Online Forecast Reconciliation in Wind Power Prediction

Chiara Di Modica*, Pierre Pinson* and Souhaib Ben Taieb†
* Technical University of Denmark, Kgs. Lyngby, Denmark (ppin@elektro.dtu.dk)
† Université de Mons, Mons, Belgium (souhaib.bentaieb@umons.ac.be)

***Abstract*—Increasing digitization of the electric power sector allows to further rethink forecasting problems that are crucial input to decision-making. Among other modern challenges, ensuring coherency of forecasts among various agents and at various aggregation levels has recently attracted attention. A number of reconciliation approaches have been proposed, from both game-theoretical and statistical points of view. However, most of these approaches make unrealistic unbiasedness assumptions and overlook the fact that the underlying stochastic processes may be nonstationary. We propose here an alternative approach to the forecast reconciliation problem in a constrained regression framework. This relies on a multivariate least squares estimator, with equality constraints on the coefficients (denoted MLSE). A recursive and adaptive version of that estimator is derived (denoted MRLSE), hence allowing to track the optimal reconciliation in a fully data-driven manner. We also prove that our methods by design guarantee the coherency property for any out-of-sample forecasts (reconciliation by design). We show the effectiveness of our forecasting methods using a Danish wind energy dataset with 100 wind farms.**

***Index Terms*—Renewable energy; Forecasting; Online Learning; Hierarchical Time-series**

I. Introduction

Large-scale deployment of renewable energy generation sources brings a wealth of opportunities and challenges. For forecasting especially, the fact that production sites are geographically distributed, in a fairly dense manner, yields an observation network that can be exploited. This eventually allows improving the accuracy of wind power forecasts by accounting for spatio-temporal dependencies in the underlying processes, e.g. [1]. This effect was also observed for the case of solar power forecasts [2], hence making the methods proposed for wind power equally relevant for solar power generation. However, other challenges that were unforeseen (or possibly considered as futile) are being identified. In fact, since many agents in power systems and electricity markets generate their own forecasts, at various aggregation levels and independently of each other, these forecasts may end up not being coherent. For example, for a portfolio composed of two wind farms, the sum of the forecasts made for these wind farms, individually, will not necessarily be equal to the forecasts readily made for the portfolio. This lack of additive coherency is a challenge when forecasts are used as input to decision-making problems in power system operation and electricity markets.

The issue of forecast reconciliation has already been identified in the statistical modelling and forecasting literature for quite some time now, with the first work related to energy applications described in [3]. Since then, a wealth of relevant works appeared, including methodological contributions and applications, e.g. [4]. Some were readily focused on the wind power forecasting application, as for the case of [5] for instance. In fact, reconciliation approaches for probabilistic forecasts were also proposed, for both electric load [6] and wind power generation [7]. Others have looked at novel approaches to temporal reconciliation for large-scale electricity consumption [8]. Distributed approaches to forecast reconciliation [9], based on the Alternating Direction Method of Multipliers (ADMM), allowed to prevent potentially sensitive information exchange between wind farm operators. However, most of these approaches make unrealistic unbiasedness assumptions and overlook the fact that the underlying stochastic processes and optimal reconciliation may be nonstationary.

As a result, our objective is to propose a new online forecast reconciliation approach which relaxes these assumptions and allows to adapt to changes in the underlying characteristics of the stochastic processes. Specifically, we make the following contributions. First, we formulate a new objective function for forecast reconciliation based on a multivariate regression problem with equality constraints on the regression parameters. This leads to a batch multivariate least squares estimator with equality constraints (MLSE). Then, we extend the MLSE estimator to the online setting, and derive a recursive and adaptive estimator inspired by recursive least squares (RLS) estimation with exponential forgetting, which we denote MRLSE. Finally, we prove that our estimators guarantee the coherency property not only in-sample but also out-of-sample. In other words, the out-of-sample forecasts will be coherent by design even though the objective function only constrains the in-sample forecasts to be coherent.

The remainder of the paper is structured as follows. The forecast reconciliation problem is described in Section II. Our proposal for forecast reconciliation is described in Section III, in both their batch and online versions. Section IV presents some experiments with Danish wind data, while conclusions and perspectives for future work are gathered in Section V.

This work was partly supported through the Smart4RES project (EU Horizon 2020, No. 864337), CITIES and Energy Collective projects (IF Denmark, and ForskEL/EUDP).

II. Forecast Reconciliation

Let $\{Y^*_{s,t}\}$ ($s = 1, \dots, m$, $t = 1, \dots, T$) be the stochastic process for wind power generation, with indices s and t for location and time, respectively, as well as corresponding realizations $y^*_{s,t}$. We denote the power observations for all m individual sites at a given time t as $\mathbf{y}^*_t = [y^*_{1,t}, \dots, y^*_{s,t}, \dots, y^*_{m,t}]^\top$.

A. Defining a Hierarchy

Individual sites are organized in a hierarchy, where quantities at upper levels are obtained by aggregating the quantities of the individual sites. The hierarchy has L levels and N total number of nodes. $\mathcal{S}$ is the set of all nodes. N_l is the number of nodes at level l, as a subset $\mathcal{S}_l \subset \mathcal{S}$, such that $N = \sum_{l=1}^{L} N_l$ and $\mathcal{S} = \bigcup_{l=1}^{L} \mathcal{S}_l$. The tuple (l, j) then uniquely identifies node j at level l. Nodes at a lower level of the hierarchy are referred to as child nodes, and those at the lowest level (the individual sites) are the bottom nodes. The number N_L of bottom nodes is equal to the number of individual sites m. An example of a 3-level hierarchy, based on 5 individual sites, is depicted in Fig. 1.

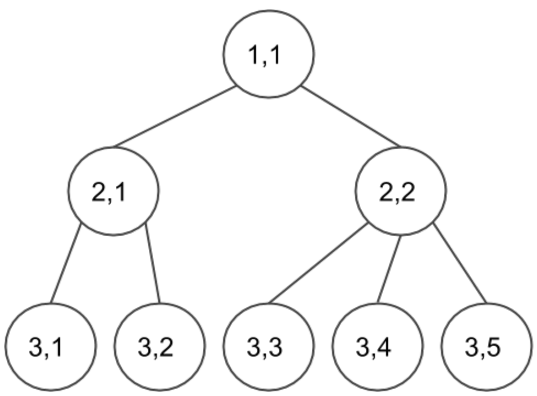

Fig. 1. Example of a 3-level hierarchy based on 5 individual sites, with $S_1 = \{(1,1)\}$, $S_2 = \{(2,1),(2,2)\}$ and $S_3 = \{(3,1),(3,2),(3,3),(3,4),(3,5)\}$.

If $\mathbf{y}^*_t$ are the observations at time t in the bottom nodes, the observations at all levels of the hierarchy $\mathbf{y}_t$ are given by

$$\mathbf{y}_t = \mathbf{S}\,\mathbf{y}^*_t, \quad \forall t\,, \tag{1}$$

where $\mathbf{S} \in \{0,1\}^{N \times N_L}$ is a summing matrix defined as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 \in \{0,1\}^{N_1 \times N_L} \\ \mathbf{S}_2 \in \{0,1\}^{N_2 \times N_L} \\ \vdots \\ \mathbf{S}_{L-1} \in \{0,1\}^{N_{L-1} \times N_L} \\ \mathbf{I}_{N_L} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_{N_L} \end{bmatrix}, \tag{2}$$

and $\mathbf{S}_l \in \mathbb{R}^{N_l \times N_L}$ is a matrix whose elements $s_{l_{ij}}$ are 1 if the j^{th} node of the bottom-level is a child (or grand-child) of the i^{th} node of level l, 0 otherwise. $\mathbf{I}_{N_L}$ is an identity matrix of dimension N_L. Thus, it has a block structure with a first block $\mathbf{A} \in \{0,1\}^{(N-N_L) \times N_L}$ for the summing operations to go up in the hierarchy and a second block being an identity matrix of size N_L to copy the elements of the bottom nodes. For the example of Fig. 1, the summing matrix reads

$$\mathbf{S} = \left[\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ \hline & & \mathbf{I}_5 & & \end{array}\right]. \tag{3}$$

In parallel, consider that given a lead time k, forecasts are issued at time t for time $t+k$. The forecasts for all individual sites are denoted by $\hat{y}^*_{s,t+k|t}$ and with $\hat{\mathbf{y}}^*_{t+k|t} = [\hat{y}^*_{1,t+k|t}, \dots, \hat{y}_{s,t+k|t}, \dots, \hat{y}^*_{m,t+k|t}]^\top$. Forecasts are also issued for all nodes of the hierarchy, individually and independently of each other, and collated in the vector of forecasts $\hat{\mathbf{y}}_{t+k|t}$.

B. Additive Coherency and Reconciliation

Many agents in power systems and electricity markets generate their own forecasts at various aggregation levels independently of each other. As a result, it is highly likely that one has

$$\hat{\mathbf{y}}_{t+k|t} \neq \mathbf{S}\,\hat{\mathbf{y}}^*_{t+k|t}, \quad \forall t, k\,, \tag{4}$$

meaning that the forecasts do not satisfy the hierarchical aggregation constraints, also called *additive coherency*.

Definition 1. *(additive coherency) The forecasts $\hat{\mathbf{y}}_{t+k|t}$ for a hierarchy defined by a summing matrix $\mathbf{S}$ are said to be additively coherent if*

$$\hat{\mathbf{y}}_{t+k|t} = \mathbf{S}\,\hat{\mathbf{y}}^*_{t+k|t} \iff \mathbf{H}^\top \hat{\mathbf{y}}_{t+k|t} = \mathbf{0}\,, \tag{5}$$

where

$$\mathbf{H}^\top = \begin{bmatrix} \mathbf{I}_{(N-N_L)} & -\mathbf{A} \end{bmatrix}. \tag{6}$$

Note that the matrix $\mathbf{H}$ naturally depends on the structure of the hierarchy through the matrix $\mathbf{A}$. As we need one equality constraint per non-bottom node, this yields $N - N_L$ equality constraints. The matrix $\mathbf{H}^\top$ therefore is a $(N - N_L) \times N$ matrix. For the specific case of the 3-level hierarchy depicted in Fig. 1, we have

$$\mathbf{H}^\top = \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 & -1 \end{array}\right]. \tag{7}$$

Given some probably incoherent forecasts $\hat{\mathbf{y}}_{t+k|t}$, the process of forecast reconciliation is defined as the transformation of the forecast vector $\hat{\mathbf{y}}_{t+k|t}$ such that it is made additively coherent (i.e., the equality is restored). For a review of the alternative approaches to forecast reconciliation, the reader is referred to [10].

Remark 1. *Contrarily to the case of forecasts, power measurements are naturally additively coherent, since measurements for upper level of the hierarchy are obtained by directly using the summing matrix $\mathbf{S}$ as in* (1).

III. Forecast Reconciliation with Multivariate Least Squares Estimation

We propose a new forecast reconciliation method which involves solving a multivariate least squares regression problem. A set of constraints on the coefficients are added to the objective function to ensure coherent forecasts. By doing so, we relax the unbiasedness assumption of existing reconciliation methods [4], and we allow to use the wealth of modern approaches for estimation in regression models including the online learning setting. We first introduce a batch version of our method, then we derive an online version based on recursive and adaptive estimation with exponential forgetting.

A. Multivariate Least Squares Estimation

We model the observations at all nodes in the hierarchy as a linear combination of the corresponding forecasts. Specifically, given lead time k, we consider the following regression model:

$$\mathbf{y}_{t+k} = \mathbf{\Theta}_k^\top \tilde{\mathbf{y}}_{t+k|t} + \boldsymbol{\varepsilon}_{t+k}, \quad \forall t\,, \tag{8}$$

where $\mathbf{\Theta}_k \in \mathbb{R}^{(N+1)\times N}$ is a matrix of regression coefficients, $\tilde{\mathbf{y}}_{t+k|t}^\top = \begin{bmatrix} 1 \; \hat{\mathbf{y}}_{t+k|t}^\top \end{bmatrix} \in \mathbb{R}^{1\times(N+1)}$, and $\boldsymbol{\varepsilon}_{t+k}$ a noise term with zero mean and finite variance.

In the batch setting, we are given a dataset composed of T pairs of forecasts and observations, for a given lead time k. With our method, this dataset is used to estimate the regression coefficients in (8). More precisely, we solve the following multivariate least squares problem with equality constraints (MLSE):

$$\hat{\mathbf{\Theta}}_k^{\text{MLSE}} = \underset{\mathbf{\Theta}}{\text{argmin}} \quad ||\mathbf{Y}_k - \hat{\mathbf{Y}}_k \mathbf{\Theta}||_2^2 \tag{9a}$$

$$\text{s.t.} \quad \hat{\mathbf{Y}}_k \mathbf{\Theta} \mathbf{H} = \mathbf{0}\,, \tag{9b}$$

where $\mathbf{Y}_k \in [0,1]^{T\times N}$ and $\hat{\mathbf{Y}}_k \in [0,1]^{T\times(N+1)}$ are given by

$$\mathbf{Y}_k = \begin{bmatrix} \mathbf{y}_{1+k}^\top \\ \vdots \\ \mathbf{y}_{T+k}^\top \end{bmatrix} \text{ and } \hat{\mathbf{Y}}_k = \begin{bmatrix} \tilde{\mathbf{y}}_{1+k|1}^\top \\ \vdots \\ \tilde{\mathbf{y}}_{T+k|T}^\top \end{bmatrix}. \tag{10}$$

The constraint $\hat{\mathbf{Y}}_k \mathbf{\Theta} \mathbf{H} = \mathbf{0}$ ensures that the reconciled forecasts $\hat{\mathbf{Y}}_k \mathbf{\Theta}$ are coherent as presented in Definition 1. After estimating $\hat{\mathbf{\Theta}}_k^{\text{MLSE}}$, when a new forecast $\hat{\mathbf{y}}_{t+k|t}$ for all nodes of the hierarchy is available, the vector of reconciled forecasts is obtained as $\left(\hat{\mathbf{\Theta}}_k^{\text{MLSE}}\right)^\top \tilde{\mathbf{y}}_{t+k|t}$.

For the MLSE problem in (9), assuming $\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k$ is invertible, a closed-form solution can be readily obtained following the developments in [12], as

$$\hat{\mathbf{\Theta}}_k^{\text{MLSE}} = \left(\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k\right)^{-1} \hat{\mathbf{Y}}_k^\top \mathbf{Y}_k \left(\mathbf{I}_{N_L} - \mathbf{C}_k\right)\,, \tag{11}$$

where $\mathbf{I}_{N_L}$ is an identity matrix of size N_L and $\mathbf{C}_k$ is a matrix whose elements depend on the structure of the hierarchy and on the variance of the forecast error, i.e.

$$\mathbf{C}_k = \mathbf{H}\left(\mathbf{H}^\top \mathbf{\Sigma}_k \mathbf{H}\right)^{-1} \mathbf{H}^\top \mathbf{\Sigma}_k\,. \tag{12}$$

The covariance matrix $\mathbf{\Sigma}_k$ needs to be estimated, possibly making some assumptions about its structure, as for some other reconciliation approaches [10]. Looking at (11), one observes that the MLSE estimator is a variant of the unconstrained multivariate Least Squares one, with a projection given by $(\mathbf{I}_{N_L} - \mathbf{C}_k)$,

$$\hat{\mathbf{\Theta}}_k^{\text{MLSE}} = \hat{\mathbf{\Theta}}_k^{\text{MLS}} \left(\mathbf{I}_{N_L} - \mathbf{C}_k\right)\,, \tag{13}$$

with

$$\hat{\mathbf{\Theta}}_k^{\text{MLS}} = (\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k)^{-1} \hat{\mathbf{Y}}_k^\top \mathbf{Y}_k. \tag{14}$$

Based on the equality constraints in (9b), coherency is imposed for all T pairs of forecasts and corresponding observations in the training dataset used to estimate the model parameters. This does not ensure that those parameters will guarantee coherency of forecasts reconciled for new data not seen in the training set (i.e., out-of-sample). The following Theorem shows that our method has the nice property of implicitly reconciling out-of-sample forecasts.

Theorem 1 (reconciliation by design). *By computing $\hat{\mathbf{\Theta}}_k^{\text{MLSE}}$ using (11), for any new forecast (out-of-sample) $\hat{\mathbf{y}}_{t+k|t}$, the reconciled forecasts given by $\left(\hat{\mathbf{\Theta}}_k^{\text{MLSE}}\right)^\top \tilde{\mathbf{y}}_{t+k|t}$ are additively coherent.*

Proof. Consider any set of forecasts $\hat{\mathbf{y}}_{t+k|t}$ for a hierarchy defined by the summation matrix $\mathbf{S}$, and corresponding matrix $\mathbf{H}$. Based on the augmented vector of forecasts $\tilde{\mathbf{y}}_{t+k|t}$, one has

$$\tilde{\mathbf{y}}_{t+k|t}^\top \hat{\mathbf{\Theta}}_k^{\text{MLSE}} \mathbf{H} = \tilde{\mathbf{y}}_{t+k|t}^\top \hat{\mathbf{\Theta}}_k^{\text{MLS}} \left(\mathbf{I}_{N_L} - \mathbf{C}_k\right)\mathbf{H}\,. \tag{15}$$

It then means that

$$\tilde{\mathbf{y}}_{t+k|t}^\top \hat{\mathbf{\Theta}}_k^{\text{MLSE}} \mathbf{H} = \tilde{\mathbf{y}}_{t+k|t}^\top \hat{\mathbf{\Theta}}_k^{\text{MLS}} \left(\mathbf{H} - \mathbf{C}_k\mathbf{H}\right)\,. \tag{16}$$

Considering the definition of $\mathbf{C}_k$ in (12), one has

$$\mathbf{H} - \mathbf{C}_k\mathbf{H} = \mathbf{H} - \mathbf{H}\left(\mathbf{H}^\top \mathbf{\Sigma}_k \mathbf{H}\right)^{-1}\left(\mathbf{H}^\top \mathbf{\Sigma}_k \mathbf{H}\right) \tag{17a}$$

$$= \mathbf{H} - \mathbf{H}\,\mathbf{I}_{(N-N_L)} = 0\,. \tag{17b}$$

This therefore yields

$$\tilde{\mathbf{y}}_{t+k|t}^\top \hat{\mathbf{\Theta}}_k^{\text{MLSE}} \mathbf{H} = 0\,, \tag{18}$$

for any forecast $\hat{\mathbf{y}}_{t+k|t}$ and whatever the chosen covariance matrix $\mathbf{\Sigma}$. □

B. Online Version of the Estimator

For most practical applications, it is beneficial to consider online estimation, i.e., involving recursive estimation based on update equations and some form of history forgetting. This has the benefit of accommodating nonstationarity of the underlying stochastic processes, while lightening the computational burden. The online version of our estimator is therefore abbreviated as MRLSE (with 'R' for recursive).

At a given time t, the MRLSE estimator is defined as

$$\hat{\mathbf{\Theta}}_{t,k}^{\text{MRLSE}} = \underset{\mathbf{\Theta}}{\text{argmin}}\; S_t(\mathbf{\Theta}) \tag{19a}$$

$$\text{s.t. } \tilde{\mathbf{y}}_{i+k|i}^\top \mathbf{\Theta}\mathbf{H} = \mathbf{0},\; \forall i \le t\,, \tag{19b}$$

where

$$S_t(\mathbf{\Theta}) = \frac{1}{2}\sum_{i\leq t}\lambda^{t-i}\left(\mathbf{y}_{t+k} - \mathbf{\Theta}^\top\tilde{\mathbf{y}}_{t+k|t}\right)^2 , \tag{20}$$

and where $0 < \lambda < 1$ is a forgetting factor, generally in the range $[0.95, 1]$. It is often more convenient to work with the equivalent number n_λ of observations instead, defined as $n_\lambda = (1-\lambda)^{-1}$.

In practice, as common for RLS estimators, the update equations for $\hat{\mathbf{\Theta}}_{t,k}^{\text{MRLSE}}$ given the previous value of the estimator, $\hat{\mathbf{\Theta}}_{t-1,k}^{\text{MRLSE}}$, and the new information available at time t, is obtained through a Newton-Raphson step. An additional projection $\pi_{\mathbf{H}}$ on the feasible space defined by (19b) ought to be used, similarly to [13]. This yields

$$\hat{\mathbf{\Theta}}_{t,k}^{\text{MRLSE}} = \pi_{\mathbf{H}}\left\{\hat{\mathbf{\Theta}}_{t-1,k}^{\text{MRLSE}} - \frac{\nabla S_t(\mathbf{\Theta}_{t-1,k})}{\nabla^2 S_t(\mathbf{\Theta}_{t-1,k})}\right\} . \tag{21}$$

After a little algebra, one obtains the update equations at time t as

$$\mathbf{R}_{t,k} = \lambda\mathbf{R}_{t-1,k} + \tilde{\mathbf{y}}_{t+k|t}\tilde{\mathbf{y}}_{t+k|t}^\top, \tag{22a}$$

$$\hat{\mathbf{\Theta}}_{t,k}^{\text{MRLSE}} = \hat{\mathbf{\Theta}}_{t-1,k}^{\text{MRLSE}} + \mathbf{R}_t^{-1}\tilde{\mathbf{y}}_{t+k|t}\left(\mathbf{y}_{t+k}(\mathbf{I}-\mathbf{C}) - \tilde{\mathbf{y}}_{t+k|t}^\top\hat{\mathbf{\Theta}}_{t-1,k}^{\text{MRLSE}}\right) . \tag{22b}$$

The MRLSE estimator naturally inherits its fundamental reconciliation property from the MLSE estimator, i.e., reconciliation by design for any new (out-of-sample) forecasts.

IV. Application and Results

We compare our new forecast reconciliation method with the state-of-the-art approaches using a real-world dataset from Denmark. After introducing our case-study, we present our forecast verification framework and some relevant benchmarks. Finally, we provide a number of results and discuss the advantages and limitations of the different forecast reconciliation methods described previously.

A. Case Study Based on a Danish Dataset

The dataset provided by the Danish Transmission System Operator, Energinet.dk, includes wind power measurements for 349 wind farms in western Denmark, for the period between January 2006 and March 2012. The measurements have a 15-minute temporal resolution. An extensive analysis of this dataset has been performed by [14], [15], [16]. These studies identified the conditional space-time dependencies of power generation at the various sites, including the nonstationarity of the underlying stochastic processes.

Only a subset of the available dataset, both in terms of number of wind farms and time period, was selected. Firstly, sites with non-negligible episodes with missing data were discarded. Out of the 250 sites left, only 100 sites were randomly selected, for simplicity. They are shown in Fig. 2.

Out of the complete dataset, a period with 70 080 time steps (2 years) was extracted for this analysis, from 2010 and 2011. The power measurement time-series for the 100 sites

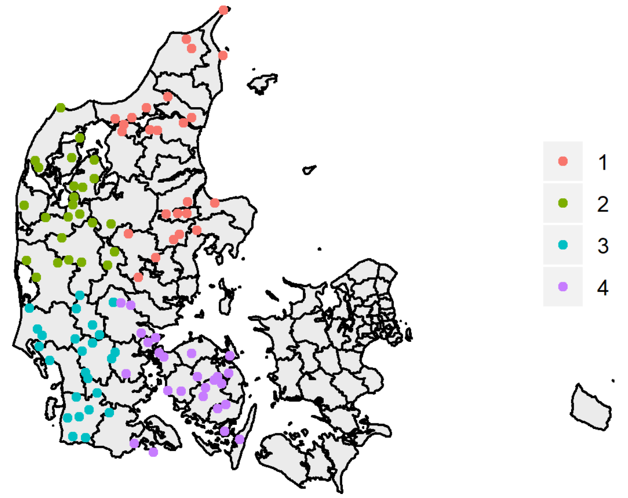

Fig. 2. The 100 Danish sites selected from the complete Danish wind power dataset, then divided into 4 regions.

were then further cleaned, considering both erroneous and suspicious data points. For each site, observations exceeding 1.5 times the quantile with nominal level 0.99 of the distribution of observations were removed. Power measurements were then normalized by the nominal capacity of that site. However, as this nominal capacity may change with time, a function computing rolling maxima was used for its estimation (package *zoo*[1]) and adaptive normalization. Rolling windows of 5 000 time steps were used. Consequently for the bottom nodes, all resulting observations take value in the unit interval [0,1].

The aggregate time-series for the various regions and the whole portfolio were obtained by using a summing matrix $\mathbf{S}$ (see Section II-A). As in the example of Fig. 1, our hierarchy has 3 levels, with bottom nodes, regions, and the overall sum (referred to as total). The 100 wind farms were grouped in 4 regions, as shown in Figure 2. Each region is composed of 25 wind farms, by dividing the Western Denmark area into 4 quadrants. Owing to this summation, power values for the region level and the whole portfolio are within [0,25] and [0,100], respectively.

Forecasts are to be generated for each and every node of the hierarchy, i.e., for the 100 bottom nodes, the 4 regions and the overall portfolio (total). These are referred to as base forecasts. For simplicity, only 1-step ahead forecasts were considered, though the methodology could be readily used to reconcile forecasts for further lead times. Following the analysis and results in [14], [15], [16], Auto-Regressive models with 2 lags - AR(2), were found sufficient to model the temporal dynamics of the time-series as input to forecasting. Thus, using the first 6 months of data as training dataset, AR(2) models were fitted through LS minimization for each node in the hierarchy. It was verified that those forecasts were competitive and their quality at the level of the state of the art for such short lead times. These could be improved by considering

[1] Available on CRAN at: https://cran.r-project.org/web/packages/zoo

more advanced models and possibly online learning, though only seen as different and possibly more accurate forecasts as input to forecast reconciliation. The following 6 months were then used as training for the batch reconciliation approaches. Specifically the online approach was initialized on the first time step of that period and then recursively updated trough the remainder of the dataset. For simplicity, the equivalent number of observations was set to $n_\lambda = 10,000$ though it could have been optimized through cross-validation. This eventually leaves the last year (2011) of data for genuine forecast verification.

B. Forecast Verification Framework and Benchmarking

Our evaluation procedure is based on current practices for the verification of wind power forecasts, as recently described in [17]. To be consistent with the least squares objective used to fit the models, i.e. the quadratic loss function, we use the Normalized Root Mean Square Error (NRMSE) as forecast verification criterion. For a set of T forecast-observation pairs for the node i of the hierarchy, the Scaled Root Mean Square Error (SRMSE) is given by

$$\text{SRMSE}_i = \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\varepsilon_{i,t+1|t}}{s_i} \right)^2 \right)^{\frac{1}{2}} \tag{23}$$

$$\text{with } s_i = \begin{cases} 100, \text{ if } i = 1 \quad \text{(total)} \\ 25, \quad \text{if } i = 2, \ldots, 5 \quad \text{(region level)} \\ 1, \quad\ \text{if } i = 6, \ldots, 105 \quad \text{(bottom level)}, \end{cases} .$$

where $\varepsilon_{i,t+1|t}$ is the one-step ahead forecast error for the forecast issued at time t for the i^{th} node of the hierarchy. Score values are commonly multiplied by 100 as if expressed as percentages (of capacity). We additionally introduce a score that combines results for all nodes of the hierarchy, accounting for the number of nodes at each level. This Weighted Root Mean Square Error (WRMSE) is defined as

$$\text{WRMSE} = \frac{1}{N_L\, L} \sum_{i=1}^{N} \text{NRMSE}_i, \tag{24}$$

where NRMSE_i naturally reflects the importance of each node since relying on different scales (directly related to the number of bottom nodes it aggregates).

Since we aim to show how forecast reconciliation contributes to both restoring coherency and improving forecast accuracy, we report improvements with respect to the base forecasts. These improvements can be interpreted as percentage decrease in SRMSE compared to the base forecasts. For a given node i and reconciliation method, this writes

$$\text{ISRMSE}_{i,\text{method}} = \frac{\text{SRMSE}_{i,\text{base}} - \text{SRMSE}_{i,\text{method}}}{\text{SRMSE}_{i,\text{base}}}, \tag{25}$$

where $\text{SRMSE}_{i,\text{base}}$ and $\text{SRMSE}_{i,\text{method}}$ are the SRMSE values for the base forecasts and reconciliation method considered, respectively. A similar criterion can defined using the WRMSE criterion.

In the following, we will consider forecast reconciliation based on our two estimators, i.e., MLSE and MRLSE, as well as the state-of-the-art MinT approach. A complete description of the MinT approach to forecast reconciliation is available in [10], while applications to wind power forecasting are described in [5], [9]. In this work, we consider the covariance matrix of the one-step-ahead forecast errors is estimated using the in-sample model residuals. More advanced shrinkage covariance estimators can also be used in high-dimensional setting. Finally, to measure the statistical significance of the differences in scores for the various reconciliation methods, we use the Diebold-Mariano (DM) test (see [17]). The differences are always found significant.

C. Results and Discussion

1) Observing the need for forecast reconciliation: To first illustrate the need for forecast reconciliation based on our case study, we look at the lack of coherence between forecasts at various levels of the hierarchy. Forecasts for the upper levels of the hierarchy (regions and total) are obtained based on the summing matrix $\mathbf{S}$ and then compared to the forecasts readily produced at these levels. These differences are therefore in the range of [-25,25] at the region level and [-100,100] at the total level. Results are depicted in Fig. 3 (in a fashion similar to the results in [5]) and support the statement made with (4). These inconsistency errors are up to 4% here, at both region and total levels. Since the various approaches we consider hereafter allow for reconciliation by design, all those inconsistencies are then removed.

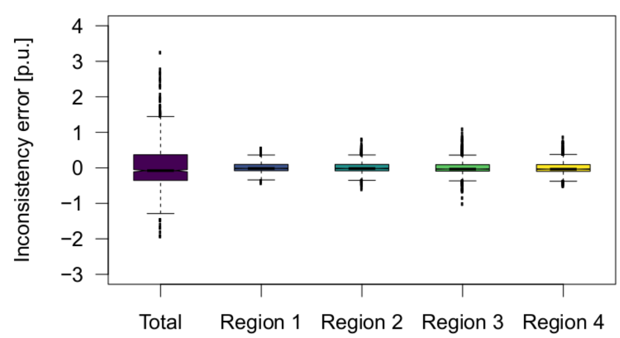

Fig. 3. Incoherency, as expressed by (4), observed in the upper levels of the hierarchy over a randomly chosen period of 2 weeks.

2) Impact on forecast quality: The literature on forecast reconciliation has regularly covered the fact that reconciliation eventually yields improvements in forecast quality. For instance already in [3], the authors made a point that their game-theoretical optimal projection approach could reconcile forecasts by design while providing a geometry-inspired proof of forecast quality improvement (under a quadratic criterion). We consequently investigate here whether forecast improvements are obtained based on the approaches we proposed, and how it compares with the existing e.g. MinT.

We first look at the score values obtained over the one-year evaluation period covering 2011. These score values are collated in Table I, using the SRMSE criterion expressed in percentage of nominal capacity (as an average for all nodes at a given level) and related improvements with the ISRMSE criterion. Scores values are lower as we go to more aggregate levels thanks to smoothing effects. All approaches yield forecast improvements, also at all levels of the hierarchy. The online forecast reconciliation approach based on MRLSE consistently gives the largest forecast improvements, those being larger as one gets towards lower levels of the hierarchy.

TABLE I
IMPACT OF FORECAST RECONCILIATION ON THE QUALITY OF THE FORECASTS, BASED ON THE SRMSE CRITERION [IN % OF NOMINAL CAPACITY] WITH RELATED ISRMSE VALUES [IN %].

		bottom (av.)	regions (av.)	total
SRMSE	base	4.90	1.31	0.703
	MinT	4.81	1.28	0.699
	MLSE	4.65	1.26	0.690
	MRLSE	**4.53**	**1.22**	**0.676**
ISRMSE	base	–	–	–
	MinT	1.84	2.29	0.57
	MLSE	5.1	3.82	1.84
	MRLSE	**7.55**	**6.87**	**3.84**

More than those average values, the distribution of improvements among bottom nodes and regions are of utmost importance. Results are qualitatively similar at these two levels of the hierarchy, hence we place emphasis on bottom nodes since relying on larger populations (100 nodes). Corresponding boxplots are depicted in Fig. 4. While forecast quality improvements are highest on average for our online forecast reconciliation approach based on the MRLSE estimator, there is also a high variability in those improvement. Those are always positive and up to more than 15% for a given site.

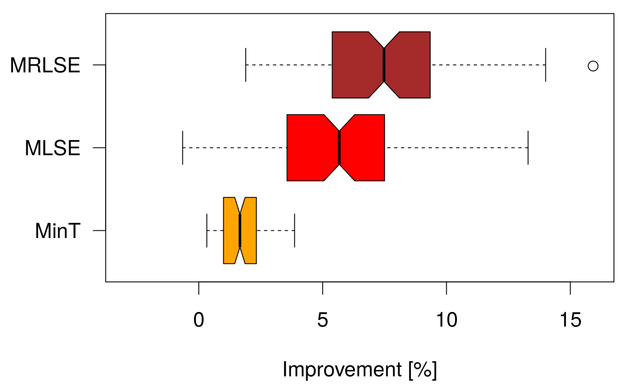

Fig. 4. Distribution of improvements (ISRMSE) for bottom nodes and for the 3 forecast reconciliation approaches.

3) Time-varying aspect of forecast reconciliation: As a motivation for the proposal of an online forecast reconciliation approach, we mentioned the fact that the underlying stochastic processes are nonstationary. As a consequence, we expect that the parameters $\mathbf{\Theta}$ evolve with time throughout the dataset. This is illustrated by Fig. 5 which show the temporal evolution of the coefficients associated to sites 25, 31, and 96 to obtain the reconciled forecast values for the total level. Their evolution combine smoother and higher-frequency fluctuations. Remember that the forgetting factor used is very large (n_λ = 10 000) hence yielding an MRLSE estimator with fairly long memory.

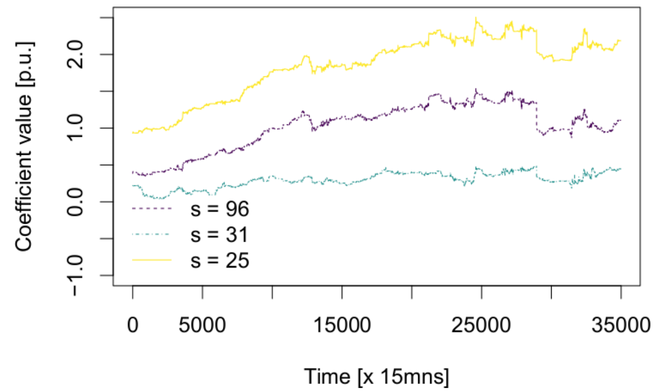

Fig. 5. Evolution of randomly chosen coefficient (for sites 25,31 and 96) contributing to obtaining the reconciled forecasts at total level.

Subsequently we look at the impact of nonstationarity on the quality of the forecasts obtained after forecast reconciliation. Figure 6 gathers monthly IWRMSE values for the 12 months of the verification period, and for the 3 reconciliation approaches considered. The MRLSE estimator, which accommodate nonstationarity, systematically performs better than the MLSE one, for which parameters are static throughout that year. There is also a trend that the improvement from MRLSE increases with time, which is consistent with the fact it is the only approach that aims to accommodate nonstationarity.

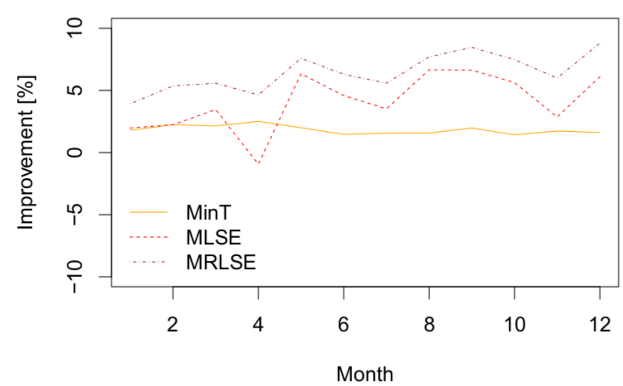

Fig. 6. IWRMSE calculated on a monthly basis through the one-year verification period.

4) Consistency among potential hierarchies: A fairly specific hierarchy was considered. Indeed, Western Denmark is specifically split into 4 quadrants, i.e., contiguous areas with the same number of wind power production sites. However, it is of interest to see how the forecast reconciliation approaches would perform if we were to consider different types of hierarchies. For simplicity, we stick to a 3-level hierarchy and the idea of having the same number (25) of wind power generation sites in each of the mid-level nodes. Consequently we perform a Monte-Carlo simulation experiment, for which instead of considering geographical information, sites are randomly assigned to the 4 regions. Strictly speaking these are not regions anymore, but geographically dispersed portfolios instead. 100 replicates of this Monte-Carlo simulation experiment are used to obtain a distribution of scores values (SRMSE, WRMSE, and related improvements) at the bottom, region and total levels. The results for the ISRMSE criterion are depicted in the form of boxplots, for the various levels and reconciliation approaches, in Fig. 7.

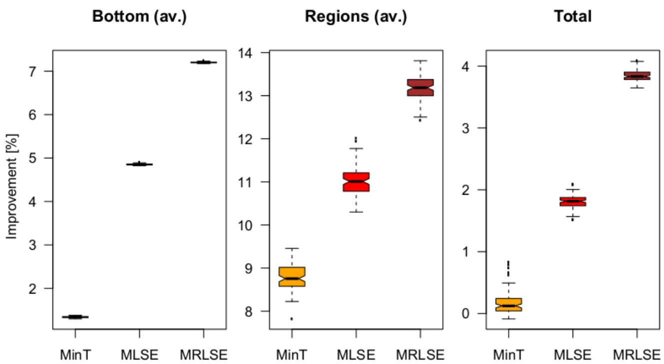

Fig. 7. Boxplots for the distribution of ISRMSE values over a Monte-Carlo experiment with 100 replicates.

At the region and total levels level, there is variability in the forecast improvements obtained, though the online forecast reconciliation through the MRSLE estimator consistently performs best. The variability is highest at the region level, since the structure of the hierarchy highly influences potential forecast improvement. Actually by comparing the results with Table I, one observes that the quadrant based hierarchy is the worst (with much lower ISRMSE values) as it is the worst hierarchy to pick, i.e., with the smallest possible smoothing effect. Such hierarchy randomization study could be extended to the case of having different number of sites per region.

V. Conclusions

A data-driven approach to forecast reconciliation was introduced, in a multivariate regression framework. The main interest of that approach is that it eventually allows for online forecast reconciliation, hence allowing to adapt to nonstationarity in the underlying stochastic processes. A proof of reconciliation by design was also provided, making that, even trained on specific past data, our approach allows to reconcile any new forecasts out of sample.

The case study application concentrated on a fairly simple setup, with 1-step ahead and short-term forecasts only, as the main focus was the reconciliation process, which is independent from the lead time, rather than the forecasting one. The approach may be readily used for multi-step ahead forecasts and day-ahead forecasting, though we expect the results to be qualitatively equivalent.

In addition, the forecast reconciliation problem is seen as centralized, but it could be readily distributed using e.g. ADMM and the likes, since consisting of a convex optimization problem. Similarly, sparsification was not considered here, while it may be clearly of interest to minimize the number of alterations to forecasts in the reconciliation process. This may be considered in the future, the same way MinT has been generalized by allowing for shrinkage. However, this will bring some complexity in the derivation of the online estimator due to the L_1-regularization which is not continuously differentiable. Finally, other types of models may be thought of in a multivariate regression framework. Although we are restricting our model to the linear setting, as done in all reconciliation literature, one could generalize it to the non linear setting, e.g. by using Support Vector Regression, Gradient Boosting or Random Forests. While clearly reconciliation properties would need to be verified in those cases, the non-linear setting would make it possible to account for conditional effects (e.g. from weather conditions and prevailing wind direction) as well as regime-switching, either explicitly or by the use of an adaptive forgetting factor scheme.

Addendum

The proposal in the above is based on the equality constraint in (9b), leading to Theorem 1 that ensures reconciliation by design (for the MLSE estimator, as well as its online version MRLSE). Actually, one can get an even more general version of that result, which does not require the equality constraint, as long as the measurements are themselves additively coherent. This leads to the following corollary to Theorem 1 (which is also valid for the online version MRLS of the MLS estimator). A proof is also given.

Corollary 1 (reconciliation by design of the MLS estimator). *By computing $\hat{\mathbf{\Theta}}_k^{MLS}$ using (14) and given that $\mathbf{Y}_k$ are additively coherent, for any new forecast (out-of-sample) $\hat{\mathbf{y}}_{t+k|t}$, the reconciled forecasts given by $\left(\hat{\mathbf{\Theta}}_k^{MLS}\right)^\top \tilde{\mathbf{y}}_{t+k|t}$ are additively coherent.*

Proof. Given the training dataset of measurements and base forecasts, respectively

$$\mathbf{Y}_k = \begin{bmatrix} \mathbf{y}_{1+k}^\top \\ \vdots \\ \mathbf{y}_{T+k}^\top \end{bmatrix} \text{ and } \hat{\mathbf{Y}}_k = \begin{bmatrix} \tilde{\mathbf{y}}_{1+k|1}^\top \\ \vdots \\ \tilde{\mathbf{y}}_{T+k|T}^\top \end{bmatrix}, \tag{26}$$

the MLS estimator is obtained by

$$\hat{\mathbf{\Theta}}_k^{\text{MLS}} = \left(\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k\right)^{-1} \hat{\mathbf{Y}}_k^\top \mathbf{Y}_k = \mathbf{\Omega}_k \mathbf{Y}_k \,, \tag{27}$$

where

$$\mathbf{\Omega}_k = \left(\hat{\mathbf{Y}}_k^\top \hat{\mathbf{Y}}_k\right)^{-1} \hat{\mathbf{Y}}_k^\top \in \mathbb{R}^{(N+1)\times T}. \tag{28}$$

Breaking down matrices $\mathbf{\Omega}_k$ and $\mathbf{Y}_k$ element-wise, and dropping index k from the element indexing to avoid clutter.

$$\hat{\mathbf{\Theta}}_k^{\text{MLS}} = \begin{bmatrix} \Omega_{1,1} & \dots & \Omega_{1,T} \\ \vdots & & \vdots \\ \Omega_{N,1} & \dots & \Omega_{T,N} \end{bmatrix} \begin{bmatrix} y_{1,1} & \dots & y_{1,N} \\ \vdots & & \vdots \\ y_{T,1} & \dots & y_{T,N} \end{bmatrix} = \tag{29}$$

$$\begin{bmatrix} \sum_{j=1}^T \Omega_{1,j} y_{j,1} & \dots & \sum_{j=1}^T \Omega_{1,j} y_{j,N} \\ \vdots & & \vdots \\ \sum_{j=1}^T \Omega_{N+1,j} y_{j,1} & \dots & \sum_{j=1}^T \Omega_{N+1,j} y_{j,N} \end{bmatrix}. \tag{30}$$

For any new forecast (out-of-sample) $\hat{\mathbf{y}}_{t+k|t}$, the product $\left(\hat{\mathbf{\Theta}}_k^{\text{MLS}}\right)^\top \tilde{\mathbf{y}}_{t+k|t}$ yields

$$\begin{bmatrix} \sum_{j=1}^T \Omega_{1,j} y_{j,1} & \dots & \sum_{j=1}^T \Omega_{N+1,j} y_{j,1} \\ \vdots & & \vdots \\ \sum_{j=1}^T \Omega_{1,j} y_{j,N} & \dots & \sum_{j=1}^T \Omega_{N+1,j} y_{j,N} \end{bmatrix} \begin{bmatrix} 1 \\ \hat{y}_{t,1} \\ \vdots \\ \hat{y}_{t,N} \end{bmatrix} = \tag{31}$$

$$= \begin{bmatrix} \sum_{j=1}^T \Omega_{1,j} y_{j,1} + \dots + \hat{y}_{t,N} \sum_{j=1}^T \Omega_{N+1,j} y_{j,1} \\ \sum_{j=1}^T \Omega_{1,j} y_{j,2} + \dots + \hat{y}_{t,N} \sum_{j=1}^T \Omega_{N+1,j} y_{j,2} \\ \vdots \\ \sum_{j=1}^T \Omega_{1,j} y_{j,N} + \dots + \hat{y}_{t,N} \sum_{j=1}^T \Omega_{N+1,j} y_{j,N} \end{bmatrix}. \tag{32}$$

The first element is equal to the sum of the others (i.e. the forecasts are reconciled) if and only if

$$\begin{cases} \sum_{j=1}^T \Omega_{1,j} y_{j,1} = \sum_{j=1}^T \Omega_{1,j} y_{j,2} + \dots \\ \qquad\qquad\qquad \dots + \sum_{j=1}^T \Omega_{1,j} y_{j,N} \\ \qquad\qquad \vdots \\ \sum_{j=1}^T \Omega_{N+1,j} y_{j,1} = \sum_{j=1}^T \Omega_{N+1,j} y_{j,2} + \dots \\ \qquad\qquad\qquad \dots + \sum_{j=1}^T \Omega_{N+1,j} y_{j,N} , \end{cases} \tag{33}$$

or equivalently,

$$y_{j,1} = y_{j,2} + \dots + y_{j,N} , \qquad \forall j = 1, \dots, T, \tag{34}$$

i.e. when the measurements are consistent. □

It is worth noting that although this is the case in theory and in some practical applications, there might be cases where this consistency is not achieved in practice. Measuring errors, missing data and subsequent imputation techniques could for example affect data coherency and justify reconciliation through $\hat{\mathbf{\Theta}}_k^{\text{MLSE}}$.

Acknowledgements

The authors would like to acknowledge Energinet.dk for providing the wind power dataset used as input to our numerical study. Additional acknowledgements are due to Pierre-Julien Trombe and Peter Nystrup for their comments on previous versions of this work.

References

[1] 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, 2014.

[2] R. J. Bessa, A. Trindade, C. S. P. Silva, and V. Miranda, "Probabilistic solar power forecasting in smart grids using distributed information," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 16-23, 2015.

[3] T. van Erven, and J. Cugliari, "Game-theoretically optimal reconciliation of contemporaneous hierarchical time-series forecasts," in *Modeling and Stochastic Learning for Forecasting in High Dimensions* (A. Antoniadis, J.-M. Poggi, eds.), pp 297-316, 2015.

[4] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization," *Journal of the American Statistical Society*, vol. 114, no. 526, pp. 804–819, 2018.

[5] Y. Zhang, and J. Dong, "Least Squares-based optimal reconciliation method for hierarchical forecasts of wind power generation," *IEEE Transactions on Power Systems*, available online.

[6] S. Ben Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical probabilistic forecasting of electricity demand with smart meter data," Working paper, under review, 2018.

[7] J. Jeon, A. Panagiotelis, F. Petropoulos, "Reconciliation of probabilistic forecasts with an application to wind power," ArXiv preprint: arXiv:1808.02635, 2018.

[8] P. Nystrup, E. Lindström, P. Pinson, and H. Madsen, "Temporal hierarchies with autocorrelation for load forecasting," *European Journal of Operational Research*, available online

[9] L. Bai, P. Pinson, "Distributed reconciliation in day-ahead wind power forecasting," *Energies*, vol. 12, art. no. 1112, 2018.

[10] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, F. Petropoulos, "Forecasting with temporal hierarchies," *European Journal of Operational Research*, vol. 262, no. 1, pp 60-74, 2016.

[11] S. Ben Taieb, B. Koo, "Regularized regression for hierarchical forecasting without unbiasedness conditions," in *Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1337-1347, 2019.

[12] L. Kubáček, "Multivariate regression model with constraints," *Mathematica Slovaca*, vol. 57, n. 3, pp. 271-296, 2007.

[13] P. Pinson, and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *Journal of Forecasting*, vol. 31, no. 4, pp. 281-313, 2012.

[14] R. Girard, and D. Allard, "Spatio-temporal propagation of wind power prediction errors," *Wind Energy*, vol. 16, no. 7, pp. 999-1012, 2013.

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

[16] J. W. Messner, and P. Pinson, "Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1485-1498.

[17] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, and I. Schicker, "Evaluation of wind power forecasts – An up-to-date view," Wind Energy, available online, 2020