affected by two parameters, i.e.,  $S_F^i$  and  $S_N^i$ . To investigate the impact of each parameter on the forecasting performance, one of them has to be fixed while the other one is varying. Firstly, the  $S_F^i$  is fixed at 25 and the average 1-step-ahead validation NRMSE of 25 wind farms and model sparsity with varying values of  $S_N^i$  are given in Fig. 6. It can be seen that the model sparsity linearly



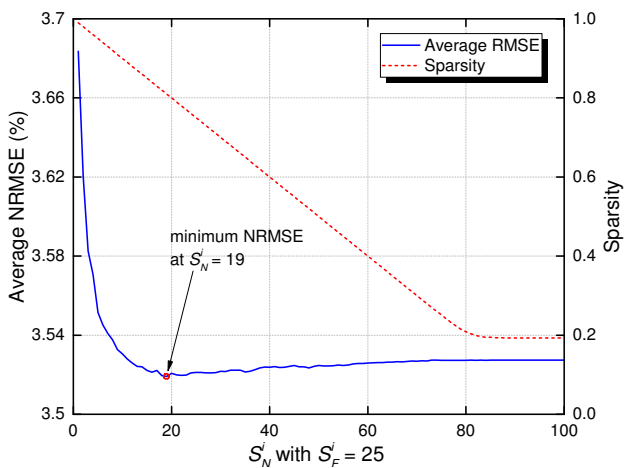


Fig. 6. The average validation NRMSE of 1-step-ahead forecasting and model sparsity of SC-VAR with varying  $S_N^i$  and fixed  $S_F^i = 25$

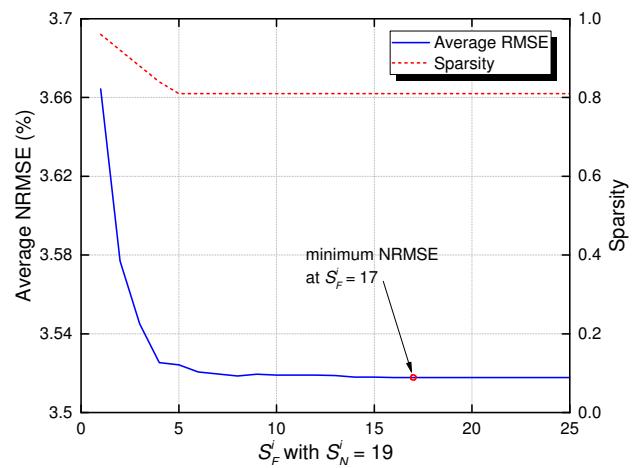


Fig. 7. The average validation NRMSE of 1-step-ahead forecasting and model sparsity of SC-VAR with varying  $S_F^i$  and fixed  $S_N^i = 19$

decreases with  $S_N^i$ . The average NRMSE decreases at the beginning and then starts to slowly increase. Both the sparsity and NRMSE finally remain unchanged after  $S_N^i$  reaches 88. The dramatic decrement at the beginning where  $1 \leq S_N^i \leq 5$  is because the wind power generation of a wind farm cannot be sufficiently explained by very few coefficients. The minimum NRMSE is achieved at  $S_N^i = 19$ .

Then the  $S_N^i$  is fixed as 19 and the average 1-step-ahead validation NRMSE and model sparsity with varying values of  $S_F^i$  are given in Fig. 7. It is found that the sparsity stops decreasing at  $S_F^i = 5$  while the NRMSE stops decreasing and reaches its minimum value at  $S_F^i = 17$ . This is because when  $S_F^i \geq 5$ , the fixed  $S_N^i = 19$  is always smaller than its upper limit  $p \cdot S_F^i$ , which means the constraint (3c) is inactive while constraint (3f) always restricts the number of wind farms to 4. Although it restricts the number of selected wind farms, constraint (3f) doesn't restrict which wind farms should be selected, so the selected wind farms may be changed when  $S_F^i$  is increasing, which makes the NRMSE change as well. Based on Fig. 6 and Fig. 7, the optimal parameters of the SC-VAR are set as  $S_F^i = 17$  and  $S_N^i = 19$ , and the corresponding minimum average NRMSE and sparsity are 3.5177% and 0.81 respectively.

As with the SC-VAR, the average 1-step-ahead validation NRMSE and sparsity of the CCSC-VAR with varying  $S_N^i$  and fixed  $\tau = 0.75$  is shown in Fig. 8. The two curves of the CCSC-VAR are similar to that of the SC-VAR in Fig. 6, but it can be seen that the NRMSE curve in Fig. 8 shows significant volatility. However, this volatility is just within a very small range of [3.5143%, 3.5163%], which is even below the minimum NRMSE of the LASSO-VAR and the SC-VAR. Therefore, the volatility doesn't affect the parameter selection. The minimum average NRMSE is achieved at  $S_N^i = 28$ .

Furthermore, the average 1-step-ahead validation NRMSE and sparsity of the CCSC-VAR with varying  $\tau$  and fixed  $S_N^i = 28$  are shown in Fig. 9. The trends of the curves in Fig. 9 can be explained in a similar way to that in Fig. 7, except that the sparsity of the SC-VAR decreases with  $S_F^i$  while the sparsity of the CCSC-VAR increases with  $\tau$ .

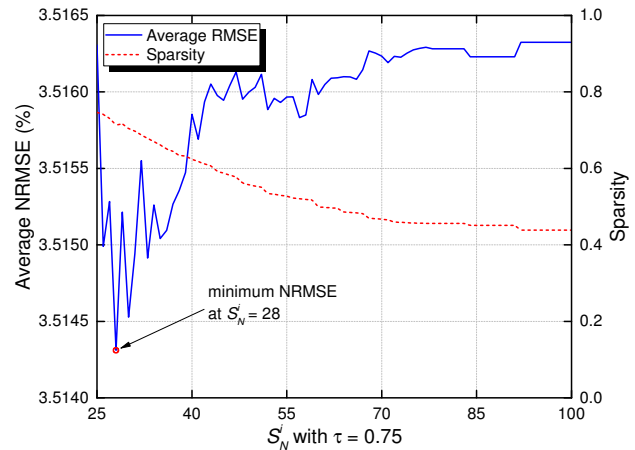


Fig. 8. The average validation NRMSE of 1-step-ahead forecasting and model sparsity of CCSC-VAR with varying  $S_N^i$  and fixed  $\tau = 0.75$

According to Fig. 8 and Fig. 9, the optimal parameters of the CCSC-VAR are set as  $\tau = 0.76$  and  $S_N^i = 28$ , and the corresponding minimum average NRMSE and sparsity are 3.5139% and 0.7264 respectively.

#### D. Results and Discussions

1) *Average Scores of Forecasting Accuracy:* The average out-of-sample testing NRMSE and testing NMAE of all 25 wind farms for different forecasting models in different time horizons are presented in Table I. All the models are configured with their optimal parameters determined in Section IV-C. The sparsity values of spatio-temporal methods are given as well.

In terms of NRMSE, all the spatio-temporal methods significantly outperform the local methods. But there is one exception, i.e., the AR is slightly better than VAR for 1-step-ahead forecasting, which reveals that a spatio-temporal model is not necessarily better than a local model due to the over-fitting problem. This is also demonstrated by the fact that all the sparse spatio-temporal methods performs better than the VAR model in all time horizons. Thus it is necessary to

TABLE I  
THE AVERAGE OUT-OF-SAMPLE TESTING NRMSE AND TESTING NMAE IN DIFFERENT TIME HORIZONS AND THE SPARSITY VALUES OF DIFFERENT FORECASTING MODELS

Indicator	Average NRMSE (%)						Average NMAE (%)						$\Theta$
	1-step	2-step	3-step	4-step	5-step	6-step	1-step	2-step	3-step	4-step	5-step	6-step	
Time horizon	1-step	2-step	3-step	4-step	5-step	6-step	1-step	2-step	3-step	4-step	5-step	6-step	n/a
Persistence	4.5967	6.4880	7.6261	8.5363	9.3276	10.0358	<b>2.7307</b>	3.9314	4.6870	5.3034	5.8380	6.3192	n/a
AR	4.5788	6.4371	7.5561	8.4541	9.2333	9.9291	2.8330	4.0214	4.7676	5.3788	5.9103	6.3907	n/a
VAR	4.5803	6.2176	7.1826	7.9625	8.6553	9.2965	2.8757	3.9423	4.5964	5.1256	5.5969	6.0317	0
LASSO-VAR	4.4382	6.1056	7.0795	7.8659	8.5683	<b>9.2135</b>	2.7820	3.8676	4.5261	5.0618	5.5387	5.9757	0.5428
SC-VAR	4.4291	6.1102	7.1036	7.9243	8.6443	9.2951	2.7730	3.8679	4.5437	5.1012	5.5889	6.0285	<b>0.8100</b>
CCSC-VAR	<b>4.4169</b>	<b>6.0842</b>	<b>7.0577</b>	<b>7.8465</b>	<b>8.5633</b>	9.2222	2.7631	<b>3.8471</b>	<b>4.5073</b>	<b>5.0437</b>	<b>5.5286</b>	<b>5.9707</b>	0.7264

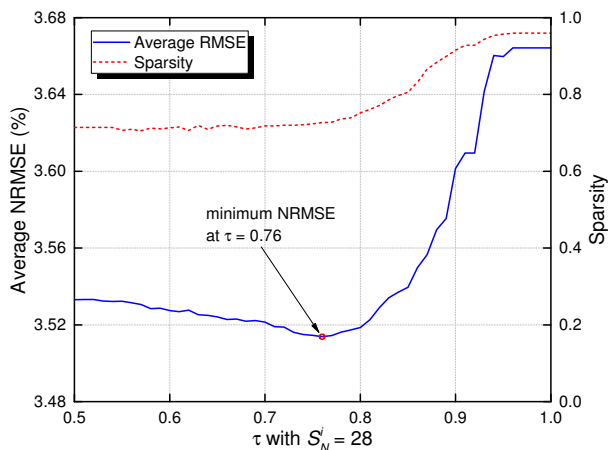


Fig. 9. The average validation NRMSE of 1-step-ahead forecasting and model sparsity of CCSC-VAR with varying  $\tau$  and fixed  $S_N^i = 28$

seek the sparsity for a high-dimensional spatio-temporal WPF model. Among the sparse models, the CCSC-VAR performs best for 1-step-ahead to 5-step-ahead forecasting. However, the LASSO-VAR is more accurate than the CCSC-VAR in 6-step horizon. This suggests one possibility to improve the accuracy of the CCSC-VAR in longer-term time horizons is to consider more time lags (both positive and negative) between time series for two wind farms when computing the cross-correlation matrix and control matrix. However, the details still need to be carefully studied.

Although the SC-VAR is better than the LASSO-VAR in the 1-step horizon, it is the worst of the sparse models in the other 5 horizons. This indicates that the SC-VAR allows controlling the sparsity in detail but can still be defeated by the black-box controlling of LASSO-VAR. Note that the SC-VAR used here is simplified version of the original SC-VAR defined in (3a)-(3i). It is possible to improve the accuracy of the SC-VAR by using all the constraints and parameters of the original SC-VAR to tune the sparsity, but this will sacrifice the implementation efficiency too much.

In terms of NMAE, the CCSC-VAR outperforms all other models for 2-step-ahead to 6-step-ahead forecasting. Notably, although the Persistence method is the worst in terms of NRMSE, it performs best for 1-step-ahead forecasting in terms of NMAE. The low MAE value is related to the flat parts of the wind-power curve model where the Persistence is almost

impossible to outperform. In cases with high wind speed, the wind power profile will be mainly constant and close to the rated power, in this case Persistence presents an almost perfect forecast performance (particularly in the 1-step horizon) in comparison with other models. This is more visible in the MAE metric since the RMSE penalizes the other situations more, i.e., the non-linear part of the power curve model. The Persistence cannot capture any ramp in wind power time series and this corresponds to the non-linear part of power curve model. This will cause large errors and contribute significantly to the RMSE due to the square of the errors.

With regard to the sparsity of sparse models, the SC-VAR acquires the highest sparsity. The LASSO-VAR has lowest sparsity, indicating that it needs more non-zero coefficients to explain the target wind farms. The CCSC-VAR has a moderate sparsity but has higher accuracy than both the SC-VAR and the LASSO-VAR. This indicates that the CCSC-VAR can extract the most useful information from a relatively small number of explanatory variables. Note that the sparsity of the LASSO-VAR in only the 1-step horizon is provided in Table I. It is found that the sparsity of the LASSO-VAR tends to decrease with the time horizons. Its sparsity in 2-step to 6-step horizons are 0.4332, 0.4, 0.3668, 0.3384 and 0.3440, respectively. Unlike the LASSO-VAR, the fixed parameter settings of the SC-VAR and the CCSC-VAR in all time horizons leads to their fixed control matrix and fixed number of explanatory wind farms. Therefore, the sparsity of the SC-VAR and the CCSC-VAR will not change with time horizon, which can partly explain why the CCSC-VAR is worse than the LASSO-VAR in 6-step horizon, since the decreasing sparsity of the LASSO-VAR implies that a sparse forecasting model may need to extract more useful information from more wind farms for longer time horizons.

To summarize, the CCSC-VAR performs best in terms of average accuracy among all the forecasting methods in this paper. This result indicates that introducing the correlation constraints to control the sparsity can improve the forecasting by exploiting the most relevant information from real data.

2) *Diebold-Mariano Test for Statistical Significance*: To assess the statistical significance of the forecast error improvement in each wind farm, Diebold-Mariano (DM) test [38], [39] is applied here to compare between the CCSC-VAR and the LASSO-VAR, which are the two most accurate methods demonstrated in previous section.

The null hypothesis of DM test is that the two methods have

TABLE II  
THE DM TEST RESULTS IN TERMS OF SQUARED LOSS FUNCTION FOR  
COMPARISON BETWEEN CCSC-VAR AND LASSO-VAR

Indicator	1-step	2-step	3-step	4-step	5-step	6-step
$N_{C>L}$	14	7	9	9	4	2
$N_{C<L}$	1	3	4	2	3	4
$N_{C=L}$	10	15	12	14	18	19

TABLE III  
THE DM TEST RESULTS IN TERMS OF ABSOLUTE LOSS FUNCTION FOR  
COMPARISON BETWEEN CCSC-VAR AND LASSO-VAR

Indicator	1-step	2-step	3-step	4-step	5-step	6-step
$N_{C>L}$	14	14	10	9	6	5
$N_{C<L}$	1	1	3	2	3	4
$N_{C=L}$	10	10	12	14	16	16

the same forecasting accuracy. The significance level used in this paper is 0.05. Three kinds of DM tests are implemented. The first one is a two-tailed test and the alternative hypothesis is that the CCSC-VAR and the LASSO-VAR have different levels of accuracy. The second one is a one-tailed test and the alternative hypothesis is that the LASSO-VAR is less accurate than the CCSC-VAR. The third test is a one-tailed test and the alternative hypothesis is that the LASSO-VAR is more accurate than the CCSC-VAR.

These tests are carried out in terms of two different loss functions, including squared error loss and absolute error loss. The tests results are summarized using three indicators and presented in Table II and Table III, where  $N_{C>L}$  is the number of wind farms for which the CCSC-VAR significantly outperforms the LASSO-VAR,  $N_{C<L}$  is the number of wind farms for which the LASSO-VAR significantly outperforms the CCSC-VAR,  $N_{C=L}$  is the number of wind farms for which the two methods have almost equivalent forecasting accuracy.

It can be seen that the DM test results are consistent with the average NRMSE and NMAE in Table I. Although the CCSC-VAR cannot defeat the LASSO-VAR for every wind farm, the  $N_{C>L}$  is greater than  $N_{C<L}$  in 1-step to 5-step horizons in terms of both squared error loss and absolute error loss, which is very significant in 1-step to 4-step horizons. Actually, it is well known that no forecasting approach provides the best results for all conditions such as different time horizons, different wind farms, different seasons, etc.

A reason for the CCSC-VAR not being better than the LASSO-VAR for some wind farms can be related to their parameter settings. The number of non-zero coefficients for each wind farm and each sparse forecasting model is shown in Fig. 10. All the sparse methods are configured with unified parameters for all wind farms, but the number of non-zero coefficients of the SC-VAR and the CCSC-VAR may be suppressed by  $S_N^i$ , while this is not the case with the LASSO-VAR. The numbers of non-zero coefficients of the SC-VAR for all wind farms are limited to  $S_N^i = 19$  and the numbers of non-zero coefficients of the CCSC-VAR for most wind farms are limited to  $S_N^i = 28$ . When the numbers of non-zero coefficients for different wind farms are restricted to one

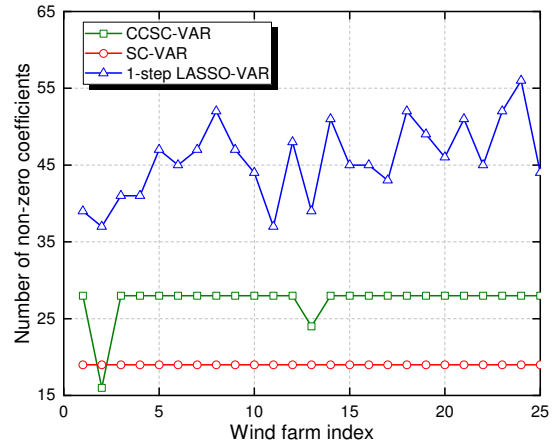


Fig. 10. The number of non-zero coefficients for each wind farm and each sparse forecasting method

unified value, the forecasting accuracy of some wind farms that need more relaxed  $S_N^i$  may be limited.

In summary, the CCSC-VAR is the overall best method when taking into account both the average scores in Table I and the DM test results.

3) *Computational Efficiency Analysis*: The computation time for training each forecasting model is provided in Table IV. All the spatio-temporal forecasting models are solved in a decomposed way by optimizing each target wind farm separately. The LASSO-VAR costs the least time because it is solved by the very efficient Coordinate Descent algorithm. It is even much faster than the local AR models. This means the spatio-temporal methods is not necessarily less efficient than simple local methods. Notice that the VAR is a special case of the LASSO-VAR ( $\mu = 0$ ), so they cost the same amount of computation time. The SC-VAR is most computationally expensive among these methods. The CCSC-VAR is much better than the SC-VAR, though it is less competitive in comparison with LASSO-VAR.

In fact, MINLPs form a particularly broad class of challenging optimization problems, as they combine the difficulty of optimizing over integer variables with the complexity of handling nonlinear functions. The simplified SC-VAR and CCSC-VAR are standard convex MINLPs [40], since they will become nonlinear convex programming if the integrality constraint on integer variables is relaxed. This is also how the solvers deal with MINLPs. Convex MINLPs can be more efficiently solved than non-convex ones. In practice, when working with MINLPs, solvers (e.g. Gurobi) might find the optimal solution (at least a very good feasible solution) quite fast but such solvers spend huge amounts of time just trying to prove the solution is optimal.

Though the simplified SC-VAR is a convex MINLP, its time consumption is still quite larger than that of other forecasting models. In practice, when solving the simplified SC-VAR, a short time limit (30 seconds in this paper) has to be imposed for each wind farm, as suggested by [33]. Otherwise, the optimization process for some wind farms may take more than ten minutes. The solving process will be stopped and

TABLE IV  
THE COMPUTATION TIME TAKEN FOR TRAINING DIFFERENT FORECASTING MODELS

Methods	Total time (s)	Average time for one wind farm (s)
AR	8.52	0.34
VAR	2.15	0.086
LASSO-VAR	2.15	0.086
SC-VAR	750	30
CCSC-VAR	10.78	0.43

the optimization results will be returned if the time limit is exceeded. The time limit is set by taking into account both the forecasting accuracy and efficiency. As for the CCSC-VAR model with better convergence property, its computation is much faster than the simplified SC-VAR. The optimization for each wind farm can be solved in a very short time and no time limit is needed, as shown in Table IV. From this point of view, the CCSC-VAR is also a significant improvement compared with the SC-VAR.

## V. CONCLUSION

This paper deals with the very-short-term WPF that makes full use of space-time dependencies among spatially distributed wind farms. The SC-VAR is employed to pursue sparse structures of forecasting models. Although the SC-VAR allows full control of sparsity, it is difficult to implement in practice due to its complicated constraints and the lack of guidelines for setting its parameters. The proposed CCSC-VAR is more efficient and accurate because it is more compact and allows incorporation of prior expert knowledge to benefit the forecasting. The proposed method is tested on 25 wind farms in Denmark and compared with several classic forecasting methods, including both local methods and spatio-temporal methods. The CCSC-VAR has best overall performance taking into account all evaluation indicators, including sparsity-control ability, sparsity, accuracy and efficiency.

Some further work could be done to improve the CCSC-VAR. The control matrix derived from the Pearson cross-correlation matrix is just one specific example to illustrate the sparsity-control and forecasting ability of the CCSC-VAR. It is flexible to choose the control matrix. Consequently, the forecasting accuracy can be improved by using other specifically designed control matrices based on reasonable expert knowledge. Furthermore, in this paper, all the forecasting methods are only discussed within the framework of batch learning mode, while an online adaptive version of the CCSC-VAR can capture and adapt to the stochastic behavior of wind by updating itself using the latest and real-time information, and thus could save computation time and improve forecasting accuracy. Moreover, in future work the proposed method will be extended to a larger spatial scale, e.g. hundreds of wind farms.

## ACKNOWLEDGMENT

The authors gratefully acknowledge Energinet.dk for providing the wind power data. The authors would also like to

thank the anonymous reviewers for their valuable comments which have improved the quality of this paper.

## REFERENCES

- [1] P. Pinson, "Wind energy: forecasting challenges for its operational management," *Statistical Science*, vol. 28, pp. 564-585, Nov. 2013.
- [2] Z. Li, L. Ye, Y. Zhao, X. Song, J. Teng and J. Jin, "Short-term wind power prediction based on extreme learning machine with error correction," *Protection and Control of Modern Power Systems*, vol. 1, no. 1, pp. 1-8, Jun. 2016.
- [3] P. Pinson, L. Mitridati, C. Ordoudis and J. Ostergaard, "Towards fully renewable energy systems: experience and trends in Denmark," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 1, pp. 26-35, March 2017.
- [4] Y. Jiang, X. Chen, K. Yu, and Y. Liao, "Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm," *Journal of Modern Power Systems and Clean Energy*, vol. 5, pp. 126-133, Jan. 2017.
- [5] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable & Sustainable Energy Reviews*, vol. 31, pp. 762-777, Mar. 2014.
- [6] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable & Sustainable Energy Reviews*, vol. 32, pp. 255-270, Apr. 2014.
- [7] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable & Sustainable Energy Reviews*, vol. 34, pp. 243-254, Jun 2014.
- [8] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H. A. Nielsen, "Spatio-temporal analysis and modeling of short-term wind power forecast errors," *Wind Energy*, vol. 14, pp. 43-60, Jan 2011.
- [9] L. Ye, Y. Zhao, C. Zeng, and C. Zhang, "Short-term wind power prediction based on spatial model," *Renewable Energy*, vol. 101, pp. 1067-1074, Feb. 2017.
- [10] D. A. Bechrakis and P. D. Sparis, "Correlation of wind speed between neighboring measuring stations," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 400-406, June 2004.
- [11] P. Sorensen et al., "Power fluctuations from large wind farms," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 958-965, Aug. 2007.
- [12] R. Girard and D. Allard, "Spatio-temporal propagation of wind power prediction errors," *Wind Energy*, vol. 16, pp. 999-1012, Oct. 2013.
- [13] G. Papaefthymiou and P. Pinson, "Modeling of spatial dependence in wind power forecast uncertainty," *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems, Rincon*, 2008, pp. 1-9.
- [14] J. Tastu, P. Pinson, P. J. Trombe and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 480-489, Jan. 2014.
- [15] M. C. Alexiadis, P. S. Dokopoulos and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, vol. 14, no. 3, pp. 836-842, Sep. 1999.
- [16] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis and P. S. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 352-361, June 2004.
- [17] T. G. Barbounis and J. B. Theocharis, "Locally recurrent neural networks for wind speed prediction using spatial correlation," *Information Sciences*, vol. 177, pp. 5775-5797, Dec. 2007.
- [18] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method," *Journal of the American Statistical Association*, vol. 101, pp. 968-979, Sep. 2006.
- [19] X. Zhu, M. G. Genton, Y. Gu, and L. Xie, "Space-time wind speed forecasting for improved power system dispatch," *Test*, vol. 23, pp. 1-25, Mar. 2014.
- [20] 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, pp. 511-520, Jan. 2014.
- [21] J. Dowell, S. Weiss, D. Hill, and D. Infield, "Short-term spatio-temporal prediction of wind speed and direction," *Wind Energy*, vol. 17, pp. 1945-1955, Dec. 2014.
- [22] 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, July 2014.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49
- [23] M. Landry, T. P. Edinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061-1066, Jul.-Sep. 2016.
- [24] R. Juban, H. Ohlsson, M. Maasoumy, L. Poirier, and J. Z. Kolter, "A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1094-1102, Jul.-Sep. 2016.
- [25] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1074-1080, Jul.-Sep. 2016.
- [26] B. M. Sanandaji, A. Tascikaraoglu, K. Poolla and P. Varaiya, "Low-dimensional models in spatio-temporal wind speed forecasting," *2015 American Control Conference (ACC), Chicago, IL, 2015*, pp. 4485-4490.
- [27] T. Hastie, R. Tibshirani and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, 2015.
- [28] A. Tascikaraoglu, B. M. Sanandaji, K. Poolla, and P. Varaiya, "Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform," *Applied Energy*, vol. 165, pp. 735-747, Mar 1 2016.
- [29] Y. Zhang, S. J. Kim and G. B. Giannakis, "Short-term wind power forecasting using nonnegative sparse coding," *2015 49th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, 2015*, pp. 1-5.
- [30] 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, March 2016.
- [31] P. Pinson, "Introducing distributed learning approaches in wind power forecasting," *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, 2016*, pp. 1-6.
- [32] L. Cavalcante, R. J. Bessa, M. Reis and J. Browell, "LASSO vector autoregression structures for very short-term wind power forecasting," *Wind Energy*, vol. 20, no. 4, pp. 657-675, April 2017.
- [33] E. Carrizosa, A. V. Olivares-Nadal and P. Ramirez-Cobo, "A sparsity-controlled vector autoregressive model," *Biostatistics*, vol. 18, no. 2, pp. 244-259, 2017.
- [34] F. Ziel, "Forecasting electricity spot prices using lasso: on capturing the autoregressive intraday structure," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4977-4987, Nov. 2016.
- [35] H. Madsen, P. Pinson, H. A. Nielsen, T. S. Nielsen, and G. Kariniotakis, "Standardizing the performance evaluation of short-term wind power prediction models", *Wind Engineering*, vol. 29, no. 6, pp. 475-489, Dec. 2005.
- [36] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions," *J. Roy. Statist. Soc. C*, vol. 61, no. 4, pp. 555-576, 2012.
- [37] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, pp. 1-22, Feb. 2010.
- [38] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 13, pp. 253-263, Jul. 1995.
- [39] D. Harvey, S. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, vol. 13, pp. 281-291, Jun. 1997.
- [40] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1-131, 2013.



**Yongning Zhao** (S'17) received his B.Sc. in electrical engineering from China Agricultural University in 2012. He is currently a Ph.D. student of the Department of Electric Power Systems, China Agricultural University (CAU), P.R. China. He is now a visiting student with Center for Electric Power & Energy, Technical University of Denmark (DTU).

His research interests are in the areas of power system operation and control, analysis of the spatio-temporal characteristics of wind power, wind power forecasting and integration.



**Lin Ye** (M'94-SM'06) received his B.Sc. from WuHan University, P.R. China in 1992 and his Ph.D. degree in 2000 from the Institute of Electrical Engineering (IEE), Chinese Academy of Sciences (CAS), all in electrical engineering. He pursued research at Karlsruhe Institute of Technology as a research fellow of Alexander von Humboldt Stiftung/Foundation (AvH) of Germany from 2000 to 2002. He joined the Interdisciplinary Research Center (IRC), Department of Engineering /Cavendish Laboratory, the University of Cambridge, as a research fellow in 2004. He received the "Hua Wei" Award from Chinese Academy of Sciences in 1999, and Research Young Investigator Award from Fok Ying Tung Education Foundation, Ministry of Education, P.R. China in 2003. He was an awardee of the Program for New Century Excellent Talents in China Universities in 2008. He is currently a full professor in electrical power engineering at the Department of Electric Power Systems, China Agricultural University (CAU), Beijing, P.R. China. He holds memberships in IEEE (USA), and European EMTP-ATP Users Group (EEUG) as well as senior member of Wolfson College at Cambridge University. He is also a core member of the IEEE Task Force on Sustainable Energy Systems for Developing Communities and an international expert in CIGRE SC B5 working group.

His research interests are electric power system analysis & control, power grids modeling & simulations, renewable energy generation & system integration and wind/solar power forecasting.



**Pierre Pinson** (M'11-SM'13) received the M.Sc. degree in applied mathematics from the National Institute for Applied Sciences (INSA Toulouse, France) and the Ph.D. degree in energetics from Ecole des Mines de Paris (France). He is a Professor at the Technical University of Denmark (DTU), Centre for Electric Power and Energy, Department of Electrical Engineering, also heading a group focusing on Energy Analytics & Markets.

His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies. Prof. Pinson acts as an Editor for the International Journal of Forecasting, and for Wind Energy.



**Yong Tang** (M'98-SM'12) received the B.S. degree of electrical engineering from the Fuzhou University, Fuzhou, China, in 1982, the M.E. and Ph.D. degree in Power System and Automation from the China Electric Power Research Institute (CEPRI), Beijing, China, in 1985 and 2002, respectively.

He is currently the Chief Engineer and a Professor in CEPRI. His research interests are in the area of power system simulation and analysis, voltage stability and control, load modeling and simulation.