International Journal of Forecasting 37 (2021) 1127–1146

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Dimensionality reduction in forecasting with temporal hierarchies

Peter Nystrup [a,b,*], Erik Lindström [a], Jan K. Møller [b], Henrik Madsen [b]

[a] Centre for Mathematical Sciences, Lund University, Sweden
[b] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

ARTICLE INFO

Keywords:
Temporal aggregation
Reconciliation
Spectral decomposition
Shrinkage
Load forecasting
Realized volatility

ABSTRACT

Combining forecasts from multiple temporal aggregation levels exploits information differences and mitigates model uncertainty, while reconciliation ensures a unified prediction that supports aligned decisions at different horizons. It can be challenging to estimate the full cross-covariance matrix for a temporal hierarchy, which can easily be of very large dimension, yet it is difficult to know a priori which part of the error structure is most important. To address these issues, we propose to use eigendecomposition for dimensionality reduction when reconciling forecasts to extract as much information as possible from the error structure given the data available. We evaluate the proposed estimator in a simulation study and demonstrate its usefulness through applications to short-term electricity load and financial volatility forecasting. We find that accuracy can be improved uniformly across all aggregation levels, as the estimator achieves state-of-the-art accuracy while being applicable to hierarchies of all sizes.

1. Introduction

Many real-life decision problems involve multiple time horizons. For example, when executing large orders in financial markets, trading is carried out throughout the day to limit price impact and secure the best possible average price (Boyd et al., 2017). In operational energy planning, production has to meet expected demand over the next hours, the next day, the next week, and even longer horizons. It is often beneficial to take into account longer-term forecasts when making short-term decisions to ensure that current decisions do not have undesirable implications for future possibilities. Reconciliation is necessary because optimal decision-making requires coherent forecasts. Decision-makers who have had to make decisions based on forecasts that were not coherent know the issues that this may cause (Nystrup, Lindström et al., 2020).

Top-down and bottom-up approaches have traditionally been used to produce coherent forecasts for a hierarchy. According to the former, forecasts are generated for the time series at the top level and then disaggregated down all the way to the bottom level, while for the latter, forecasts are generated at the very bottom level and then aggregated up (Athanasopoulos et al., 2009; Gross & Sohl, 1990). These approaches ensure that forecasts add up across a hierarchy but do not take advantage of information differences by combining forecasts constructed at different temporal aggregation levels.

1.1. Benefits of reconciliation

Temporal aggregation. Temporal aggregation has been studied since the seminal work by Amemiya and Wu (1972) and Tiao (1972) (see Silvestrini & Veredas, 2008, for a literature review). Different temporal aggregations can reveal important information about the underlying data-generating process. In financial applications, it is common to aggregate intraday data to improve estimates of daily volatility (Ghysels et al., 2006; Nystrup, Kolm

* Correspondence to: Centre for Mathematical Sciences, Lund University, Box 118, 221 00 Lund, Sweden.
E-mail address: peter.nystrup@matstat.lu.se (P. Nystrup).

https://doi.org/10.1016/j.ijforecast.2020.12.003

et al., 2020). When temporal aggregation is applied to a time series, it can strengthen or attenuate different features. Non-overlapping temporal aggregation is a filter of high-frequency components. At an aggregate view, low-frequency components, such as trend and cycle, will dominate. The opposite is true for disaggregate data, where short-term seasonality may be visible. Hence, temporal aggregation can be seen as a tool to better understand and model the data at hand.

Forecast combination. Several studies have shown that combining forecasts from multiple aggregation levels can lead to improvements in forecast accuracy, while overcoming the need to select a single optimal level (Kourentzes et al., 2014, 2017; Petropoulos & Kourentzes, 2015). The implied combination mitigates model uncertainty, which is a contributory cause to this improvement. Forecast combination is generally beneficial, leading to a reduction of forecast error variance (see, e.g., Clemen, 1989; Hall & Mitchell, 2007). Ways of combining forecasts have been investigated extensively, resulting in various advanced weighting methods (Makridakis et al., 2020). Yet, simple approaches, such as the unweighted average, are often found to perform as well as more sophisticated methods (Timmermann, 2006).

Coherency. Temporal hierarchies for forecasting can be constructed for any time series by means of non-overlapping temporal aggregation. Rather than building one complex model that captures all temporal attributes (see, e.g., Clements et al., 2016; Livera et al., 2011; Nystrup et al., 2017), forecasts for different horizons can be made with different (simple) methods using temporal hierarchies. Forecasts that are produced by different approaches and are based on different information will most likely not be coherent. This incoherency can lead to decisions that are not aligned, or even conflicting. Deploying the framework proposed by Athanasopoulos et al. (2017), forecasts from different aggregation levels can be combined to yield temporally reconciled, accurate, and robust forecasts, independently of forecasting models (Nystrup, Lindström et al., 2020).

1.2. Contribution

Dimensionality. There is a strong incentive to consider ever-larger hierarchies by adding data with higher and higher resolution, because the biggest improvements typically occur at the highest level of a hierarchy, where information from lower levels (i.e., higher-resolution data) is aggregated up (Athanasopoulos et al., 2017, 2011; Rostami-Tabar et al., 2013). Furthermore, the dimension of the hierarchy grows proportionally to the forecast horizon, as explained in Section 3.1. Given that the purpose of temporal aggregation is to harness information in a time series at different frequencies, auto- and cross-covariance information in the forecast errors should be beneficial when reconciling forecasts. However, the size of the cross-covariance matrix quickly becomes an issue as the dimension of the temporal hierarchy grows. In practice it can be necessary to reduce the dimension of the hierarchy by shortening the forecast horizon, omitting aggregation levels, or settling on a simple approximation of the covariance structure that neglects auto-, cross-covariances, or both. Unfortunately, it is difficult to know a priori which parts of the hierarchy or error structure are most important.

Eigendecomposition. We propose a novel estimator for reconciling forecasts in a temporal hierarchy based on an eigendecomposition of the cross-correlation matrix. Our proposal is to reduce the dimension of the temporal correlation matrix, rather than the dimension of the hierarchy, by only keeping its most important structure. To the best of our knowledge, we are the first to propose a general approach to dimensionality reduction for forecast reconciliation. By decomposing the cross-correlation matrix and estimating its inverse based on a small subset of its eigenvectors plus a regularization term, our estimator extracts as much information as possible from the error structure given the data available. Our estimator facilitates information sharing between aggregation levels, overcomes the problem of inverting a potentially singular estimate of the cross-covariance matrix, and is robust to noise and high dimensionality.

Results. We document the usefulness of the proposed estimator through applications to short-term electricity load and financial volatility forecasting. We show that incorporating information about the cross-covariance structure significantly improves forecast accuracy, compared with traditional approaches and compared with the diagonal covariance-matrix approximations proposed by Athanasopoulos et al. (2017). To better understand the differences and advantages of our estimator compared with those proposed by Nystrup, Lindström et al. (2020), we compare their performance in a simulation study based on load data. The simulation study shows that it is possible to significantly reduce the dimension of the temporal covariance matrix without loss of accuracy such that it can handle much larger hierarchies. When limited data are available compared with the dimension of the temporal hierarchy, forecast accuracy can still be improved by considering dependencies between aggregation levels.

Outline. We discuss related work in Section 2. In Section 3, we outline the forecast reconciliation problem and its relation to generalized least-squares estimation. We outline our proposal for reducing the dimension of the cross-correlation matrix using spectral decomposition in Section 4. Results from an application to short-term load forecasting are presented in Section 5. The simulation study is presented in Section 6. In Section 7, we consider an application to intraday volatility forecasting. Finally, we discuss the results in Section 8 before concluding in Section 9.

2. Related work

Regression. Hyndman et al. (2011) used ordinary least squares (OLS) to compute reconciled forecasts for a structural hierarchy, thereby avoiding the problem of estimating the covariances of base forecasts. Subsequently, Hyndman et al. (2016) proposed weighted least squares

(WLS) for forecast reconciliation, taking account of the variances on the diagonal of the covariance matrix but ignoring the offdiagonal covariances. Wickramasuriya et al. (2019) studied the generalized least-squares (GLS) estimator and found the inclusion of covariance information in the reconciliation procedure to be beneficial for forecast accuracy when combined with a linear shrinkage estimator.

Optimization. Van Erven and Cugliari (2015) formulated the forecast reconciliation problem as an optimization problem instead of a regression model. Their formulation included only a diagonal weight matrix and, thus, no information sharing between base forecasts. They proved that the reconciled forecasts are at least as good as the base forecasts for any loss function that is based on a Bregman divergence. Based on geometric interpretations, Panagiotelis et al. (2021) established the connection between the regression and optimization formulations and generalized the proof to any monotonic function of distance. Wickramasuriya et al. (2020) proposed algorithms for solving the least-squares minimization problem with constraints that ensure that the coherent forecasts are nonnegative. The reconciliation problem has also been extended to probabilistic forecasts (Gamakumara et al., 2018; Jeon et al., 2019; Taieb et al., 2017, 2020).

Temporal hierarchy. Temporally aggregated time series can be represented as a hierarchical time series. This led Athanasopoulos et al. (2017) to show that it is possible to use the reconciliation framework proposed by Hyndman et al. (2011) to produce temporally coherent forecasts. They proposed three diagonal approximations of the sample covariance matrix. By construction, these approximations share the property that they ignore any autocorrelation in the forecast errors. Taieb (2017) extended their work by introducing regularization terms to obtain sparse and smooth adjustments that satisfy aggregation constraints and minimize forecast errors.

Spatio–temporal. Both structural and temporal forecast reconciliation have been successfully applied to a number of energy-related time series (Yang, Quan, Disfani, Liu, 2017; Yang, Quan, Disfani, Rodríguez-Gallegos, 2017; Zhang & Dong, 2018). Recently, several attempts have been made to combine structural and temporal hierarchies. Kourentzes and Athanasopoulos (2019) suggested a cross-temporal approach to reconciling forecasts across both geographical divisions and planning horizons for Australian tourist flows. Meanwhile, Yagli et al. (2019) and Spiliotis et al. (2020) used sequential approaches to spatio–temporal reconciliation of solar-power and electricity-load forecasts, respectively.

Cross-covariances. Our work is most closely related to that of Nystrup, Lindström et al. (2020), who extended the work of Athanasopoulos et al. (2017) by proposing four different estimators that take into account the auto- and cross-covariance structure when reconciling short-term electricity-load forecasts in a temporal hierarchy. They showed that by taking account of auto- and cross-covariances when reconciling forecasts, accuracy could be significantly improved uniformly across all aggregation levels. The reported accuracy gains showed the potential value of the information embedded in the error structure, which motivated us to analyze this in greater detail and extend their work in this article by proposing a novel estimator based on an eigendecomposition of the cross-correlation matrix that can be applied to hierarchies of much larger dimension.

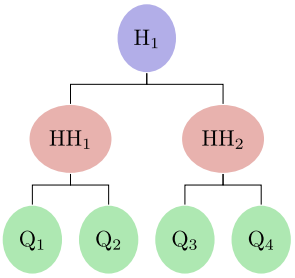


Fig. 1. Temporal hierarchy for quarter-hourly, half-hourly, and hourly series.

3. Forecast reconciliation

3.1. Temporal hierarchies

Hierarchies are defined by linear constraints. Given n individual *base* forecasts stacked in a column vector $\hat{y} \in \mathbb{R}^n$, we are interested in finding *reconciled* forecasts $\tilde{y} \in \mathbb{R}^n$ that are coherent. By introducing a matrix

$$G = \left[0_{m\times(n-m)} \middle| I_m \right] \tag{1}$$

of order $m \times n$ that extracts the m bottom-level forecasts, we can write the reconciliation constraint(s) as

$$\tilde{y} = SG\tilde{y}. \tag{2}$$

Reconciliation is needed when base forecasts $\hat{y}$ do not satisfy this constraint.

The matrix SG projects the base forecasts onto the coherent subspace spanned by the summation matrix S. While there are multiple ways to define a summation matrix, in all cases its columns will span the same coherent subspace, which is unique (Panagiotelis et al., 2021). For the hierarchy illustrated in Fig. 1 we define

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

as quarter-hourly forecasts should reconcile to half-hourly forecasts, which should reconcile to hourly forecasts.

In general, if there are $k \in \{k_1, \ldots, k_K\}$ aggregation levels, where k is a factor of m, with $k_1 = m$, $k_K = 1$, and m/k is the number of observations at aggregation level k, then the summation matrix is given by

$$S = \begin{bmatrix} I_{m/k_1} \otimes 1_{k_1}^T \\ \vdots \\ I_{m/k_K} \otimes 1_{k_K}^T \end{bmatrix}. \tag{3}$$

Here $\otimes$ denotes the Kronecker product, $I_{m/k}$ is an identity matrix of order m/k, and 1_k is a k-vector of ones. For example, the hierarchy illustrated in Fig. 1 has aggregation levels $k_1 = 4$, $k_2 = 2$, and $k_3 = 1$ with $m = 4$ and $n = 7$.

Note that the dimension of the hierarchy grows proportionally to the forecast horizon when keeping the same aggregation levels. For example, for the hierarchy illustrated in Fig. 1, if we were to forecast two hours ahead with the same aggregation levels, the dimension would be 14 rather than seven, since we would then have eight quarter-hourly, four half-hourly, and two hourly base forecasts. We could also add a two-hourly aggregation level, in which case the dimension of the hierarchy would increase to 15. This assumes a direct, as opposed to a recursive, approach to multi-step forecasting, which is consistent with using a temporal hierarchy in the first place.

3.2. Relation to generalized least squares

Hyndman et al. (2011) and Athanasopoulos et al. (2017) formulated the structural and temporal reconciliation problems, respectively, as a linear regression model. The reconciled forecasts can be found as the *generalized* least-squares estimate:

$$\begin{aligned} &\text{minimize} && \left(\hat{y}-\tilde{y}\right)^T \Sigma^{-1} \left(\hat{y}-\tilde{y}\right) \\ &\text{subject to} && \tilde{y} = SG\tilde{y}, \end{aligned} \tag{4}$$

where $\tilde{y} \in \mathbb{R}^n$ is the variable and the parameter $\Sigma \in \mathbb{R}^{n\times n}_{++}$ is the covariance matrix for the *coherency errors* $\varepsilon = \hat{y} - \tilde{y}$, which are assumed to be multivariate Gaussian and unbiased, i.e., have zero mean. The reconciled forecasts are optimal in that the base forecasts are adjusted by the least amount (in the sense of least squares) so that these become *coherent* (Nystrup, Lindström et al., 2020).

In formulation (4), the inverse covariance matrix Σ^{-1} is used to scale deviations from the base forecasts. It is more expensive to adjust base forecasts with a higher precision, which leads to the lowest error on average (Panagiotelis et al., 2021). Hyndman et al. (2011) and Van Erven and Cugliari (2015) argued for selecting uniform weights to increase the importance of forecasting the aggregate. The unweighted case $\Sigma = I$ corresponds to *ordinary* least-squares estimation. Unweighted implies that each aggregation level is assigned the same total weight. However, due to the different scales of the various levels, this means that OLS emphasizes the highest aggregation levels.

If Σ were known, the solution to (4) would be given by the GLS estimator

$$\tilde{y} = S\left(S^T\Sigma^{-1}S\right)^{-1}S^T\Sigma^{-1}\hat{y}. \tag{5}$$

Hyndman et al. (2016) showed how the computations that are required to evaluate (5) can be handled efficiently by exploiting the sparse structure of the summation matrix. Wickramasuriya et al. (2019) derived an alternative representation that is less demanding in terms of computation.

3.3. Optimal covariance-matrix estimator for forecast reconciliation

Wickramasuriya et al. (2019) showed that, in general, Σ is not known and is not identifiable. Assuming that the base forecasts are unbiased, they showed that the variance of the *reconciled forecast errors* $\tilde{e} = y - \tilde{y}$ is given by

$$\operatorname{var}\left[y-\tilde{y}\right] = SPBP^TS^T, \tag{6}$$

where $B = \operatorname{var}\left[y-\hat{y}\right]$ is the covariance matrix of the *base forecast errors* $\hat{e} = y - \hat{y}$, for any P such that $SPS = S$, i.e., SP is a projection matrix (Panagiotelis et al., 2021).

Wickramasuriya et al. (2019) showed that the optimal reconciliation matrix that minimizes the sum of the variances of the reconciled forecast errors $\operatorname{tr}\left[SPBP^TS^T\right]$, such that $SPS = S$, is given by

$$P = \left(S^TB^{-1}S\right)^{-1}S^TB^{-1}. \tag{7}$$

This is the GLS estimator (5) with a different covariance matrix. Hence, they provided theoretical justification for using the covariance matrix for the base forecast errors as a proxy for the unidentifiable covariance matrix for the coherency errors. Although it does not suffer from a lack of identifiability, it can still be challenging to estimate. Essentially, optimal forecast reconciliation boils down to finding a good estimator of the covariance matrix of the base forecast errors.

4. Dimensionality reduction using spectral decomposition

Nystrup, Lindström et al. (2020) extended the ideas of Wickramasuriya et al. (2019) to a temporal hierarchy by proposing four different estimators of the cross-covariance matrix based on the in-sample base forecast errors. As the purpose of temporal aggregation is to exploit important information about a time series at different frequencies, potential information in the correlation structure of the forecast errors should be useful for improving the accuracy of the reconciled forecasts.

Their first estimator was based on estimating the *autocovariance* matrix within each aggregation level, while ignoring correlations between aggregation levels. Their second estimator assumed that forecast errors could be approximated by a stationary first-order *Markov* process, leading to a Toeplitz structure within each aggregation level. Their third proposal was to estimate a sparse representation of the inverse cross-correlation matrix using the graphical least absolute shrinkage and selection operator (*GLASSO*). Their fourth proposal was to consider a Stein-type *shrinkage* estimator of the sample cross-correlation matrix.

The first three estimators assume independence, sparsity, or a specific model for the forecast errors. The spectral scaling approach suggested in this article is different in that it is a data-driven, *model-free* approach that does not assume a specific structure in the cross-correlation matrix for the base forecast errors. Imposing structure reduces the variance of the result; however, if the assumed structure is misspecified, the estimator can be arbitrarily bad. Below, we outline our proposal for reducing the dimension of the temporal correlation matrix, and we account for its connection with the shrinkage estimator proposed by Nystrup, Lindström et al. (2020).

4.1. Eigendecomposition

We consider the cross-correlation rather than the cross-covariance matrix to avoid problems with heteroscedasticity. In most cases, forecast errors from different temporal aggregation levels have very different variances. Doing the eigendecomposition of the cross-covariance matrix would cause the eigenvectors corresponding to the largest eigenvalues to align with the errors from the highest aggregation levels in order to retain as much variance as possible.

Correlations between base forecast errors from the various aggregation levels can be readily estimated using the standard formula for empirical correlation. The ij-th element of the empirical cross-correlation matrix R is

$$r_{ij} = \frac{\sum_{t=1}^{T}\left(\hat{e}_{t,i} - \bar{\hat{e}}_i\right)\left(\hat{e}_{t,j} - \bar{\hat{e}}_j\right)}{\sqrt{\sum_{t=1}^{T}\left(\hat{e}_{t,i} - \bar{\hat{e}}_i\right)^2}\sqrt{\sum_{t=1}^{T}\left(\hat{e}_{t,j} - \bar{\hat{e}}_j\right)^2}}, \quad \forall i, j \in 1, \dots, n. \tag{8}$$

Even though base forecast errors from different aggregation levels are measured at different frequencies, for each time period t (one hour for the hierarchy in Fig. 1) we have one observation of each base forecast error that can be used to estimate the cross-correlation matrix.

We propose to use spectral decomposition to extract the most important structure from the cross-correlation matrix of the base forecast errors. This is done through an eigendecomposition of the empirical cross-correlation matrix

$$R = V \Lambda V^T, \tag{9}$$

where V is an orthogonal $n \times n$ matrix whose columns are the eigenvectors of R in order of decreasing eigenvalue, and Λ is the diagonal matrix whose elements are the corresponding eigenvalues, $\Lambda_{ii} = \lambda_i \geq 0$. This is the same decomposition that is the core of principal component analysis (Tipping & Bishop, 1999). The ith eigenvalue is equal to the variance explained by the ith eigenvector, with $\mathrm{tr}(R) = \mathrm{tr}(\Lambda) = n$.

4.2. Linear shrinkage

Nystrup, Lindström et al. (2020) considered a linear shrinkage estimator of the sample cross-correlation matrix:

$$R_{\text{shrink}} = (1 - \nu) R + \nu I_n, \tag{10}$$

where $0 \leq \nu \leq 1$ is a regularization parameter that controls the degree of shrinkage toward the identity matrix. The shrinkage estimate is always positive definite and well conditioned when $\nu > 0$. The shrinkage estimator shrinks all cross-correlations uniformly toward zero. In the Appendix, we show that this is equivalent to constraining the size of the adjustments made to the bottom-level base forecasts, as in generalized ridge regression.

Ledoit and Wolf (2003) showed that the value of the shrinkage parameter that minimizes the mean squared difference between the shrinkage estimator and the true correlation matrix is

$$\nu = \frac{\sum_{i \neq j} \mathrm{var}\left[r_{ij}\right]}{\sum_{i \neq j} r_{ij}^2}, \tag{11}$$

where r_{ij} is the ij-th element of the empirical cross-correlation matrix R. The optimal value of the shrinkage parameter decreases with the variance of the empirical cross-correlation estimate, which is readily calculated from the standardized base forecast errors. Therefore, with increasing sample size, the influence of the shrinkage target I_n diminishes, as it is assigned less weight. Furthermore, the optimal value of the shrinkage parameter decreases as the size of the correlations increases to protect the shrinkage estimate against a misspecified target.

The linear shrinkage estimator of the cross-correlation matrix (10) is equal to[1]

$$R_{\text{shrink}} = V \Lambda_{\text{shrink}} V^T, \tag{12}$$

where Λ_{shrink} is the diagonal matrix whose elements are the corresponding shrunken eigenvalues $\lambda_i^{\text{shrink}} = (1 - \nu)\lambda_i + \nu > 0$, which are always positive when $\nu > 0$. This means that shrinking the elements of R to the target I_n with intensity ν is equivalent to keeping the eigenvectors of R and shrinking its eigenvalues to the target (i.e., one) with the same intensity ν.

As shown by Ledoit and Wolf (2004), sample eigenvalues are more dispersed around their true mean than true ones, and the excess dispersion is equal to the error in the sample correlation matrix. Excess dispersion implies that the largest sample eigenvalues are biased upward and the smallest ones downward. Therefore, we can improve upon the sample correlation matrix by shrinking its eigenvalues toward one, while retaining the sample eigenvectors. This can be extended by applying individual shrinkage intensities to each of the eigenvalues, as proposed by Ledoit and Wolf (2012).

4.3. Spectral scaling

Based on the eigendecomposition of the cross-correlation matrix of the base forecast errors (9), we propose to reconstruct its inverse (i.e., a filtered precision matrix) using the n_{eig} first eigenvectors, assuming $\lambda_i = \sigma^2$ for $i = n_{\text{eig}} + 1, \dots, n$:

$$Q_{\text{spectral}} = \left(WAW^T + \sigma^2 I_n\right)^{-1}, \tag{13}$$

where W is an $n \times n_{\text{eig}}$ matrix whose columns are the n_{eig} first eigenvectors (i.e., the first n_{eig} columns of V), and $A = \mathrm{diag}(\lambda_1^{\text{shrink}} - \sigma^2, \dots, \lambda_{n_{\text{eig}}}^{\text{shrink}} - \sigma^2)$ is the diagonal matrix whose elements are the corresponding n_{eig} largest shrunken eigenvalues subtracted σ^2. The noise variance σ^2 is the average of the $n - n_{\text{eig}}$ smallest shrunken eigenvalues. By selecting $n_{\text{eig}} < n$ we keep only the parts of R that contain the most information, without imposing additional structure on the estimation problem.

[1] $R_{\text{shrink}} = V\Lambda_{\text{shrink}}V^T = V\left((1-\nu)\Lambda + \nu I_n\right)V^T = (1-\nu)V\Lambda V^T + \nu V I_n V^T = (1-\nu)R + \nu I_n$.

Note that even without shrinkage, the spectral estimate (13) corrects the smallest eigenvalues by replacing them with the average value σ^2 (Tipping & Bishop, 1999). However, without shrinkage the largest eigenvalues would remain biased, which is highly undesirable given that the spectral estimate emphasizes these.

Based on the filtered precision matrix (13), we propose a new estimator, which we refer to as *spectral* scaling:

$$\Sigma^{-1}_{\text{spectral}} = D^{-1/2}_{\text{hvar}} Q_{\text{spectral}} D^{-1/2}_{\text{hvar}}. \tag{14}$$

Here, D_{hvar} is a diagonal matrix whose elements are the variances of the base forecast errors. Note that we could introduce a structural, series, and hierarchy version of spectral scaling by substituting D_{struc} or D_{svar} for D_{hvar}, as defined by Athanasopoulos et al. (2017); but we find that using the hierarchy variances always leads to the best result.

Selecting the number of eigenvectors n_{eig} to use for the compression is somewhat equivalent to selecting the regularization parameter ν for GLASSO and shrinkage scaling. When $n_{\text{eig}} = n$, it is equivalent to scaling by the shrinkage estimate of the sample cross-covariance matrix. We use the value of the shrinkage intensity parameter given by (11).

Dimensionality reduction. The number of parameters in a correlation matrix of dimension n is $\frac{n(n-1)}{2}$. In the spectral domain, the parameters to be estimated are the n_{eig} eigenvectors that each have n elements and the n_{eig} eigenvalues. Orthogonality imposes $n_{\text{eig}}(n_{\text{eig}}+1)/2$ constraints, namely n_{eig} normalization constraints on the eigenvectors and $n_{\text{eig}}(n_{\text{eig}} - 1)/2$ orthogonality constraints, one for each pair of eigenvectors. That leaves $n_{\text{eig}}n - n_{\text{eig}}(n_{\text{eig}} - 1)/2$ parameters. Thus, when $n_{\text{eig}} \ll n$ the number of parameters is approximately reduced by a factor of $\frac{n(n-1)}{2n_{\text{eig}}n - n_{\text{eig}}(n_{\text{eig}}-1)} \approx \frac{n}{2n_{\text{eig}}}$. For example, retaining $n_{\text{eig}} = 15$ eigenvectors reduces the number of parameters by a factor of two, whereas keeping $n_{\text{eig}} = 5$ eigenvectors leads to a reduction in the number of parameters by a factor of six when $n = 60$. In summary, through eigendecomposition it is possible to greatly reduce the number of parameters that need to be estimated.

Complexity. Finding the n_{eig} largest eigenvalues and eigenvectors using an iterative procedure has complexity $O(n_{\text{eig}}n^2)$. The inverse (13) can be computed efficiently using the matrix inversion lemma. Thereby we only need to invert an $n_{\text{eig}} \times n_{\text{eig}}$ matrix rather than a full $n \times n$ matrix with complexity $O(n^3)$. The speedup can be substantial when n is much larger than n_{eig}. The GLASSO estimator also estimates the precision matrix directly; but solving the GLASSO optimization problem with complexity $O(n^3)$ per iteration is significantly more demanding computationally, as the number of iterations needed to converge to an accurate solution can be very large. Hence, the spectral estimator is at a computational advantage compared to both the GLASSO and shrinkage estimators. Moreover, we note that the eigendecomposition can be updated efficiently in an incremental fashion, which makes the spectral estimator advantageous for online applications (Zhao et al., 2006).

Comparison. Similar to the GLASSO and shrinkage estimators, the spectral estimator allows for *information sharing* between aggregation levels by accounting for correlations between forecast errors from different aggregation levels. Furthermore, it overcomes the problem of inverting a potentially singular estimate of the cross-covariance matrix while being robust to noise and high dimensionality, due to the inherent dimensionality reduction. The potential for dimensionality reduction is the most significant advantage of the spectral estimator, because this makes it feasible to consider hierarchies of much larger dimension. Whereas the GLASSO estimator tries to achieve sparsity in the partial cross-correlation structure, the goal of spectral decomposition is to achieve dimensionality reduction. Compared to shrinkage scaling, the spectral estimator has the advantage that by estimating the inverse directly it avoids inverting the potentially large cross-correlation matrix. Where the shrinkage estimator (10) shrinks all cross-correlations uniformly toward zero, the spectral estimator (13) attempts to separate information from noise.

Number of eigenvectors. The main disadvantage of spectral scaling is the need to select a number of eigenvectors. Several criteria have been proposed in the literature for selecting the number of eigenvectors based on information theoretic criteria (Wax & Kailath, 1985), Bayesian model selection (Minka, 2001), hypothesis testing, and random matrix theory (Bun et al., 2017). When the number of observations available is less than what is required to estimate the full cross-covariance matrix, it might be desirable to simply choose the maximum number of eigenvectors that can be estimated to extract as much information as possible. In Sections 6 and 5.5 we report the results from applying spectral scaling with different numbers of eigenvectors to gain an understanding of its impact on the reconciliation result.

5. Load forecasting

Automated short-term load forecasting is needed to support transactions in electricity markets and for efficient operation of power systems (Fan & Hyndman, 2012; Hahn et al., 2009). For day-ahead prediction, a model based on weather forecasts is typically used. Weather forecasts are less important for shorter horizons, as weather variables tend to change relatively smoothly over short intervals of time. Moreover, weather forecasts are sometimes not available or only available with a delay. This prompts consideration of modeling approaches that use only historical load data (Nystrup, Lindström et al., 2020; Taylor, 2010, 2012).

In short-term load forecasting, seasonal Holt–Winters exponential smoothing is a common choice for modeling seasonality (Gould et al., 2008; Livera et al., 2011; Taylor, 2003, 2010, 2012). Holt–Winters exponential smoothing was extended by Taylor (2003) to accommodate intraday and intraweek cycles in intraday data. It is well suited for electricity-demand forecasting, as demand has both daily and weekly seasonalities. We do not consider more complicated, state-of-the-art load forecasting models (Hong et al., 2019), because the focus of this article is on the new

Fig. 2. Map showing the Nord Pool region. Norway currently has five electricity price areas. Sweden is divided into four price areas. Denmark is divided into western and eastern Denmark. Finland, Estonia, Lithuania, and Latvia are undivided.

method for dimensionality reduction when reconciling forecasts. In Section 8, we discuss the importance of the accuracy of the base forecasts.

Letting p_1 and p_2 denote the periods of the two seasons, the additive double-seasonal exponential smoothing method from Taylor (2012) can be written on state-space from:

$$y_t = l_{t-1} + s^{(1)}_{t-p_1} + s^{(2)}_{t-p_2} + \phi e_{t-1} + \varepsilon_t, \tag{15a}$$

$$e_t = y_t - \left(l_{t-1} + s^{(1)}_{t-p_1} + s^{(2)}_{t-p_2}\right), \tag{15b}$$

$$l_t = l_{t-1} + \alpha e_t, \tag{15c}$$

$$s^{(1)}_t = s^{(1)}_{t-p_1} + \gamma_1 e_t, \tag{15d}$$

$$s^{(2)}_t = s^{(2)}_{t-p_2} + \gamma_2 e_t, \tag{15e}$$

where $\varepsilon_t \sim N\left(0, \sigma^2\right)$ and σ^2 is a constant variance; y_t is the load; l_t and $s^{(1)}_t$ are the state variables for the level and intraday cycle, respectively; $s^{(2)}_t$ is the state variable for the intraweek cycle remaining after $s^{(1)}_t$ is removed; α, γ_1, and γ_2 are smoothing parameters; and the term involving ϕ is an autoregressive adjustment for first-order residual autocorrelation, which Taylor (2010) found to greatly improve forecast accuracy. The method assumes the same intraday cycle for all days of the week and does not include a trend component, as short-term electricity demand mostly does not have a trend.

5.1. Data

We consider hourly load data for 2018 and 2019 from Nord Pool.[2] Nord Pool Spot is the power market for Norway (NO), Sweden (SE), Finland (FI), Denmark (DK), Estonia (EE), Latvia (LV), and Lithuania (LT). The day-ahead market is an auction, where power is traded for delivery each hour of the next day. The Nord Pool markets are divided into several bidding areas, as shown in Fig. 2. We consider the load in GWh in each of the seven countries in the Nord Pool area. We use data from 2018 for in-sample training, with the first two weeks used for initialization, and data from 2019 for out-of-sample testing.

Fig. 3 shows the autocorrelation functions (ACFs) for three different aggregation levels of the load data for Norway, Sweden, and Finland in 2018. The load data is strongly autocorrelated across all aggregation levels. In Fig. 3(a), the ACFs for the daily load reveal a clear weekly pattern, which is much stronger in Sweden than in Norway. In Fig. 3(b), the ACFs for the six-hourly load display both daily and weekly variation. Once again, the magnitude of the seasonal variation differs by country, being significantly more pronounced in Sweden than in Norway. The daily and weekly cycles are also evident from Fig. 3(c), although the weekly cycle is less pronounced at the hourly level.

5.2. Base forecasts

We model the entire temporal hierarchy from the daily level at the top to the hourly level at the bottom. With aggregation levels $k = 24, 12, 8, 6, 4, 3, 2, 1$, the total dimension of the hierarchy is 60.

Base forecasts at the daily level are generated based on exponential smoothing with a weekly seasonality using the automatic forecasting procedure implemented by Hyndman and Khandakar (2008). If multiple years of data were available for estimation, the models could be extended to accommodate yearly seasonality. Forecasts at all other levels are generated using the additive double-seasonal exponential smoothing method from Taylor (2012). All model parameters are fitted by minimizing the mean squared error of the in-sample one-step-ahead forecasts using the data from 2018.

We focus on squared errors to be consistent with the objective function in (4). Table 1 shows the root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{m/k}\sum_{h=1}^{m/k}\left(y_{t,h} - \hat{y}_{t,h}\right)^2} \tag{16}$$

and the root mean squared percentage error

$$\text{RMSPE} = 100\sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{m/k}\sum_{h=1}^{m/k}\left(\frac{y_{t,h} - \hat{y}_{t,h}}{y_{t,h}}\right)^2}, \tag{17}$$

where m/k is the sampling frequency per day, for base forecasts of one-day-ahead power consumption in each country at the different aggregation levels.

At the daily level, this corresponds to one-step-ahead forecasts, whereas at the hourly level, it corresponds to 24-steps-ahead forecasts, once per day at the end of each day. For example, at the hourly level the RMSE is the square root of the average mean squared error across the 24 steps each day across all days in the sample.

The RMSE is much larger for Norway and Sweden compared to the other countries, because consumption is larger. The RMSPE is better suited for comparing the

[2] https://www.nordpoolgroup.com/historical-market-data/.

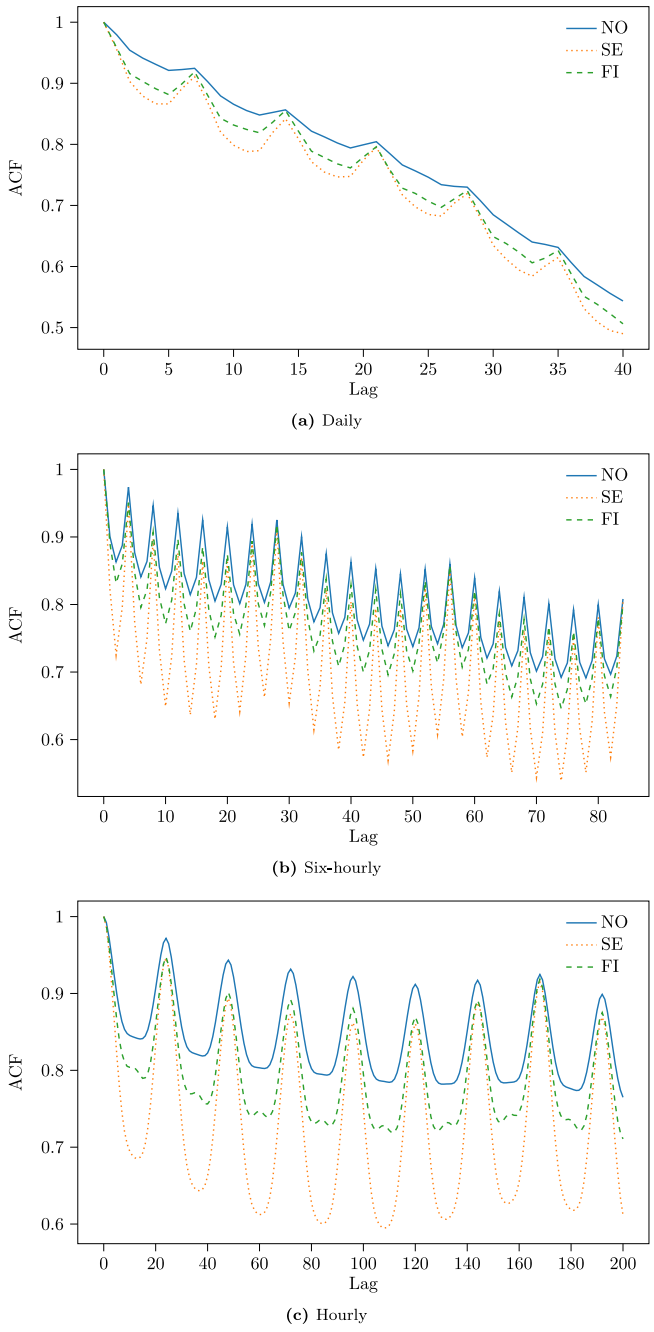


Fig. 3. Autocorrelation functions for three aggregation levels of load data for Norway, Sweden, and Finland in 2018.

forecast accuracy for the different countries, since it has the advantage of being scale-independent. The RMSPE is easier to interpret than scaled errors, but can only be applied to time series that are strictly positive. The RMSPE is higher in Sweden and Denmark compared with the other countries at most levels.

The load forecasts for Norway and Finland are the more accurate, relatively speaking, compared with the other countries. This is true both in and out of sample. The RMSPE for Norway and Finland is relatively similar across the different aggregation levels, whereas it increases significantly, for example in Latvia, when going from the daily to the hourly level. We emphasize that this does not mean that the one-day-ahead forecasts based on hourly data are worse than those based on daily data, as evident from the tables below. The RMSPE generally increases when moving down the hierarchy from the daily level toward the hourly level, with a few exceptions.

Table 1
In- and out-of-sample RMSE and RMSPE for base forecasts of one-day-ahead power consumption for the countries in the Nord Pool area at different aggregation levels.

	In sample (2018)							Out of sample (2019)						
	NO	SE	FI	DK	EE	LV	LT	NO	SE	FI	DK	EE	LV	LT
Daily														
RMSE	10.1	13.0	5.84	2.88	0.69	0.56	0.98	10.7	13.9	7.18	2.94	0.82	0.67	1.08
RMSPE	2.56	3.27	2.37	3.26	2.97	2.85	3.08	2.84	3.62	2.99	3.34	3.73	3.66	3.41
Twelve-hourly														
RMSE	5.48	7.17	2.98	1.60	0.37	0.32	0.52	5.28	6.88	3.42	1.46	0.41	0.33	0.53
RMSPE	2.72	3.53	2.40	3.54	3.15	3.19	3.25	2.81	3.59	2.83	3.29	3.66	3.60	3.37
Eight-hourly														
RMSE	3.52	4.77	1.96	1.14	0.26	0.22	0.37	3.44	4.59	2.25	1.04	0.28	0.25	0.38
RMSPE	2.63	3.56	2.38	3.70	3.23	3.30	3.35	2.75	3.57	2.79	3.43	3.76	3.84	3.48
Six-hourly														
RMSE	2.61	3.59	1.47	0.89	0.20	0.18	0.29	2.58	3.53	1.69	0.83	0.22	0.20	0.30
RMSPE	2.61	3.59	2.38	3.81	3.37	3.44	3.49	2.73	3.65	2.80	3.59	3.87	4.05	3.66
Four-hourly														
RMSE	1.71	2.40	0.99	0.61	0.14	0.12	0.20	1.66	2.38	1.13	0.56	0.15	0.13	0.21
RMSPE	2.57	3.63	2.42	3.92	3.53	3.58	3.65	2.66	3.75	2.83	3.64	3.95	4.21	3.80
Three-hourly														
RMSE	1.31	1.82	0.75	0.46	0.11	0.09	0.16	1.27	1.85	0.85	0.43	0.11	0.11	0.16
RMSPE	2.62	3.68	2.45	3.98	3.64	3.70	3.77	2.71	3.88	2.85	3.74	4.06	4.49	3.96
Two-hourly														
RMSE	0.88	1.23	0.51	0.31	0.08	0.06	0.11	0.84	1.22	0.57	0.29	0.08	0.07	0.11
RMSPE	2.68	3.72	2.51	4.04	3.80	3.79	3.87	2.72	3.85	2.85	3.81	4.11	4.73	4.03
Hourly														
RMSE	0.46	0.62	0.28	0.16	0.04	0.04	0.05	0.42	0.60	0.28	0.15	0.04	0.04	0.05
RMSPE	2.81	3.76	2.72	4.16	3.84	4.20	3.90	2.75	3.76	2.84	3.84	4.17	5.41	3.98

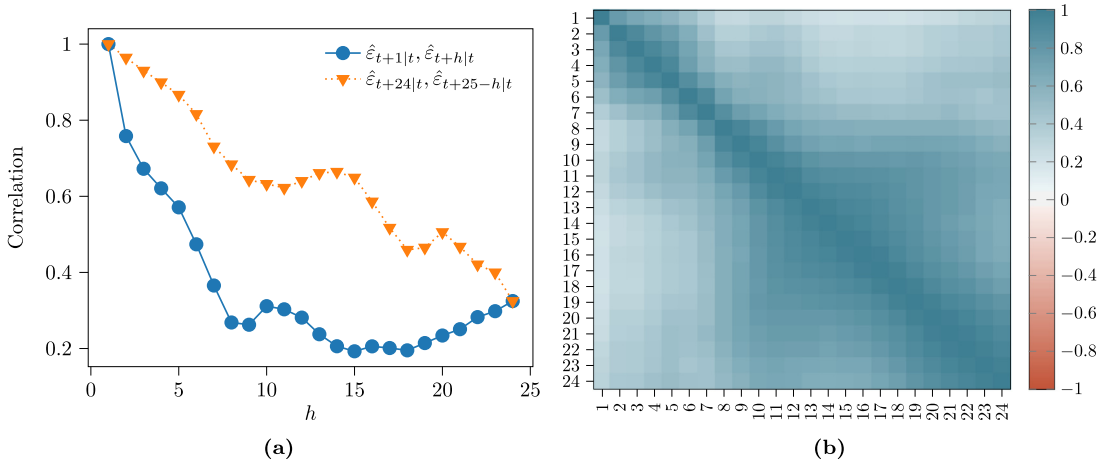
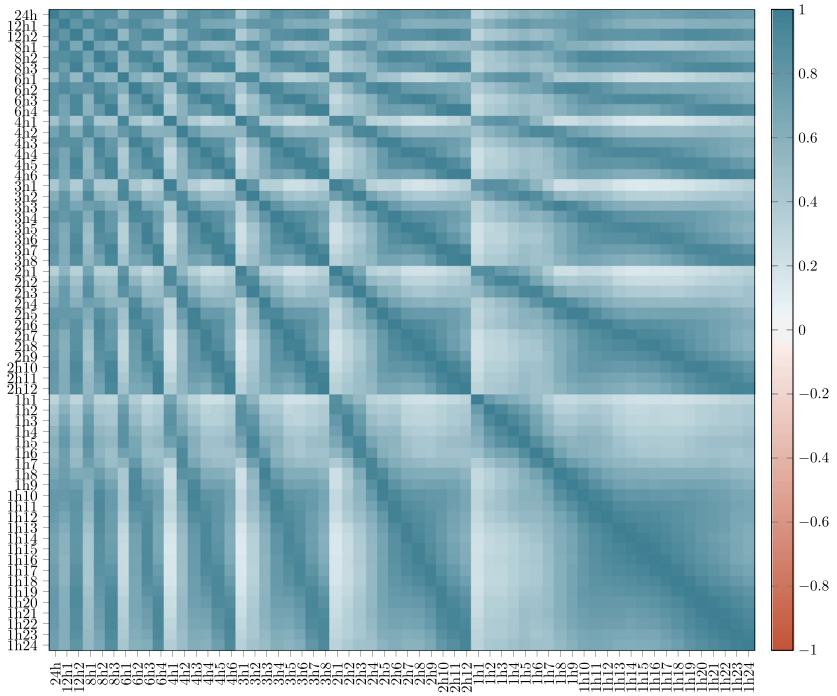


Fig. 4. Correlations between hourly base forecast errors for Norway in 2018. (a) Correlation between the one-step-ahead and h-steps-ahead and between the 24-steps-ahead and $(25-h)$-steps-ahead hourly base forecast errors. (b) Correlation heat map for hourly base forecast errors.

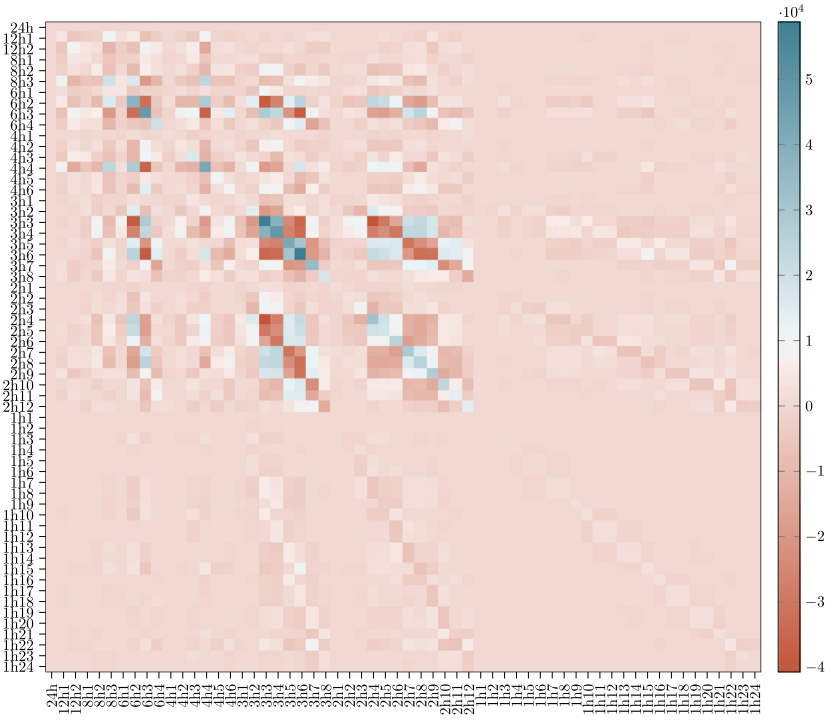
5.3. Residual autocorrelation

Fig. 4 shows the correlations between the hourly base forecast errors for Norway in 2018. When forecasting 24 steps ahead using double-seasonal exponential smoothing (15), the forecast errors are strongly correlated. The correlation between forecast errors is stronger the further out in time that the forecasts are made; for example, the correlation between the errors for the 23rd step and the 24th step are stronger than the correlation between the errors for the one-step-ahead and the two-steps-ahead forecasts, as evidenced by Fig. 4(a).

It is important to emphasize that the strong correlations in Fig. 4 are not a sign that the model is misspecified. On the contrary, when simulating data from the double-seasonal exponential smoothing model (15) using the parameters that were estimated for Norway and forecasting multiple steps ahead using the same parameter values, the correlations are even stronger and more persistent as a function of the number of steps ahead that the forecasts are made. It is indeed well known that, even for a correctly specified autoregressive model, multi-step forecast errors are autocorrelated. In other words, autocorrelation in temporal hierarchies arises even when forecasts are made using the *true* model.



(a) Cross-correlation matrix

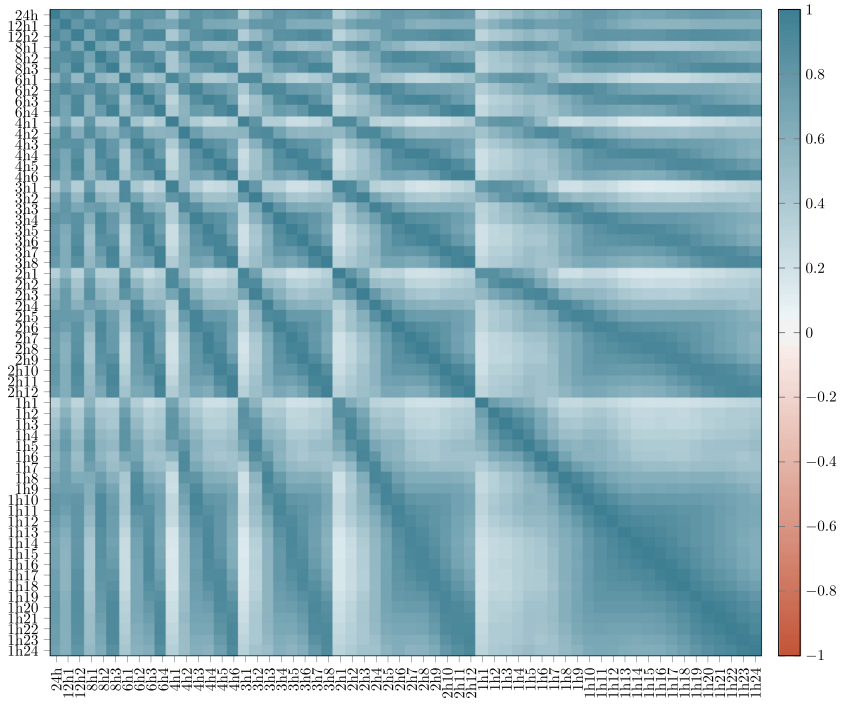


(b) Precision matrix

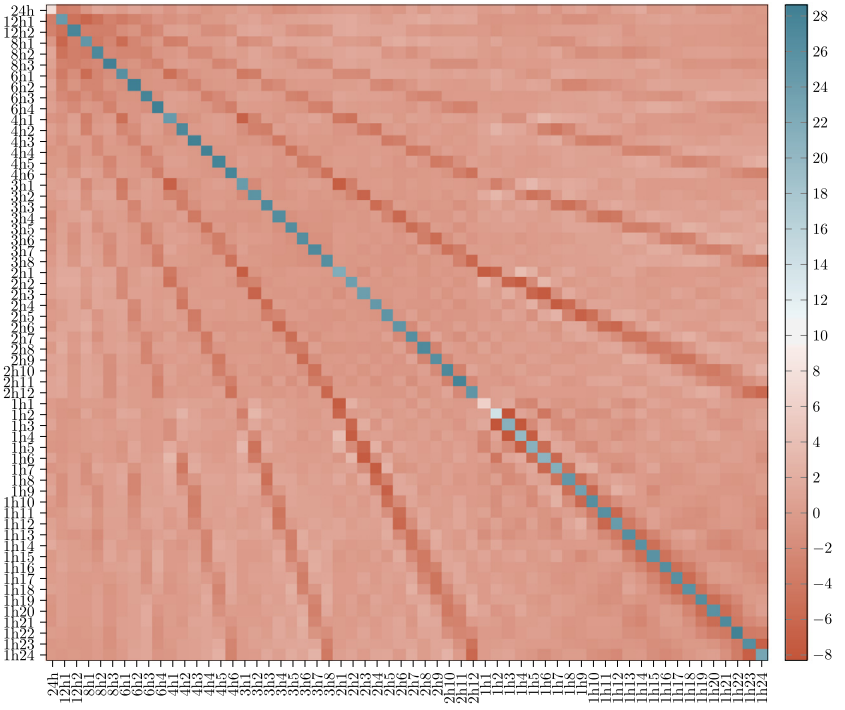
Fig. 5. Cross-correlation and precision matrix of the in-sample base forecast errors for Norway.

5.4. In-sample cross-correlation and precision matrix

Fig. 5 shows the cross-correlation and precision matrix of the in-sample base forecast errors for Norway. It is evident that there is a structure in the cross-correlation matrix with the strongest correlations occurring between base forecasts that are directly linked through the summation matrix. For example, the errors of the first three



(a) Reconstructed cross-correlation matrix



(b) Reconstructed precision matrix

Fig. 6. Reconstructed cross-correlation and precision matrix of the in-sample base forecast errors for Norway based on the first 15 eigenvectors.

hourly base forecasts are strongly correlated with the error of the first three-hour base forecast. The same plot for the other countries looks fairly similar.

Despite the clear structure in the cross-correlation matrix, there is limited structure in the estimated precision matrix. Intuitively, the precision matrix should

display the same "traces" as the cross-correlation matrix, clearly showing which errors are directly linked through the summation matrix. Presumably, conditional correlations for base forecast errors that are not directly linked through the structure of the hierarchy should be close to zero, which would make the structure even more evident. The missing structure suggests that the precision matrix is poorly estimated, despite being based on a full year of data.

5.4.1. Reconstruction

Fig. 6 shows the reconstructed cross-correlation and precision matrix of the in-sample base forecast errors for Norway based on the first 15 eigenvectors. The cross-correlation matrix shows the same structure as in Fig. 5, although the lines are a bit more blurry. The blurriness is the result of the reconstruction error from using only a subset of the eigenvectors and could easily be reduced by including more eigenvectors.

The reconstructed precision matrix, on the other hand, has a much clearer structure compared with Fig. 5. From the reconstructed precision matrix in Fig. 6 it is easy to identify which base forecasts errors are directly linked through the structure of the hierarchy. Moreover, it is evident that the conditional correlations are strongest between base forecasts from adjacent levels of the hierarchy. The conditional correlations get progressively weaker as the distance to the main diagonal increases. Similar to the spectral estimator, the shrinkage and GLASSO estimators also enhance the structure of the estimated precision matrix.

5.4.2. Variance explained

Fig. 7 shows the cumulative percentage of variance in the cross-correlation structure of the in-sample base forecast errors for Norway that can be captured as a function of the number of eigenvectors retained. As noted by Ledoit and Wolf (2015), since the estimates of the largest eigenvalues are biased upward, the variance that can be explained by the corresponding eigenvectors is also overestimated. The solution is to consider the shrunken eigenvalues instead of the sample eigenvalues.

Using just the first five eigenvectors, 91% of the total variance in the cross-correlation structure can be explained, according to the shrunken eigenvalues. With 15 eigenvectors, 97% of the variance can be explained, and with 30 eigenvectors, 99% of the variance can be explained. Evidently it is possible to obtain significant dimensionality reduction using spectral decomposition while maintaining high explanatory power. In the following sections we report the results from applying spectral scaling with different numbers of eigenvectors to gain an understanding of its impact on the reconciliation result.

5.5. Empirical results

Similar to Athanasopoulos et al. (2017) and Nystrup, Lindström et al. (2020), we follow the recommendation of Hyndman and Koehler (2006) by considering the percentage relative improvement in average loss (PRIAL), defined as

$$\text{PRIAL} = 100 \times \left(1 - \frac{\text{RMSE}}{\text{RMSE}^{\text{base}}}\right), \tag{18}$$

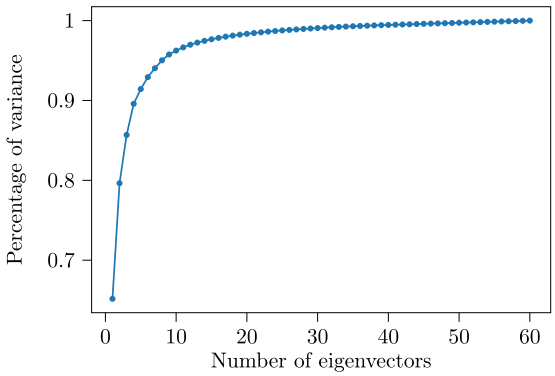


Fig. 7. Cumulative percentage of variance explained as a function of the number of eigenvectors.

when comparing the accuracy of reconciled and base forecasts. A positive entry shows a percentage decrease in RMSE relative to the base forecast, i.e., improved accuracy. By definition, the PRIAL of base forecasts is zero.

Table 2 summarizes the PRIAL of reconciled over base forecasts for each country in and out of sample. The in- and out-of-sample results are very similar, with the exception of scaling by the sample cross-covariance matrix. The largest improvements in accuracy are achieved for Norway and the three Baltic countries.

Table 3 summarizes the out-of-sample PRIAL of reconciled over base forecasts for each aggregation level and across aggregation levels. The daily forecasts improve the most, while the smallest improvements occur at the hourly level.

Bottom up. The bottom-up approach, where forecasts made at the hourly level are aggregated to all other levels, in general, is not very successful, though it does lead to large improvements of the daily forecasts out of sample.

Identity scaling. The OLS approach, where all levels are given the same weight to increase the importance of forecasting the aggregate, on average, does not improve the base forecasts out of sample.

Structural scaling. Structural scaling, where the variance is assumed to be proportional to the number of forecast errors contributing to each aggregation level, leads to slightly better results than identity scaling.

Variance scaling. Scaling by the pooled multi-step-ahead variance estimated at each aggregation level of the series in most cases yields slightly better results than structural scaling. The results for hierarchy variance scaling—i.e., scaling by the non-pooled variances at each aggregation level of the hierarchy—are almost identical.

Markov and autocovariance scaling. The Markov and autocovariance scaling estimators proposed by Nystrup, Lindström et al. (2020) both ignore correlations between aggregation levels. The results when using Markov or autocovariance scaling are similar to when using structural and variance scaling but with less variation from in sample to out of sample, suggesting that there is little benefit to accounting explicitly for autocorrelation.

Table 2
In- and out-of-sample PRIAL between reconciled and base forecasts of one-day-ahead power consumption in each country across the aggregation levels.

	In sample (2018)							Out of sample (2019)						
	NO	SE	FI	DK	EE	LV	LT	NO	SE	FI	DK	EE	LV	LT
Bottom up	3	3	−3	−2	−3	−11	−3	10	7	8	3	3	−16	2
Identity	−1	1	1	3	5	5	5	−5	−1	−3	0	0	4	2
Structural	3	3	5	2	4	4	3	3	3	3	2	4	3	3
Series variance	5	5	5	1	3	2	1	8	5	7	3	5	−1	3
Hierarchy variance	5	5	5	1	3	3	2	8	6	7	3	5	−1	3
Markov	4	4	5	2	5	5	4	5	4	4	2	4	5	4
Autocovariance	4	4	5	2	5	5	5	5	4	4	2	4	5	4
GLASSO	15	14	6	8	13	17	14	11	6	5	2	12	13	17
Shrinkage	18	15	12	8	16	17	16	16	8	8	3	11	14	15
Spectral 5	−3	7	7	−1	4	3	7	−5	−2	4	−9	2	5	5
Spectral 15	16	15	12	8	16	18	16	13	7	8	3	11	14	16
Spectral 30	18	16	12	8	16	17	16	16	8	8	3	11	14	15
Cross-covariance	26	23	20	17	24	24	25	12	4	10	5	11	15	16

GLASSO scaling. We implement the hierarchy GLASSO estimator proposed by Nystrup, Lindström et al. (2020) by doing a grid search over the values $\nu = 0,\ 10^{-5},\ 10^{-4},\ \ldots,\ 0.01,\ 0.02$ of the regularization parameter that controls the degree of sparsity, and comparing the average in-sample value of the Bayesian information criterion (Yuan & Lin, 2007). We find that the optimal value is $\nu = 0.01$. GLASSO scaling, on average, yields a 9% improvement in out-of-sample RMSE, which is a substantial improvement over the simpler approaches.

Shrinkage scaling. We use the value of the regularization parameter given by (11) rather than the fixed value $\nu = 0.05$ that Nystrup, Lindström et al. (2020) used. The results of shrinkage scaling are slightly better than those using GLASSO scaling. The improvement that can be obtained using either GLASSO or shrinkage scaling compared with the simpler estimators shows the value of capturing dependencies between forecast errors from different aggregation levels.

Spectral scaling. The results for spectral scaling are shown for different numbers of eigenvectors to gain an understanding of its impact on the reconciliation result. With five eigenvectors, the result is worse than GLASSO and shrinkage scaling. With 15 or 30 eigenvectors, the results are similar to shrinkage scaling. With a whole year of hourly observations available for estimation, there is more information than can be extracted using just five eigenvectors. On the other hand, the improvement when going from 15 to 30 eigenvectors is small compared with the increase in dimension. With 15 eigenvectors and less than half of the parameters, it is possible to obtain the same, state-of-the-art accuracy out of sample using spectral scaling as using cross-covariance scaling.

Cross-covariance scaling. Scaling by the full cross-covariance matrix is the best performing approach in all countries in sample, but slightly worse than shrinkage and spectral scaling out of sample. The greatest improvements in RMSE out of sample are at the daily level, where they range from 11% to 31%. The PRIAL is 10% out of sample, with the greatest improvements occurring in the three Baltic countries. Cross-covariance scaling is only able to contend with the regularized estimators because of the large number of observations available. In the simulation study in Section 6, we compare the estimators when different sample lengths are available for estimation.

5.6. TSO forecasts

To study the impact of the temporal correlation structure on the reconciliation results, we repeat the empirical study, only this time we replace our hourly base forecasts with the hourly consumption prognosis for each country that is available from the Nord Pool website. The hourly forecasts for the next day are made by the different national transmission system operators (TSOs) at different times during the day. In some countries the forecasts are made already before noon the day before, while in other countries the forecasts are updated up until a few hours before midnight the day before.

Fig. 8 shows the cross-correlation matrix of the in-sample base forecast errors for Norway when the hourly forecasts are replaced by the consumption prognosis from the Nord Pool website. It is evident that the TSO forecasts are very different from our forecasts, with correlations between the base forecast errors at the hourly level and all other levels being significantly weaker compared with Fig. 5. The TSO forecasts are undoubtedly based on a different model and additional information, such as weather data.

Table 4 summarizes the PRIAL of reconciled over base forecasts for each country in and out of sample when the TSO forecasts are included. The accuracy gains through reconciliation are significantly larger compared with Table 2. Once again the in- and out-of-sample results are very similar, with the exception of scaling by the sample cross-covariance matrix. The largest improvements in accuracy are achieved for Denmark and the smallest for Estonia.

In all cases, the TSO forecasts are based on slightly older load information than our base forecasts. Therefore, the TSO forecasts do not improve upon our forecasts for all countries, as evidenced by the results of bottom-up reconciliation. In fact, in several of the countries, aggregation of the hourly TSO forecasts leads to a significantly higher RMSE compared with our base forecasts. This means that the significant accuracy improvements

Table 3
Average out-of-sample PRIAL of reconciled over base forecasts of one-day-ahead power consumption in the Nord Pool countries for each aggregation level. The last column shows the average across the aggregation levels.

	Out of sample (2019)								
	24	12	8	6	4	3	2	1	Average
Bottom up	12	4	1	1	−1	1	1	0	2
Identity	9	2	−1	−2	−3	−2	−2	−3	0
Structural	13	6	2	2	0	1	1	0	3
Series variance	14	7	3	3	1	3	2	1	4
Hierarchy variance	14	7	4	3	1	3	2	2	4
Markov	14	6	3	3	1	2	2	1	4
Autocovariance	14	6	3	3	1	2	2	1	4
GLASSO	19	12	9	8	6	8	7	6	9
Shrinkage	21	13	10	10	8	9	8	8	11
Spectral 5	10	2	−2	−1	−4	−2	−2	−3	0
Spectral 15	20	13	10	9	7	8	8	7	10
Spectral 30	21	13	10	10	8	9	9	8	11
Cross-covariance	21	13	10	9	7	9	8	7	10

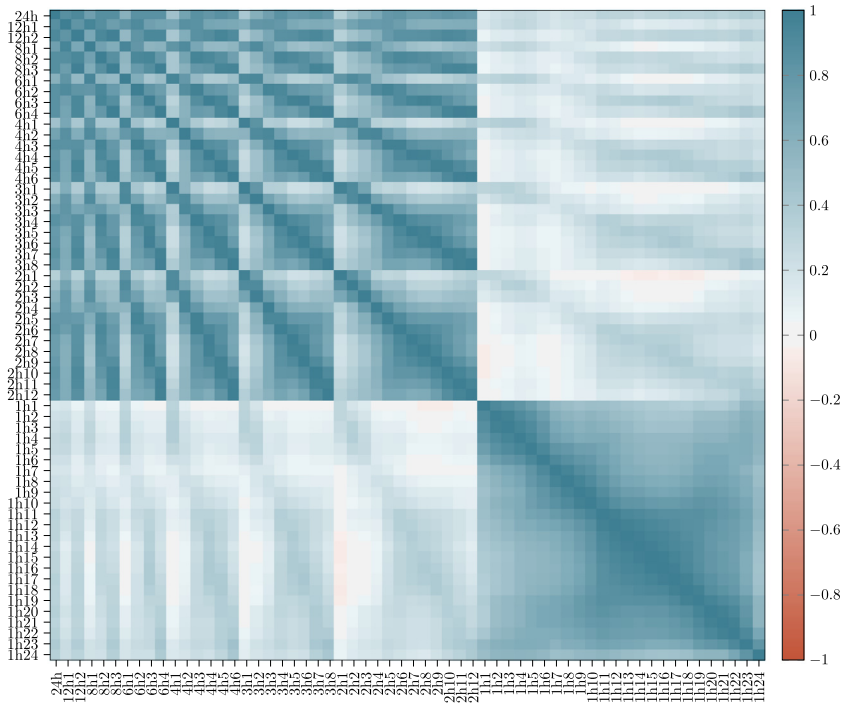


Fig. 8. Cross-correlation matrix of the in-sample base forecast errors for Norway when the hourly forecasts are replaced by the TSO consumption prognosis.

compared with Table 2 are not caused by the TSO forecasts being more accurate than our hourly base forecast. Rather, the improvements are driven by the weaker correlations in Fig. 8 combined with a greater need for reconciliation, because the forecasts are based on different information and different models.

Table 5 summarizes the out-of-sample PRIAL of reconciled over base forecasts for each aggregation level and across aggregation levels. The daily forecasts improve the most, while the smallest improvements are at the hourly level, but even the given TSO forecasts can be improved by up to 29%. Presumably, if the TSO forecasts had not been based on older load information than our base forecasts, they would have been harder to improve. In practical applications, however, information such as weather forecasts is typically only available with a two- to four-hour delay. It is also not uncommon that aggregate data is available sooner than disaggregate data.

With the additional information embedded in the temporal correlation structure when the TSO forecasts are included, the differences between the estimation approaches are bigger. The Markov and autocovariance estimators still do not improve the results compared to variance scaling. Meanwhile, the GLASSO, shrinkage, and spectral

Table 4
In- and out-of-sample PRIAL of reconciled over base forecasts of one-day-ahead power consumption in each country across the aggregation levels when TSO forecasts are included.

	In sample (2018)							Out of sample (2019)						
	NO	SE	FI	DK	EE	LV	LT	NO	SE	FI	DK	EE	LV	LT
Bottom up	−7	17	−32	66	−69	10	−149	3	29	−4	66	−108	23	−47
Identity	1	0	4	−17	9	5	12	−3	−5	−1	−19	7	1	8
Structural	10	9	13	−6	8	11	11	8	6	11	−4	7	9	14
Series variance	18	26	18	62	3	24	10	16	25	18	63	0	24	11
Hierarchy variance	19	27	19	63	4	32	11	17	26	20	64	1	31	11
Markov	13	16	20	59	9	29	13	11	13	15	60	10	26	10
Autocovariance	15	18	20	60	9	30	26	12	15	17	61	10	25	24
GLASSO	34	36	36	70	16	44	40	28	36	33	69	12	42	41
Shrinkage	34	37	37	70	16	43	34	29	36	34	69	13	41	35
Spectral 5	14	27	27	66	−6	34	28	4	27	24	66	−13	38	29
Spectral 15	33	36	37	69	16	43	34	27	35	33	69	13	41	35
Spectral 30	34	37	37	70	16	43	34	29	36	34	69	13	41	35
Cross-covariance	38	42	40	71	24	46	50	26	30	33	68	12	40	37

Table 5
Average out-of-sample PRIAL of reconciled over base forecasts of one-day-ahead power consumption in the Nord Pool countries for each aggregation level when TSO forecasts are included. The last column shows the average across the aggregation levels.

	Out of sample (2019)								
	24	12	8	6	4	3	2	1	Average
Bottom up	−1	−7	−8	−6	−8	−6	−6	0	−5
Identity	10	3	0	0	−2	0	0	−23	−2
Structural	18	11	8	8	7	8	8	−11	7
Series variance	32	26	22	22	20	21	21	15	23
Hierarchy variance	34	27	24	24	22	23	22	17	24
Markov	32	25	21	20	18	19	19	11	21
Autocovariance	34	27	23	23	21	22	22	14	23
GLASSO	48	41	38	37	35	36	35	29	37
Shrinkage	47	40	37	37	35	35	34	29	37
Spectral 5	36	29	26	25	23	23	23	19	25
Spectral 15	46	40	37	36	34	34	33	28	36
Spectral 30	47	40	37	37	35	35	34	29	37
Cross-covariance	45	39	36	35	33	34	33	27	35

estimators yield substantial improvements in forecast accuracy for all countries and across all aggregation levels.

6. Simulation study

To better understand the differences and advantages of the proposed spectral estimator compared to the estimators proposed by Athanasopoulos et al. (2017) and Nystrup, Lindström et al. (2020), we compare their performance in a simulation study.

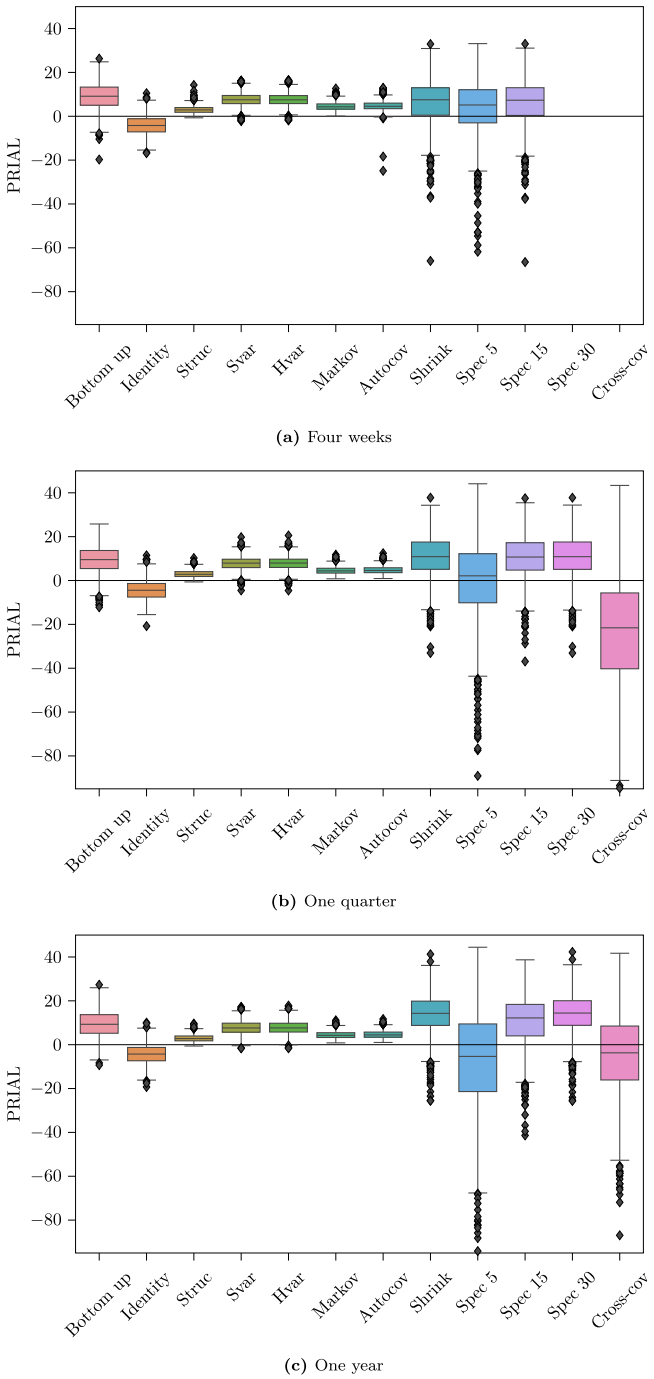
6.1. Setup

In the simulation study we focus on the load data and our base forecasts for Norway, since these are representative for most of the countries. First, we sample a number of days from the 2018 data and base forecasts with replacement, to be used for estimation. Subsequently, we sample 14 days without replacement from the 2019 data and base forecasts, to be used for testing. We compare the performance of the estimators when forecasting one day ahead for 14 days using the test data. We repeat this process 1000 times for each training sample length.

6.2. Results

Box plots of the PRIAL of reconciled over base forecasts across aggregation levels and across the two-week test period are shown in Fig. 9. Results are not shown for the GLASSO estimator, since this is considerably more time consuming to compute. The spectral methods improve both the median level and variance compared with scaling by the empirical cross-covariance matrix. Meanwhile, across training sample lengths, the variation in PRIAL of the methods that estimate a full cross-covariance matrix is much larger than for the methods that impose a simpler structure on the cross-covariance matrix. Future work should explore this tradeoff between median improvement and variability in greater detail.

Four weeks. Four weeks of training data are not enough to estimate the full cross-covariance matrix nor the spectral estimate based on 30 eigenvectors. With only four weeks of observations, the bottom-up approach leads to the largest median improvements in RMSE, closely followed by the diagonal variance and shrinkage estimators. The results when using the diagonal variance or Markov estimators are clearly less dispersed than the results using the more complex estimators. This is true across the training sample lengths. The results when using the spectral



(a) Four weeks

(b) One quarter

(c) One year

Fig. 9. Box plots of PRIAL of reconciled over base forecasts across aggregation levels and across the two-week test period based on 1000 Monte Carlo simulations for each training sample length.

estimator based on 15 eigenvectors are similar to those using shrinkage scaling, and better than those based on five eigenvectors. This means that when only four weeks of data are available for estimation, the dimension of the matrix can be reduced significantly by only retaining the first 15 eigenvectors, without affecting the results.

One quarter. When 13 weeks of training data are available for estimation, the largest median improvement is obtained using the spectral estimator based on 30 eigenvectors, closely followed by that based on 15 eigenvectors, the shrinkage estimator, and the bottom-up approach. With this much data available, there is a big difference

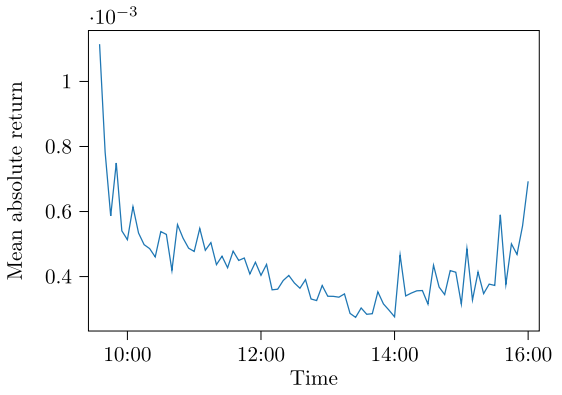


Fig. 10. Average intraday volatility periodicity for the S&P 500 index.

between the spectral estimators based on five and 15 eigenvectors. On the other hand, there is almost no difference between using 15 and 30 eigenvectors. That is, the dimension of the cross-covariance matrix can still be reduced significantly by only retaining the first 15 eigenvectors, without affecting the results. It is also worth noting that the spectral estimators yield much better results than scaling by the sample cross-covariance matrix.

One year. One year of training data is equivalent to the setting in Section 5.5. If the worst-case outcome is more important than the median improvement, then the Markov or autocovariance estimator is the best choice. With a full year of data available, the largest median improvements are obtained when using the spectral estimator based on 30 eigenvectors or shrinkage scaling. With this much training data, there is a difference between the spectral estimators based on different numbers of eigenvectors. Both the shrinkage estimator and the spectral estimators based on 15 and 30 eigenvectors yield much better results than scaling by the sample cross-covariance matrix. This is due to the adaptive nature of the shrinkage intensity parameter (11). Even when a full year of training data is available, the dimension of the cross-covariance matrix can be reduced significantly by only retaining the first 15 eigenvectors, without a big deterioration in accuracy.

7. Forecasting realized variance

In this section, we consider a data set consisting of five-minute returns for the S&P 500 stock index from February 19 through July 16, 2019, comprising 7956 observations. It is widely documented that volatility in financial markets varies systematically over the trading day and that this pattern is highly correlated with the intraday variation of trading volume and bid–ask spreads (Andersen & Bollerslev, 1997). Fig. 10 shows the mean absolute five-minute return of the S&P 500 index as a function of the time of the day. As seen from the figure, the average realized volatility at the market opening and close is three to four times the average volatility around lunch.

The open-to-close realized variance is given by the sum of squared intraday returns. If p_τ is the log price of the index at time τ, then

$$RV_t = \frac{1}{n_t}\sum_{\tau=1}^{n_t}(p_\tau - p_{\tau-1})^2\,, \tag{19}$$

where n_t is the number of observations within day t. It is common to measure realized variance at multiple time scales to mitigate microstructure noise (Zhang, 2006).

Reconciling forecasts of realized variance at different frequencies using a temporal hierarchy is an alternative to combining them in mixed data sampling (MIDAS) regressions to predict future volatility (Ghysels et al., 2006). We forecast realized intraday variance at aggregation levels of 5, 10, 15, 30, 65, 130, and 195 min using the standard additive Holt–Winters' seasonal method. The open-to-close realized variance over the entire 390-minute-long trading day is forecasted using a so-called high-frequency-based volatility (HEAVY) model (Shephard & Sheppard, 2010) specified as

$$\mathbb{E}\left[RV_{t+1}\right] = y_{t+1} = \omega + \alpha RV_t + \beta y_t, \quad \omega, \alpha, \beta \geq 0$$
$$\alpha + \beta < 1. \tag{20}$$

Table 6 summarizes the PRIAL of reconciled over base forecasts of realized intraday variance of the S&P 500 index for each aggregation level. Given the limited number of observations relative to the dimension of the hierarchy, the realized variance results are entirely in sample. With aggregation levels $k = 78, 39, 26, 13, 6, 3, 2, 1$, the total dimension of the temporal hierarchy being 168, we do not have nearly enough observations to estimate the full cross-covariance matrix. Financial volatility is known to be non-stationary, meaning that even if we had a longer time series of intraday returns, it is questionable whether this would be useful for estimating a constant cross-covariance matrix.

The largest improvements in forecast accuracy occur at the highest aggregation levels. The reconciled forecasts of the open-to-close realized variance over the entire 390-minute-long trading day are up to 4.5% more accurate than the base forecasts from the HEAVY model, which is a common benchmark. We also tried using a MIDAS regression to forecast the open-to-close realized variance, but the reconciled forecasts are more accurate. In addition to improved accuracy, neither MIDAS regression nor HEAVY models offer a way to reconcile forecasts across multiple aggregation levels. We note that if we had a way to produce more accurate base forecasts, we could simply include them in the hierarchy.

Although the improvements in forecast accuracy are smaller compared with the load forecasting applications, they are still positive for all aggregation levels. The results show that there is a benefit to considering both auto- and cross-correlation information when reconciling the realized variance forecasts. The largest PRIAL is obtained using the spectral estimator based on 15 eigenvectors. Estimating the inverse cross-correlation matrix based on 15 eigenvectors reduces the number of parameters that need to be estimated by a factor of almost six. The realized variance application illustrates the potential of the proposed dimensionality reduction approach in high-frequency settings.

Table 6
PRIAL of reconciled over base forecasts of realized intraday variance of the S&P 500 index for each aggregation level. The last column shows the average across the aggregation levels.

	390	195	130	65	30	15	10	5	Average
Bottom up	3.2	3.6	0.2	1.3	−0.1	−0.9	0.2	0	0.9
Identity	0.5	3.0	−1.5	0.3	−0.5	−0.3	0.1	−0.7	0.1
Structural	2.1	4.3	1.0	1.9	0.6	0.4	0.9	0.5	1.5
Series variance	3.3	4.1	1.3	2.0	0.5	0.0	0.8	0.5	1.6
Hierarchy variance	3.6	4.2	1.3	2.0	0.6	0.1	0.8	0.3	1.6
Markov	3.7	4.2	1.4	2.1	0.7	0.2	0.9	0.4	1.7
Autocovariance	4.0	4.4	1.5	2.2	0.8	0.6	1.2	1.0	2.0
Shrinkage	4.5	4.5	1.2	2.4	1.1	0.8	1.5	1.9	2.2
Spectral 5	4.3	4.4	1.3	2.2	0.8	0.5	1.0	1.0	2.0
Spectral 15	4.5	4.6	1.3	2.4	1.2	0.8	1.5	1.7	2.3

8. Discussion

The improvement in accuracy that can be achieved by considering auto- and cross-correlation when reconciling forecasts in a temporal hierarchy naturally depends on the data-generating process and the models used for forecasting. If all forecast errors are strongly positively correlated, then the potential for accuracy improvements through reconciliation is limited. On the other hand, as evidenced by the results in Section 5.6, weakly correlated forecast errors can lead to substantial accuracy gains. Even though the TSO forecasts were not more accurate, on average, than our base forecasts, their inclusion led to large accuracy improvements, because they were only weakly correlated with the other base forecasts. This showed the potentially very large accuracy gains that can be obtained when reconciling forecasts from different models that are based on different information.

The inclusion of the TSO forecasts clearly showed that it is not the accuracy of the base forecasts that determines the potential for accuracy improvements. If base forecasts are very accurate, then even a small improvement will be large in relative terms. The potential for accuracy improvements is determined by the error correlation structure and the degree of incoherence. Recall that the objective of the reconciliation problem (4) is to adjust the base forecasts by the least amount so that these become coherent. If the base forecasts are close to coherent, then the adjustments made will be minor, and hence the potential accuracy improvement will also be small.

For dimensionality reduction to be most effective, the correlations have to be strong. The stronger the correlations, the more the dimension of the hierarchy can be reduced without affecting the accuracy. As seen in the simulation study, the shorter the time series that is available for estimation, the more the dimension can be reduced without affecting accuracy. The advantage of dimensionality reduction through eigendecomposition compared with the other estimators is that no additional structure is imposed on the estimation problem. Rather, the dimension of the temporal correlation matrix is reduced by only keeping its most important structure. Reducing the dimension by omitting aggregation levels would require prior knowledge about which levels are most important and could potentially impact accuracy negatively. Adding aggregation levels comes at the cost of increasing the dimension of the estimation problem, which is why our proposed model-free dimensionality reduction approach, insofar as it is capable of handling hierarchies of very large dimension, is an important contribution.

9. Conclusion

We proposed a novel estimator for reconciling forecasts in a temporal hierarchy. The estimator is based on an eigendecomposition of the temporal correlation matrix. By decomposing the correlation matrix and estimating its inverse based on a small subset of its eigenvectors, our estimator extracts as much information as possible from the error structure given the data available to achieve state-of-the-art accuracy on hierarchies of all sizes.

We demonstrated the usefulness of the proposed estimator in a simulation study and through applications to short-term load and financial volatility forecasting. The simulation study showed that it is possible to significantly reduce the dimension of the temporal correlation matrix without loss of accuracy, while the best performing estimator depends on the number of observations available for estimation. Even in settings with limited data compared with the dimension of the temporal hierarchy, forecast accuracy could be improved by considering correlations.

The spectral estimator was always at least as good as the best estimators in terms of accuracy, while offering dimensionality reduction and computational advantages. By taking account of the correlation structure when reconciling forecasts, accuracy could be improved uniformly across all temporal aggregation levels. The most accurate forecasts were obtained when capturing dependencies in forecast errors between aggregation levels. The resulting coherent and significantly more accurate forecasts are likely to yield better multi-horizon decisions.

Even with hourly data, when forecasts are made more than one day ahead in time, the sample cross-correlation matrix quickly becomes infeasible to estimate. In applications where limited data are available—or where limited data are relevant due to stationarity concerns—compared with the dimension of the temporal hierarchy, the spectral estimator can greatly reduce the dimension of the estimation problem while yielding accurate reconciled forecasts. Dimensionality reduction through eigendecomposition should be particularly useful for future research considering temporal hierarchies with high-frequency

forecasts or spatio–temporal hierarchies where dimensionality is inevitably a challenge. The computational advantages of the spectral estimator could also be essential for online forecast reconciliation applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Vergstiftelsen and the Centre for IT-Intelligent Energy Systems (CITIES) project funded in part by Innovation Fund Denmark under Grant No. 1305-00027B. We thank an Associate Editor and two anonymous reviewers for helpful comments.

Appendix. Covariance shrinkage and generalized ridge regression

The solution to the forecast reconciliation problem with a penalty on the size of the deviations between reconciled and base forecasts

$$\begin{aligned} \text{minimize} \quad & (1-\nu)\left(\hat{y}-\tilde{y}\right)^T \Sigma^{-1}\left(\hat{y}-\tilde{y}\right) + \nu\left\|\hat{y}-\tilde{y}\right\|_2^2 \\ \text{subject to} \quad & \tilde{y} = SG\tilde{y}, \end{aligned} \tag{21}$$

where $\tilde{y} \in \mathbb{R}^n$ is the variable and the parameter $\Sigma \in \mathbb{R}^{n\times n}_{++}$ is the covariance matrix for the coherency errors, is given by

$$\tilde{y} = S\left(S^T \Sigma^{-1}_{\text{shrink}} S\right)^{-1} S^T \Sigma^{-1}_{\text{shrink}} \hat{y}. \tag{22}$$

Here, the shrinkage estimate of the inverse covariance matrix is

$$\Sigma^{-1}_{\text{shrink}} = (1-\nu)\,\Sigma^{-1} + \nu I_n. \tag{23}$$

This has the same limiting behavior as the linear shrinkage estimator of the covariance matrix (10), i.e., $\lim_{\nu\to 0} \Sigma_{\text{shrink}} = \Sigma$ and $\lim_{\nu\to 1} \Sigma_{\text{shrink}} = I_n$.

The penalized reconciliation problem (21) is equivalent to

$$\begin{aligned} \text{minimize} \quad & \left(\hat{y}-\tilde{y}\right)^T \Sigma^{-1}\left(\hat{y}-\tilde{y}\right) \\ \text{subject to} \quad & \tilde{y} = SG\tilde{y} \\ & \left\|\hat{y}-\tilde{y}\right\|_2^2 \le \tau, \end{aligned} \tag{24}$$

since there is a one-to-one correspondence between ν in (21) and τ in (24), as shown by Hastie et al. (2009, Chapter 3.4) in the case of ridge regression. In other words, constraining the size of the adjustments made to the bottom-level base forecasts is equivalent to linear shrinkage of the covariance estimate.

References

Amemiya, T., & Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, *67*(339), 628–632.

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*(2–3), 115–158.

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, *25*(1), 146–166.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844.

Boyd, S., Busseti, E., Diamond, S., Kahn, R. N., Koh, K., Nystrup, P., & Speth, J. (2017). Multi-period trading via convex optimization. *Foundations and Trends in Optimization*, *3*(1), 1–76.

Bun, J., Bouchaud, J.-P., & Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, *666*, 1–109.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Clements, A., Hurn, A., & Li, Z. (2016). Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research*, *251*(2), 522–530.

Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, *27*(1), 134–141.

Gamakumara, P., Panagiotelis, A., Athanasopoulos, G., & Hyndman, R. J. (2018). *Probabilisitic forecasts in hierarchical time series*: *Working Paper 11/18*, Monash University.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1–2), 59–95.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*(1), 207–222.

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, *9*(3), 233–254.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, *199*(3), 902–907.

Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, *23*(1), 1–13.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer.

Hong, T., Xie, J., & Black, J. (2019). Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, *35*(4), 1389–1399.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579–2589.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

Hyndman, R. J., Lee, A. J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, *97*, 16–32.

Jeon, J., Panagiotelis, A., & Petropoulos, F. (2019). Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*, *279*(2), 364–379.

Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, *75*, 393–409.

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, *30*(2), 291–302.

Kourentzes, N., Rostami-Tabar, B., & Barrow, D. K. (2017). Demand forecasting by temporal aggregation: using optimal or multiple aggregation levels?. *Journal of Business Research*, *78*, 1–9.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603–621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Ledoit, O., & Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, *139*, 360–384.

Livera, A. M. D., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.

Minka, T. P. (2001). Automatic choice of dimensionality for PCA. In *Advances in neural information processing systems* (pp. 598–604).

Nystrup, P., Kolm, P. N., & Lindström, E. (2020). Greedy online classification of persistent market states using realized intraday volatility features. *Journal of Financial Data Science*, *2*(3), 25–39.

Nystrup, P., Lindström, E., Pinson, P., & Madsen, H. (2020). Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, *280*(3), 876–888.

Nystrup, P., Madsen, H., & Lindström, E. (2017). Long memory of financial time series and hidden Markov models with time-varying parameters. *Journal of Forecasting*, *36*(8), 989–1002.

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, *37*(1), 343–359.

Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, *66*(6), 914–924.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics*, *60*(6), 479–498.

Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *25*(2), 197–231.

Silvestrini, A., & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, *22*(3), 458–497.

Spiliotis, E., Petropoulos, F., Kourentzes, N., & Assimakopoulos, V. (2020). Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy*, *261*, Article 114339.

Taieb, S. B. (2017). Sparse and smooth adjustments for coherent forecasts in temporal aggregation of time series. In O. Anava, A. Khaleghi, M. Cuturi, V. Kuznetsov, & A. Rakhlin (Eds.), *Proceedings of machine learning Research*: *Vol. 55*, *Proceedings of the time series workshop at NIPS 2016* (pp. 16–26).

Taieb, S. B., Taylor, J. W., & Hyndman, R. J. (2017). Coherent probabilistic forecasts for hierarchical time series. In D. Precup, & Y. W. Teh (Eds.), *Proceedings of machine learning research*: *Vol. 70*, *Proceedings of the 34th international conference on machine learning* (pp. 3348–3357).

Taieb, S. B., Taylor, J. W., & Hyndman, R. J. (2020). Hierarchical probabilistic forecasting of electricity demand with smart meter data. *Journal of the American Statistical Association*, 1–17. http://dx.doi.org/10.1080/01621459.2020.1736081.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*(8), 799–805.

Taylor, J. W. (2010). Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. *International Journal of Forecasting*, *26*(4), 627–646.

Taylor, J. W. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, *27*(1), 458–464.

Tiao, G. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, *59*(3), 525–531.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting, Vol. 1* (pp. 135–196). Elsevier: Amsterdam.

Tipping, M. E., & Bishop, C. M. (1999). Mixtures of probabilistic principal component analyzers. *Neural Computation*, *11*(2), 443–482.

Van Erven, T., & Cugliari, J. (2015). Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In A. Antoniadis, J.-M. Poggi, & X. Brossat (Eds.), *Lecture notes in statistics*: *Vol. 217*, *Modeling and stochastic learning for forecasting in high dimensions* (pp. 297–317). Cham: Springer.

Wax, M., & Kailath, T. (1985). Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *33*(2), 387–392.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, *114*(526), 804–819.

Wickramasuriya, S. L., Turlach, B. A., & Hyndman, R. J. (2020). Optimal non-negative forecast reconciliation. *Statistics and Computing*, *30*(5), 1167–1182.

Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*, *179*, 391–397.

Yang, D., Quan, H., Disfani, V. R., & Liu, L. (2017). Reconciling solar forecasts: Geographical hierarchy. *Solar Energy*, *146*, 276–286.

Yang, D., Quan, H., Disfani, V. R., & Rodríguez-Gallegos, C. D. (2017). Reconciling solar forecasts: Temporal hierarchy. *Solar Energy*, *158*, 332–346.

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, *12*(6), 1019–1043.

Zhang, Y., & Dong, J. (2018). Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation. *IEEE Transactions on Power Systems*, 1–1. http://dx.doi.org/10.1109/tpwrs.2018.2868175.

Zhao, H., Yuen, P. C., & Kwok, J. T. (2006). A novel incremental principal component analysis and its application for face recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *36*(4), 873–886.