

# Reconciling temporal hierarchies of wind power production with forecast-dependent variance structures

Mikkel L. Sørensen, Jan K. Møller and Henrik Madsen

*For electricity grid operators, the planning of grid operations depends on having accurate models for forecasting wind power production on a number of different time resolutions. Together, these time resolutions create what has become known as a temporal hierarchy. Previous studies have considered methods of reconciling forecast hierarchies inspired by least squares methodologies that produce coherent and more accurate forecasts. In this study, we highlight some challenges in the established approach when applied to wind power production, and consider methods which more appropriately take into account the full conditional probability densities. We suggest methods using maximum likelihood techniques to estimate the prediction variance ahead of time. Using base forecasts from a commercial forecast provider together with simpler forecasting models, we test the modified approach against the established reconciliation approach on data from Danish wind farms. The results show significant improvements in accuracy when compared to both the state-of-the-art commercial forecasts and the simpler models.*

## Introduction

As part of the ongoing green transition, the share of renewable energy being incorporated into the electricity grid is increasing. With this comes the challenge of renewable energy sources, such as wind power, being stochastic in nature, i.e., power production depends on, e.g., the wind speed. Thus, in order to plan and control the electricity grid, transmission system operators (TSOs) require reliable and accurate forecasts of power production. With targets such as the European Union having 70% of power production stemming from renewable energy sources by 2050,<sup>1</sup> this challenge is only becoming greater. Hence, the importance of accurate and reliable wind power forecasts is more crucial now than ever before.

The main goal in this respect is thus to reduce the uncertainty of the forecasted power production as much as possible. One of

the more recent developments within forecasting, which has been shown to reduce the forecast uncertainty, is the method of hierarchical forecast reconciliation. The method in its current form was suggested by [7], with further developments by [8, 14, 17]. Hierarchical forecast reconciliation utilizes the hierarchical structures that are often naturally present in data by ensuring that forecasts on different resolutions are coherent. Coherency refers to the fact that the layers should add up throughout the hierarchy in a natural way. An example of this comes from the study on Australian domestic tourism by [1], where forecasts are produced at the national scale, state scale, zone scale, and regional scale. These forecasts form a hierarchy, where one would expect that all the regional forecasts within a zone add up to the zone forecast, all the zone forecasts within a state add up to the state forecast, etc. When forecasts add up in this manner, the forecasts are said to be coherent. However, when the forecasts on each layer are produced independently of each other this will not be the case. The purpose of hierarchical forecast reconciliation is thus to produce a set of reconciled forecasts that have the coherency property based on a set of incoherent base forecasts.

Hierarchies can be either spatial, as was the case for the example on Australian domestic tourism, temporal, as has been explored by [2] and [14], or spatio-temporal, e.g., [9]. In a spatial hierarchy one might examine regional forecasts, see, e.g., [7]. In a temporal hierarchy, which will be the case for this study, the forecasts should add up based on the timescale such that, e.g., the three monthly forecasts within a quarter add up to quarterly forecasts, and the four quarterly forecasts add up to yearly forecasts.

Forecast reconciliation of temporal hierarchies has shown great promise in reducing the uncertainty of the base forecasts and producing more accurate forecasts at all layers [14]. However, the majority of studies on temporal hierarchies only show such improvements using simple models to produce the base forecasts, such as the ARIMA and exponential smoothing models by [2]. These models also often require constant data availability, which under real-time operational conditions cannot always be guaranteed. The established reconciliation approach also assumes that the covariance structure between the forecasts is constant in time,

<sup>1</sup>IRENA, Global energy transformation: A roadmap to 2050; <https://www.irena.org/publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition>.

which for many applications may not be the case. For wind power production it is well-known that the forecast uncertainty is highly dependent on, e.g., the wind speed [13]. Furthermore, wind power production is naturally bounded between zero and the installed capacity, which poses a challenge when reconciling, since the established method of reconciling assumes unbounded data. These challenges call for more refined methods to generate reconciled forecasts.

In the literature, forecast reconciliation has been applied previously to cases of both wind power forecasting and solar power forecasting, which both face similar issues. The paper [16] provides a brief overview of the literature on this. In [6] the authors test some of the current methods for forecast reconciliation in a spatial hierarchy with a similar dataset to our case study and are able to achieve significant improvements in accuracy. They do not, however, account for the issue of the non-constant variance structure. The need for varying and mean-dependent variance when reconciling forecasts has also been highlighted by [3] for the case of heat load forecasts, although the approach therein mostly addresses slower seasonal variations. Additionally, [18] demonstrates a method of changing the forecast reconciliation methodology to account for non-negativity.

The purpose of the present study is thus:

- To use real measurements and weather forecasts to examine whether improvements to commercial state-of-the-art wind power production forecasts can be achieved through forecast reconciliation.
- To propose adjustments to the established reconciliation approach (MinT-Shrink by [17]), which addresses the level and time-dependent covariance structure in the wind power framework.

This study is structured such that Section 1 introduces the concept of forecast reconciliation. Section 2 presents the wind power production data used in the study. Section 3 then examines how base forecasts are produced, and introduces the proposed adjustments to the reconciliation process. Section 5 shows the results, and lastly Section 6 concludes on the findings and gives some further perspectives.

## 1 Forecast reconciliation

For the temporal hierarchy in this study, prediction horizons up to 24 hours will be examined. So, to introduce the concepts of reconciling forecasts based on temporal hierarchies, a simple example hierarchy is constructed within the 24-hour prediction frame with 6-hour, 12-hour and 24-hour forecast resolutions. This example could, e.g., cover the total wind power production in the stated intervals. The hierarchical structure of this example is illustrated in Figure 1.

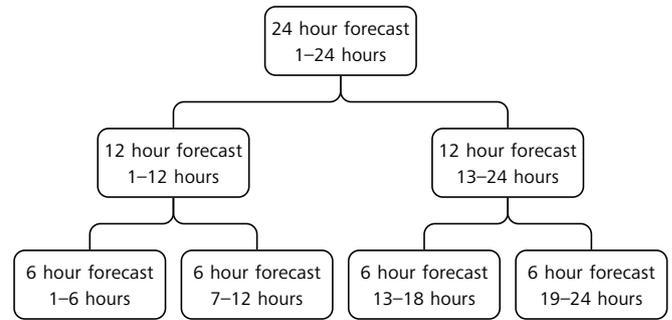


Figure 1. Temporal hierarchy with three levels covering a 24-hour period.

As illustrated by the figure, for such a hierarchy of forecasts to be coherent, the bottom-level 6-hour forecasts should add up pairwise to the second level 12-hour forecast, and the two 12-hour forecasts should add up to the highest level daily forecast. This summation property of the hierarchy is described by a summation matrix  $S \in \mathbb{R}^{n \times m}$ , where  $n$  is the number of individual base forecasts (in this case 7) and  $m$  is the number of forecasts on the lowest resolution (in this case 4). For the hierarchy in Figure 1, the summation matrix is given by:

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The rows of the summation matrix define how the four bottom-level forecasts add up to each level of the hierarchy, so, e.g., the top-level forecast should be equal to the sum of all the four bottom-level forecasts. This also implies that the base forecasts ( $\hat{y} \in \mathbb{R}^n$ ) are ordered from top to bottom, i.e.

$$\hat{y} = \begin{bmatrix} \hat{y}_{24h}^{24h} \\ \hat{y}_{24h}^{12h} \\ \hat{y}_{12h}^{12h} \\ \hat{y}_{24h}^{6h} \\ \hat{y}_{18h}^{6h} \\ \hat{y}_{12h}^{6h} \\ \hat{y}_{6h}^{6h} \end{bmatrix}. \quad (1)$$

For clarity, the notation in this study uses superscripts for naming or other distinctions and subscripts for indices, so, e.g.,  $\hat{y}_{18h}^{6h}$  is the forecast on the 6-hour resolution 13 to 18 hours ahead.

With the summation matrix and the base forecasts as defined above, the most commonly used method of forecast reconciliation, the MinT-Shrink [17], can be introduced.

The method is based on a generalized least squares approach, which minimizes the coherency errors, i.e., the difference between the reconciled forecast ( $\tilde{y} \in \mathbb{R}^n$ ) and the base forecast ( $\hat{y}$ ). This implies solving the following optimization problem

$$\min_{\tilde{y}} (\tilde{y} - \hat{y})^T \Sigma (\tilde{y} - \hat{y}) \quad \text{subject to } \tilde{y} = \mathbf{S} \mathbf{G} \tilde{y},$$

where  $\Sigma \in \mathbb{R}^{n \times n}$  is the covariance matrix of the coherency errors  $\tilde{y} - \hat{y}$ , and  $\mathbf{G} = [\mathbf{0}_{n \times (n-m)} \quad \mathbf{I}_m]$ . If  $\Sigma$  is assumed to be known, the problem has a closed form solution, given by

$$\tilde{y} = \mathbf{S} (\mathbf{S}^T \Sigma^{-1} \mathbf{S})^{-1} \mathbf{S}^T \Sigma^{-1} \hat{y}. \quad (2)$$

In [17] the authors showed that  $\Sigma$  is not identifiable; instead, they propose the trace minimization approach (MinT), which uses the covariance matrix of the base forecast errors, i.e.,

$$\Sigma = V[\varepsilon],$$

where  $\varepsilon = y - \hat{y}$  with  $y \in \mathbb{R}^n$  being the observations arranged similarly to the base forecasts in equation (1). This approach has shown great promise in increasing forecast accuracy across all levels of the hierarchy. MinT-Shrink, also proposed by [17], and extended to temporal hierarchies by [14], is an extension of the MinT approach which applies a shrinkage estimator to the correlation.

The method estimates the covariance  $\Sigma^{\text{shrink}}$  from the correlation matrix of the base forecast errors  $\mathbf{R}$  using the shrinkage parameter  $\lambda$ :

$$\mathbf{R}^{\text{shrink}} = (1 - \lambda) \mathbf{R} + \lambda \mathbf{I}_n, \quad (3)$$

$$\Sigma^{\text{shrink}} = \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{R}^{\text{shrink}} \mathbf{\Lambda}^{\frac{1}{2}}, \quad (4)$$

where  $\mathbf{\Lambda}$  is the so-called hierarchical variance, which is a diagonal matrix consisting of the empirical variances of the base forecast errors at each level. So, for the example hierarchy in Figure 1, the hierarchical variance is

$$\mathbf{\Lambda} = \begin{bmatrix} (\sigma_{24h}^{24h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (\sigma_{24h}^{12h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (\sigma_{12h}^{12h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (\sigma_{24h}^{6h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (\sigma_{18h}^{6h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & (\sigma_{12h}^{6h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (\sigma_{6h}^{6h})^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is usually assumed that both  $\mathbf{\Lambda}$  and  $\mathbf{R}$  are either constant or slowly varying (see, e.g., [3]) in time. However, as mentioned in the introduction, this might not be a reasonable assumption for wind power forecasts, since the uncertainty is directly coupled to the wind power generation through the power curve, as discussed by, e.g., [10]. A similar relation between the uncertainty and the prediction is probably also needed in other cases, like for solar power forecasting.

## 2 Wind power data

The hierarchy that will be explored in this study has eight layers ranging from the 1-hour resolution up to a 24-hour resolution, i.e., the layers will have resolutions 1, 2, 3, 4, 6, 8, 12, and 24 hours. To perform the reconciliation, base forecasts need to be obtained for each temporal resolution in the hierarchy.

The data that these forecasts should be generated from consists of hourly measurements of wind power production from on-shore wind turbines, as well as numerical weather predictions (NWP) of wind speed and wind direction. These data are available for 15 areas of the DK1 region in western Denmark (Jutland and Funen), as seen in Figure 2. The areas have very different capacities for power production, ranging from as low as 39.3 MW to 746.7 MW, and each of the areas will be forecasted and reconciled individually using temporal hierarchies. The same region has also been examined by [6] using a spatial hierarchy.

Data were available from the beginning of January 2018 and through December 2019, and hence the data were split such that 2018 data were used for training base forecast models and estimating shrinkage parameters, while 2019 data are used for testing the reconciliation out of sample. Out of sample testing will be based on a rolling window containing 12 months of previous measurements and forecasts. This window rolls forward in time, one day at a time, as the data would become available. Each day the coefficients of the base forecast models are re-estimated, and new base forecasts for the following day are produced and reconciled. This mimics how operational wind power forecasting takes place.

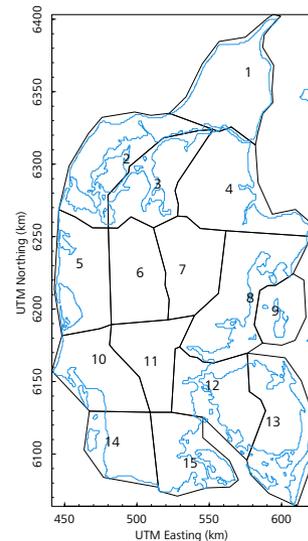


Figure 2. Map of the 15 subareas of the DK1 price area in western Denmark. Coastline in blue.

Area	Rated power [MW]
1	428.1
2	400.9
3	253.3
4	360.8
5	746.7
6	328.0
7	86.9
8	100.4
9	39.3
10	201.6
11	219.7
12	97.3
13	194.4
14	230.2
15	75.6

Table 1. Rated power production for each of the 15 subareas in the DK1 price area given in MW.

From each of the areas depicted, hourly measurements of wind power production, as well as forecasts of wind speed and direction, are available. The power production data is courtesy of Energinet, the TSO that operates the Danish high-voltage transmission grids. The weather forecasts were made by the Danish Meteorological Institute (DMI).

Commercial state-of-the-art forecasts of hourly wind power production were available for the 1-hour resolution (courtesy of ENFOR A/S). These were produced using their wind power forecasting tool WindFor,<sup>2</sup> and will constitute the lowest level of the temporal hierarchy.

Table 2 below summarizes which data are available along with the notations in this study.

Variable	Unit	Description
$y$	kWh	Measurements of power production
$\hat{y}$	kWh	Forecasts of power production
$\widehat{WSpd}$	m/s	Forecasted wind speed
$\widehat{WDir}$	degrees	Forecasted wind direction

Table 2. Data descriptions for variables and the used units. These are all on 1-hour time resolution for each of the 15 areas.

The measurements  $y$  are settlement measurements, meaning that they are generally available 8–10 days after the measured production has occurred. This fact will have to be accounted for when modelling the power production, since not having real-time data means that some models, such as the classic ARIMA models, are not applicable. The measurements also include periods of missing data. For shorter periods ( $\leq 6$  hours), measurements were linearly interpolated, while longer periods were left missing, meaning that models could not be updated here.

The forecasts in this study are generated every night at midnight (00:00) and cover the following 24-hour period. The forecasts of wind speed and direction are generated every six hours and are available a few hours later. To generate forecasts at midnight, the NWP for the following day were taken from the weather forecast generated at 18:00. This also means that the forecast horizon follows the time of day.

The time series in Figure 3 shows some of the volatility in the data, as changes from low production to high production and vice versa can happen within a couple of hours.

The relation between wind speed and power production seen in Figure 3 shows the classic power curve structure, with power production increasing slowly at first until about 4 m/s, then rapidly increasing above the 4 m/s mark, and flattening out when the fore-

casted wind speed approaches 8 m/s. This power curve structure is also what should be mimicked when modelling the individual layers of the hierarchy.

Figure 3 also illustrates how the variance in power production depends on the wind speed, and hence on the predicted wind power generation. This means that the wind power production cannot be assumed to be Gaussian, since the variance would then be independent of the power production.

Lastly, it is worth noting that the power production does not reach the rated production in the available data. This is likely due to the fact that the turbines within the areas are positioned differently in the terrain, making it very rare for all of them to generate according to their rated power simultaneously.

## 2.1 Aggregation

In order to construct forecasts for the other levels of the hierarchy, the data had to be aggregated up to the 24-hour resolution. Thus, datasets of power production and weather predictions need to be constructed for the 2, 3, 4, 6, 8, 12, and 24-hour resolutions. This is done by aggregating the 1-hour measurements and NWP. Power production measurements are aggregated by simply adding up measurements, while wind speed and direction forecasts are aggregated by averaging the forecasts. For wind speed, this is a simple arithmetic average, while for the direction, the average has to be computed using the circular mean. This implies converting the angular directions to points on the unit circle, taking the arithmetic mean of the points, with the resulting angle being the circular mean.

At the 1-hour level, the 12 months of training data contain 8760 measurements of power production and as many forecasted values of wind speed and direction. As a result of the aggregation, the 2-hour level has 4380 data points, the 3-hour level has 2920 data points, and so on, up to the 24-hour level, which has 365 data points. This reduction in data quantity puts a limit on how well the higher levels can be modelled. Consequently, if a larger temporal hierarchy should be attempted, more data would be required.

## 3 Base forecasts

When modelling the levels above the 1-hour resolution, i.e., 2, 3, 4, 6, 8, 12, and 24 hours, it was decided to use beta regression models. These purely statistical models are chosen mainly to address the challenge of power measurements not being available in real time. However, this is also a well-established method of predicting wind energy, see, e.g., [15]. Additionally, choosing such simple models can also show if improvements to already high-quality commercial forecasts are possible using forecast reconciliation, even without employing state-of-the-art forecasts for the other levels.

The beta regression models used in this study are constructed using the `betareg` package in R [4]. The following brief intro-

<sup>2</sup> <https://enfor.dk/services/windfor/>

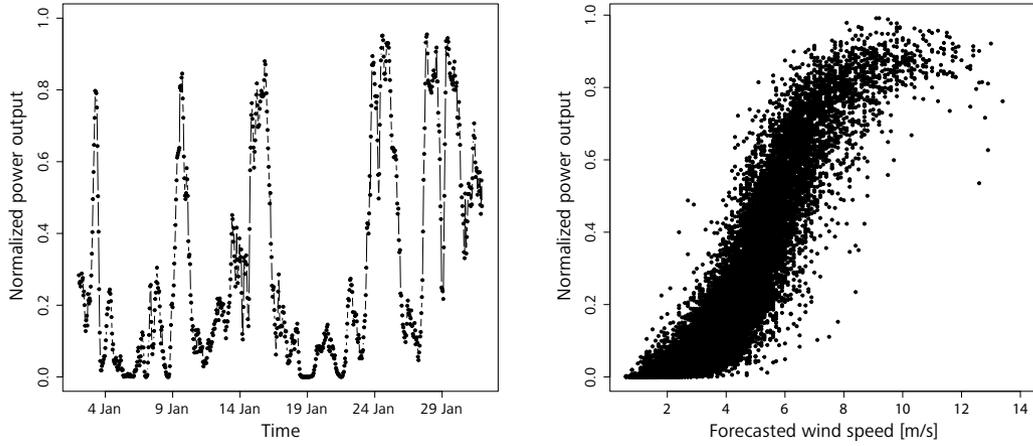


Figure 3. Left: Time series plot of the wind power production in area no. 3 in northern Jutland from January 2018. Right: Relation between forecasted wind speed for all horizons and their corresponding normalized power productions in area no. 3.

duction leans heavily on the documentation for the package, so for a more in-depth walk-through of these methods, see the documentation by [4].

We model the normalized power production as a beta distribution  $Y_t \sim \mathcal{B}(\mu_t, \phi_t)$ , where  $\mu_t \in (0, 1)$  is the mean value ( $E[Y_t] = \mu_t$ ) and  $\phi_t > 0$  is the precision. Therefore, the variance of the power production at time  $t$  can be expressed as

$$V[Y_t] = \frac{\mu_t(1 - \mu_t)}{1 + \phi_t}. \quad (5)$$

In order to be able to introduce explanatory variables the parameters  $\mu_t$  and  $\phi_t$  are modelled individually by

$$g_1(\mu_t) = x_t^T \beta, g_2(\phi_t) = z_t^T \gamma.$$

Here  $\beta$  and  $\gamma$  are regression coefficients, and  $x_t$  and  $z_t$  are regressors;  $g_1$  and  $g_2$  are link functions [11], where  $g_1$  was chosen to be the logit function and  $g_2$  to be the natural logarithm. Given the assumed density, the coefficients  $\beta$  and  $\gamma$  are estimated by maximum likelihood estimation.

When constructing the models for the hierarchy, a method of forward selection was used based on the Wald test with significance level  $p = 0.05$ . For modelling the mean  $\mu_t$ , the initial model consisted only of the forecasted wind speed. The two following modelling steps then added the second and third power of the wind speed. The fourth power did not provide a significant improvement and was thus dropped. Next, the contribution of the wind direction was considered by adding the sine and cosine of the wind direction. However, only the cosine was significant, resulting in the following final model for the conditional mean:

$$g_1(\mu_{t+h|t}) = \theta_0 + \beta_1 \widehat{WSpd}_{t+h|t} + \beta_2 \widehat{WSpd}_{t+h|t}^2 + \beta_3 \widehat{WSpd}_{t+h|t}^3 + \beta_4 \cos(\widehat{WDir}_{t+h|t}),$$

where  $\widehat{WSpd}_{t+h|t}$  is the forecasted wind speed from the NWP available at time  $t$  for horizon  $h$  and  $\widehat{WDir}_{t+h|t}$  is similarly the forecasted wind direction. Since the forecasts are produced daily at midnight,  $t$  changes in steps of 24 hours, i.e.,  $t = 0, 24, 48, 72, \dots$

When modelling the precision  $\phi$ , a very similar procedure was used. The wind speed terms were significant up to the second order, and only the sine of the wind direction was significant. Additionally, the prediction horizon of the weather forecast was added as a regressor. This was added to address the general tendency that the uncertainty is increasing with the prediction horizon. In the end, the precision is modelled as

$$g_2(\phi_{t+h|t}) = \nu_0 + \gamma_1 \widehat{WSpd}_{t+h|t} + \gamma_2 \widehat{WSpd}_{t+h|t}^2 + \gamma_3 \sin(\widehat{WDir}_{t+h|t}) + \gamma_4 h, \quad (6)$$

where  $h$  is the prediction horizon of the weather forecast.

The power curve (Figure 4) resulting from fitting the model to the 2018 training set seems to fit well with what was expected from the data (Figure 3). However, it seems that the wind direction plays a very small role, as the four wind directions (N, E, S, W) are indistinguishable (Figure 4, top).

With forecasted weather variables, the uncertainty of the NWP is translated to power by the non-linear power curve. As nicely illustrated by [10], this explains the form of the standard deviation in Figure 4 (bottom), and this is also in line with what was seen by [12]. Furthermore, the standard deviation clearly depends on the prediction horizon.

Examining the residuals of the fitted models for the area on each resolution, the accuracy can be assessed by the *root-mean-square error* (RMSE), which is computed as

$$\text{RMSE}_{\text{area},l} = \sqrt{\frac{\sum_{t=1}^{N_l} (y_{l,t} - \hat{y}_{l,t})^2}{N_l}},$$

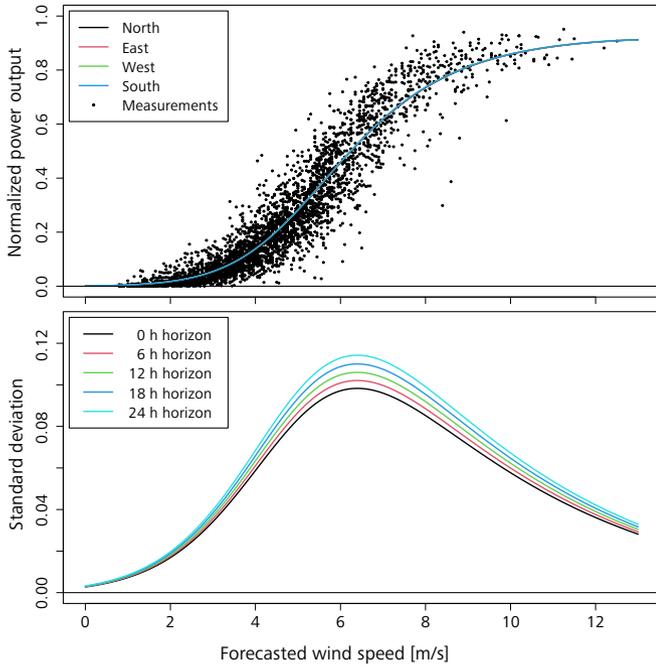


Figure 4. Top: Power curve resulting from fitting the beta regression model to the aggregated 2-hour data for area no. 3. Bottom: Standard deviation of fitted values from beta regression model on aggregated 2-hour data for area no. 3, given the wind direction is fixed in a northerly direction.

where  $y_t$  are the wind power measurements at time  $t$ ,  $\hat{y}_t$  are the predicted values at time  $t$ , and  $N_l$  is the amount of data points at resolution  $l$ , i.e. for the 2018 training set  $N_{1h} = 8760$ ,  $N_{2h} = 4380$  etc.

While the RMSE values as seen in Table 3 increase gradually with the temporal resolution, it is seen that relative to the rated power all forecasts perform well. It should, however, be noted that this is not a fair comparison between the 1-hour forecast and the

Model resolution	RMSE[MW]	% of rated power
1 hour	15.50	6.12
2 hours	33.39	6.59
3 hours	48.04	6.32
4 hours	62.26	6.14
6 hours	88.89	5.85
8 hours	112.58	5.55
12 hours	155.79	5.12
24 hours	265.12	4.36

Table 3. Accuracy of the different models for area no. 3 on the 2018 training data. 1-hour predictions were made by a state-of-the-art commercial model, while those for 2 hours and beyond are from the beta regressions.

other resolutions, as the forecasts on the 1-hour resolution were made out of sample in real time, while the forecasts on the other resolutions are in sample.

#### 4 Forecast reconciliation methodology

With the base forecasts that have been generated, the forecasts in the temporal hierarchy can be reconciled. With the MinT-Shrink method, this implies computing the hierarchical variance  $\Lambda$  and the correlation matrix  $R$ , inserting them in the equations (3) and (4), and finally computing the reconciled forecast by equation (2).

This on its own is likely to improve the accuracy of the base forecasts, as seen in, e.g., [5] for wind power. However, we propose an alternative which better addresses some of the inherent challenges calling for a functional relationship between the uncertainty (conditional variance) and the predicted wind power generation (conditional mean).

One such challenge is that the boundedness of wind power production should be accounted for, since wind power is inherently bounded both from below at zero and from above at the rated power production. This means that the covariance will depend on the forecasted power production. This dependence is not accounted for in the MinT-Shrink method, where the covariance is assumed to be constant.

To address this challenge, we propose a method for linking the hierarchical variance  $\Lambda$  to the predicted power production.

In each time step  $t$ ,  $\Lambda_t$  will be set based on a parametric estimate of the prediction uncertainty. In the case of the example hierarchy from Figure 1, at time  $t$ ,  $\Lambda_t$  becomes

$$\Lambda_t = \text{diag} \begin{pmatrix} (\hat{\sigma}_{t+24h}^{24h})^2 \\ (\hat{\sigma}_{t+24h}^{12h})^2 \\ (\hat{\sigma}_{t+12h}^{12h})^2 \\ (\hat{\sigma}_{t+24h}^{6h})^2 \\ (\hat{\sigma}_{t+18h}^{6h})^2 \\ (\hat{\sigma}_{t+12h}^{6h})^2 \\ (\hat{\sigma}_{t+6h}^{6h})^2 \end{pmatrix}.$$

This forecast-dependent hierarchical variance  $\Lambda_t$  then replaces the constant hierarchical variance  $\Lambda$ . The shrinkage equations thus become

$$\mathbf{R}^{\text{Var}} = (1 - \lambda^{\text{Var}})\mathbf{R} + \lambda^{\text{Var}}\mathbf{I}_n, \quad \Sigma_t^{\text{Var}} = \Lambda_t^{\frac{1}{2}}\mathbf{R}^{\text{Var}}\Lambda_t^{\frac{1}{2}},$$

which will be denoted MinT-Var.

This approach adapts the variance to the forecast, but not the correlation, which is still estimated from the response residuals. The next step is to allow the correlation structure to change with the forecast. One way of doing this is to compute the correlation from Pearson residuals instead, which are normalized by the forecast-dependent variance.

Additionally, it is well-known from the theory of generalized linear models that Pearson residuals will be closer to normal distributions than the original response residuals, see e.g. [11].

For a set of forecasts at time  $t$ , the Pearson residuals may be expressed as

$$\varepsilon_t^p = \mathbf{\Lambda}_t^{-1/2}(y_t - \hat{y}_t) \sim N(0, \mathbf{\Sigma}^p).$$

The second proposal is thus to use both the correlation of the Pearson residuals  $R^p$  and a forecast-dependent variance, i.e.

$$\mathbf{R}^{\text{PVar}} = (1 - \lambda^{\text{PVar}})\mathbf{R}^p + \lambda^{\text{PVar}}\mathbf{I}_n, \quad \mathbf{\Sigma}_t^{\text{PVar}} = \mathbf{\Lambda}_t^{\frac{1}{2}}\mathbf{R}^{\text{PVar}}\mathbf{\Lambda}_t^{\frac{1}{2}},$$

which will be denoted MinT-PVar.

These approaches lead to a flexible and more correct description of the uncertainty than the original method, while also addressing the challenges imposed when reconciling wind power forecasts. The performance of the methods will, however, depend on how well the prediction uncertainty is estimated. This then raises the question of how one should estimate the prediction uncertainty.

In our case, where the beta regressions are used to generate base forecasts, the variance of the predictions is given by equation (5). Conversely, the variance on the 1-hour resolution is not known, since the underlying model is not available. Therefore, for the 1-hour resolution the prediction uncertainty will have to be estimated by another method. Here, a well-known and effective way of doing this is through maximum likelihood estimation, which also allows us to describe how the conditional variance depends on the conditional mean.

Assuming that residuals are normally distributed, a likelihood function can be established for each prediction horizon  $h$ . For the set of  $k$  observations  $\mathbf{y}_{t+h} = \{y_{t-24k+h}, \dots, y_{t-24+h}\}$  and the corresponding set of forecasts  $\hat{\mathbf{y}}_{t+h} = \{\hat{y}_{t-24k+h}, \dots, \hat{y}_{t-24+h}\}$ , the log-likelihood will be computed as

$$\begin{aligned} \ell(\hat{\sigma}_{t+h}^2; \hat{\mathbf{y}}_{t+h}, \mathbf{y}_{t+h}) &= \sum_{i=1}^k -\frac{k}{2} \log(\hat{\sigma}_{t+h}^2) \\ &+ \left( -\frac{1}{2\hat{\sigma}_{t+h}^2} (y_{t-24i+h} - \hat{y}_{t-24i+h})^2 \right). \end{aligned}$$

This requires a parametrization of the variances  $\sigma_t^2$ , for which inspiration can be drawn from the beta regression models for the precision as well as the relation between variance and precision given in equation (5). This results in the following parametrization of the variance:

$$\begin{aligned} \hat{\sigma}_{t+h}^2 &= \exp(\alpha_{0,h}) \\ &+ \frac{\hat{y}_{t+h}(1 - \hat{y}_{t+h})}{1 + \exp(\alpha_{1,h} + \alpha_{2,h}\widehat{\text{WSpd}}_{t+h} + \alpha_{3,h}\widehat{\text{WDir}}_{t+h})}. \end{aligned}$$

This parametrization represents a slight simplification of the precision model (equation (6)), as the second-order term is omitted. This simplification is justified by examining the *Akaike information criterion* (AIC) of the likelihood functions, with the full precision

model and with the simplified model both fitted on the training data. The simpler model had a lower AIC, and hence it was chosen instead of the full precision model. Furthermore, this parametrization has an added constant term determined by  $\alpha_{0,h}$ . The addition of this term improved the robustness of the optimization and lowered the AIC significantly. Moreover, it is in line with the behaviour of the variance seen in Figure 4 not going fully to zero when there is no wind. Figure 5 shows contour plots of the estimated prediction uncertainty with parameters estimated in sample for area no. 3.

The uncertainty seems to be dependent linearly on the wind speed, while the uncertainty varies sinusoidally with the wind direction, topping at around  $270^\circ$ , corresponding to a westerly wind direction. This is in line with the behaviour observed by, e.g., [13]. The uncertainty related to power output quite clearly follows the polynomial relation from the variance model, with maximum at around  $\hat{y}_{t+h} = 0.5$ . For the prediction horizon (Figure 5, right column) the picture is less clear. The first few hours have quite low uncertainty, increasing gradually before spiking around the 10-hour horizon, then dropping down again for a few hours before again spiking around the 19–21-hour mark.

To find the coefficients  $\alpha_h$  of the model, the log-likelihood is numerically optimized in each time step  $t$  for each horizon  $h$  using a window of past data. The resulting variance estimate can then be used in the reconciliation of the forecast.

Lastly, to estimate the shrinkage intensities, i.e.,  $\lambda$ ,  $\lambda^{\text{Var}}$ ,  $\lambda^{\text{PVar}}$  we follow [14] and optimize them by doing a grid search, discretizing the regularization parameter on a grid of  $\lambda = 0, 0.01, 0.02, \dots, 1$ . The parameters are then chosen based on the greatest in-sample improvements to RMSE for each area. The resulting optimal parameter is then used for the out-of-sample tests.

## 5 Results

In this section, the out-of-sample performance of the three different methods for reconciling will be examined, those being the MinT-Shrink as proposed by [17], and the two proposed in this study, MinT-Var and MinT-PVar. For testing, a rolling window will move one day at a time. Each day, the coefficients of the base forecast models are re-estimated, and new base forecasts for the following day are produced. Then the variances of the base forecasts are estimated. In the case of the beta regressions, this is done directly using equation (5), and for the commercial forecasts on the 1-hour resolution, maximum likelihood estimation is performed using the windowed data to estimate coefficients. Then the base forecasts are reconciled using each of the three methods. Lastly, the covariances of the residuals are computed, so they can be used for the reconciliation of the following day's data.

To evaluate the forecasts for individual areas, the RRMSE% (*percentage relative root-mean-square error*) is examined as a way

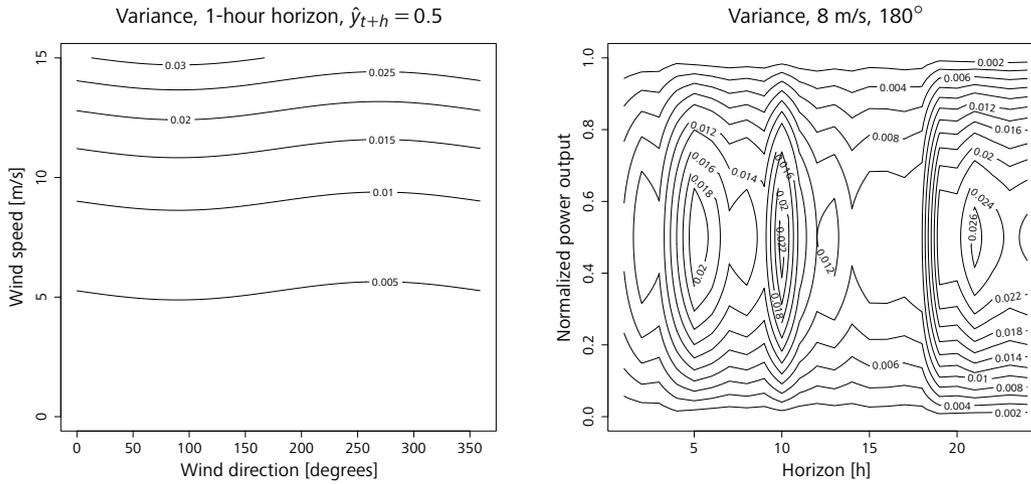


Figure 5. Left: Contour plot of variance for wind speed and direction; 1-hour prediction horizon and normalized power production of 0.5. Right: Contour plot of variance for prediction horizon and power output; 8 m/s wind speed and 180° wind direction.

to evaluate the performance and measure the accuracy of the proposed methods. For each area and for each resolution ( $l$ ), the RRMSE% is given as

$$\text{RRMSE}\%_{\text{area},l} = 100\% \left( \frac{\text{RRMSE}_{\text{area},l}}{\text{RRMSE}_{\text{area},l}^{\text{base}}} - 1 \right).$$

When evaluating the forecasts for the entire region as a whole, we use the total RRMSE%, since the areas have very different rated power productions. Hence for each resolution ( $l$ )

$$\text{RRMSE}\%_l^{\text{total}} = 100\% \left( \frac{\sum_{\text{area}=1}^{15} \text{RRMSE}_{\text{area},l}}{\sum_{\text{area}=1}^{15} \text{RRMSE}_{\text{area},l}^{\text{base}}} - 1 \right).$$

Similarly, we will examine the performance for the entire region based on forecast horizon. We previously introduced the  $\text{RMSE}_{\text{area},l}$  which sums over all prediction horizons. Therefore, we introduce the hourly RMSE as well, which based on the 12 months of results, for a given area, horizon  $h$  and resolution  $l$ , is given as

$$\text{RRMSE}_{\text{area},l,h} = \sqrt{\frac{\sum_{i=1}^{365} (y_{h-24i} - \hat{y}_{h-24i})^2}{365}}.$$

Again, summing across areas, we get

$$\text{RRMSE}\%_{l,h}^{\text{total}} = 100\% \left( \frac{\sum_{\text{area}=1}^{15} \text{RRMSE}_{\text{area},l,h}}{\sum_{\text{area}=1}^{15} \text{RRMSE}_{\text{area},l,h}^{\text{base}}} - 1 \right).$$

For all of these scores, if the RMSE of the reconciled forecast is lower than the RMSE of the base forecast, the RRMSE% will be negative, indicating an increase in the accuracy of the forecast. Alternatively, if the accuracy is decreased, the RRMSE% will be positive.

To get a general picture of the performance of the three methods for reconciling wind power forecasts, we examine the  $\text{RRMSE}\%_l^{\text{total}}$  for all the different temporal resolutions in Table 4.

Table 4 shows that forecasts on all temporal resolutions are improved, with greater improvements in accuracy for the higher temporal resolutions. As expected, for the 2–24-hour resolutions significant improvements in accuracy are observed when reconciled with the commercial state-of-the-art 1-hour forecasts. However, for the 1-hour forecast resolution, improvements are also seen. With the exception of the 24-hour resolution, the two proposed methods both outperform MinT-Shrink on all resolutions, with MinT-PVar showing the best improvements on all resolutions. The difference between MinT-Var and MinT-PVar is quite minor, which tells us that the forecast-dependent variance structure is the main source of the improvements. However, adapting the correlation structure also clearly gives an improvement.

Since the base forecasts on the 1-hour resolution are state-of-the-art and used commercially, improvements here are significantly more valuable for operational purposes than the other resolutions. Therefore, these improvements are examined in further detail in

Temporal resolution [h]	MinT-Shrink	MinT-Var	MinT-PVar
24	-19.34	-19.11	-19.43
12	-17.12	-17.43	-17.74
8	-14.96	-15.13	-15.38
6	-13.71	-13.84	-14.05
4	-13.73	-14.04	-14.26
3	-13.47	-13.85	-13.99
2	-12.87	-13.29	-13.46
1	-1.46	-1.95	-2.08

Table 4.  $\text{RRMSE}\%_l^{\text{total}}$  of the three methods for each temporal resolution. Greatest reductions are highlighted in bold.

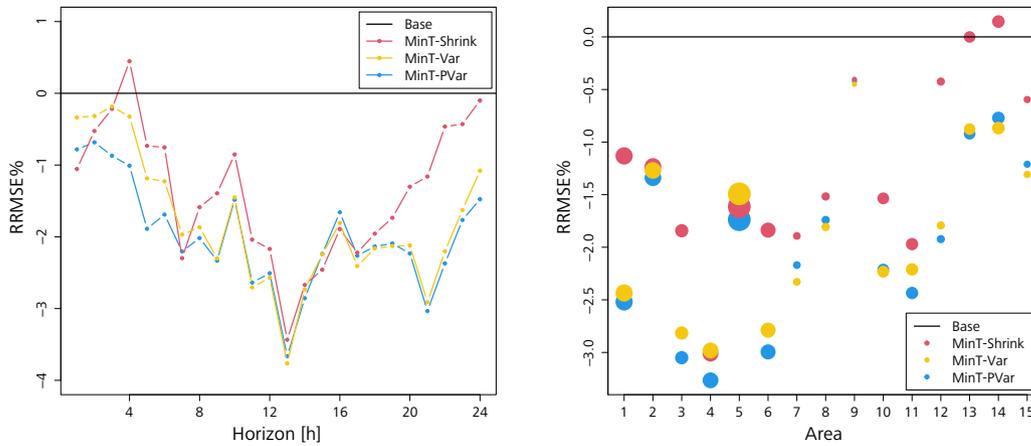


Figure 6. Left:  $RRMSE\%_{i,1h}^{total}$  for each of the three methods across the 24-hour prediction period. Right:  $RRMSE\%_{area,1h}$  for each method on each of the 15 areas. The size of the points has been scaled relative to the rated power production of the area.

the rest of the section. Figure 6 shows how the improvements on the 1-hour resolution are distributed across the 24-hour prediction horizon and across the 15 areas.

The three methods that have been tested are able to improve the accuracy of the base forecast across almost all horizons. Only MinT-Shrink falls below the base forecast in accuracy, and only for a single horizon. There seems to be a pattern to the accuracy improvements for all methods, where the RRMSE drops until the 13-hour horizon and then rises until the end of the 24-hour cycle.

Looking at how the methods perform for each area, MinT-Var and MinT-PVar once again come out as the best methods. In all areas tested, either MinT-Var or MinT-PVar had the best performance. MinT-Var falls behind MinT-Shrink in areas no. 4 and 5, while MinT-PVar is consistently better than MinT-Shrink. This highlights the importance of adapting the correlation to the forecast, as for some areas only adapting the variance is clearly not enough.

Altogether the results show that a reconciliation based on temporal hierarchies can improve even already state-of-the-art commercial wind power forecasts. Although the average improvement of approximately 2% does not seem like much, such improvements are still significant for what are already high-quality forecasts. Further applying the variance estimation techniques proposed in this study results in additional improvements in accuracy, especially when combined with the use of Pearson residuals for correlation estimation.

## 6 Discussion and conclusion

The investigations performed in this study have shown that, even when using rather simple base forecasts for all the aggregated levels, commercial state-of-the-art base forecasts can be improved

using forecast reconciliation based on temporal hierarchies. This finding is very promising, as integrating the method of forecast reconciliation into commercial forecasts can thus help meet the demand for increasingly accurate forecasts. The additional coherency property will also be advantageous for the TSOs, as the planning across multiple time horizons becomes simpler and more coherent.

As discussed throughout the study, reconciling forecasts of wind power production poses some challenges in terms of the non-constant variance. The variance will depend greatly on, e.g., the wind speed. Introducing a method which adapts the variance structure to the weather forecast thus helps in addressing this challenge. This more rigorous approach to reconciliation reflecting the conditional variance dependency on the prediction has resulted in better accuracy of the reconciled forecast.

For cases from other fields where data are close to Gaussian, it is more reasonable to assume that the covariance is constant or at least independent of the mean. Hence, we expect less difference between our proposed method and the MinT-Shrink in these cases.

There is, however, still much that can be done to build upon the results found here, and as such these results can be seen as a step towards an improved methodology for modelling the required covariance structure for reconciling forecasts in cases with non-constant variance. Furthermore, this can help build toward a more general understanding of the uncertainty in forecasts when using hierarchies.

Other authors have also developed methods for handling, e.g., bounded data [18]. It would therefore be interesting to examine how this performs compared to our proposed adjustments, or if these methods could be combined in a meaningful way. Due to the limits imposed by the data in this study, further testing on different data would also be beneficial to more confidently show which method is preferable when reconciling forecasts of wind power production.

*Acknowledgements.* We extend a sincere thank you to Krzysztof Burnecki for kindly inviting us to write this article.

*Funding.* This work is supported by *SEM4Cities* (Innovation Fund Denmark, no. 0143-0004), *ARV* (H2020-101036723), *ELEXIA* (EU HE-101075656), *IEA Wind Task 51* (EUDP nr. 134-22015), and finally the projects *PtX*, *Sector Coupling and LCA* and *DynFlex*, which both are part of the Danish Mission Green Fuel portfolio of projects.

## References

- [1] G. Athanasopoulos, R. A. Ahmed and R. J. Hyndman, [Hierarchical forecasts for Australian domestic tourism](#). *Int. J. Forecast.* **25**, 146–166 (2009)
- [2] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes and F. Petropoulos, [Forecasting with temporal hierarchies](#). *Eur. J. Oper. Res.* **262**, 60–74 (2017)
- [3] H. G. Bergsteinsson, J. K. Møller, P. Nystrup, Ó. P. Pálsson, D. Guericke and H. Madsen, [Heat load forecasting using adaptive temporal hierarchies](#). *Appl. Energy*. **292**, article no. 116872 (2021)
- [4] F. Cribari-Neto and A. Zeileis, [Beta regression in R](#). *J. Stat. Software* **34**, 1–24 (2010)
- [5] C. Di Modica, P. Pinson and S. Ben Taieb, [Online forecast reconciliation in wind power prediction](#). *Electr. Power Syst. Res.* **190**, article no. 106637 (2021)
- [6] M. E. Hansen, P. Nystrup, J. K. Møller and H. Madsen, [Reconciliation of wind power forecasts in spatial hierarchies](#). *Wind Energy*. **26**, 615–632 (2023)
- [7] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos and H. L. Shang, [Optimal combination forecasts for hierarchical time series](#). *Comput. Statist. Data Anal.* **55**, 2579–2589 (2011)
- [8] R. J. Hyndman, A. J. Lee and E. Wang, [Fast computation of reconciled forecasts for hierarchical and grouped time series](#). *Comput. Statist. Data Anal.* **97**, 16–32 (2016)
- [9] N. Kourentzes and G. Athanasopoulos, [Cross-temporal coherent forecasts for Australian tourism](#). *Ann. Tourism Res.* **75**, 393–409 (2019)
- [10] M. Lange, [On the uncertainty of wind power predictions – Analysis of the forecast accuracy and statistical distribution of errors](#). *J. Sol. Energy Eng.* **127**, 177–184 (2005)
- [11] H. Madsen and P. Thyregod, [Introduction to general and generalized linear models](#). CRC Press, Boca Raton, FL (2011)
- [12] J. K. Møller, M. Zugno and H. Madsen, [Probabilistic forecasts of wind power generation by stochastic differential equation models](#). *J. Forecast.* **35**, 189–205 (2016)
- [13] H. A. Nielsen, H. Madsen and T. S. Nielsen, [Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts](#). *Wind Energy*. **9**, 95–108 (2006)
- [14] P. Nystrup, E. Lindström, P. Pinson and H. Madsen, [Temporal hierarchies with autocorrelation for load forecasting](#). *Eur. J. Oper. Res.* **280**, 876–888 (2020)
- [15] P. Pinson, [Wind energy: Forecasting challenges for its operational management](#). *Stat. Sci.* **28**, 564–585 (2013)
- [16] M. L. Sørensen, P. Nystrup, M. B. Bjerregård, J. K. Møller, P. Bacher and H. Madsen, [Recent developments in multivariate wind and solar power forecasting](#). *WIREs Energy Environ.* **12**, article no. e465 (2023)
- [17] S. L. Wickramasuriya, G. Athanasopoulos and R. J. Hyndman, [Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization](#). *J. Am. Stat. Assoc.* **114**, 804–819 (2019)
- [18] S. L. Wickramasuriya, B. A. Turlach and R. J. Hyndman, [Optimal non-negative forecast reconciliation](#). *Stat. Comput.* **30**, 1167–1182 (2020)

---

Mikkel Lindstrøm Sørensen has an MSc in mathematical modelling and computing from the Technical University of Denmark. He is currently a PhD student working on renewable energy forecasting and hierarchical forecast reconciliation, under the supervision of Jan Møller and Henrik Madsen.

[mliso@dtu.dk](mailto:mliso@dtu.dk)

Jan Kloppenborg Møller is an associate professor in stochastic dynamical systems at the Technical University of Denmark. He holds an MSc of applied mathematics (2006) and a PhD in engineering from DTU (2011). His research is concentrated on modelling and forecasting of (continuous- or discrete-time) stochastic dynamical systems.

[jkmo@dtu.dk](mailto:jkmo@dtu.dk)

Henrik Madsen received his PhD from the Technical University of Denmark (DTU) in 1986. Since then he has been employed by DTU, first as assistant professor, then associate professor from 1989, and he has been a full professor in stochastic systems since 1999. In June 2016, he was appointed as Knight of the Order of Dannebrog by Her Majesty the Queen of Denmark, and he was appointed as Doctor h.c. at Lund University in June 2017. His research areas are primarily modelling, forecasting and control, with a special focus on applications related to dynamics and flexibility in energy systems.

[hmad@dtu.dk](mailto:hmad@dtu.dk)