Received: 29 August 2022 | Revised: 12 March 2023 | Accepted: 19 March 2023

DOI: 10.1002/we.2819

RESEARCH ARTICLE

WILEY

Reconciliation of wind power forecasts in spatial hierarchies

Mads E. Hansen[1,2] | Nystrup Peter[1,3] | Jan K. Møller[1] | Madsen Henrik[1]

[1]DTU Compute, Technical University of Denmark, Lyngby, Denmark

[2]ENFOR A/S, Holte, Denmark

[3]QUENT ApS, Hellerup, Denmark

Correspondence
Mads E. Hansen
Email: s174434@win.dtu.dk

Present address
Mads E. Hansen, Danmarks Tekniske Universitet, Asmussens Alle, Building 303B, 2800 Kgs., Lyngby, Denmark.

Funding information
ARV, Grant/Award Number: 101036723; ebalance-plus, Grant/Award Number: 864283; Flexible Energy Denmark (FED), Grant/Award Number: IFD 8090-00069B; IEA Wind Task 36, Grant/Award Number: EUDP 64018-0515; IEA Wind Task 51, Grant/Award Number: EUDP 134-22015; Sustainable plus energy neighbourhoods (syn.ikia), Grant/Award Number: 869918

Summary

We consider reconciliation of wind power forecasts in a spatial hierarchy with three aggregation levels. We produce base forecasts for the bottom level consisting of 407 substations (connection points for local groups of wind turbines). State-of-the-art forecasts from a commercial forecast provider are available for the middle and top levels, which consist of 15 regions and the entire Western Denmark (DK1), respectively. We find that the accuracy of the total forecast can be improved through spatial reconciliation, even with a relatively simple model used at the lowest level of the hierarchy. Computing the base forecasts for the substations using wind speed as the only predictor, the RMSE of the DK1 forecast is reduced by 20.5%, while the RMSE of the regional forecasts is reduced by 4.7%, on average, through reconciliation. The increase in accuracy is partly due to reduced errors in the individual regional forecasts and partly due to reduced residual correlation between the reconciled regional forecasts. We test adaptive estimation of the covariance matrix of the base forecast errors and find that it has a limited impact on the accuracy, hinting toward a time-stable covariance structure.

KEYWORDS
forecast reconciliation, forecasting, spatial hierarchy, wind power

1 | INTRODUCTION

The development of renewable energy sources is increasing all over the world. In Denmark, the majority of electricity already comes from renewable energy sources. The Danish transmission system operator (TSO) Energinet expects wind power alone to account for 59% of the electricity production in Denmark in 2022.[1] In order to manage a power system where more than half of the power generation is dependent on the weather, the TSO requires accurate and reliable production forecasts. Spodniak[2] showed that there is a direct link between wind power forecast errors and spreads in electricity prices in Denmark. Consequently, the development of new methods to improve power production forecasts is necessary to ensure affordable and clean energy.

An efficient implementation of the future low-carbon energy system requires electricity demand to follow the weather-driven energy production at all scales of the power system. This implies that forecasting techniques will play an important role for both market participants and system operators. The trend toward more decentralized production, which is often integrated at low-voltage levels, means that distribution system operators (DSOs) need to have more focus on forecasting renewable production, for example, to keep the voltage within appropriate limits or to ensure that the temperature of transformers do not exceed critical limits. The future calls for more coordination between the low- and high-voltage system operators, and consequently, there is a need for coherence between actions taken by TSO and DSOs, who operate at different spatial scales.

Anyone who has dealt with forecasts in decision-making knows that in many situations there can be multiple different forecasts for the same variable. This comes down to the way the forecasts were generated. It leads to a natural desire to reconcile forecasts, such that they become coherent. In the context of forecast reconciliation, *coherency* refers to forecasts from different levels of a hierarchy being consistent with each other.[3] For example, if we have a wind power forecast for a certain region, the sum of the forecasts for the individual turbines within that region should align with the forecast for the region as a whole. *Reconciliation* is the process of adjusting or combining forecasts in order to ensure coherency across all levels of a hierarchy. This can involve adjusting the individual forecasts to align with the aggregate forecasts, or combining multiple forecasts in a way that takes into account the hierarchical structure.

A *hierarchy* is a way of organizing information into different levels, with each level representing an aggregation of information from the level below. We focus on forecasts that can be divided into a hierarchical structure defined by linear constraints. Such structure can be either spatial/structural, temporal, or spatio-temporal. A hierarchical structure is useful when information is available with multiple resolutions. Even when only one specific resolution is of interest, by utilizing information from other aggregation levels, it might be possible to obtain better forecasts across all levels.

Grunfeld[4] argued that the best approach to solving a forecasting problem that can be divided into a hierarchical structure is to generate a model for the grand total, that is, the top level of the hierarchy, and divide that into subtotals using a top-down approach. Another approach is to generate individual forecasts at the lowest level and aggregate them. Using the bottom-up method, forecasts for higher aggregation levels are the sum of the relevant bottom-level forecasts. Many argue for the superiority of the bottom-up approach because information is lost when looking directly at the higher aggregation levels. Schwarzkopf[5] looked at the differences between top-down and bottom-up and found that bottom-up in general is more robust, at least if the data at the lowest aggregation level is reliable.

It can be difficult to decide whether to use a top-down or bottom-up approach. Depending on the scenario, forecasts at different aggregation levels might capture different phenomena in the data. Hyndman[6] argued that generating forecasts independently across all aggregation levels and reconciling them afterwards produces optimal forecasts. Specifically, they minimized the total squared coherency error with equal weighting across all levels. They found that this method is better than both top-down and bottom-up. Mathematically this method corresponds to solving an ordinary least squares (OLS) problem.

Athanasopoulos[7] extended this work. They argued that solving an OLS problem might not lead to the optimal solution. They presented different ideas for how to weight the coherency errors, effectively solving a weighted least squares (WLS) problem. One possibility is to use the observed variance of the base forecast residuals at the different aggregation levels as weights when reconciling.

Nystrup[8] formulated the reconciliation problem as a generalized least squares (GLS) problem. They argued that the OLS and WLS approaches are both special cases of GLS, where all off-diagonal elements of the covariance matrix are zero. They showed that it is worthwhile to include residual correlation within the aggregation levels and cross-correlation between forecast errors from different aggregation levels.

Bergsteinsson[9] used adaptive estimation of the covariance matrix to improve heat load forecasts through reconciliation. They were the first to include commercial state-of-the-art forecasts for the bottom level and base forecasts based on numerical weather predictions (NWPs) for the other levels of their temporal hierarchy.

Several studies have considered reconciliation of wind power forecasts. Zhang[10] proposed a least-squares-based reconciliation method for wind power forecasts that can be divided into aggregates based on geography and network structures. They constructed a spatial hierarchy and formulated an optimization problem to minimize RMSE of the coherency errors, that is, the differences between the reconciled forecasts that respect the hierarchical constraints and the original base forecasts. They used AR models to generate base forecasts, as their focus was on very short forecast horizons (10–60 min). They found that reconciliation using the full covariance matrix produced the best results.

Jeon[11] introduced a method for reconciling probabilistic forecasts in temporal hierarchies. Their focus was on reconciliation and evaluation of probabilistic wind power forecasts for two wind farms in Greece. They considered forecast horizons up to 24 h ahead and produced base forecasts with data-driven time series models such as ARMA and GARCH. They showed the benefits of temporal reconciliation of probabilistic forecasts with the biggest improvements occurring at the highest levels of the hierarchy.

Modica[12] proposed a recursive and adaptive multivariate least squares estimator for online reconciliation of wind power forecasts. Similar to our study, they evaluated their method using wind power measurements from wind farms in Western Denmark. They divided 349 wind farms into four regions with 25 wind farms in each region, which is different from the spatial hierarchy that we consider. Furthermore, since they focused on online reconciliation, they only considered a 15-min forecast horizon and produced base forecasts for all nodes in the hierarchy using AR(2) models.

In this paper, we consider an application of reconciliation to state-of-the-art hierarchical forecasts of wind power production intended for the day-ahead market. Our focus is on reconciling wind power production forecasts on horizons beyond 24 h, which presents a different challenge from previous studies. Bai[13] also used state-of-the-art wind power production forecasts for spatial reconciliation; however, their focus was on the optimization part itself and, specifically, how to perform reconciliation using the Alternating Direction Method of Multipliers in a distributed setup. In contrast, we consider an offline setup where base forecasts are generated based on NWPs rather than time series models without exogenous explanatory variables.

We consider wind power production from onshore wind turbines in Western Denmark, which consists of 15 regions and 407 substations. Our goal is to leverage data for the substations to improve commercial, state-of-the-art forecasts at the regional and total levels. Fitting complex, individual models to each substation would be cumbersome given the number of substations. Furthermore, this would make the computation of the regional and total forecasts much more demanding.

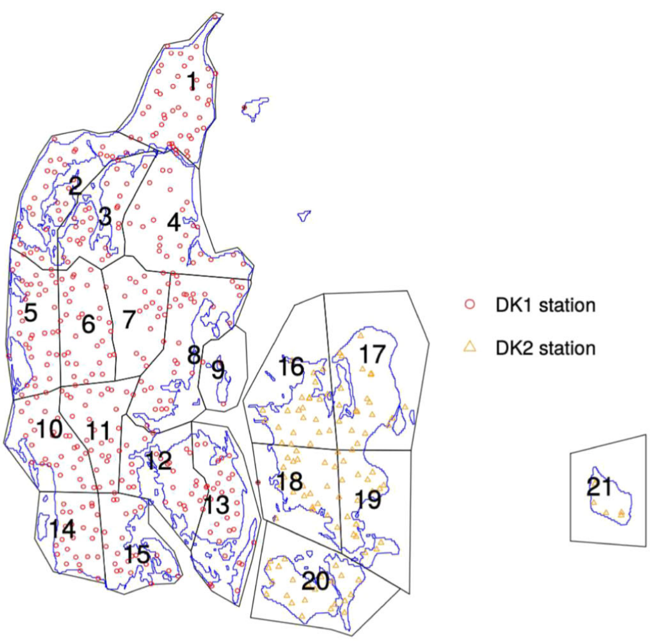
Our contribution is to present a method that allows us to utilize localized substation data efficiently and meaningfully with a low computational burden. For this purpose, we propose to construct simple base forecasts for the individual substations based on local NWPs and reconcile them with the state-of-the-art regional and total forecasts in a three-level spatial hierarchy. We are among the first to investigate spatial reconciliation of hierarchical wind power production forecasts that are based on NWPs. Moreover, we are the first to include and compare with state-of-the-art wind power production forecasts in the context of forecast reconciliation. We implement and investigate methods for recursive and adaptive estimation of the reconciliation weights. Finally, we perform a residual analysis of the reconciled forecasts and discuss the accuracy improvements and their origin.

The outline of this paper is as follows. In Section 2, we present the data for the spatial hierarchy; we show examples of the state-of-the-art base forecasts and produce base forecasts for the lowest level of the hierarchy. In Section 3, we introduce reconciliation of hierarchical forecasts, discuss several ways to weight the elements in the hierarchy, and present different ways to estimate the covariance of the base forecasts errors. In Section 4, we show the performance improvements achieved by forecast reconciliation and perform residual analysis. We discuss our findings in Section 5 and conclude the paper in Section 6.

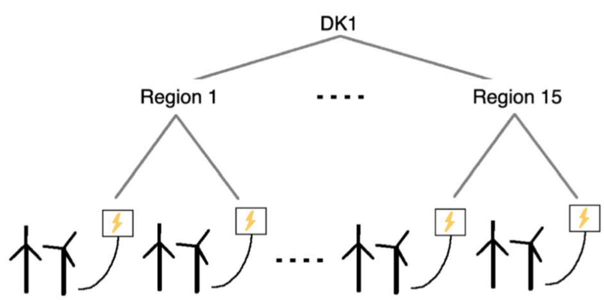
2 | DATA AND BASE FORECASTS

We consider hourly settlement data for wind power production from onshore wind turbines in Western Denmark (DK1). The data were made available by ENFOR A/S, which is a Danish company that provides forecasts to the Danish TSO and some DSOs. We have settlement data for all substations (small local groups of wind turbines) in DK1, regional forecasts (partly aggregated), and state-of-the-art total and regional forecasts. By *state-of-the-art* we mean widely used operational forecasts. For an extensive review of what this entails, see, for example, Sørensen.[14] Additionally, we have hourly high-resolution ($9\text{km} \times 9\text{km}$) NWPs from the European Centre for Medium-Range Weather Forecasts (ECMWF). Specifically, we have wind speed and temperature forecasts at 100 m. The operational delay in the settlement data is around 10 days, meaning that the data are available 10 days after the settlement date. The delay in NWPs is expected to be 7–9 h. We focus on forecasts for the day-ahead (also known as the spot) market. Thus, before noon on any given day, we want to produce hourly forecasts for the next day.

Figure 1A shows the location of the electricity substations and their regional division. The red dots in the figure represent the substations in DK1. We do not consider the substations in the price area in Eastern Denmark (DK2). We consider a three-level spatial hierarchy consisting of



(A) Regional map of Denmark with locations of substations.



(B) Spatial hierarchy for DK1 with three aggregation levels.

FIGURE 1 (A) Map of Denmark showing the locations of the individual power substations and regions. (B) Illustration of the spatial hierarchy for the western price area DK1 with 15 regions at the middle level and 407 individual substations at the bottom level.

the total DK1, the 15 regions, and the 407 substations. Figure 1B shows a graphical illustration of this hierarchy. The total and regional forecasts are from the commercial provider, whereas we construct base forecasts for the substations.

2.1 | Total

We consider the total forecast for DK1 to be the most important and will emphasize the accuracy of this in our study. The reason is that the residuals are directly linked to the power deficit/surplus in the price area, which is important to market participants because it can lead to up or down regulation. The total forecasts for DK1 are generated using publicly available data[*] for wind power production and ECMWF high-resolution NWPs as input variables. We have these forecasts from mid-2018 to end-2020.

Figure 2 compares measurements, base forecasts, and aggregates at the different levels of the hierarchy for a one-week period in the spring of 2019. The data varies both in terms of scale and time series features across the levels of the hierarchy. The total forecasts for DK1 in Figure 2A follow the measurements relatively well, except for the peak in the afternoon on March 6 and March 12 where the aggregated regional and station forecasts seem to do better.

The fundamental difference between these forecasts is the granularity of the data they are based on. The bottom-level forecasts can capture more fine-grained variations in the data but may be more prone to noise and errors. The aggregated forecasts may miss some of the fine-grained variations but are generally more stable. At the most aggregate total level, the measurements and base forecasts are more smooth compared to the lower levels. Naturally, as we move down the hierarchy, the signal-to-noise ratio decreases. Yet, there is a fairly strong resemblance between the data for DK1, Region 1, and even the two substations. This indicates that the weather conditions are fairly similar throughout the area, which makes sense given the limited geographical extent of Western Denmark.

2.2 | Regions

The DK1 price area is divided into 15 regions by the TSO, as seen in Figure 1A. Region 1 is the most northern while regions 14 and 15 are on the southern boarder to Germany. Although the individual regional deviations do not lead to up or down regulation in the price area, they do serve a meaningful purpose for the TSO. As implied by Figure 1A, the wind turbines are not uniformly distributed across the area, nor is the electricity consumption. Hence, there can be distributional challenges within the price area. Accurate regional forecasts are helpful in alleviating or preventing some of these distributional challenges. This becomes increasingly relevant as the fraction of power production coming from renewables increases.

For each region, we have hourly settlement data, that is, the confirmed production. There are no onshore wind turbines outside these 15 regions in DK1. Furthermore, we have state-of-the-art hourly regional forecasts generated using ECMWF's high-resolution NWPs. Both settlement data and forecasts are available from mid-2018 to end-2020.

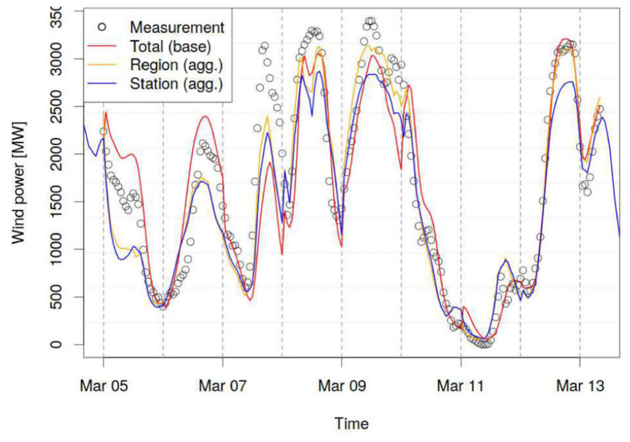
Figure 2B shows measurements and base forecasts for the first of the 15 regions in DK1. We see that the forecasts follow the measurements relatively well, although the deviations are slightly larger than what we saw for DK1. Specifically, the normalized root mean square error (NRMSE) for DK1 is 0.058, while it is 0.100 for Region 1. Both estimates are based on the entire period and not just the period shown in the figures.

2.3 | Stations

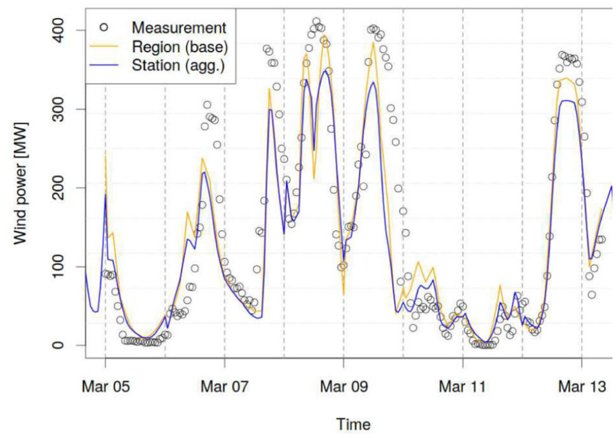
Finally, we have data for all of the 407 substations that connect wind turbines to the electricity grid in DK1. For each substation, we have settlement data and the exact coordinates of the station from which it can be assigned to a region. These data are available from end-2017 to end-2020. We are not provided with forecasts for the individual substations and will therefore make these ourselves.

When generating station forecasts, we train a model using data from the first half of 2018 and use this model to generate forecasts for the second half of 2018. We then train a model using all data from 2018 and use this to generate forecasts for 2019. Finally, we train a model on all data from 2019 and use this model to generate forecasts for 2020. Figure 3 shows a graphical representation of when the data are available and what it is used for. Note that the burn-in period is related to the estimation of the covariance of the base forecast errors described in Section 4.

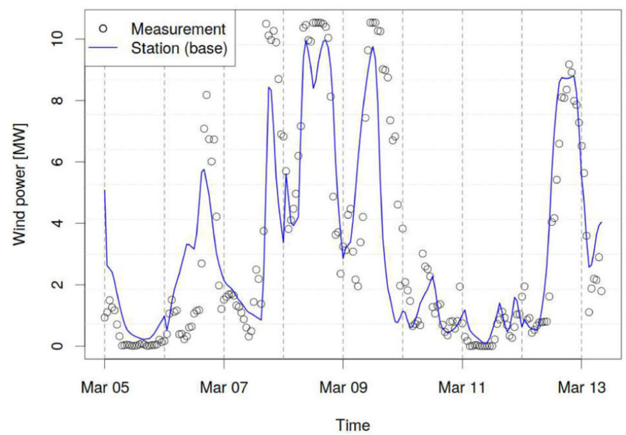
[*]https://www.energidataservice.dk/tso-electricity/productionconsumptionsettlement.



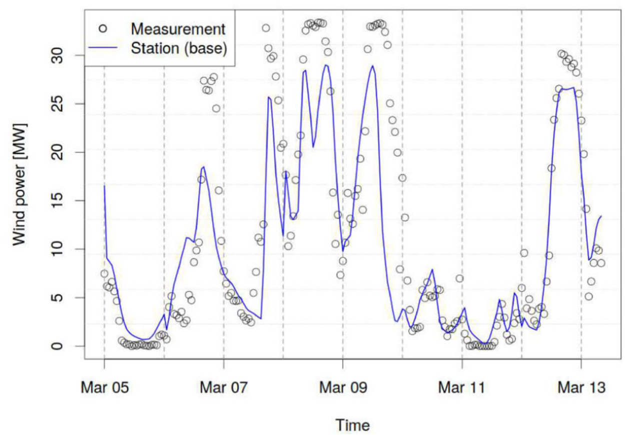
(A) Total wind power production, base forecast, aggregated regional and station forecasts for DK1, which consists of regions 1–15.



(B) Wind power production, base forecast, and aggregated station forecasts for Region 1.



(C) Wind power production and base forecast for Station 1 in Region 1 (LPR using wind speed only).



(D) Wind power production and base forecast for Station 2 in Region 1 (LPR using wind speed only).

FIGURE 2 Wind power production and base forecasts at the different levels of the hierarchy for a week in the spring of 2019. The dashed vertical lines show the shift between days.

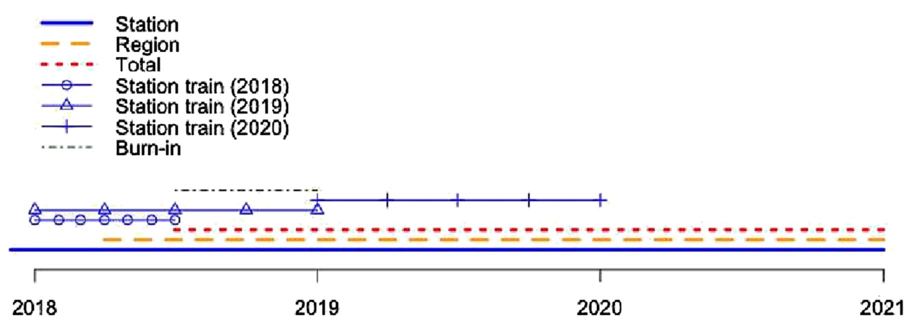


FIGURE 3 Timeline of data and forecast availability. “Station train (2018)” denotes the training data used for fitting the model used for generating forecasts for year 2018 and so on. The burn-in period is the data used for the initial covariance estimates.

2.3.1 | Station model

Giebel,[15] Lange,[16] and Pinson[17] discussed how to forecast wind power production. The consensus is that a *physical* model that predicts wind power based on weather predictions is a key component. Wind power production is related to wind speed by a nonlinear function called the *power curve*. Further, they all discussed the use of *statistical* models, that is, models that only rely on power production data, such as ARIMA models. They found that the statistical models can be beneficial, especially on short horizons (30 min to 6 h). For the day-ahead horizon that we consider, it generally makes sense to use a combination of *statistical* and *physical* models according to, for example, Costa[18] and Soman.[19]

Pinson[17] discussed how to predict the output from a wind turbine based on NWPs by modeling it as a physical system. We know from physics that the total amount of energy that flows by a wind turbine can be expressed as

$$P_{total} = \frac{1}{2}\rho\pi R^2 v^3, \tag{1}$$

where ρ is the density of the air, R is the specific gas constant for air, and v is the wind speed. However, the turbine cannot extract all of the energy from the air. In fact, the theoretical upper limit is given by Betz's limit at $C_p = \frac{16}{27}$. In practice, due largely to cost-efficiency, this number is much lower for commercial wind turbines. We note that the air density, ρ, through the ideal gas law is proportional to the inverse of the temperature, T. This leads to the physical relation

$$P_{turbine} \propto \frac{v^3}{T}. \tag{2}$$

In reality, the power production from wind turbines does not follow this relation as nicely as we would want. First, wind turbines need a certain amount of wind to be able to produce any power, known as the "cut-in" speed. Second, they have a maximum capacity that is limited by many factors, including the design of the generator and other parts of the turbines. Third, if the wind speed gets too high, the wind turbine will come to a complete stop to prevent it from being damaged; this is known as its "cut-out" speed.[20] Pinson[17] discussed the advantages of using nonparametric regression to estimate the power curve. They found that this approach is often advantageous compared to a linear model with input given by the physical relation. Xu[21] further discussed this. Specifically, they suggested the use of local polynomial regression (LPR). Nielsen[22] were among the first to model the power curve rather then using the physical relation given in Equation (2). In an offline setup, they used weighted least squares estimates with a weighting defined by a tricube kernel, that is, LPR.

We have NWPs whose computations are initialized at 00:00. These are available to us shortly after the computations are completed, which usually takes around 7.5 h. Since our focus is day-ahead, we use the NWPs initialized at 00:00 to generate forecasts 24–48 h into the future from that point. As we are modeling an offline setup and our forecast horizon is more than 6 h, we focus on time-invariant models using NWPs, as discussed by Costa,[18] Soman,[19] and later by Pinson.[17] As NWPs, we use ECMWF HRES at the grid point that is closest (i.e., has the minimal Euclidean distance) to the substation.

Regarding model selection, we have tested a number of different approaches including simple linear models. Based on the performance of these methods, and since this is not the main focus of this paper, we limit ourselves to using LPR as described in Madsen[23] and specifically for wind power forecasting by Nielsen.[22] The LPR estimates are given by

$$Y_t = P(X_t, \theta(X_t)) + \varepsilon_t, \tag{3}$$

where P is a polynomial of arbitrary degree and θ are the locally estimated parameters that depend on the value of X_t. The parameters are found as

$$\hat{\theta}(X_t) = \arg\min_{\theta(X_t)} \frac{1}{N}\sum_{s=1}^{N} W(X_t, X_s)(Y_s - P(X_s, \theta(X_s)))^2, \tag{4}$$

where the kernel W describes how the residuals are weighted and N is the number of observations in the training data.

We consider the following three models for forecasting using NWPs:

Model 1: LPR on wind speed only.
Model 2: LPR on wind speed and horizon.
Model 3: LPR on wind speed, horizon, and temperature.

We use the `R`-function `loess` for LPR. We use it with a tricube kernel, additive (no mixed terms) second degree polynomials, and a nearest neighbor bandwidth (width of kernel) with span parameter of 0.4. This means that for each new point, second degree polynomials are fitted with weights given by the tricube kernel with a bandwidth corresponding to 40% of the data being used in all dimensions.

Other potential covariates that could be included in the model include wind direction, air density, and atmospheric pressure. These covariates were not included in this forecast model, because the focus of this paper is to investigate the effects of spatial hierarchies and not to find the best possible station model.

Table 1 shows the NRMSE for the three models. It also shows the relative differences to model 1, with differences smaller than zero indicating a more accurate model. The differences are larger in sample than out of sample. The most complex model has a NRMSE that is 4.9% lower than the simplest model out of sample.

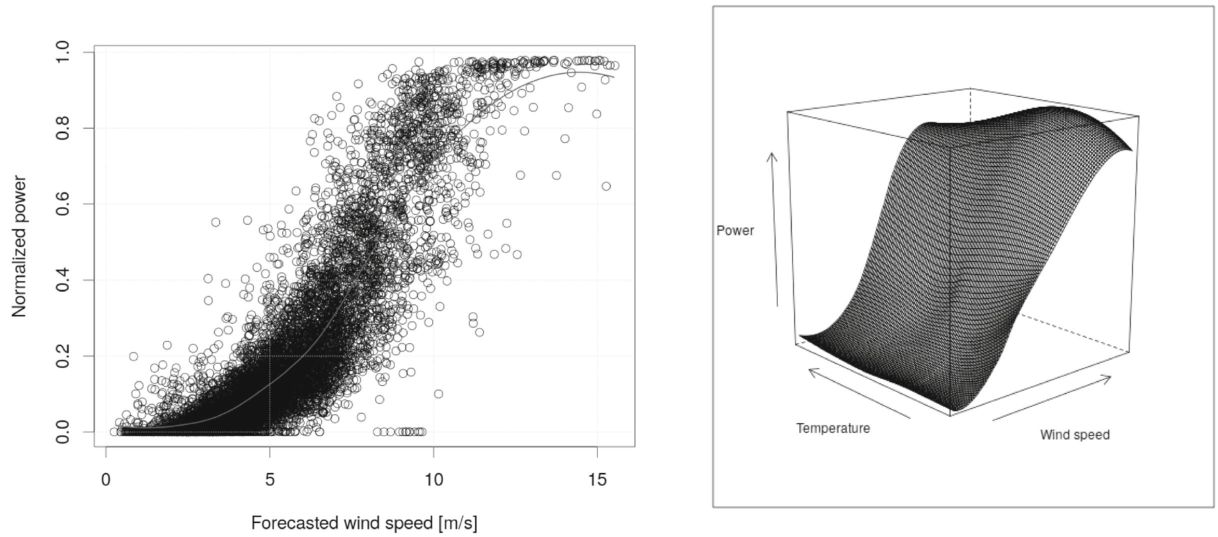
Figure 4 shows examples of fits using the `loess` function with one and two input parameters. Figure 4B shows a 2d fit using LPR. Along the wind speed direction, the curve resembles the power curve in Figure 4A. Air is less dense at higher temperatures, which primarily impacts the curve at high wind speeds. The power curve in Figure 4A with wind speed as the only predictor corresponds to Model 1. Figure 2C,D shows examples of base forecasts for stations 1 and 2 using this model. If we compare the forecast accuracy with the regional and total forecasts, we see that these have the largest residuals relatively speaking, that is, the largest normalized errors. The base forecasts for the two substations look very different, with the base forecasts for Station 1 being much more smooth than that for Station 2. This is because they are not located in the same place within the region, and hence, the NWPs are different. Additionally, differences between substations are due to differences between specific types of turbines and differences in the surrounding terrain.

3 | FORECAST RECONCILIATION

We follow the notation proposed by Athanasopoulos.[7] We define base forecasts as forecasts made directly on the data from the same aggregation level. Let $\hat{y}_t^{[s_i]}$ denote the base forecast for station i at time t, $\hat{y}_t^{[r_j]}$ the regional forecast for region j at time t, and $\hat{y}_t^{[tot]}$ the total forecast for DK1 at time t. Let us then define a vector, $\hat{y}_t$, given by

TABLE 1 Accuracy of substation models using local polynomial regression and relative differences out of sample.

Model	In-sample NRMSE	Out-of-sample NRMSE	Out-of-sample relative difference
Model 1	0.1047	0.1124	0%
Model 2	0.1030	0.1112	−1.1%
Model 3	0.1017	0.1069	−4.9%



(A) Power curve using NWP of wind speed as the only predictor.

(B) 2D fit with wind speed and temperature as inputs.

FIGURE 4 Illustrations of local polynomial regression with one and two input parameters.

$$\hat{y}_t = [\hat{y}_t^{[tot]}, \hat{y}_t^{[r_1]}, \hat{y}_t^{[r_2]}, \hat{y}_t^{[s_1]}, ..., \hat{y}_t^{[s_5]}]^T. \tag{5}$$

Drawing on work done by Hyndman,[6] we define a *summation matrix* with columns representing the lowest aggregation level in the hierarchy and rows the aggregation given in the same order as in $\hat{y}_t$. Let n denote the number of base forecasts, that is, the sum of the number of stations and regions plus one (for the total). Let m denote the number of stations. Let us consider an example that corresponds to a hierarchy with two regions with three and two stations, respectively, then $n=8, m=5$, and

$$S_{(n\times m)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

We define a reconciled forecast, $\tilde{y}_t$, to be a forecast that respects the hierarchical structure; that is, the forecast for Region 1 equals the sum of the forecasts for the first three stations, and so on. Next, we define an auxiliary matrix, $G_{(m\times n)}$, that extracts the forecasts for the bottom level from $\tilde{y}_t$:

$$G_{(m\times n)} = [0_{(m\times(n-m))} | I_{(m\times m)}]. \tag{7}$$

The structural constraint imposed by the hierarchy can now be written as

$$\tilde{y}_t = SG\tilde{y}_t. \tag{8}$$

Reconciliation is needed when base forecasts $\hat{y}_t$ do not satisfy this constraint. Naturally, this can be done in many ways and comes down to assumptions about the errors in the individual forecasts. In this paper, we assume that the coherency errors, $\varepsilon_t = \tilde{y}_t - \hat{y}_t$, have mean zero and are homoscedastic. Therefore, it makes sense to view the reconciliation problem as a regression problem

$$\hat{y}_t = S\beta_t + \varepsilon_t, \tag{9}$$

where the parameter β_t is the bottom-level reconciled forecasts, $[\tilde{y}_t^{[s_1]}, ..., \tilde{y}_t^{[s_5]}]^T$.

Nystrup[8] used generalized least squares (GLS) to solve this regression problem; notice this implies potential correlation between coherency errors:

$$\underset{\tilde{y}_t}{\operatorname{argmin}} (\tilde{y}_t - \hat{y}_t)^T \Sigma^{-1} (\tilde{y}_t - \hat{y}_t) \tag{10}$$

$$\text{where}\, \tilde{y}_t = SG\tilde{y}_t. \tag{11}$$

Here, Σ describes the variance of the coherency error. Wickramasuriya[3] showed that in general Σ is not identifiable. However, they also provided theoretical justification for using the base forecast errors, $e_t = y_t - \hat{y}_t$, as a proxy. For reasonable models, this is always identifiable, but it can be demanding to compute. Note that this regression approach aims to adjust the base forecast in such a way that the hierarchical structure is respected while maximizing the likelihood of the the necessary adjustments, where the likelihood of the adjustments is found under the assumption of normality in the base forecast errors.

Nystrup[8] showed that the closed-form solution to the optimization problem is given by

$$\tilde{y}_t = S(S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1}\hat{y}_t. \tag{12}$$

This allows us to easily compute the reconciled forecast that minimizes the objective function, Equation (10), without the need for any complicated iterative approach. Boyd[24] showed that when the dimension of the problem is very large, it can be faster to solve it using an iterative method. Wickramasuriya[3] argued that due to the structure of Σ, it will often be numerically close to singular. As a consequence, they argued that using the Monroe pseudo inverse instead of the inverse would be advantageous. Additionally, since we are required to use the inverse of Σ, it can be advantageous to update the inverse directly using the matrix inversion lemma, depending on the frequency of the updates.

3.1 | Structure of Σ

The problem of forecast reconciliation comes down to finding the best estimate of the covariance of the coherency errors, Σ. Unfortunately, this is not necessarily a trivial task. The simplest estimate is to assume an identity matrix, $\Sigma = I$. This was proposed by Hyndman[6] and is equivalent to assuming that the coherency errors are independent with uniform variance across the aggregation levels.

Athanasopoulos[7] proposed three different approaches, which correspond to three variations of the diagonal estimates of Σ. The first approach, known as *structural variance scaling*, is to weigh the lowest aggregation level by one, and each higher aggregation level by the number of units it consists of. In other words, Σ is approximated as the row sums of the summation matrix, S. In our small example from before, this would be

$$\Sigma_{\text{struc}} = diag(5,3,2,1,\ldots,1). \tag{13}$$

The advantage of this method is that it does not require any estimation of Σ. In the specific case of this paper, this approach would most likely produce bad results, since it assumes uniformity in the coherency errors of the substations. To adjust for differences in the sizes of the substations, you would need to rescale by the number of turbines or the installed capacity of each substation.

The second approach is known as *series variance scaling*. The idea here is to use pooled variance estimates for each aggregation level as weights:

$$\Sigma_{\text{series}} = diag(\sigma_{tot}^2, \sigma_r^2, \sigma_r^2, \sigma_s^2, \ldots, \sigma_s^2). \tag{14}$$

Again, it is unlikely that this approach would produce good results for the spatial hierarchy for DK1, since the variance of, for example, the substations, which are of different sizes, would be pooled. However, it could make sense to compute the normalized errors and pool these, and then rescale by the installed capacity afterwards.

The third approach is known as *hierarchical variance scaling* and is an extension of series variance scaling, where the individual variance estimates are used instead of pooling them by aggregation level:

$$\Sigma_{\text{hier}} = diag(\sigma_{tot}^2, \sigma_{r_1}^2, \sigma_{r_2}^2, \sigma_{s_1}^2, \ldots, \sigma_{s_5}^2). \tag{15}$$

All the diagonal approximations assume independence between forecast errors. Naturally, this assumption is not always reasonable. Therefore, Nystrup[8] proposed methods with nonzero values for some/all of the off-diagonal elements as well.

The first idea is to assume that the errors follow a Markov process. In temporal hierarchies, this seems quite natural. In spatial hierarchies, it corresponds to assuming that any forecast error is independent from all other stations than its immediate neighbors. Based on this idea, it is possible to formulate a Markovian correlation structure.

A slightly more advanced approach is to use block estimates of the covariance matrix. A natural block structure is to estimate the covariance within the aggregation levels. This would result in the covariance structure given below:

$$\Sigma_{\text{within}} = \begin{bmatrix} \sigma_{tot}^2 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_{r_1}^2 & \sigma_{r_1,r_2} & 0 & 0 & \ldots & 0 \\ 0 & \sigma_{r_1,r_2} & \sigma_{r_2}^2 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_{s_1}^2 & \sigma_{s_1,s_2} & \ldots & \sigma_{s_1,s_5} \\ 0 & 0 & 0 & \sigma_{s_1,s_2} & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \sigma_{s_4}^2 & \sigma_{s_4,s_5} \\ 0 & 0 & 0 & \sigma_{s_1,s_5} & \ldots & \sigma_{s_4,s_5} & \sigma_{s_5}^2 \end{bmatrix}. \tag{16}$$

Finally, we can use the entire cross-covariance matrix

$$\Sigma_{\text{cross}} = \begin{bmatrix} \sigma^2_{tot} & \sigma_{tot,r_1} & \sigma_{tot,r_2} & \sigma_{tot,s_1} & \sigma_{tot,s_2} & \cdots & \sigma_{tot,s_5} \\ \sigma_{tot,r_1} & \sigma^2_{r_1} & \sigma_{r_1,r_2} & \sigma_{r_1,s_1} & \sigma_{r_1,s_2} & \cdots & \sigma_{r_1,s_5} \\ \sigma_{tot,r_2} & \sigma_{r_1,r_2} & \sigma^2_{r_2} & \sigma_{r_2,s_1} & \sigma_{r_2,s_2} & \cdots & \sigma_{r_2,s_5} \\ \sigma_{tot,s_1} & \sigma_{r_1,s_1} & \sigma_{r_2,s_1} & \sigma^2_{s_1} & \sigma_{s_1,s_2} & \cdots & \sigma_{s_1,s_5} \\ \sigma_{tot,s_2} & \sigma_{r_1,s_2} & \sigma_{r_2,s_2} & \sigma_{s_1,s_2} & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \sigma^2_{s_4} & \sigma_{s_4,s_5} \\ \sigma_{tot,s_5} & \sigma_{r_1,s_5} & \sigma_{r_2,s_5} & \sigma_{s_1,s_5} & \cdots & \sigma_{s_4,s_5} & \sigma^2_{s_5} \end{bmatrix}. \tag{17}$$

Zhang[10] found that using the full covariance structure, Σ_{cross}, for forecast reconciliation of wind power production on short horizons yielded the best results. Additionally, Nystrup[8] also argued for the superiority of using a more complex covariance structure in a temporal setting. Therefore, we expect Σ_{cross} to yield the best results. For computational purposes, it might still be worthwhile to consider using some block structure, especially if the reconciliation is done in multiple dimensions, such as in a spatio–temporal hierarchy.

3.2 | Estimating Σ

Before we can solve the regression problem given in Equation (10), we need to estimate the covariance of the coherency errors, Σ. In cases where we have many variables and many data points per variable, it quickly becomes computationally expensive to compute. Let $\hat{\sigma}^t_{ij}$ denote the estimated covariance between elements i and j at time t. As given by Pitman,[25] the covariance between any two elements in the cross-covariance matrix can be computed using

$$cov(\varepsilon^t_i, \varepsilon^t_j) = E[\varepsilon^t_i \varepsilon^t_j] - E[\varepsilon^t_i]E[\varepsilon^t_j], \tag{18}$$

as

$$\hat{\sigma}^t_{ij} = \frac{1}{t}\sum_{l=1}^{t} e^l_i e^l_j - \frac{1}{t^2}\left(\sum_{l=1}^{t} e^l_i\right)\left(\sum_{l=1}^{t} e^l_j\right). \tag{19}$$

If we have the cross-covariance matrix at time t and wish to estimate it at $t+k$, the naive approach is simply to recompute it for every observation:

$$\hat{\sigma}^{t+k}_{ij} = \frac{1}{t+k}\sum_{l=1}^{t+k} e^l_i e^l_j - \frac{1}{(t+k)^2}\left(\sum_{l=1}^{t+k} e^l_i\right)\left(\sum_{l=1}^{t+k} e^l_j\right). \tag{20}$$

3.2.1 | Recursive estimation

Recomputing the cross-covariance matrix each time we update the model is computationally expensive. Hence, we want to estimate it recursively. By changing the index of the summation and rewriting, we get

$$\hat{\sigma}^{t+k}_{ij} = \frac{1}{t+k}\sum_{l=1}^{t} e^l_i e^l_j + \frac{1}{t+k}\sum_{l=t+1}^{t+k} e^l_i e^l_j - \frac{1}{(t+k)^2}\left(\sum_{l=1}^{t+k} e^l_i\right)\left(\sum_{l=1}^{t+k} e^l_j\right). \tag{21}$$

If we reorder the original expression a bit, we get

$$\hat{\sigma}^t_{ij} + \frac{1}{t^2}\left(\sum_{l=1}^{t} e^l_i\right)\left(\sum_{l=1}^{t} e^l_j\right) = \frac{1}{t}\sum_{l=1}^{t} e^l_i e^l_j \tag{22}$$

$$t \cdot \left(\hat{\sigma}_{ij}^{t} + \frac{1}{t^2} \left(\sum_{l=1}^{t} e_i^l \right) \left(\sum_{l=1}^{t} e_j^l \right) \right) = \sum_{l=1}^{t} e_i^l e_j^l. \tag{23}$$

By insertion, we get

$$\hat{\sigma}_{ij}^{t+k} = \frac{t}{t+k} \cdot \left(\hat{\sigma}_{ij}^{t} + \frac{1}{t^2} \left(\sum_{l=1}^{t} e_i^l \right) \left(\sum_{l=1}^{t} e_j^l \right) \right) + \frac{1}{t+k} \sum_{l=t+1}^{t+k} e_i^l e_j^l - \frac{1}{(t+k)^2} \left(\sum_{l=1}^{t+k} e_i^l \right) \left(\sum_{l=1}^{t+k} e_j^l \right). \tag{24}$$

Since we are dealing with forecast residuals, we assume that the true mean is zero, that is, that the forecasts are unbiased. Using this assumption and thereby omitting the mean estimates from the expression, we get

$$\hat{\sigma}_{ij}^{t+k} = \frac{t}{t+k} \cdot \hat{\sigma}_{ij}^{t} + \frac{1}{t+k} \sum_{l=t+1}^{t+k} e_i^l e_j^l. \tag{25}$$

This way of estimating the covariance prevents doing unnecessary work and keeps the computation time for updating the covariance matrix low.

3.2.2 | Adaptive

In the recursive estimation, we choose the weight of the previous value to ensure that the weight given to each residual is equal. However, this idea can easily be expanded such that the most recent values get the highest weight.

Nielsen[26] proposed to use an adaptive variance estimate through exponential weighting

$$\hat{\sigma}_{ij}^{t+k} = \lambda \hat{\sigma}_{ij}^{t} + (1-\lambda) \hat{\sigma}_{ij}^{t+1:t+k}, \tag{26}$$

where $\hat{\sigma}_{ij}^{t+1:t+k}$ is the covariance estimate for the period $t+1$ to $t+k$ and $\lambda \in (0,1)$. It is the weighted sum of the previous covariance with the covariance of newly attained data. This is also under the assumption that the true mean of the residuals is zero. The effective number of observations, also referred to as the memory (length) is given by

$$\frac{1}{1-\lambda} \cdot k, \tag{27}$$

where k is block size, that is, the number of new observations in each update.

This way of adaptively estimating the covariance is computationally less expensive than the naive approach. Furthermore, it is a way of allowing the covariance structure to adapt over time following the data.

4 | RESULTS

In this section, we apply the method of forecast reconciliation outlined in Section 3 to a spatial hierarchy for wind power production in Western Denmark. Specifically, we construct the hierarchy shown in Figure 1B, with 15 regions and 407 stations, such that $n = 423$ and $m = 407$. The focus is offline forecasts for the day-ahead electricity market. Since our focus is an offline setup where the limiting factor is the accuracy of the NWPs, the timestamps of the forecasts denote when the ECMWF computations were initialized. All timestamps that denote intervals are right bounded, such that 01:00 means 00:00–01:00, and so forth. Hence, we want to predict 25–49 h into the future (from NWP initialization). This is consistent with forecasts for the day-ahead market.

The forecasts for the hierarchy are reconciled by solving the optimization problem given in Equation (11) for each hour. The covariance estimates are marginal estimates that are made for all time horizons (25–49 h) with equal weighting. Hence, we assume that the covariance structure of the forecast residuals does not change with respect to the forecast horizon. As discussed by Møller[27] and Nielsen,[28] wind power forecasts are decoupled from the observed values after approximately 12 h. Therefore, the covariance structure for horizons greater than 12 h will be assumed stationary. Additionally, assuming a constant covariance structure allows us to utilize all data points for its estimation. This is expected to result in a better estimate of the covariance structure in general.

Due to computational constraints, the covariance matrix is only updated and inverted once every 30 days during the 2-year period 2019 and 2020. The delay in data of 10 days is respected, also with regards to the estimation of Σ. When estimating the covariance matrix, we use the first 6 months as a burn-in period. Since the total forecasts are available from mid-2018, this corresponds to using 2018 as training data. No data from this period are included when evaluating the forecast accuracy across the hierarchy. Initial tests show that using a full cross-covariance matrix, Σ_{cross}, yields the best results. Therefore, we do not consider any block structures of Σ.

Evaluation of model performance after reconciliation showed a small difference in accuracy between the three station-level models described in Section 2. Therefore, we only consider the simplest model in the following, that is, the station model using NWPs of wind speed as the only predictor. Results for the other models can be found in Appendix A. Generally, the results show that while there is a difference in accuracy for the individual substations, as discussed in Section 2, there is little to no difference in the accuracy of the total forecasts after reconciliation.

As performance statistics, we consider root mean square error (RMSE), mean absolute error (MAE), bias, and symmetric mean absolute percentage error (SMAPE):

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} |e_t|^2}, \tag{28}$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T} |e_t|, \tag{29}$$

$$\text{BIAS} = \frac{1}{T}\sum_{t=1}^{T} e_t, \tag{30}$$

$$\text{SMAPE} = \frac{1}{T}\sum_{t=1}^{T} \frac{|e_t|}{(y_t + \tilde{y}_t)/2}, \tag{31}$$

where T is the number of residuals, which is slightly less than $2 \cdot 365 \cdot 24 = 17520$ because of missing data points. Since the installed wind power capacity changes significantly during the evaluation period, we normalize the performance statistics. We normalize with respect to the full capacity in the individual steps; that is, the forecasts for each day are normalized with respect to the installed capacity for that given day, such that $\text{NRMSE}_{\text{DK1}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left|\frac{e_{\text{DK1},t}}{C_{\text{DK1},t}}\right|^2}$, and so forth. This ensures that the accuracy of the model does not depend on the capacity of the system. When comparing two models, we consider the relative forecast accuracy as advocated by Hyndman,[29] for example, relative root mean square error (RRMSE), given by

$$\text{RRMSE} = 100\% \cdot \frac{\text{RMSE}_{\text{rec}} - \text{RMSE}_{base}}{\text{RMSE}_{\text{base}}}, \tag{32}$$

where $\text{RMSE}_{\text{base}}$ and RMSE_{rec} are the RMSE of the base and reconciled forecasts, respectively.

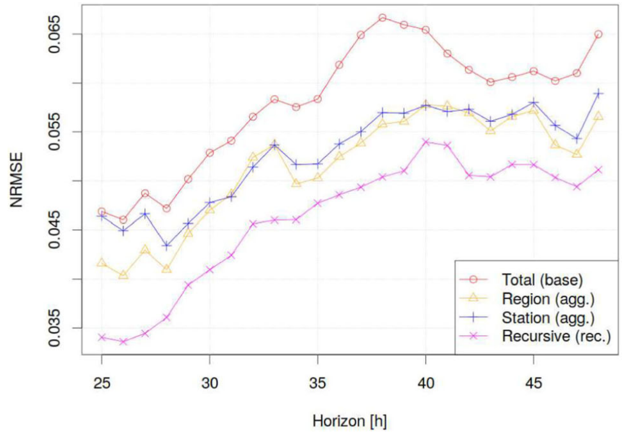
4.1 | Accuracy of total forecast for DK1

Table 2 compares the accuracy of reconciled and aggregated forecasts for the highest level of the hierarchy, that is, the total forecast for DK1. The covariance matrix is re-estimated once every 30 days; hence, the forgetting factor is per 30 days.

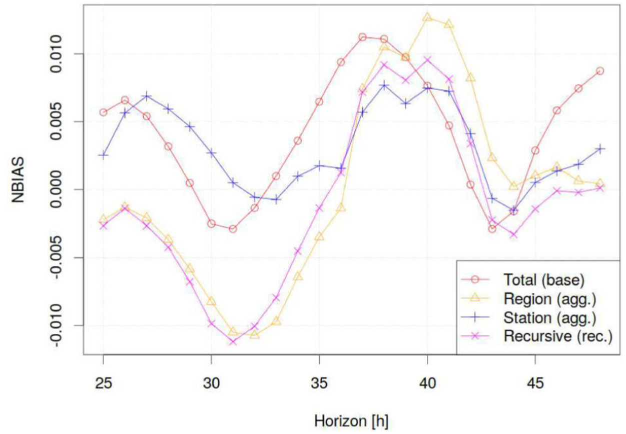
TABLE 2 Accuracy of the total DK1 forecast made using the different reconciliation methods and by aggregating the regional and station forecasts, respectively. The station forecasts only use wind speed as predictor. More results can be found in Table A1.

Model	NRMSE	NMAE	SMAPE	RRMSE	RMAE	RSMAPE
Total (base)	0.0581	0.0400	0.2981	0%	0%	0%
Region (agg.)	0.0514	0.0356	0.2694	−11.5%	−11.0%	−9.63%
Station (agg.)	0.0528	0.0367	0.2808	−9.1%	−8.3%	−5.80%
Recursive (rec.)	0.0462	0.0333	0.2612	−20.5%	−16.8%	−12.38%
Adaptive (rec.)[a]	0.0462	0.0333	0.2611	−20.5%	−16.8%	−12.41%

[a] $\lambda = \frac{5}{6}$ corresponding to an effective memory of 180 days, because the covariance matrix is estimated every 30 days.



(A) NRMSE.



(B) NBIAS.

FIGURE 5 Performance statistics as a function of the forecast horizon for the total DK1 forecast using station model 1 at the lowest level of the hierarchy. Notice "Region" and "Station" are both aggregated forecasts.

Aggregating either the regional base forecasts or the station base forecasts produces two alternatives for the total forecast. Both are more accurate than the total base forecast. This is consistent with previous studies of bottom-up versus top-down methods. Aggregating the regional base forecasts yields a more accurate forecast for DK1 than aggregation of the station base forecasts. This underlines that the regional base forecasts are state-of-the-art forecasts from a much more complex model fitted by a professional forecast provider.

The total forecasts for DK1 produced through forecast reconciliation using recursive and adaptive covariance estimation both outperform the other forecasts quite substantially. Forecast reconciliation reduces the RMSE by 20.5% compared to the total base forecasts, which is 10% points better than aggregating the regional state-of-the-art base forecast. For MAE and SMAPE, the reductions were 16.8% and 12.4%, respectively. The accuracy improvement is stable with respect to the adaptivity of the covariance estimator. In other words, the accuracy is very similar for recursive estimates, estimates with low memory, and estimates with high memory. The almost nonexisting difference in performance between using adaptive and recursive estimates of the covariance matrix implies a time-stable covariance structure.

Figure 5A,B shows the RMSE and BIAS as a function of the forecast horizon across all levels of the hierarchy (possibly aggregated) and the reconciled total. The reconciled forecast is obtained via recursive estimates of Σ in these figures; however, the equivalent figures using the adaptive estimates are extremely similar. From the figures, it is evident that the reconciled forecasts outperform all the individual forecasts in terms of RMSE on all horizons. The bias of the reconciled total forecast is similar to that of the aggregated regional forecast.

RMSE increase with the forecast horizon. This largely comes down to errors in the weather predictions. The accuracy of the NWPs decreases with forecast horizon, which in turn yields less accurate base forecasts. The effect of forecast reconciliation seems to be independent of the forecast horizon as the distance between the lines in Figure 5A is more of less constant across horizons.

Figure 6 shows a boxplot of the residuals of the total base forecast and the reconciled total forecast using recursive estimation of the covariance. It is evident that reconciliation significantly reduces the number of very large residuals, which suggests that the reductions in RMSE and MAE come somewhat from fewer extreme errors. This implies that the reconciled total forecast is more robust forecast than the total base forecast.

4.2 | Accuracy of regional forecasts

To better understand the increase in performance made by reconciling the forecasts, we will assess the change in performance at the regional level. From Table 2, we can see that the regional forecast seems to produce the most accurate total forecast when aggregated. Hence, we will compare the forecast at the regional level produced by the hierarchy with the original regional forecasts.

Figure 7 shows the change in performance on a regional level between the regional base forecast and the reconciled regional forecast. Notice the average decrease in RMSE and MAE is 4.7% and 0.9%, respectively, which is substantially lower than the change seen at the highest level of the hierarchy. The explanation for the discrepancy between the performance improvement on the regional and total levels lies in the residual correlation between the regions.

Figure 8A shows the residual correlation between the regions for the original regional forecast. From this, it is obvious that there is a positive correlation between most of the regions. In practice, this means that the variance of the forecast residuals for the "Total" model will be larger than the sum of the variances of the forecast residuals for the regional forecasts. Additionally, we see some grouping structures, for example, regions

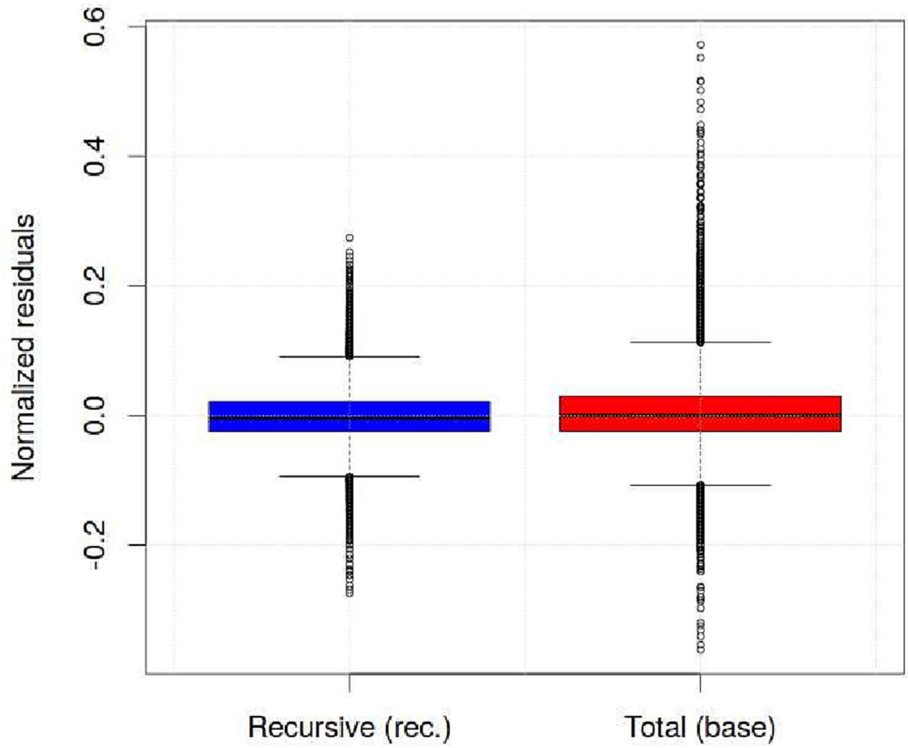


FIGURE 6 Boxplot of the residuals of the reconciled total forecast and the residuals of the base total forecast.

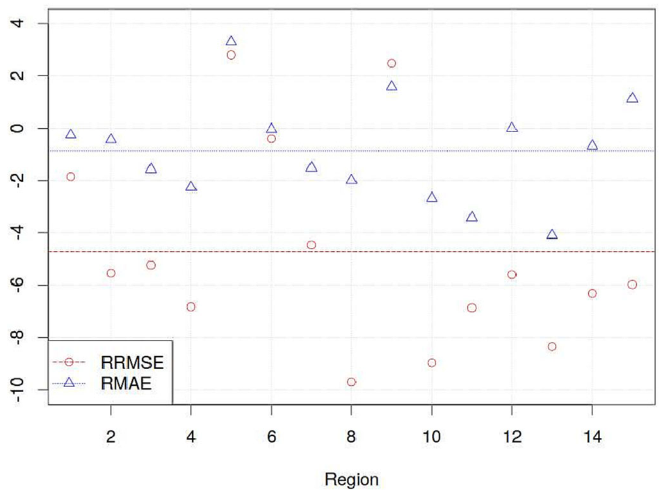


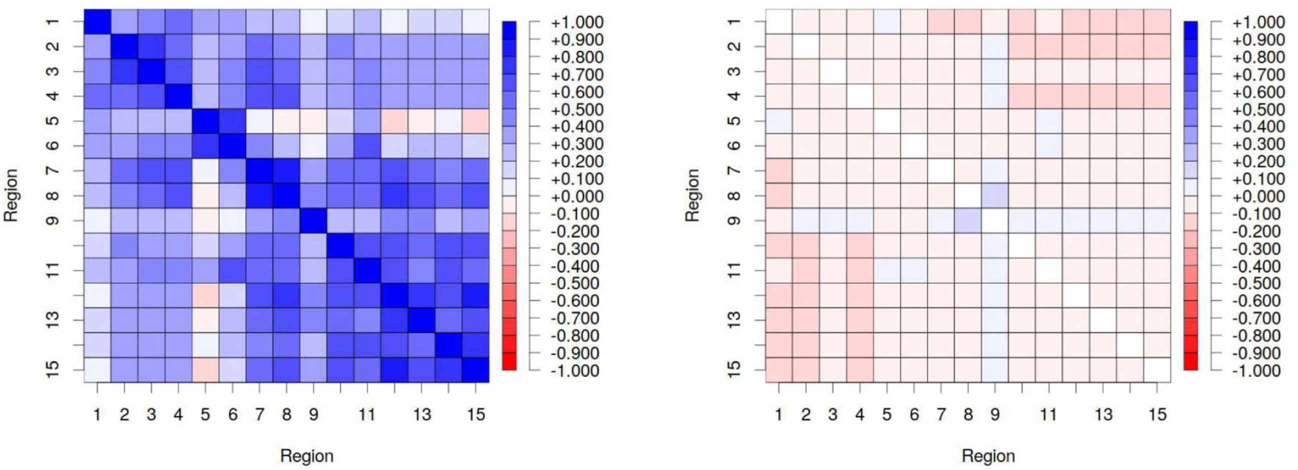
FIGURE 7 Change in RMSE and MAE between the regional base forecast and the reconciled regional forecast.

$\{2,3\}$, $\{5,6\}$, and $\{10,11,12,13,14,15\}$. Comparing this with the regional map (Figure 1A) indicates that some of these groups could be based on some geographical phenomena.

Figure 8B shows the change in residual correlation when reconciling. From this figure, it is clear that the residual correlation between the regions is decreased (≈ 0.05 on average), which in turn results in better performance on the total aggregate.

4.3 | Performance of station forecasts

Table 3 shows that while forecast reconciliation may be effective at an aggregate or regional level, it does not appear to improve forecasts at the level of individual stations. This could be due to the fact that the assumption that all residuals are Gaussian, which is made during reconciliation,



(A) Regional base forecast errors.

(B) Change in correlation when reconciling (reconciled - base).

FIGURE 8 Correlation between regional base forecast errors and the change in correlation after reconciliation.

TABLE 3 Performance of base and reconciled station forecasts.

Model	RMSE	MAE	NRMSE	NMAE	SMAPE
Station	2459	897	0.128	0.082	0.701
Reconciled	3055	1308	0.333	0.127	0.892

may not hold true at the station level, particularly for smaller stations. Additionally, it can be seen that the mean of the performance metrics that are normalized by capacity are particularly affected by this limitation. This is likely because the errors introduced by the hierarchical structure are given more weight for stations with smaller capacities, and in some cases, this may result in, for example, negative reconciled forecasts.

It is important to note that while the reconciled forecasts for individual stations may not have improved, we saw substantial improvements on the total and regional level, as demonstrated above. The fact that the station-level forecasts did not improve is unlikely to have a significant impact on the overall accuracy of the hierarchy, given the relatively small contribution of individual stations to the total forecast. Therefore, while the limitations of forecast reconciliation at the station level should not be overlooked, they do not negate the benefits of the approach for forecasting wind power production at higher levels of the hierarchy.

5 | DISCUSSION

We have seen that using spatial reconciliation for hierarchical forecast of wind power in DK1 produces improvements to the commercial forecasts. To maintain comparability, the same producer of NWPs (ECMWF) has been used for all the considered forecasts. The methods used by the commercial forecast provider to generate base forecasts were largely the same as we used for producing base forecasts for the stations; see, for example, Nielsen.[22] Combined with the fact that the same weather provider has been used for all models, this ensures that any potential improvements to performance are not a result of including additional information about the weather, but it is only a result of increased robustness via the structural information, that is, the hierarchy.

In this paper, the focus was on how to improve forecasts of wind power production using information from different spatial levels. After the spatial reconciliation of the forecasts, we saw that the performance of the total forecast was improved substantially. Further investigation showed that these improvements were partly due to improvements to the regional forecasts but also due to decorrelation of the residuals of the regional forecasts. After analysis of the individual stations, we did see that on average the individual station forecasts were getting worse after reconciliation. This is largely because of the assumption of normality in the base residuals does not hold and reconciliation might lead to nonsensible results (e.g., negative station forecasts). If individual station forecasts are of importance, then one should consider additional constraints to the optimization problem given in Equation (11) and apply, for example, non-negative reconciliation, as described by Wickramasuriya.[30]

The presented methods will lead to coherent forecasts across all spatial aggregation levels. The coherent forecasts imply that key network operators, such as the TSO and DSO, have a better basis for taking coordinated actions, which is of increasing importance for the future low-carbon power system. As discussed in Section 2, the improved accuracy of the regional forecasts are helpful in alleviating or preventing distributional challenges. As mentioned in the introduction, forecast accuracy has an impact on the price of electricity. It would be interesting to investigate the effect the improved accuracy of the DK1 forecast would have on a trading strategy, particularly considering that much of the improvement was seen on the largest forecast errors.

We found that using adaptive estimation of the covariance matrix produced very similar results to the nonadaptive. This suggests that the covariance structure is somewhat stable over time and does not change greatly on a monthly basis. However, we cannot say anything about what would happen if we were to estimate the covariance matrix much more frequently, for example, every day. We were simply not able to do so, partly due to a relatively long delay in data availability, and partly due to computational limitations when back-testing.

For future studies, it might be interesting to test the effect of updating the covariance matrix more frequently. Especially regarding adaptive estimation, since this relies on capturing local structural changes. Additionally, it would be interesting to investigate potential relations between wind direction and covariance structure. It would also be interesting to expand the hierarchy and include a larger geographical area. Initially, it could be expanded to include the entirety of Denmark. In an increasingly connected Europe, coherent and accurate forecasts are becoming increasingly relevant everywhere.

It is important to note that the focus of the current study was on achieving improvements of existing forecasts with a low additional computational burden. There may be benefits to exploring improved station forecasts through the use of more detailed models that could include more covariates; see, for example, Sørensen.[14] Including additional covariates in the model may lead to more accurate forecasts, but it would come at the cost of increased computational resources. It would be interesting to conduct further research to compare the performance of models that include all relevant covariates with models that prioritize computational efficiency. Such research could help identify the trade-off between performance and computational resources and guide future forecasting efforts.

6 | CONCLUSION

The goal of this paper was to implement a spatial hierarchy in the context of day-ahead forecasts of wind power production based on NWPs. Specifically, we wanted to investigate the possibility of getting meaningful improvements by utilizing localized substation data in a computationally inexpensive and simple way that is easy to implement in an operational setting.

When modeling, we saw that correcting for both horizon and temperature yielded improvements for station models. However, when comparing the reconciled forecast of the different stations model, the performance improvement was diminished. Implying that most of the benefits at the total level can be achieved using a simple model at the station level. This alleviates some computational burden and demonstrates that a spatial hierarchy can be an easy way of incorporating data from substations into an already existing setup.

We found that by reconciling the spatial hierarchy and comparing it with operationally used commercial state-of-the-art forecasts at the highest level of the hierarchy, RMSE and MAE were reduced by 20.5% and 16.8%, respectively. Especially, the number and magnitude of the most extreme residuals were reduced.

Using adaptive estimation of the covariance matrix produced results very similar to the nonadaptive covariance structure, implying a very stable covariance structure for the wind power stations and their regions. Assessing the effect reconciliation has on the regional forecasts, it is evident that there are performance improvements in the individual regions. With that being said, the performance improvement we saw for the total forecast is partly due to lower variance in the individual regions but also due to a quite significant decrease in interregional residual correlation.

Hierarchical forecasting implies a transfer of information between aggregation levels, which typically leads to improved forecasts on the individual levels. Furthermore, forecasting using spatial hierarchies will ensure an improved consistency between aggregated forecasts on different aggregation levels, and the coherent forecasts will help facilitating a better coordination between various operators in the grids.

ACKNOWLEDGEMENTS

The authors acknowledge the obtained funding from *ebalance-plus* (H2020 No. 864283), *Sustainable plus energy neighbourhoods (syn.ikia)* (H2020 No. 869918), *ARV* (H2020 No. 101036723), *ELEXIA* (HE No. 101075656), *Flexible Energy Denmark (FED)* (IFD 8090-00069B), *DynFlex* (IFD Mission Green Fuels), and *IEA Wind Task 51* (EUDP 134-22015).

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

FINANCIAL DISCLOSURE

None reported.

DATA AVAILABILITY STATEMENT

Due to the nature of this research, the data used are confidential, so supporting data are not available.

ORCID

Mads E. Hansen https://orcid.org/0000-0001-5782-2230

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/we.2819.

REFERENCES

1. Energinet. Energi data portal.
2. Spodniak P, Ollikka K, Honkapuro S. The impact of wind power and electricity demand on the relevance of different short-term electricity markets: the Nordic case. *Appl. Energy*. 2021;283:116063.
3. Wickramasuriya SL, Athanasopoulos G, Hyndman RJ. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J Am Stat Assoc*. 2019;114(526):804-819.
4. Grunfeld Y, Griliches Z. Is aggregation necessarily bad? *Rev Econ Stat*. 1960;42(1):1-13.
5. Schwarzkopf AB, Tersine RJ, Morris JS. Top-down versus bottom-up forecasting strategies. *Int J Prod Res*. 1988;26(11):1833-1843.
6. Hyndman RJ, Ahmed RA, Athanasopoulos G, Shang HL. Optimal combination forecasts for hierarchical time series. *Comput Stat Data Anal*. 2011;55(9): 2579-2589.
7. Athanasopoulos G, Hyndman RJ, Kourentzes N, Petropoulos F. Forecasting with temporal hierarchies. *Eur J Oper Res*. 2017;262(1):60-74.
8. Nystrup P, Lindström E, Pinson P, Madsen H. Temporal hierarchies with autocorrelation for load forecasting. *Eur J Oper Res*. 2020;280(3):876-888.
9. Bergsteinsson HG, Møller JK, Nystrup P, Ólafur P. Pálsson, Guericke D, Madsen H. Heat load forecasting using adaptive temporal hierarchies. *Appl Energy*. 2021;292:116872.
10. Zhang Y, Dong J. Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation. *IEEE Trans Power Syst*; 2018. doi:10.1109/tpwrs.2018.2868175
11. Jeon J, Panagiotelis A, Petropoulos F. Probabilistic forecast reconciliation with applications to wind power and electric load. *Eur J Oper Res*. 2019; 279(2):364-379.
12. Di Modica C, Pinson P, Taieb SB. Online forecast reconciliation in wind power prediction. *Electr Power Syst Res*. 2021;190:106637.
13. Bai L, Pinson P. Distributed reconciliation in day-ahead wind power forecasting. *Energies*. 2019;12(6):1112.
14. Sørensen ML, Nystrup P, Bjerregard MB, Møller JK, Bacher P, Madsen H. Recent developments in multivariate wind and solar power forecasting. *WIREs Energy Environ*. 2023;12(2):e465.
15. Giebel G, Kariniotakis G. Wind power forecasting—a review of the state of the art. In: Kariniotakis G, ed. *Renewable Energy Forecasting: From Models to Applications*, Woodhead Publishing Series in Energy: Woodhead Publishing; 2017:59-109.
16. Lange M, Focken U. *Physical Approach to Short-term Wind Power Prediction*. 1st ed. Berlin, Heidelberg: Springer; 2006.
17. Pinson P, Tastu J. Discussion of "prediction intervals for short-term wind farm generation forecasts" and "combined nonparametric prediction intervals for wind power generation". *IEEE Trans Sustain Energy*. 2014;5(3):1019-1020.
18. Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. *Renew Sustain Energy Rev*. 2008;12(6):1725-1744.
19. Soman SS, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. In: North American Power Symposium 2010, Arlington, TX, USA; 2010:1-8. doi:10.1109/NAPS.2010.5619586
20. Lin J, Sun Y, Cheng L, Gao W. Assessment of the power reduction of wind farms under extreme wind condition by a high resolution simulation model. *Appl Energy*. 2012;96:21-32.
21. Xu M, Pinson P, Lu Z, Qiao Y, Min Y. Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting. *Wind Energy*. 2016;19(12):2321-2336.
22. Nielsen TS, Madsen H, Nielsen HA, Landberg L, Giebel G. Zephyr—the prediction models. In: WIP-Renewable Energies/ETA; 2001:1248.
23. Madsen H, Holst J. *Modelling Non-linear and Non-stationary Time Series*: Department of Applied Mathematics and Computer Science, Technical University of Denmark; 2006.
24. Boyd S, Vandenberghe L. *Convex Optimization*: Cambridge University Press; 2004.
25. Pitman J. *Probability*. 1st ed., Springer; 1993.
26. Nielsen HA, Nielsen TS, Madsen H, Pindado MJSI, Marti I. Optimal combination of wind power forecasts. *Wind Energy*. 2007;10(5):471-482.
27. Møller JK, Zugno M, Madsen H. Probabilistic forecasts of wind power generation by stochastic differential equation models. *J Forecast*. 2016;35(3): 189-205.
28. Madsen H, Nielsen HA, Nielsen TS. A tool for predicting the wind power production of off-shore wind plants. In: Proceedings of the Copenhagen Offshore Wind Conference & Exhibition; 2005.
29. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. 2006;22(4):679-688.
30. Wickramasuriya SL, Turlach BA, Hyndman RJ. Optimal non-negative forecast reconciliation. *Stat Comput*. 2020;30(5):1167-1182.

How to cite this article: Hansen ME, Peter N, Møller JK, Henrik M. Reconciliation of wind power forecasts in spatial hierarchies. *Wind Energy*. 2023;26(6):615-632. doi:10.1002/we.2819

APPENDIX A

TABLE A1 Accuracy of the total DK1 forecast made using the different reconciliation methods and by aggregating the regional and station forecasts, respectively.

Model	NRMSE	NMAE	SMAPE	RRMSE	RMAE	RSMAPE
Total (base)	0.0581	0.0400	0.2981	0%	0%	0%
Region[a]	0.0514	0.0356	0.2694	−11.53%	−11.00%	−9.63%
Station 1 (wind)[a]	0.0528	0.0367	0.2808	−9.12%	−8.25%	−5.80%
Station 2 (wind + horizon)[a]	0.0524	0.0366	0.2760	−9.81%	−8.50%	−7.41%
Station 3 (wind + horizon + temp)[a]	0.0522	0.0365	0.2760	−10.15%	−8.75%	−7.41%
Recursive 1 (wind)[b]	0.0462	0.0333	0.2612	−20.48%	−16.75%	−12.38%
Recursive 2 (wind + horizon)[b]	0.0461	0.0333	0.2598	−20.65%	−16.75%	−12.85%
Recursive 3 (wind + horizon + temp)[b]	0.0461	0.0333	0.2599	−20.65%	−16.75%	−12.81%
Adaptive $\lambda = \frac{11}{12}$ (wind)[b]	0.0463	0.0333	0.2612	−20.31%	−16.75%	−12.38%
Adaptive $\lambda = \frac{5}{6}$ (wind)[b]	0.0462	0.0333	0.2611	−20.48%	−16.75%	−12.41%
Adaptive $\lambda = \frac{1}{2}$ (wind)[b]	0.0463	0.0332	0.2866	−20.31%	-17.00%	−3.86%

[a]Aggregated forecast.
[b]Reconciled forecast.