Heat load forecasting using adaptive temporal hierarchies

Bergsteinsson, Hjörleifur G.; Møller, Jan Kloppenborg; Nystrup, Peter; Pálsson, Ólafur Pétur; Guericke, Daniela; Madsen, Henrik

Published in:
Applied Energy

Link to article, DOI:
10.1016/j.apenergy.2021.116872

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Bergsteinsson, H. G., Møller, J. K., Nystrup, P., Pálsson, Ó. P., Guericke, D., & Madsen, H. (2021). Heat load forecasting using adaptive temporal hierarchies. Applied Energy, 292, [116872]. https://doi.org/10.1016/j.apenergy.2021.116872

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

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

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Heat load forecasting using adaptive temporal hierarchies

Hjörleifur G. Bergsteinsson a, Jan Kloppenborg Møller a, Peter Nystrup a, b, Ólafur Pétur Pálsson c, Daniela Guericke c, Henrik Madsen a, d

a Technical University of Denmark, Denmark
b Lund University, Sweden
c University of Iceland, Iceland
d Norwegian University of Science and Technology, Norway

A B S T R A C T

Heat load forecasts are crucial for energy operators in order to optimize the energy production at district heating plants for the coming hours. Furthermore, forecasts of heat load are needed for optimized control of the district heating network since a lower temperature reduces the heat loss, but the required heat supply at the end-users puts a lower limit on the temperature level. Consequently, improving the accuracy of heat load forecasts leads to savings and reduced heat loss by enabling improved control of the network and an optimized production schedule at the plant. This paper proposes the use of temporal hierarchies to enhance the accuracy of heat load forecasts in district heating. Usually, forecasts are only made at the temporal aggregation level that is the most important for the system. However, forecasts for multiple aggregation levels can be reconciled and lead to more accurate forecasts at essentially all aggregation levels. Here it is important that the auto- and cross-covariance between forecast errors at the different aggregation levels are taken into account. This paper suggests a novel framework using temporal hierarchies and adaptive estimation to improve heat load forecast accuracy by optimally combining forecasts from multiple aggregation levels using a reconciliation process. The weights for the reconciliation are computed using an adaptively estimated covariance matrix with a full structure, enabling the process to share time-varying information both within and between aggregation levels. The case study shows that the proposed framework improves the heat load forecast accuracy by 15% compared to commercial state-of-the-art operational forecasts.

1. Introduction

Energy forecasting has become an essential method in the green transition due to the increased complexity of energy systems required to achieve high energy efficiency. This is highlighted in Hong et al. [1], who give an extensive historical overview of energy forecasting, including current trends. The authors point to the significant growth of research related to renewable energy forecasts during the past decade. Energy systems that rely on renewable energy sources, e.g., wind and solar, to produce either electricity or heat need an accurate prediction of the energy demand for the upcoming days to ensure that they can fulfill the demand and optimize their operation accordingly. Planning needs to take into account that wind and solar as energy sources cannot always provide the required energy demand due to their weather dependency. Therefore, they are integrated with other energy sources to ensure that the energy demand is met. However, wind and solar are an important part of the green transition. The share of renewable energy in the European Union is expected to increase to 70% by 2050 [2]. Therefore, accurate forecasts are becoming increasingly valuable for the energy sector.

Many different forecasting models have been proposed in recent years for the energy sector. From classical methods, e.g. Taylor [3] using double-seasonal Holt–Winters exponential smoothing to forecast electricity demand, to more complicated models, e.g. Nielsen and Madsen [4] using a grey-box model to forecast heat consumption. However, there is probably not one unique forecasting model that can provide the best forecast in all situations for every energy system. Therefore, frequent studies suggest using multiple models, and combining forecasts typically gives better overall forecasts than using only one model [1,5]. There has also been a significant increase in the number of studies using temporal hierarchies to improve forecast accuracy by utilizing the information between different aggregation levels and by optimally combining them with reconciliation [6,7]. Using temporal hierarchies could be a beneficial approach to improve the accuracy of energy forecasts because they use both short-term and long-term forecasts to
improve forecast accuracy and consequently give coherent forecasts on all aggregation levels. For example, district heating plants need short-term forecasts for operational optimization, such as supply temperature control, while long-term forecasts are used for planning, e.g. biomass supply planning. The district heating sector can gain heavily from the temporal hierarchies, as they need accurate and coherent short-term and long-term forecasts for control and planning.

1.1. District heating

District heating is an efficient way to provide heat to buildings in densely populated areas. Due to its flexibility, district heating has become a crucial part of the agenda to reach a renewable, non-fossil heat supply in the future. Additionally, district heating increases the flexibility of the overall integrated energy system by storing energy. For example, excess wind power during low-electricity-demand periods can be used to heat water with heat pumps or electric boilers that can be stored to reduce pressure on the heating system during peak hours [8]. The change from traditional fossil fuels to renewable district heating embedded in a smart energy system is referred to as 4th generation district heating [9]. To pave the way, district heating systems are currently undergoing a digital transformation. Along with this transformation, the district heating systems are continuously improving the efficiency of their operations. An important prerequisite for operating a district heating system efficiently is the accuracy of heat load forecasts. Being able to look accurately into the future enables the operators to run the system with more precision, resulting in a more efficient operation, hence lowering the system costs.

Lowering the supply temperature in the district heating network using control methods has substantial potential for cost savings, since lower supply temperatures lead to lower heat production costs as well as reduced heat losses in the transmission and distribution network [10,11]. The supply temperature from the plant and flow in the network are controlled by using feedback from the critical points in the network. The feedback is typically based on an adaptively estimated model describing the time-varying time-delay in the network, as well as the heat loss between the plant and the critical points in the network [4,10,12]. The controller uses predictions of the future heat load to adjust the supply temperature in order to reach the lowest possible temperature which still ensures that the temperatures are as required at the critical points. The flow is typically regulated to match the required heat demand while keeping the supply temperature as low as possible [10,13].

The European Union has set a target of 5% for the share of solar thermal production in district heating systems in 2050 [14]. Tschopp et al. [15] provide an extensive review of the performance and future of solar thermal production in district heating. Current developments in the performance of collectors and control strategies make the use of solar thermal units more attractive for the district heating sector. District heating systems generally design the size of solar farms to cover the heat demand in summer periods without production from other units. By using thermal storage and appropriate control strategies at a solar farm, district heating operators have the flexibility to store heat from surplus production and use this stored heat in periods with low solar radiation [16]. The control strategies for a solar farm require accurate heat load forecasts for optimal utilization of solar heat.

Furthermore, the overall production optimization of district heating systems is highly dependent on forecasts of the heat load [17]. Since production is optimized for the heat load forecast, accurate forecasts result in reliable optimal production planning for the system, with a larger potential for savings and utilization of green energy. Without an accurate forecast, the determined production schedule will be inefficient, costly, and maybe even infeasible. To summarize, the presented scenarios on how heat load forecasts are used to operate district heating more efficiently exemplify why improved heat load forecasts are highly desired and beneficial for the district heating sector.

1.2. Heat load forecasting

Heat load forecasting has been frequently studied, and several methods have been proposed along with a selection of variables that influence heat demand as inputs to the forecasting model. Dotzauer [18] suggest using a simple regression model including forecasts of outdoor temperature and heat demand of consumers. This simple regression model gives an accurate prediction of the future heat load by creating a piecewise linear function linking the heat demand to the outdoor temperature and using seasonal profiles of the heat demand. Dahl et al. [11] use the weather forecast uncertainty and an ARX (autoregressive exogenous) model to forecast the heat load using outside temperature, wind speed, and solar irradiance as inputs. The input variables are determined using model selection of different input variables. Their results show that the time-varying uncertainties improve the supply temperature control of a heat exchanger, thereby lowering the supply temperature and increasing savings by reducing heat losses to the surroundings in the transportation pipes.

Nielsen and Madsen [4] model heat consumption using the grey-box modeling approach to take advantage of combining physical and statistical modeling of district heating system. They also propose which input variables to use for heat load forecasting, and in some cases they suggest filtering the input variables due to e.g. the thermal inertia of buildings. The forecasting models presented in this paper are based on the methodology proposed in Bacher et al. [19]. Bacher et al. [19] propose an adaptive and recursive method to forecast heat load for single-family houses. Their method uses recursive least squares (RLS) to estimate the coefficients in the forecasting model. This allows the model coefficients to change over time and adapt to changes. The model is therefore self-calibrating, which is important because heat load is a nonlinear process, highly dependent on weather conditions. Therefore, the coefficients of the model change as the heat load dynamics change over time along with the weather. Bacher et al. [20] use the same RLS scheme for solar-power forecasting of a PV system, and Rasmussen et al. [21] use it to forecast electrical load for supermarket refrigeration after demonstrating how to incorporate non-linearity into the linear regression model.

1.3. Temporal hierarchies

Temporal hierarchies have not been applied to heat load forecasts, although such hierarchies have shown promising accuracy improvements in other areas, such as tourism and electrical load forecasting [5,7]. In temporal hierarchies, the hierarchy levels are different, non-overlapping temporal aggregations. For example, the total heat load of a single-family house over a week can be disaggregated into the total heat load for each day of the week. The heat load for each day can then be disaggregated into the demand per hour of the day. Here, the top level of the temporal hierarchy is the total weekly heat load, which is the aggregation of the daily heat load for the days in the week, which themselves are aggregations of the hourly heat load for each hour of the day.

Athanasopoulos et al. [6] state that forecasting different temporal aggregation levels reveals different information in the data. They demonstrate that on a lower resolution frequency, e.g. the weekly heat load, the trend in the load could be more dominant, while on a higher frequency, e.g. hourly heat load, seasonality can be more dominant. Hierarchical prediction models on different aggregation levels can capture this different behavior in the data. The information at each level is shared between levels using forecast reconciliation. Forecast reconciliation is the process of optimally combining hierarchical forecasts to yield coherent forecasts. A coherent forecast fulfills the constraints defined by the temporal hierarchy framework proposed by Athanasopoulos et al. [6], i.e. any forecast at an aggregate level is equal to the sum of the respective subaggregate forecasts from the level below.
The bottom-up method is a simple approach to generate coherent forecasts where the forecasts from the lowest aggregation level are aggregated to fit the hierarchical structure created beforehand. The top-down method disaggregates forecasts from the highest aggregation level using probabilities or past experiences to divide the forecasts into lower levels. Neither of these methods takes the relationship between aggregation levels into consideration. Therefore, information is lost between the levels, leading to a sub-optimal result. A detailed summary of traditional hierarchical forecasting is given in Athanasopoulos et al. [22].

Recent studies propose different methods to optimally reconcile forecasts in a temporal hierarchy. Hyndman et al. [23] create an independent forecast for each level, which is referred to as the base forecast and can be created from any model. The base forecasts are then reconciled to yield coherent forecasts. The reconciled forecasts are linear combinations of the base forecasts computed using generalized least squares (GLS) based on the covariance matrix $\Sigma$. Wickramasuriya et al. [24] show that the covariance matrix estimated from the coherency errors is nonidentifiable and therefore impossible to estimate. Hyndman et al. [25] propose using a diagonal covariance matrix estimated from the base forecast errors leading to a weighted least squares estimate of the reconciled forecasts. Wickramasuriya et al. [24] provide theoretical justification for using the empirical covariance matrix for the base forecast errors, $\hat{\Sigma}$, as an estimator for $\Sigma$. Nystrup et al. [7] demonstrate that using the full covariance matrix for the base forecast errors results in significantly more accurate reconciled forecasts than assuming no auto- and cross-correlation between aggregation levels.

Temporal hierarchies have been applied for several types of energy forecasting. Nystrup et al. [7] use temporal hierarchies to improve short-term electricity load forecasts. Jeon et al. [26] apply the temporal hierarchy framework to ensure coherence of probabilistic forecasts of wind power production, using a cross-validation method to find the weights in the reconciliation process. Yagüi et al. [27] and Yang et al. [28] use temporal hierarchies to improve the accuracy of solar power production forecasts for PV plants. These results demonstrate a promising potential of applying a temporal hierarchy to improve heat load forecasts.

The temporal hierarchy frameworks in literature frequently do not take the covariance into account and thereby result in lower accuracy improvements when dismissing the connection between aggregation levels [7]. There have also been no attempts to extend the framework to have adaptive and recursive updates of the covariance matrix. In this paper, we suggest a method to overcome these issues.

### 1.4. Contribution

This paper proposes a novel framework based on temporal hierarchies to improve the accuracy of heat load forecasts. This is done by combining forecasts from multiple temporal aggregation levels using an adaptive forecast reconciliation method. We propose three different approaches to update the covariance matrix, and hence the weights in the reconciliation process change over time to handle the time-varying dynamics of the heat load. The first method recursively updates the covariance matrix without any forgetting, the second method uses a rolling window with fixed width and equal weights to forget past information, and the third method uses exponentially decaying weights on past information to increase the importance of the most recent observations. All three covariance matrices use a shrinkage estimate of the full matrix structure based on the in-sample prediction errors. This enables the framework to share information both within and between aggregation levels.

We demonstrate the usefulness of the framework by improving the accuracy of state-of-the-art operational heat load forecasts for the Danish district heating planner, Varmelast. The operational forecasts were provided to Varmelast by a commercial forecast provider. The objective is to increase the accuracy of the hourly heat load forecasts (i.e., the operational heat load forecasts) for day-ahead operational planning using the proposed framework. Forecasts of the heat load for all aggregation levels for the next 24 h are issued at 23:00 every night. The base forecasts are used in the reconciliation process to generate reconciled forecasts. These forecasts are then used to demonstrate the accuracy improvements for the Varmelast case study. The usage of the commercial state-of-the-art forecasts is discussed by comparing them to our own simple forecasting model and the benefit of improving the forecasts using the proposed framework.

The contribution of this article is fourfold: (1) To the best of our knowledge, this is the first article to apply a temporal hierarchy to improve heat load forecasts. (2) We propose three adaptive and recursive methods to estimate the covariance matrix, which is used to reconcile the forecasts across the hierarchy. By allowing for time-varying weights, the approach is able to handle non-stationary processes by adapting to changing dynamics when required. This improves the potential for achieving more accurate forecasts in practical applications where stationary is often an issue. (3) We shrink the covariance matrix before reconciling the heat load forecasts. In order to do so, we extend the Ledoit and Wolf [29] closed-form solution for the optimal shrinkage intensity parameter by deriving a recursive estimation method. (4) This is the first article to include a commercial state-of-the-art operational forecast in the hierarchy and demonstrate accuracy improvements while using simple forecasts on the other temporal aggregation levels. Our results show that the suggested approach is able to improve state-of-the-art forecasts that have a direct influence on the operation of the system through optimal control of the network supply temperature. In addition to improving the current state-of-the-art forecast, our method yields coherent forecasts at multiple temporal aggregation levels which are useful for planning, e.g. purchases of biomass for a combined heat and power plant [7,30].

The remainder of this article is organized as follows. The data is presented in Section 2. In Section 3, the base forecast models used to generate forecasts for the aggregation levels above the operational forecast are presented. Section 4 outlines the theory of forecast reconciliation and proposes three different covariance estimators. The results are presented in Section 5 and discussed in Section 6. The paper is concluded in Section 7 with a summary.

### 2. Data

The data for this study is the heat load in the Greater Copenhagen area in Denmark. The data was provided by Varmelast, who deliver heat production planning for the district heating plants within this area. The data is the hourly heat load from 1 January 2016 to 31 December 2019, denoted by

$$\{Y_t; t = 1, \ldots, N\},$$

with the total number of observations $N = 35064$. The measurements are visualized in Fig. 1 and the seasonal dynamics of the heat load can clearly be seen. The top plot visualizes the heat load over the four years, showing the yearly seasonality. The yearly dynamics are the result of increased heating demand in the winter when the ambient temperature decreases and decreased heating demand in the summer when the ambient temperature increases. The lower plot shows one winter week to demonstrate the weekly and daily seasonality in the heat load. The weekly and daily patterns can be explained by consumer behavior, as the demand peaks in the morning and evening when consumers leave for work and return home, respectively. The weekends have a different pattern than the weekdays as the morning peaks disappear, which can be explained by fewer people going to work on the weekend.
2.1. Numerical weather prediction

The numerical weather predictions (NWPs) used as input to the forecasting models were provided by MetFor™ from the commercial forecast provider. The NWPs consist of climate variables with an hourly resolution that are updated every hour and are available to forecast the heat load, $\hat{Y}_{t+k}$. An example of an NWP for the $k$th forecast horizon is the predicted ambient temperature [in °C] denoted by

$$\left\{ T_{t+k|t}, t = 1, \ldots, N, k = 1, \ldots, 24 \right\}. \quad (2)$$

2.2. Operational heat load forecasts

Varmelast uses forecasts with an hourly resolution for multiple steps ahead to create the operational plan for the district heating plants. It is crucial to increase the accuracy of these forecasts, since this will allow more accurate planning. The commercial forecast provider supplies the hourly heat load forecasts to Varmelast. The heat load forecasts are supplied by HeatFor™. The heat load forecast, $\bar{Y}_{t+k|i}$, is updated every hour for $k$-steps ahead, as shown below,

$$\left\{ Y_{t+k|i}, t = 1, \ldots, N, k = 1, \ldots, 24 \right\}. \quad (3)$$

The objective of the paper is to improve the accuracy of the day-ahead hourly heat load forecasts using forecast reconciliation. The forecasts made at 23:00 each night for the next 24 h will be used to see if the proposed method can improve the accuracy of the hourly forecasts. Thus, the predictions horizon of interest in improving the hourly forecast accuracy is $k = 1, 2, \ldots, 24$.

3. Base forecasts

Base forecasts for each aggregation level are required for the proposed method. We will create base forecasts using the method proposed in Bacher et al. [19], i.e., recursive and adaptive regression-based models, and use some of the proposed input variables from Nielsen and Madsen [4], including their suggestions for filtering of some input variables. The filtering compensates, for instance, the thermal inertia of buildings, where only slow variations in the outdoor temperature are reflected in the heat needed to maintain a particular indoor temperature. The hourly heat load forecasts are supplied by the commercial forecast provider, but forecasts on the other temporal aggregation levels are needed. As focus is on the day-ahead operation, the temporal hierarchy will span from the daily level at the top level to the hourly level at the bottom. The aggregation levels are $\ell = \{24, 12, 6, 4, 3, 2, 1\}$. A total of eight different base forecast models will be created, including a simple model for the hourly level to demonstrate the difference between forecasts from a simple model and the commercial state-of-the-art forecasts.

The models will use regression form, for example

$$Y_{t+k} = \theta_0 + \theta_1 T_{t+k|t} + \epsilon_{t+k}, \quad (4)$$

where $Y$ is the heat load, $\theta$ are the coefficients, and $\epsilon$ is the residual. The predictors in this model are the intercept and the ambient temperature from the NWP. Therefore, the regressor at time $t$ is

$$X_t^T = (1, T_{t+k|t}), \quad (5)$$

and the parameter vector,

$$\theta^T = (\theta_0, \theta_1). \quad (6)$$

Hence, the model can be written as

$$Y_{t+k} = X_t^T \theta_{t+k} + \epsilon_{t+k}. \quad (7)$$

The base forecasts are generated using adaptive methods where the coefficients of the model are time-varying, i.e., they are updated every time a new observation becomes available. In addition, the model uses a forgetting factor to discount old information and increase the importance of the newest observations. This type of forecasting model is referred to as a recursive least squares model with a forgetting factor $\lambda$ [31]. The solution at time $t$ can be written in a recursive form. The RLS updates the coefficients when new information becomes available using:

$$\hat{\theta}_{k,i} = \hat{\theta}_{k,i-1} + R_{k,i}^{-1} X_{t+k|i} \left[ Y_t - X_{t+k|i}^T \hat{\theta}_{k,i-1} \right], \quad (8)$$

$$R_{k,i} = \lambda X_{t+k|i-1} + X_{t+k|i}^T X_{t+k|i}. \quad (9)$$

That is, the coefficients are recursively updated by weighted least squares estimation with the weights exponentially decaying over time. The rate of decay is determined by the forgetting factor, $\lambda$. $X_{t+k|i}$ is the regressor vector, $\hat{\theta}_{k,i}$ is the coefficient vector, and $Y_t$ is the dependent variable. The subscript $k$ represents the unique coefficient estimates for each $k$-step prediction model. The regressor vector only uses information available at time $t - k$ to forecast the dependent variable at time $t$ when updating the coefficients.

Each aggregation level has its own unique model that is estimated independently of the other aggregation levels. This is due to the fact that each aggregation level has unique dynamics. Thereby, each model structure is made in the regression form with the $k$-step prediction at time $t$ being

$$\hat{Y}_{t+k|i} = X_t^T \tilde{\phi}_{k,i}. \quad (10)$$

where the $\ell$ superscription refers to the target aggregation level of the forecast.

The regression vector, $X$, is created using filtered and non-filtered inputs to describe the heat consumption dynamics adequately for each aggregation level. Table 1 shows the models for each aggregation level with the inputs used and the prediction steps. The ambient temperature forecast, $T_{t+k|i}$, is filtered using a low-pass filter with a stationary gain of one. The systematic heat load peaks in the data are modeled using
function in optim() minimizing the root mean square error (RMSE) using the
is a parameter which corresponds to the time constant between the heat
number of harmonics, and \( \alpha \)
\( q \)-hourly level, forecasting three steps ahead is necessary to match
For example, the top level is the daily aggregation and the prediction
steps for each aggregation level are shown in the last column in Table 1.
The base forecast models for all aggregation levels and their corresponding inputs and
inputs to the base forecast models have an hourly resolution, which
is the total heat load for the given time resolution, i.e.
\( \hat{Y} \)
\( \alpha \)
\( T \)
\( T \)
\( q \)-hourly
Three-hourly
Two-hourly
Hourly
\( H \)
\( q \)
\( a \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
\( T \)
The base forecast models are created independently for each aggregation level. They are defined as \( \hat{Y}_{t+4|t} \) for the temporal hierarchy. The number of steps ahead, \( k \), of the prediction is the vector of \( 1, \ldots, m/\ell \). The forecasts are generated at time \( t \) for the hierarchy. The levels \( \ell = [24, 12, 4] \) would result in Fig. 2. Thus, we have nine base forecasts in total: six for \( \ell = 4h \), two for \( \ell = 12h \), and one for \( \ell = 24h \) at these timestamps. The base forecasts are not necessarily coherent, i.e. the sum of the first three four-hour forecasts is not necessarily equal to the first 12-hour forecast. However, the reconciled forecasts, \( \hat{Y}_{t+k|t} \) are coherent.

The vector \( \hat{Y}_{t+k|t} \) consists of all base forecasts for all aggregation levels. An example of this using the same hierarchy from Fig. 2 is

\[
\hat{Y}_{t+1|t} = \begin{bmatrix}
\hat{y}^{24h}_{t+1|t} \\
\hat{y}^{12h}_{t+1|t} \\
\hat{y}^{4h}_{t+1|t}
\end{bmatrix}
\]

which contains predictions from the three aggregation levels. The order of the vector is that the top row is the highest aggregation level and the corresponding prediction step; in this case, the daily and the one-step ahead prediction. Then come the lower aggregation levels with their \( k \) predictions until the lowest level, which is the four-hourly aggregation with the one-step to six-steps ahead predictions. The same holds for the reconciled forecasts, \( \hat{Y}_{t+k|t} \). In this paper, the reconciled forecasts are only computed daily at 23:00 for the next day to demonstrate the accuracy improvements. Therefore, the base forecasts generated at 23:00 are used in the reconciliation process. Consequently, \( \hat{Y}_{t+k|t} \) and \( \hat{Y}_{t+k|t} \) are updated daily when the base forecasts are generated and reconciled.

There are multiple ways to make hierarchical forecasts coherent. The simplest ways are the bottom-up and top-down methods, where forecasts are made either at the lowest or highest aggregation level, respectively. The bottom-up method is defined as

\[
\hat{Y} = SG\hat{Y}.
\]

where \( \hat{Y} \) is the base forecast made at the lowest level, \( S \) is the summation matrix matching the hierarchy structure, and \( G \) is a matrix of order \( m \times n \), which extracts the \( n \) bottom-level forecasts, transforming it to the correct form for the summation matrix. The matrix \( SG \) can be seen as the projection matrix. This method does not use any information between different aggregation levels, but can be used to demonstrate the reconciliation process. The process, as shown in Eq. (16), is a linear combination of the base forecasts where \( G \) extracts and combines the base forecasts into a vector of size \( m \) of disaggregated forecasts. The summation matrix, \( S \), then creates the reconciled forecasts from the disaggregated vector. In an example using the hierarchical structure in Fig. 2, the bottom-up method would only extract the four-hourly base forecasts from \( \hat{Y} \). The reconciled forecasts, \( \hat{Y} \), would then be created using the summation matrix from the four-hourly forecasts. Base forecasts from the other two aggregation levels would therefore not be used, and the reconciliation process would not use information from the higher levels. However, this can be changed by modifying the projection matrix. Hence, every reconciliation process can be written using this form with different \( G \).

Van Erven and Cugliari [34] propose a game-theoretical approach to estimate the optimal reconciled forecasts in two independent steps. They prove that their reconciliation method improves any base forecast or is at least as good as the base forecast. The steps are (1) creating the best possible predictions without coherency constraints and (2) mapping them to new predictions that are coherent. They formulate the problem of computing the reconciled forecasts as a minimax optimization problem that can be solved using convex optimization. Unfortunately, the problem does not have a closed-form solution. Hyndman et al. [23] and Athanasopoulos et al. [6] propose a closed-form solution using linear regression to estimate the reconciled forecasts for structural and temporal hierarchies using generalized least squares estimation. The base forecasts are written on the regression form, \( \hat{Y}_{t+k|t} = S\hat{y}(k) + \epsilon(k) \), (18)

where \( S \) is the appropriate summation matrix; \( \hat{y}(k) = E(Y_{t+k|t} | Y_{t+1}, \ldots, Y_t) \) is the unknown conditional mean of the future values of the most disaggregated observed series, i.e. the reconciled forecasts; and \( \epsilon(k) \) represents the error between the base forecasts and their expected value, i.e. the coherency error \( \hat{Y} - \hat{y} \). The error \( \epsilon(k) \) is assumed to have zero mean and covariance matrix, \( \Sigma \). Hence, the generalized least squares estimation of \( \hat{y}(k) \) in Eq. (18). If \( \Sigma \) is assumed to be known and the base forecasts are unbiased, the reconciled forecasts can be estimated by

\[
\hat{y} = S(S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1}\hat{y},
\]

where the matrix \( G = (S^T\Sigma^{-1}S)^{-1}S^T\Sigma^{-1} \).

It has been shown that \( \Sigma \) is never known nor identifiable [24]. Therefore, it is proposed to estimate it from the in-sample base forecast errors while imposing additional structure on the matrix. Numerous different methods have been proposed to use the \( \Sigma \) estimator instead of \( \Sigma \). Hyndman et al. [23] suggest using ordinary least squares (OLS), i.e. an identity matrix, which results in the equal weighting of all base forecasts. In Hyndman et al. [25], weighted least squares is used as an estimator by using the variance of the base forecasts to create the weights. In Athanasopoulos et al. [6], three different weighted least squares estimators, \textit{Hierarchy Variance Scaling}, \textit{Series Variance Scaling}, and \textit{Structural Scaling} are presented. These are naive methods for creating the estimator as the off-diagonal entries are zero, thus dismissing any cross-correlation between levels. Wickramasuriya et al. [24] propose the minimum trace (MinT) estimator, which uses all information between levels by using a full covariance matrix of the base forecast errors. Nystrup et al. [7] present four covariance estimators that account for the cross-correlation between levels.

In this study, the full covariance matrix will be used to generate the reconciled forecasts. The matrix shares information within levels and between levels. Recent results have shown that using the auto- and cross-covariance leads to significant improvements in forecast accuracy [7,24]. The covariance matrix is estimated from the past base forecast errors,

\[
\hat{\Sigma}_t = \frac{1}{l} \sum_{i=1}^{t} \epsilon_i \epsilon_i^T,
\]

where \( \epsilon \) is the vector of the base forecast error for each aggregation level and step-ahead prediction at each time step. It is assumed that the forecast errors are unbiased, i.e., \( E[\epsilon] = 0 \).

It is known that heat load follows seasonal patterns as the ambient temperature changes and space heating increases or decreases. The heat load is also more stationary over the summer period as there is no space heating. Therefore, two of the covariance estimators proposed include a forgetting factor. This allows them to forget past errors as the dynamics of the heat load changes between seasons and throughout the years. The third estimator is the full covariance matrix that uses all of the past available base forecast errors.

### 4.1 Method 1: Expanding window covariance matrix

The first method proposed is the approach where the covariance matrix estimator is based on all past available errors. The estimator includes both the auto- and cross-covariance between all aggregation levels. Nystrup et al. [7] applied this estimator of the covariance matrix and concluded that using all information between levels improves the forecast accuracy on all aggregation levels.

This estimator will be recursively updated every day when new observations are available and the error of the forecast can be computed,

\[
\hat{\Sigma}_{t} = \frac{1}{t} \epsilon_t \epsilon_t^T + \frac{l-1}{l} \hat{\Sigma}_{t-1}.
\]

Initially, this will update the covariance matrix quickly as new information becomes available, and the updating will be slowed down later.
4.2. Method 2: Rolling window covariance matrix

The second proposal is to estimate the covariance matrix using a rolling window. The estimation is performed on a fixed window of past errors where they all have equal weight. The rolling window adds new errors to the estimation when they become available and removes the oldest errors from the window. The length of the window or the memory is optimized by finding which memory yields the highest accuracy improvements of the operational heat load forecasts over the in-sample period.

The rolling window estimator is given by

$$\hat{\Sigma}_t = \frac{1}{t-j} \sum_{j=1}^{t} e_i e_i^T,$$

where the index $j$ is the length of the rolling window, i.e., the memory.

4.3. Method 3: Exponential smoothing covariance matrix

The exponential smoothing covariance matrix is a recursive and adaptive estimator. It is updated when a new observation is available, and every observation is weighted differently using a forgetting factor, $\lambda$. The weights are normalized to ensure that the sum of the weights is one [31]. The exponential smoothing estimator is given by

$$\hat{\Sigma}_t = \lambda \hat{\Sigma}_{t-1} + (1 - \lambda) e_i e_i^T.$$  \hspace{1cm} (23)

When $\lambda \to 1$, the covariance is not updated, and when $\lambda \to 0$, the covariance is highly influenced by the newest observations. The initial covariance matrix for $\hat{\Sigma}_0$ is usually initialized by computing the sample covariance over an initialization period in an offline setting. The forgetting factor and the initial period for the covariance are optimized to maximize the improvement in the accuracy of the operational heat load forecast as for the rolling window method.

4.4. Shrinkage

The fourth proposal is to shrink the estimated covariance matrices. Shrinkage of the covariance matrix has been shown to improve forecast accuracy considerably [7,24]. The shrinkage method used is the scale and location invariant shrinkage proposed by Ledoit and Wolf [29]:

$$\hat{\Sigma}_t^* = \lambda_{\text{shrink}}^* \hat{\Sigma}_t + (1 - \lambda_{\text{shrink}}^*) \hat{\Sigma}_t,$$  \hspace{1cm} (24)

where $\lambda_{\text{shrink}}^*$ is the diagonal entries from the covariance matrix $\hat{\Sigma}_t$. The shrinkage target is the diagonal variance of the levels since the off-diagonal elements of the covariance matrix are shrunk towards zero as $\lambda_{\text{shrink}}^*$ increases. Ledoit and Wolf [29] derived a closed-form solution for the optimal value of $\lambda_{\text{shrink}}^*$ by minimizing the mean squared error. This shrinkage estimator of the covariance matrix is ideal for a small number of data points with a large number of parameters. If the variance is assumed to be constant, then the optimal shrinkage parameter is given by

$$\lambda_{\text{shrink}}^* = \frac{\sum_{i,j} \hat{\text{Var}}(\hat{\theta}_{ij})}{\sum_{i,j} \hat{\theta}_{ij}^2},$$  \hspace{1cm} (25)

where $\hat{\theta}_{ij}$ is the $i$th element of the covariance matrix from the base forecast errors. The variance of the estimated covariance, $\delta_{ij}$, from the covariance matrix $\hat{\Sigma}_t$, is computed as shown in Appendix A in Schäfer and Strimmer [35].

We extend the work by Ledoit and Wolf [29] and Schäfer and Strimmer [35] such that the variance of the empirical covariance matrix is estimated recursively. A recursive estimate of the variance of the covariance matrix is found to be

$$\hat{\text{Var}}(\hat{\Sigma}_t) = \lambda (1 - \lambda) \bar{e}_i^2 (\bar{e}_i^2)^T - \hat{\Sigma}_t^2 + \lambda^2 \hat{\text{Var}}(\hat{\Sigma}_{t-1}).$$  \hspace{1cm} (26)

The shrinkage parameter can then be recursively updated at each time step using

$$\lambda_{\text{shrink}, t}^* = \frac{\sum_{i,j} \hat{\text{Var}}(\hat{\Sigma}_{t+1})_{ij}}{\sum_{i,j} (\hat{\Sigma}_{t+1})_{ij}^2}.$$  \hspace{1cm} (27)

The proof is given in the Appendix and the algorithmic solution is used in a recursive scheme. The recursive shrinkage will be imposed only on the exponential smoothing estimator. The other two estimators will be shrunk using Eq. (25). Hence, the estimate will be updated at each time step and shrunk before being used to reconcile the forecasts.

4.5. Optimization of memory parameter

For two of the covariance estimation methods, the rolling window and exponential smoothing, the memory parameter needs to be optimized to give the largest improvements. The expanding window estimator does not have a memory parameter. The initialization period for the sample covariance matrix also needs to be determined.

The initialization period and the memory parameter will be optimized to maximize the accuracy improvements of the operational heat load forecasts. The metric used to measure the accuracy improvements is the relative root mean squared error (RRMSE) that is frequently used to compare accuracy improvements of reconciled forecasts [6,7]. It was recommended by Hyndman and Koehler [36] due to the interpretability of the relative measure when considering accuracy improvements. The RRMSE is defined as

$$\text{RRMSE} = \frac{\text{RMSE}}{\text{RMSE}_{\text{base}}} - 1.$$  \hspace{1cm} (28)

where negative values describe a percentage improvement of the reconciled forecasts over the base forecasts. In this paper, the RRMSE is only computed for each aggregation level. The RMSE is computed as the average error for all prediction steps for the corresponding aggregation level:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \frac{1}{m/\ell} \sum_{k=1}^{m/\ell} (Y_{t+k} - \hat{Y}_{t+k})^2},$$  \hspace{1cm} (29)

where $T$ is the number of days forecasts are generated, $m$ is the number of forecasts at the bottom level, $\ell$ is the aggregation level, $k$ is the prediction horizon, $Y$ is the heat load observation, and $\hat{Y}$ is the heat load forecast.

For the rolling window, the memory is defined as the number of past days. These past days will be used to estimate the covariance at each time step and shrunk before being used to reconcile the forecasts.

The rolling window estimator is given by

$$\hat{\Sigma}_t = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{m/\ell} \sum_{k=1}^{m/\ell} (Y_{t+k} - \hat{Y}_{t+k})^2.$$  \hspace{1cm} (30)

where $N_{\text{eff}}$ is the effective memory for the exponential smoothing. The memory for the exponential smoothing is not a fixed-length window as for the rolling window due to exponential decay.

The initial sample covariance matrix needs to be invertible such that the reconciled forecasts can be estimated as shown in Eq. (19). Thus, an optimal initial period needs to be defined to make this feasible. The shrinkage will also ensure that the covariance is invertible when $\lambda_{\text{shrink}}^* > 0$ [29].

4.6. Workflow of forecast reconciliation

The workflow of the method proposed in this paper is illustrated in Fig. 3 and described in the following. Before the workflow is started, the base forecast models and the covariance matrix are initialized.

1. Generate base forecasts. The base forecasts in this paper are generated from the RLS regression models as introduced in Section 3 using the corresponding input variables for each aggregation level. A base
forecast vector at time $t$, $\hat{Y}_{t+1|t}$, is created which includes every aggregation level forecast with the corresponding forecast horizon as shown in Eq. (15). When the commercial state-of-the-art forecasts are used, they are combined with the higher aggregation base forecasts, and the authors’ hourly base forecasts are removed before the reconciliation process.

2. Update the covariance matrix. A covariance matrix is needed for the reconciliation process to combine the base forecasts. In this paper, the covariance matrix is updated recursively using one of the three methods proposed. As the covariance matrix is recursively updated, the previous covariance matrix needs to be kept for the update when a new observation becomes available.

3. Shrinkage. The covariance matrix is shrunk before the reconciliation process using either of the two shrinkage methods suggested in Section 4.4. One of the methods recursively updates the shrinkage parameter using Eqs. (26) and (27) while the other estimates it directly from the current covariance matrix using Eq. (24).

4. Generate reconciled forecasts. When the base forecasts and the covariance matrix are ready, the reconciled forecasts are computed using Eq. (19).

5. Compute forecast error when observations become available. When new observations become available, the base forecast errors, $e_t = Y_t - \hat{Y}_{t|t-1}$, are computed and the covariance matrix is updated. The updated covariance matrix is then used to generate new reconciled forecasts.

5. Results

In this section, the results from the proposed method, using three different covariance estimators, are presented and evaluated. The results will be discussed in terms of improving the accuracy of the operational forecast using the temporal hierarchy. The benefits of having a state-of-the-art forecast in the proposed method will be discussed by comparing the accuracy result with the result of using the authors’ simple model proposed in this article.

The optimization of the hyperparameters for the covariance estimators is presented in Section 5.1. The estimators are then used with the selected hyperparameters to create reconciled forecasts from the base forecast for the Varmelast case study. The results for the reconciled forecasts are presented in Section 5.2.

5.1. Hyperparameter optimization

As described above, the memory and initialization period for the covariance estimators need to be optimized. They are determined by minimizing the RRMSE. The year 2016 will be used as initialization period and years 2017 and 2018 as in-sample training period to estimate the optimal initial period and memory. Data from year 2016 is only available from 17 January until the end of the year, as the first 16 days are used for initialization of the base forecasts.

Fig. 4 shows the RRMSE of the operational heat load forecasts for different memories and initial periods for all three methods in the two in-sample years. The rolling window and exponential smoothing methods test three different initialization periods of 100, 200, and 350 days. The initialization period does not seem to affect the results for the rolling window, as it results in the same accuracy improvements for the three initial periods. However, the initialization period affects the exponential smoothing, as the plot for 2017 demonstrates. When comparing the result between the two years, it can be seen that the results are sensitive to the initialization period and the memory length. The 2017 test year is closer to the initial sample covariance, $\hat{\Sigma}_0$, which results in three different curves based on the initial period while the curves for 2018 are more similar. When the memory is longer than what is optimal, the improvements seem to decrease again. The rate of deterioration after the optimal point depends on the initial period for which the sample covariance is estimated. Fig. 5 demonstrates the effects between the memory and the initialization period for the exponential smoothing method. The plot shows the influence of the older observations, depending on the memory. The longer the memory is, the more influence older observations have. In other words, the initial sample covariance matrix will influence the current estimator, and the magnitude of the influence depends on the memory. Hence, when the memory is long and the initial period is 100 days, then the initial sample covariance has a considerable impact on the covariance matrix many months later. For example, the influence of the initial sample covariance on the current covariance after one year is close to 0.16 with a memory of 200 days, as the green curve in Fig. 5 shows.

The rolling window does not outperform the other two methods as the memory increases. Fig. 4 shows that the rolling window converges...
to the same accuracy improvements as the expanding window method with increasing memory size. This is not the case for the exponential smoothing method, as it seems to have an optimal point where it yields the maximum improvement. Based on these results, the initial period was set to the first 350 days, i.e. the entire year of 2016 after removing the first 16 days for the base forecast initialization to compute the initial sample covariance. The memory for the exponential smoothing was set to 365 days, i.e. a full year, as these parameters are hyperparameters and have a flat curve around the optimal point. The memory for the rolling window is the same as for exponential smoothing.

5.2. Empirical results

The results for the three years, 2017, 2018, and 2019 are presented in Table 2 along with the result for all three years combined. The 2017 and 2018 data were used to optimize the hyperparameters as shown in Section 5.1, while out-of-sample data from 2019 is used to validate the results. The hourly results in the table are for the commercial state-of-the-art forecasts. The hourly improvements are highlighted in Table 2 to emphasize that they are of most importance in this case. They are the operational heat load forecasts used to operate the district heating. The higher aggregation levels are included in the result for completeness. The RMSE of the base forecasts is shown in the first column for each year followed by the RRMSE for the three proposed covariance estimators: expanding window, rolling window, and exponential smoothing. The table shows that, for all aggregation levels and for both the in-sample and out-of-sample years, the reconciled forecasts using any of the estimation methods improve the base forecasts. The improvement of the operational heat load forecasts is consistently around 15% in all years. The improvement of the other levels varies more between years and the improvements are even higher in the out-of-sample year. The reconciliation approach performs at least as well out-of-sample as in-sample, which highlights its robustness. Each level has an improvement ranging from 10% to 50%, with the six- and eight-hourly aggregation levels having the largest improvements. These large improvements are due to the same accuracy improvements as the expanding window method with increasing memory size. This is not the case for the exponential smoothing method, as it seems to have an optimal point where it yields the maximum improvement. Based on these results, the initial period was set to the first 350 days, i.e. the entire year of 2016 after removing the first 16 days for the base forecast initialization to compute the initial sample covariance. The memory for the exponential smoothing was set to 365 days, i.e. a full year, as these parameters are hyperparameters and have a flat curve around the optimal point. The memory for the rolling window is the same as for exponential smoothing.

Rolling Window. The covariance matrix with fixed window size and equal weighting performs the worst out of the three methods. Optimization of the hyperparameters showed that the rolling window will converge to the same result as the expanding window with increasing memory, but it never outperforms the expanding window method.

Exponential Smoothing. The adaptive and recursively updated covariance that exponentially weights past errors yields the largest accuracy improvements. Having an adaptive and recursive covariance matrix results in the best performing estimator, as it always gives the largest accuracy improvements for the operational heat load forecasts. It also gives the largest improvements on the other levels in nearly all cases. Across all three years, it outperforms on all aggregation levels.

Fig. 6 shows the heat load observations, base forecasts, and reconciled forecasts using the exponential smoothing estimator for four of the aggregation levels. The gray dashed vertical lines represent the hour 23:00 every day when the models are updated and forecasts are generated for the next day. The plots show that the reconciled forecasts are frequently better than the base forecasts. There are, however, a few time periods where the reconciled forecasts are worse. Yet, overall the reconciled forecasts result in higher accuracy, as the RMSE in the lowest plot shows for the operational level. It is difficult to understand where the improvements at the lowest aggregation level come from. The improvements could come from truncation in the AR process and different memory in the forecasting models for the higher aggregation levels. The higher the aggregation level, the shorter the memory, and therefore it can react more rapidly. For example, rapid changes in the heat load or the outdoor temperature are shared with the lower levels.

The eight-hourly plot shows that the base forecast is only able to capture the trend, not the daily seasonality. However, the reconciled forecasts capture both the trend and seasonality. The reconciled forecasts are able to share information with the lower aggregation levels, where the seasonality is more dominant. The forecasts for the higher aggregation levels benefit from sharing information with the lower
Table 2
RMSE for the base forecasts and RRMS for the reconciled forecasts for daily heat load in the Greater Copenhagen area. Results are shown for three different years and the whole period from 2017 to 2019. The years 2017 and 2018 were used to optimize the hyperparameters while 2019 is the out-of-sample year. The result from the hourly forecasts is highlighted using gray-background to emphasize that the objective is to improve the hourly forecasts.

| Year      | Base RMSE | Expanding Window | Rolling Window | Exp. Smoothing |
|-----------|-----------|------------------|---------------|---------------|
| Daily     | 0.5585    | −17.88           | −17.7         | −18.1         |
| Twelve-hourly | 0.3151    | −17               | −16.72        | −17.19        |
| Eight-hourly  | 0.3333    | −39.78           | −39.61        | −39.9         |
| Six-hourly   | 0.2628    | −41.07           | −40.77        | −41.17        |
| Four-hourly  | 0.1715    | −35.24           | −34.86        | −35.34        |
| Three-hourly | 0.1273    | −31.98           | −31.62        | −32.09        |
| Two-hourly   | 0.0846    | −29.07           | −28.64        | −29.16        |

| Year      | Base RMSE | Expanding Window | Rolling Window | Exp. Smoothing |
|-----------|-----------|------------------|---------------|---------------|
| Daily     | 0.6058    | −29.6            | −30.21        | −30.41        |
| Twelve-hourly | 0.3602    | −28.91           | −28.54        | −29.1         |
| Eight-hourly  | 0.3759    | −49.82           | −49.36        | −50.01        |
| Six-hourly   | 0.3106    | −49.67           | −49.36        | −49.87        |
| Four-hourly  | 0.1852    | −40.74           | −40.22        | −41.05        |
| Three-hourly | 0.141     | −36.36           | −35.56        | −36.54        |
| Two-hourly   | 0.0924    | −32.94           | −32.02        | −33.13        |
| Hourly     | 0.0389    | −14.67           | −13.3         | −14.83        |

| Year      | Base RMSE | Expanding Window | Rolling Window | Exp. Smoothing |
|-----------|-----------|------------------|---------------|---------------|
| Daily     | 0.5628    | −17.88           | −17.7         | −18.1         |
| Twelve-hourly | 0.3151    | −17               | −16.72        | −17.19        |
| Eight-hourly  | 0.3333    | −39.78           | −39.61        | −39.9         |
| Six-hourly   | 0.2628    | −41.07           | −40.77        | −41.17        |
| Four-hourly  | 0.1715    | −35.24           | −34.86        | −35.34        |
| Three-hourly | 0.1273    | −31.98           | −31.62        | −32.09        |
| Two-hourly   | 0.0846    | −29.07           | −28.64        | −29.16        |

Bold values represent the best performance for each aggregation.

levels when they are not able to capture the seasonality. The lower aggregation levels benefit from sharing information with higher levels as they can capture the trend more accurately. Thus, temporal hierarchies using the covariance matrix with auto- and cross-covariance, which shares information both within each level and between levels in the reconciliation process, significantly improves the accuracy of all forecasts, including the operational heat load forecasts.

Fig. 7 shows four different error measures for each prediction horizon for the hourly day-ahead forecasts. The top plot shows the RMSE in 2019 for each prediction step. Recalculating the forecasts using the proposed exponential smoothing covariance improves accuracy for each step. The second plot shows the difference in RMSE between the base and reconciled forecasts for each prediction horizon. The plot shows that the reconciled forecasts improve the base forecast for all horizons. In general, the differences are quite similar in size except for the peak around $k = 6$. The third plot shows the root mean scaled squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2},$$

(31)

for each prediction step. Similar results are seen when comparing the RMSE and RRMS plots. The reconciled forecasts improve accuracy for all prediction horizons, as shown in the bottom plot, where the RRMSSE is plotted for each prediction step $k$. The plot also demonstrates that the accuracy improvements using the proposed method are consistent over the three years. Generally, the error should increase with the forecast horizon. This is not always the case, as the two top plots show some variation over the horizons due to the diurnal variation of heat load. This can be seen in the RMSE and RRMSSE plots, where the shape of the error follows the daily heat load profile, i.e., the error peaks in the mornings and afternoons when the heat load starts to increase. The horizon accuracy from the NWPs used as inputs to the models could also be a contributing factor here, combined with the daily profile. Comparing the RMSE of the $k = 1$ horizon between the years shows that accuracy is improved the least in 2019. For the first six steps ahead, the RMSE varies between the years, while the RRMSSE for the longer horizons is more consistent across the years. A possible explanation for this is that the first six hourly forecasts are the heat load between 23:00 and five in the morning when the heat load is relatively stationary.

The heating season can be split into three different periods. The first period is the first four months of the year, when the heat load is quite high and the transition from winter to summer occurs. The second period is the summer months from May through the end of August, when the heating demand in the Greater Copenhagen area is low. The third period is the last four months of the year, when the heat load starts to increase again as outdoor temperature decreases. Fig. 8 shows the cumulative squared error of the operational heat load forecasts for the year 2019 split into these three periods. The three plots show that the temporal hierarchy improves the accuracy in all three periods. Hence, the methods can perform even though the heat demand changes concurrently with the outdoor temperature. The largest accuracy improvements occur around the times when the heat load changes the most. For example, after October the heat load starts to increase as shown in Fig. 1, and that is also when a significant change in accuracy improvement happens. The same occurs in the transition period in the spring. In the summer period, when the heat load is stationary, there are only small improvements using the proposed method. All of the three covariance estimators outperform the commercial state-of-the-art base forecast in all three periods, with the exponential smoothing method performing the best.

To demonstrate how powerful temporal hierarchies are, we compare reconciled forecasts based on our own simple base forecasts from Section 3 for all aggregation levels to the state-of-the-art base forecast from the commercial provider. The simple model is the hourly model from Table 1. Note that this model uses accurate weather forecasts provided by the forecast provider as input. Fig. 9 shows that the state-of-the-art base forecast is significantly better than our own hourly base forecast. Yet, when applying the proposed adaptive and recursive covariance estimator to our base forecasts, the resulting reconciled forecasts outperform the commercial base forecasts slightly. Thus, creating forecasts from simple models at multiple aggregation levels can compete with state-of-the-art operational heat load forecasts in terms of accuracy. As expected, the reconciled forecasts based on the commercial state-of-the-art base forecasts are even more accurate than those based entirely on our own base forecasts. In other words, using more accurate base forecasts results in more accurate reconciled forecasts.

6. Discussion

In this work, we have suggested recursive and adaptive methods to update the covariance matrix used in forecast reconciliation, and we have shown that the suggested methods lead to significant improvements of the accuracy of operational heat load forecasts. The covariance estimator in the reconciliation process uses the auto- and cross-covariance to connect the base forecasts from all aggregation levels, and this improves the forecasts by sharing information within
Fig. 7. Error measures for the different prediction horizons from commercial hourly aggregation forecasts are illustrated. The plots show four different measures, the RMSE, the difference in RMSE between base and reconciled forecasts, RMSSE, and RRMSE for each prediction horizon. The reconciled forecast improves the accuracy for all horizons in all three years.

Fig. 8. Cumulative squared error of the base forecasts and the reconciled forecasts from the three covariance estimators. The three plots divide the year 2019 into three different periods with respect to heating demand and transition period.

Fig. 9. Cumulative squared error comparing state-of-the-art hourly forecasts from the commercial provider and forecasts created using the authors' simple model for the operational level. More accurate base forecasts lead to more accurate reconciled forecasts.

and between the hierarchy levels. The covariance matrix was made recursive and adaptive as the heat load is a time-varying process due to changes in the outdoor temperature. The results show that the covariance method needs an entire year of in-sample data for initialization and a long memory to yield the largest improvements. However, it only needs to store the previous covariance matrix and current errors for the reconciliation procedure.

We focused solely on improving the base forecasts for the next day that were generated at 23:00 every night. The proposed method could easily be applied to recursively reconcile new base forecasts generated every hour for the next 24 h using the same hierarchy. If only interested in one-hour ahead forecasts a different temporal hierarchy would be more beneficial. The strength of the proposed method is its adaptive property for the covariance matrix and the proposed recursive estimation of the shrinkage parameter. The weights for the reconciliation process and the shrinkage parameter are able to change over time when new information becomes available. This is highly relevant for systems that are governed by non-stationary processes.

Regression models were used to generate base forecasts for all aggregation levels above the operational hourly level for district heating. However, any forecast model could have been used to generate the base forecasts. Information is shared between the base forecasts through the covariance matrix, and different levels contribute with different information, e.g. the trend from higher levels and seasonality from the lower levels, to achieve accuracy improvements. Therefore, the temporal hierarchy structure needs to be chosen based on the application and the dynamics of the system. Commercial base forecasts were used to demonstrate the ability of the proposed method to achieve large accuracy improvements, even when combining state-of-the-art forecasts with simple forecasts on higher levels.

Similar findings of accuracy improvements when forecasting for different temporal levels and using auto- and cross-covariance matrices to compute the reconciled forecasts are reported in other applications. Wickramasuriya et al. [24] propose using the full covariance matrix estimated from the base forecast errors and the same shrinkage as used in this paper in the forecast reconciliation process. They benchmark different structures of the covariance matrix to demonstrate the importance of using all of the information in the covariance matrix by sharing it within and between aggregation levels. Nystrup et al. [7] has suggested various structures and regularization for the auto- and cross-correlation to maximize the accuracy improvements when reconciling forecasts in a temporal hierarchy.

Improving commercial state-of-the-art operational forecasts will result in more reliable operational planning of the district heating system. It will help to find the optimal production schedule of the units and improve control of the heat delivered to the network. Blanco et al. [17] propose a method that optimizes the production and creates bids for the day-ahead and balancing electricity markets using stochastic programming that needs an accurate heat load forecast. Thus, by improving operational heat load forecasts, the system can be optimized more accurately and the overall system cost can be reduced by profiting from the electricity market.
7. Conclusion

We demonstrate how temporal hierarchies can be used to improve the accuracy of operational heat load forecasts. The heat load forecasts are generated each day at 23:00 for the next 24 h. State-of-the-art heat load forecasts from a commercial energy forecasting provider are included to see if accuracy improvements are possible.

The empirical results are based on four years of heat load data. The first year is used to initialize the method and the next two years to find the optimal hyperparameters for the covariance estimators. The last year is used to show the accuracy improvement on out-of-sample data.

Three covariance estimators are presented based on the base forecast errors from the different aggregation levels. The first estimator is recursively updated and uses an expanding window, meaning that all past errors are used with the same weight. The second estimator uses a rolling window with a fixed number of past errors having the same weight. The third estimator uses exponential smoothing to put weights on past errors. This is the estimator that we propose for the reconciliation process to improve the accuracy of heat load forecasts. The covariance matrix estimate is shrunk before it is used to compute the reconciled forecasts. The shrinkage parameter is recursively updated using the method proposed.

The hyperparameters for the estimators are selected by minimizing the RRMSE over two years. The three different estimators are tested and compared using RRMSE. The exponential smoothing estimator yields the largest accuracy improvements in the reconciliation process in our case study. This estimator improves the accuracy of the commercial state-of-the-art operational heat load forecasts by, on average, 15% over the three years. The proposed method improves the base forecasts at all aggregation levels and for all forecast horizons considered. When using only simple base forecasts as inputs at all aggregation levels, the reconciled forecasts have similar accuracy to the commercial state-of-the-art base forecast. However, using the state-of-the-art base forecast as input, the reconciled operational heat load forecast becomes significantly better. Hence, it is necessary to include a state-of-the-art forecast to get maximum accuracy.

There are several possible directions for future research on improving heat load forecast accuracy using temporal hierarchies. One possibility is to investigate the optimal number of aggregation levels in greater detail, including an analysis of how longer-term forecasts (e.g., one-week ahead) would perform. Another option is to improve the covariance estimator used in the reconciliation process, e.g. through dimensionality reduction or by imposing additional structure on it. Finally, it is possible that the accuracy of the reconciled operational heat load forecasts could be further improved by improving the accuracy of the base forecasts.

CRediT authorship contribution statement

Hjörleifur G. Bergsteinnson: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. Jan Kloppenborg Møller: Conceptualization, Methodology, Writing - review & editing, Supervision. Peter Nystrup: Methodology, Software, Writing - review & editing, Ölafur Pétur Pálsson: Writing - review & editing. Daniela Guericke: Writing - review & editing. Henrik Madsen: Writing - review & editing, Supervision, Funding acquisition.

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 is funded by Innovation Fund Denmark through the projects Heat 4.0 (8090-00046B), CITIES (1305-00027B), Flexible Energy Denmark (8090-00069B), the IDASC, Denmark (18012745) project funded by Region H, and the Norwegian FME-zen project funded by the ZEN partners and the Research Council of Norway.

We thank the energy forecasting provider ENFOR, especially Henrik Aalborg Nielsen and Torben Skov Nielsen, for their support and for supplying the forecasts & information. We would also like to thank Varmelast for making their heat load data available.

Appendix. Recursive shrinkage

Computing the optimal estimated shrinkage intensity, \( \lambda_{\text{shrink}} \), as Schäfer and Strimmer [35] demonstrate in Appendix, while doing recursive and adaptive estimation of the covariance matrix is infeasible. Here we show how to recursively estimate the empirical variance and covariance of the individual entries of the recursive and adaptive covariance matrix \( \Sigma_t \).

The recursive and adaptive covariance matrix is defined as

\[
\Sigma(t+1) = (1 - \lambda) \sum_{i=0}^{t} \lambda_i e_{t,i} e_{t,i}^T + \lambda (\Sigma(t))
\]

where \( \lambda \) is the forgetting factor and \( e_{t,i} \) is the base forecast error vector at time \( t - i \). We will assume that the error is normally distributed with zero mean and covariance, \( \Sigma \). Then, we can assume that Eq. (A.1) comes from the Wishart distribution. Therefore, the matrix from the Wishart distribution computes the variance of the individual entries.

The variance and covariance of the individual entries of \( \Sigma_t \) are computed by taking the variance of each term in Eq. (A.1),

\[
\begin{align*}
X_{t,i,j} &= e_{t,i} e_{t,j}^T, \\
\text{Var}(\Sigma_{i,j}) &= \text{Var} \left[ (1 - \lambda) \sum_{i=0}^{t} \lambda_i X_{t,i,j} \right],
\end{align*}
\]

assuming independence,

\[
\begin{align*}
\text{Var}(\Sigma_{i,j}) &= (1 - \lambda)^2 \sum_{i=0}^{t} \lambda^2 \text{Var}[X_{t,i,j}], \\
\text{Var}(X_{t,i,j}) &= \mathbb{E}[(X_{t,i,j} - \mathbb{E}[X_{t,i,j}])^2], \\
&= \mathbb{E}[X_{t,i,j}^2 - 2 \mathbb{E}[X_{t,i,j}]X_{t,i,j} + \mathbb{E}[X_{t,i,j}]^2], \\
&= \mathbb{E}[X_{t,i,j}^2] - \mathbb{E}[X_{t,i,j}]^2, \\
&= \mathbb{E}[X_{t,i,j}^2] - \text{Var}[X_{t,i,j}], \\
&= \mathbb{E}[X_{t,i,j}^2] - \sigma_{i,j}^2,
\end{align*}
\]

where with this we have an analytical solution to estimate the variance and covariance of the individual entries of \( \Sigma_t \). We also present an algorithmic solution of Eq. (A.10) where the objective is to have the result in a matrix. The method is shown here

\[
\begin{align*}
\text{Var}(\Sigma_t) &= (1 - \lambda)^2 \sum_{i=0}^{t} \lambda^2 (\mathbb{E}[X_{t,i,j}^2] - \sigma_{i,j}^2), \\
\text{Var}(\Sigma_t) &= (1 - \lambda)^2 \sum_{i=0}^{t} \lambda^2 (\mathbb{E}[X_{t,i,j}^2] - \sigma_{i,j}^2) + \lambda^2 (1 - \lambda)^2 \sum_{i=0}^{t-1} \lambda^2 (e_{t,i}, (e_{t,i})^T - \Sigma_i^2),
\end{align*}
\]
\[ \bar{V}(\Sigma_{-1}) = (1 - \lambda^2) \sum_{i=0}^{\infty} 2x_i^2 (e_i^2 - \delta_i^2), \quad (A.13) \]

\[ \bar{V}(\Sigma) = (1 - \lambda^2) \bar{\lambda} (e_i^2 - \delta_i^2) + \lambda^2 \bar{V}(\Sigma_{-1}) \quad (A.14) \]

References

[1] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: A review and outlook. IEEE Open Access J Power Energy 2020;7:376–88. http://dx.doi.org/10.1109/OAJPE.2020.3029979.

[2] International Renewable Energy Agency (IRENA). Global energy transformation: A roadmap to 2050. 2018.

[3] Taylor J. Short-term electricity demand forecasting using double seasonal exponential smoothing. J Oper Res Soc 2003;54(8):799–805. http://dx.doi.org/10.1057/palgrave.jors.2610589.

[4] Nielsen H, Madsen H. Modelling the heat consumption in district heating systems using a grey-box approach. Energy Build 2006;38(1):63–71. http://dx.doi.org/10.1016/j.enbuild.2005.05.002.

[5] Clemen RT. Combining forecasts: A review and annotated bibliography. Int J Forecast 1989;5(4):559–83. http://dx.doi.org/10.1016/0169-2070(89)90012-5.

[6] Van Erven T, Cugliari J. Game-theoretically optimal reconciliation of contemporaneous time series. Stat Appl Genet Mol Biol 2008;7(4):article 53. http://dx.doi.org/10.1016/j.ijforecast.2020.12.003.

[7] Ledoit O, Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J Emp Financ 2003;10(5):603–21. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[8] Hyndman RJ, Ahmed R, Athanasopoulos G, Shang H. Optimal combination forecasts for hierarchical and grouped time series. J Amer Stat Assoc 2011;106(5):289–99. http://dx.doi.org/10.1016/j.joarev.2018.09.005.

[9]ipient for hierarchical and grouped time series. Int J Forecast 2009;25(1):146–66. http://dx.doi.org/10.1016/j.ijforecast.2008.07.004.

[10] Hyndman R, Ahmed R, Athanasopoulos G, Shang H. Optimal combination forecasts for hierarchical and grouped time series. J Amer Stat Assoc 2011;106(5):289–99. http://dx.doi.org/10.1016/j.joarev.2018.09.005.

[11] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[12] Madsen H, Sejling K, Seggaard HT, Palsson OP. On flow and supply temperature control in district heating systems. Heat Recov Syst CHP 1994;14(6):613–20. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[13] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[14] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[15] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[16] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[17] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.

[18] Dahl M, Brun A, Andersen G. Using ensemble weather predictions in district heating systems. Proc Inst Mech Eng A 1993;207(A3):157–63. http://dx.doi.org/10.1016/S0927-5398(03)00078-8.