Heat load forecasting using adaptive spatial hierarchies

Bergsteinsson, Hjörleifur G.; Sørensen, Mikkel Lindstrøm; Møller, Jan Kloppenborg; Madsen, Henrik

Published in:
Applied Energy

Link to article, DOI:
10.1016/j.apenergy.2023.121676

Publication date:
2023

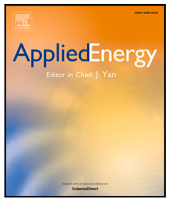
Document Version
Publisher's PDF, also known as Version of record

Applied Energy 350 (2023) 121676

Contents lists available at ScienceDirect

Applied Energy

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



Heat load forecasting using adaptive spatial hierarchies

Hjörleifur G. Bergsteinsson *, Mikkel Lindstrøm Sørensen, Jan Kloppenborg Møller, Henrik Madsen

Technical University of Denmark, Denmark

ARTICLE INFO

Keywords:
Forecast reconciliation
Adaptive and recursive estimator
Enhancing load forecasts
Spatial hierarchy
District heating

ABSTRACT

District heating is an efficient method of distributing heat in densely populated areas at a low cost. The heat is usually produced at central production plants and then distributed to consumers through large networks of pipes. However, district heating is gradually becoming more decentralised with additional heat sources, e.g. heat pumps, solar thermal farms, and industrial waste heat connected to the network. Therefore, the system is changing from a system with centralised heat sources to a more decentralised system with several different heat sources within the network, including also still a large production area. Operationally this is more complex than the previous setup, especially in terms of temperature optimisation. Typically, the temperature must be adjusted for each area in order to work efficiently with the decentralised heat sources, so a forecast of the local heat load is required. It is relatively easy to make a forecast for each area, but they are usually made independently and are therefore not necessarily coherent. In this paper, we propose a methodology to spatially reconcile hierarchies of individual localised heat load forecasts with a coherency constraint. This results in coherent reconciled forecasts. Enhancing forecast accuracy and making them coherent are essential for future decentralised systems as temperature and production optimisation need accurate information to yield optimal operation. We will use two different case studies to illustrate the proposed method. One case study has a few areas, while the other case study will have more areas, and here it is proposed to add a new level of aggregation to the hierarchy to increase accuracy. The results in this paper show that the reconciled forecast, where information is shared between areas through the spatial hierarchy, improves forecast accuracy by 1% to 20%, depending on the prediction horizon.

1. Introduction

Supply temperature in district heating networks needs to be reduced to increase efficiency and maximise the flexibility potential of district heating [1]. Improving the efficiency of district heating and maximising flexibility are important as they can be seen as a solution to the occasional electricity surplus that is expected to grow over time with the increase in the share of intermittent renewable energy sources (RES) [2]. District heating can increase the efficiency of the overall energy system due to its unique ability to store heat over long periods of time and its high efficiency in converting electricity into heat, e.g. through heat pumps [3]. In addition, due to the increasing number of intermittent heat sources, more thermal energy storage (TES) systems are being installed [4,5]. These new changes with additional heat sources and TES systems lead to more complex systems than traditional district heating systems. Future district heating systems will become more decentralised, with more RES sources and a closer coupling with the power sector.

However, to make this transition more feasible and efficient, the supply temperature of district heating networks needs to be lowered to make the integration of the new heat sources, e.g. heat pumps, into the district heating system more effective [3,6]. In addition, more detailed knowledge of the system needs to be acquired as district heating becomes more decentralised with heat units and TES in the network. For instance, if the district heating system consists of a large production area with transmission lines and two heat exchangers supplying two distribution networks for heat. The distribution networks could be equipped with heat pumps and TES. Therefore, the supply temperature for each distribution network area and the transmission line needs to be optimised precisely by taking all factors into account in order for the system to work efficiently. Temperature optimisation requires information about future demand to obtain optimal supply temperature set points [7,8]. It takes time to send one unit of hot water from a production plant to the users (the time delay). Therefore, information about the future heat load and network characteristics (e.g., time delay, temperature loss) are needed to ensure that sufficient temperature

* Correspondence to: Anker Engelunds vej 1, Building 101A, 2800 Kongens Lyngby, Denmark.
E-mail address: hgbe@dtu.dk (H.G. Bergsteinsson).

https://doi.org/10.1016/j.apenergy.2023.121676
Received 26 March 2023; Received in revised form 29 June 2023; Accepted 25 July 2023
Available online 11 August 2023

reaches the consumers [9]. Hence, an accurate heat load forecast must be available. In practice, however, district heating companies usually only produce a forecast for a large area, or they distribute a forecast for a large area to smaller areas by scaling it according to a heat consumption ratio determined in the past [10]. This practice introduces a significant error and bias in the heat load forecast, as the dynamics of the heat load can change rapidly due to weather or social changes.

Accurate forecasts of the heat load are important for district heating utilities for controlling the supply temperature. A total heat load forecast is also needed for district heating production optimisation and planning. Therefore, forecasts for the total heat load (transmission) and the local areas (distribution) are needed. Ideally, these forecasts must be coherent, i.e. the aggregation of the local forecasts must be consistent with the total forecast. However, this is not guaranteed if the forecasts are produced independently of each other. Coherence can be achieved by reconciling forecast hierarchies [11]. Hierarchy constraints ensure that forecasts are coherent but have also been shown to lead to more accurate forecasts as they share information between them through the hierarchy [12]. Hierarchy constraints are used in the reconciliation process, where incoherent forecasts are used as inputs and produce coherent forecasts as output, which are called reconciled forecasts. This not only solves the problem of incoherent forecasts but also improves forecast accuracy [13].

1.1. Heat load characteristics and forecasting

Energy forecasting is evolving rapidly, especially for renewable energy sources [14]. Energy forecasts are valuable because they are needed to optimise future strategies for the sectors. For instance, in a district heating system, it could be decided if the TES system is charged when electricity prices are expected to be low during low heat load or if the Combined Heat and Power (CHP) plant is operated at full capacity when electricity prices are expected to be high. Even if the heat demand is low during the operation of the CHP, the TES system could be charged. Therefore, a heat load forecast is needed to support decision-making and production optimisation when selecting heat units [15], charging/discharging the TES system [4,16], or temperature optimisation [8,17,18].

Heat load forecast is therefore often studied, and its popularity has increased in recent decades due to the growing importance of unlocking the flexibility of district heating for energy systems with a high share of fluctuating renewables. It is important to understand the characteristics of heat in order to build an appropriate and robust forecasting model. Heat load can be divided into two main categories, space heating and domestic hot water usage with the addition of heat losses in the systems. Space heating is about heating the consumer's home to maintain thermal comfort, while domestic hot water usage is about meeting the consumer's other needs, such as showering. The heat load can also be divided into a physical and a social heat load, as described in Gadd and Werner [19]. The physical part is determined by the weather and the thermal insulation of the building, where the ambient air temperature is a significant variable for driving the heat consumption [20]. The insulation of the building envelope acts as a resistance in the heat transfer between the ambient air and indoor temperature, and the heat consumption is used to maintain a certain indoor temperature. Other weather variables also influence heat consumption, e.g. solar radiation and wind speed. Solar radiation reduces heat consumption as the solar beams enter through the window and heat the floor. The effect depends, for example, on the strength of the radiation, the angle of the rays and the size of the window. The wind influences the natural ventilation in buildings, and a higher wind speed increases heat consumption. The social effect affects heat consumption when hot water is used, e.g. when taking a shower in the morning before work. The heat load is not stationary, as heat consumption also changes due to weather changes. For example, as the ambient air temperature rises at the beginning of summer, less space heating is then needed until a certain threshold temperature is reached, at which space heating is no longer required. Social behaviour also changes over time, e.g. during the summer holidays when fewer people are at home. More detailed information about physical and social heat load dependencies can be found in Nielsen and Madsen [20] and Gadd and Werner [19].

Knowledge as outlined above can then be used to identify significant input variables for the model and how they should be treated (e.g., instantaneous or transformed). Several prediction model methods have been proposed that can be used to predict heat load. Dotzauer [21] propose a linear regression model where the relationship between the heat and the input variables is investigated, e.g. the dependence on the ambient air temperature is treated by creating a piecewise linear function. Dahl et al. [22] also uses a linear regression model with seasonal lags. The inputs are the weather forecasts, where the aim was to investigate the gain from uncertainty in heat load forecast by using an ensemble of weather forecasts as inputs. Grosswindhagera et al. [23] proposes using Seasonal Autoregressive Integrated Moving Average (SARIMA) formulated in a state-space form to produce online forecasts using the Kalman filter. These methods are simple and neglect the non-linearity and non-stationarity of the heat load. In Dahl et al. [24] and Idowu et al. [25], linear regression, neural network and support vector regression (SVR) are compared, which in Idowu et al. [25] also is compared with regression tree. The SVR method provides the best predictive performance with the lowest prediction error. The SVR and the neural network can deal with non-linearity because they are nonlinear models, whereas the SVR can use nonlinear kernel values to model the relationship, and the neural network has nonlinear activation functions. However, these models have problems with the non-stationary heat load and would therefore often need to be recalibrated frequently. Nielsen and Madsen [26] propose to use the grey-box method to predict the heat load where physical insights and statistical methods are used to find an optimal model. A linear regression model is used for predicting heat load, where the inputs are transformed to deal with non-linearity. Also, the coefficients are estimated using recursive least squares (RLS) with exponential forgetting to deal with non-stationarity, allowing the coefficients to adapt to changes as new observations become available and exponentially down-weighting the older ones. Recurrent neural networks (RNN) and convolutional neural networks, and long short-term memory (CNN-LSTM) have been proposed to extend the neural network to deal with non-stationarity, as shown in Kato et al. [27] and Song et al. [28].

1.2. Hierarchical forecasting and reconciliation

Reconciliation is the process of making forecasts coherent according to their hierarchical structure. The individual forecasts are usually incoherent, and in many applications, these forecasts need to be coherent as specified by their hierarchy. Energy production planning, for example, needs coherent forecasts to make optimal decisions based on forecasts for different horizons [29]. The reconciliation process not only makes the individual base forecast coherent but has also been shown to increase the accuracy of the forecasts [12,13,29]. Van Erven and Cugliari [30] prove that reconciliation forecasts should perform at least as well as the base forecast on average. However, this depends on the quality of the weights in the reconciliation process, as shown in Nystrup et al. [31], where it is demonstrated that no improvements can be achieved if the errors of the base forecasts, which are used to estimate the weights, are too highly correlated. Therefore, the independent base forecast at the different levels in the hierarchy cannot come from the same model using the same information (e.g. input variables) at all levels, as they have nothing to exchange between levels. Hollyman et al. [32] show that the reconciliation process is a special case of a combination forecasting method by reformulating it into a combination of direct forecasts using linear coherent constraints.

Several recent studies have demonstrated the improvement in accuracy using the reconciliation process, and the benefits of coherent

forecasts in the energy sector [33]. Nystrup et al. [29] demonstrates the improvement for short-term electricity load forecasts. Jeon et al. [34] demonstrate the reconciliation process for probabilistic forecasting of wind power and electric load to ensure coherence, resulting in higher accuracy of the forecasts. Bergsteinsson et al. [12] demonstrate that using temporal hierarchies with the hierarchy of all-natural levels from one hour ahead to daily time resolutions leads to 15% higher accuracy compared with state-of-the-art hourly heat load forecasting.

Therefore, reconciling an independent base forecast of heat load between different areas using spatial hierarchy inside the district heating network can result in accuracy improvements for all forecasts inside the hierarchy. Hence, it will simultaneously make the forecasts across the spatial hierarchy coherent and sharing information between levels will improve forecasts and have a positive effect on the operation of district heating. As mentioned, a district heating network usually consists of a transmission system and multiple distribution systems, these systems need individual temperature optimisation such that the heat demand is fulfilled for every customer within each distribution while keeping the supply temperature as low as possible. Temperature optimisation requires heat load forecast due to the time it takes to send a unit of water with the correct temperature over the network before it reaches the customer. Hence, simultaneously sharing and improving the forecasts through the spatial hierarchy will lead to higher accuracy in forecasting resulting in increased precision in temperature optimisations. The forecast improvements will then lead to higher cost savings for the operation of the system since a more precise forecast will improve the production optimisation, temperature optimisation and the operation of the decentralised heat sources, e.g. heat pumps. It will also increase the possibility of reducing supply temperature in the network, reducing cost, and reducing heat losses in the system [8]. This will also increase the efficiency of the heat sources in the system, e.g. cogeneration plant [35]. More importantly, it will increase the feasibility of power-to-heat units, and the flexibility of district heating is then leading to a flexibility of the power system [3].

1.3. Contribution

The purpose of this study is to develop a spatial hierarchical framework that handles the dynamics of heat load to improve the accuracy of individual forecasts for each area in the hierarchy for district heating load. The proposed method makes individual base forecasts coherent. It improves accuracy by using the reconciliation process with an empirical covariance matrix estimator estimated from the base forecast errors. Due to the nature of the heat load observations, it is proposed to use a recursive and adaptive covariance estimator, i.e. the estimator can be easily updated and weights down previous observations to give more importance to new observations. We propose using exponential smoothing with a forgetting factor to estimate the covariance matrix as suggested in Bergsteinsson et al. [12].

We propose to estimate an individual covariance matrix for each prediction horizon and use an optimal forgetting factor for each horizon. We document the accuracy improvements by using a covariance matrix estimated for each horizon compared to only using forecast errors from one-step-ahead predictions. Also, the accuracy difference between using the same forgetting factor for all horizons and the optimal factor for each horizon is also investigated. These results are demonstrated using two different case studies, one with few areas and another with many areas. For the case study with many areas, it was possible to add a new level, and it is shown that adding a new level to the hierarchy increases the accuracy. An operational state-of-the-art heat load forecasting system is used to compare and illustrate the accuracy improvements possible through the reconciliation process. The paper also includes a simulation study on heat load forecasting to investigate the accuracy improvements when the spatial hierarchy is applied in the reconciliation process.

The contribution of this article is to demonstrate the accuracy improvement achieved using spatial hierarchies with recursive and adaptive methods for heat load forecast. Also, the proposed method results in coherent forecasts that will be essential for future decentralised district heating systems. A decentralised system will introduce more complexity to the operation of the system and having forecasts from all individual areas to match the total aggregation will reduce the complexity for operators. Thus, continuous effort on enhancing forecast is needed for improving decision making which should yield lower costs and make the system more robust.

The paper is organised as follows. The data from the two case studies are presented in Section 2. Section 3 discusses the base forecasting model used by the authors to forecast heat load. The spatial hierarchies and the reconciliation process are also presented. The results are presented in Section 4 and discussed in Section 6. The paper is then concluded in Section 7.

2. Data

The data used in this study is heat load from two district heating utilities in Denmark; *Fjernvarme Fyn* and *Brønderslev Forsyning*. Both utilities produce and deliver heat to their consumers in Fyn and Brønderslev, respectively. *Fjernvarme Fyn*'s heat load data consists of total heat consumption and the twelve areas into which the total heat consumption is divided, while *Brønderslev Forsyning* consists of total consumption and three areas. The data for both utilities have an hourly resolution from 1 January 2019 to 1 December 2021 for Fyn, while Brønderslev has data from 1 February 2020 to 1 June 2022. The heat load is denoted by

$$\left\{Y_t^{\mathrm{i}};\ t=1,\dots,T,\ i=\mathrm{Total},\mathrm{Area}_1,\dots,\mathrm{Area}_{\mathrm{g}}\right\}, \tag{1}$$

where T is the total number of observations, and g is the corresponding total number of areas. The number of areas for *Fjernvarme Fyn* is 12 while for *Brønderslev Forsyning* it is three.

The heat load measurements for the Total and the three areas in Brønderslev are shown in Fig. 1. It shows the usual dynamics of the heat load in Denmark with an annual seasonal dependence on the ambient air temperature. When the ambient air temperature decreases, more heat is used for space heating (i.e. to achieve the desired thermal comfort at each consumer). However, when the ambient air temperature rises, almost no space heating is needed and heat is then only used for domestic hot water consumption (e.g. showers). The dynamics and scale between the areas in Brønderslev are quite similar, as Fig. 1 shows.

The heat load measurements for *Fjernvarme Fyn* are visualised in Fig. A.18. The heat load dynamics for the areas in Fyn show differences between the areas and also in magnitude. Fig. 2 shows the map of the area layout where *Fjernvarme Fyn* distributes the heat. From the map it can be seen that some areas (e.g. 6, 7, 8, 11, 12) are outside the cluster and therefore away from the more densely populated areas. However, the authors do not know what types of consumers (e.g. apartments, single-family houses, industry) and what number of consumers are present in each area. They only know that *Fjernvarme Fyn* supplies heat to the city of Odense and the surrounding towns, with the centre of the map showing Odense.

The heat consumption from both utilities shows seasonal variations in heat load as described in Gadd and Werner [19], correlating with ambient air temperature, i.e. high in cold periods and low in warm periods. The daily heat load profile also correlates with ambient air temperature, but the social behaviour of each area also shapes the profile. In addition, there are some missing values in all series. The magnitude of load is not the same in all areas, and the volatility is also different. The heat load dynamics in the areas at Fyn, as seen in Fig. A.18, differ more than the heat load dynamics between the areas in Brønderslev, as seen in Fig. 1. The areas in Brønderslev have similar dynamics with slight differences in magnitude but almost identical. The

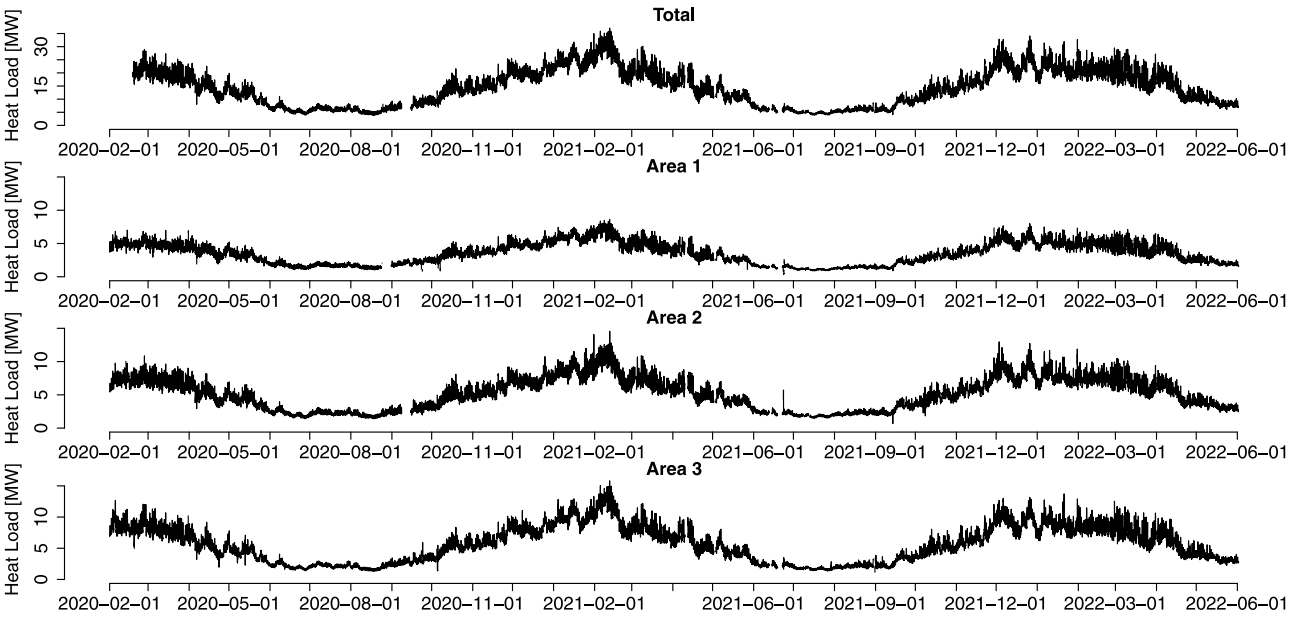


Fig. 1. Time series plot of the heat load over the one and half year period for each group inside the *Brønderslev Forsyning* system.

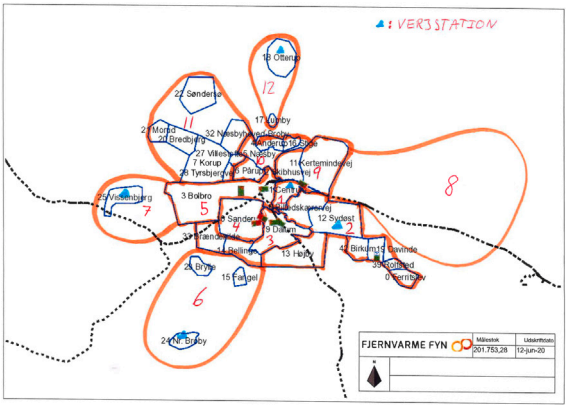


Fig. 2. Layout of the areas at *Fjernvarme Fyn*.

areas on Fyn have both different dynamics and magnitudes however, some areas are similar and could therefore possibly be grouped together based on the heat load measurements and using the map in Fig. 2.

The total heat load in this study is not the total heat load for the utilities, as the heat loss is discarded here. This approach is used since otherwise, heat loss would also have to be included as an area in the hierarchy, and a forecasting model would have to be created to predict heat loss. It can be difficult to accurately predict heat loss as it varies depending on the pipe's supply temperature and flow and the pipes' surrounding temperature. Hence, it was decided not to consider this in this study for simplicity. Therefore, the total heat load in this study is only the aggregated load from all areas.

2.1. Operational heat load forecasts

Both *Brønderslev Forsyning* and *Fjernvarme Fyn* use forecast with an hourly resolution for several forecast horizons, which are used to optimise their operations. A commercial forecast provider delivers the hourly heat load forecasts to both utilities. The HeatFor™[1] solution provides the heat load forecasts. The heat load forecast, $\hat{Y}^{\mathrm{i}}_{t+k|t}$, is updated every hour for k steps in advance for each group i, as shown below,

$$\left\{\hat{Y}^{\mathrm{i}}_{t+k|t}, t = 1,\dots,T,\ k = 1,\dots,K, i = \mathrm{total}, \mathrm{area}_1,\dots,\mathrm{area}_{\mathrm{g}}\right\}. \tag{2}$$

State-of-the-art operational forecasts are referred to throughout the text as *operational* base forecasts. This work aims to improve the accuracy of the one-step to 24-step forecast. Thus, the forecast horizon of interest for improving hourly forecast accuracy is $k = 1, 2, \dots, 24$.

2.2. Numerical weather prediction

The numerical weather predictions (NWPs) used as input to the forecast models were provided by the MetFor™[2]; i.e. by the same commercial forecast provider as for the heat load. The NWPs consist of climate variables with an hourly resolution, updated every hour and available for forecasting heat load $\hat{Y}^{\mathrm{i}}_{t+k|t}$.

An example of an NWP for the kth prediction horizon is the predicted ambient temperature [in °C] denoted by

$$\left\{T^{\mathrm{a,NWP}}_{t+k|t}, t = 1,\dots,T,\ k = 1,\dots,K\right\}. \tag{3}$$

The NWPs are used to produce the heat load forecasts required in addition to the *operational* base forecast.

3. Methods

This section introduces the methods used for generating the forecasts for the study. In Section 3.1, the forecasting methodology used in creating the base forecasts is presented, which are needed additionally with the *operational* base forecasts. A grey-box method is used to generate the base forecasts, which will be referred to as *simple* base forecast throughout the text. Section 3.2 presents the methods of hierarchical forecasting and the linear constraints they impose on spatial aggregation. Lastly, Section 3.3 defines the reconciliation process, which is done when the base forecast does not satisfy the coherency constraints by the hierarchy.

[1] https://enfor.dk/services/heatfor/

[2] https://enfor.dk/services/metfor/

3.1. Heat load base forecast

Since the *operational* base forecast is not available for each level of aggregation, additional forecasts for heat load are needed. The forecasting method used here is a linear regression model where the coefficients are estimated using adaptive RLS with a forgetting factor, where past observations are exponentially weighted downwards as suggested by Ljung and Söderström [36]. This forecasting model has shown promising results for several energy forecast studies. For instance, Bacher et al. [37] and Bacher et al. [38] use the proposed method to predict the electricity generation from PV systems and heat load for single-family homes, Bergsteinsson et al. [12] to predict the heat load for a district heating system to be used in temporal hierarchical forecasting, Rasmussen et al. [39] to predict the electrical load for supermarket refrigeration, and Jónsson et al. [40] to predict electricity spot prices.

The forecasts are made using the R- package, onlineforecast [41]. The package provides an ideal forecasting framework for heat load forecasting as it provides the tools to deal with the non-stationarity and non-linearity of the heat load time series. Thus, account for the time-varying dynamics and the non-linear relationship between the load and input variables, such as NWP and social behaviour. The forecasting methodology of the package is only briefly introduced here. For a more detailed introduction, see Bacher et al. [41] and Bergsteinsson et al. [42]. The package is based on a regression model that models the output variable as a linear combination of the input variables. However, it also contains a possibility for mapping the input variables to handle non-stationarities and non-linearities of the output and input variables. The method consists of a two-stage modelling procedure as proposed in [39,43]. This procedure consists of a *Transformation Stage* and a *Regression Stage*. In the *Transformation Stage*, the input variables are transformed either directly by a function or non-parametrically, e.g. using splines. A linear model is then created using the transformed data to predict the heat load. Subsequently, in the *Regression Stage*, the model's coefficients are estimated recursively using the RLS method. Hence, the onlineforecast package can handle the non-linearity and non-stationarity using the proposed two-stage modelling procedure with the RLS method to estimate the coefficients.

3.2. Hierarchical time series forecasting

Hierarchies for time series define the structure from the most granulated series to the most aggregated. Hierarchies are divided into three types; temporal hierarchies (see e.g. [44]), spatial hierarchies, in literature sometimes called cross-sectional hierarchies (see e.g. [45]), and spatio-temporal hierarchies which combines the two [46].

In *spatial* hierarchies, the structure is related to either spatial aggregation or some grouping aggregation, e.g. Yang et al. [47] uses a geographical hierarchy to make PV forecasts coherent across the whole energy system from the inverters on the PV, subsystems, PV plants and all the way up to the transmission zones. All these different groups require individual base forecasts to operate their system and are therefore not necessarily coherent. They argue that if the forecasts are constrained to be coherent, then the decision-makers in the electricity grid can improve their planning of the grid due to the intermittent individual power injection to the grid. Hence, a more detailed and coherent overview of when and where the power will enter the grid is obtained.

Temporal hierarchies consider aggregation in the time domain. An example is described in Bergsteinsson et al. [12] reconciles temporal hierarchies forecasts to make different heat load resolution forecasts coherent. This allows for better alignment of the decision-making for production scheduling which has a higher temporal resolution (e.g. day ahead), while temperature optimisation needs heat load forecasts on lower resolution (e.g. hourly). Yagli et al. [48] demonstrates a spatio-temporal hierarchy of PV systems, where the focus is on simultaneously

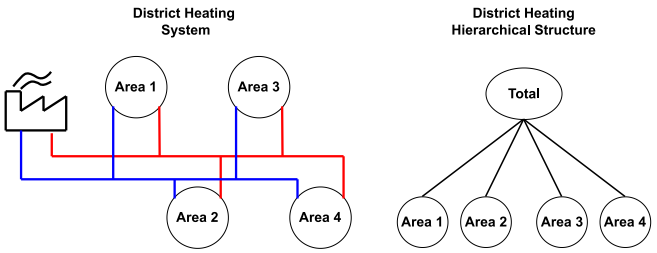


Fig. 3. Example of a simple spatial hierarchy structure for heat load forecasting.

making the forecasts spatially and temporally coherent. Petropoulos et al. [49] discuss this more in detail and refer to multiple studies that use spatial, temporal or both in their research.

This study considers spatial hierarchies of the areas within district heating networks. In these systems, the total heat consumption should be coherent with the area forecasts, i.e. the aggregation of area forecasts should equal the forecast total consumption. For example, a district heating system that has one production plant, one transmission line and four distribution systems. Assuming no heat loss, the heat produced at the production plant is the total heat consumption, and the four distribution systems are the bottom areas. Fig. 3 illustrates this type of district heating system on the left and shows the corresponding hierarchy on the right. Thus, the total of the four areas should aggregate to the total load.

A summation matrix $\boldsymbol{S}$ is used to describe the structure of a hierarchy. The general definition of the summation matrix for any balanced hierarchy structure (as shown in Fig. 4) is given in Nystrup et al. [29] as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{I}_{m/\ell_1} \otimes 1_{\ell_1}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{I}_{m/\ell_L} \otimes 1_{\ell_L}^{\mathrm{T}} \end{bmatrix} \tag{4}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{I}_{m/\ell}$ is an identity matrix of order m/ℓ, and 1_ℓ is an $\ell-$vector of ones. The aggregation levels are a factor of m, which is the sampling frequency of the lowest level. In the example above, $\ell_1 = m$, $\ell_L = 1$, and m/ℓ is the number of observations at the aggregation level ℓ. Using the hierarchy structure in Fig. 3 to illustrate this, the hierarchy has aggregation levels $\ell_1 = 4$ and $\ell_2 = 1$ with $m = 4$ and the dimension of the base forecasts in the structure is $n = 5$. This results in the following summation matrix corresponding to Fig. 3

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

It is also possible to add additional layers to the hierarchy. This adds to the complexity of the problem, but assuming that the forecast errors are not too highly correlated, then the added information should improve the accuracy of the reconciled forecasts [31]. Fig. 4 shows the hierarchy from Fig. 3 with an additional layer.

Here a layer has been added between the area forecasts and the total forecast, which aggregates the area forecasts in pairs; however non-symmetrical hierarchies are equally valid. Adding a layer in this manner results in the following summation matrix corresponding to Fig. 4.

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

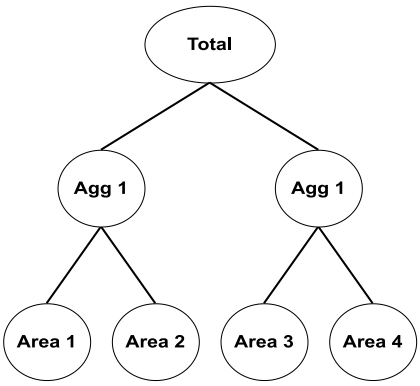


Fig. 4. Example of hierarchy with added aggregation layer between area and total forecasts.

This requires $m = 4$ bottom-level forecasts and $n = 7$ base forecasts.

3.3. Reconciling forecasts

The reconciliation process describes the transformation of individual time series to be coherent according to the defined hierarchical structure. In relation to heat load forecasts, it results in making them coherent such that the aggregation of the lower levels match the higher levels and vice versa. Hence, the process is defined by linear constraints. Thus, for the reconciliation process, a projection matrix is needed that projects the individual base forecast to a coherent subspace as defined by the linear constraints [50]. For example, making independent base forecast reconciled using the bottom-up method,

$$\tilde{Y} = \boldsymbol{SG}\hat{Y}, \tag{7}$$

$$\boldsymbol{G} = \left[0_{m\times(n-m)}|I_m\right], \tag{8}$$

where $\tilde{Y}$ is the reconciled forecast, $\boldsymbol{G}$ is a extraction matrix of order $m \times n$, which extracts the m bottom-level forecasts, $\hat{Y}$ is the base forecasts. Here, $\boldsymbol{G}$ extracts the bottom forecasts and aggregates them up according to the summation matrix and the $\boldsymbol{SG}$ is the projection matrix in the reconciliation process. The $\boldsymbol{G}$ matrix is the mapping of the base forecasts to the bottom-level reconciled forecasts therefore the accuracy of the reconciled forecasts depends on $\boldsymbol{G}$. Simply choosing $\boldsymbol{G}$ according to the bottom-up method is inefficient as it disregards all information in the higher levels.

Therefore, a more optimal approach is needed where all information in the hierarchy is shared between the forecasts using the $\boldsymbol{G}$ and hence the projection matrix to reduce the error of the reconciled forecast. Hyndman et al. [11] propose a regression approach to estimate the mapping matrix $\boldsymbol{G}$ using generalised least square estimation, minimising the coherency errors, i.e. the error between the base forecast and the reconciled forecast, subject to the coherency constraint. In this formulation, the base forecasts are then written in regression form,

$$\hat{Y}_{t+k|t} = \boldsymbol{S}\beta(k) + \epsilon(k), \tag{9}$$

where $\beta(k) = \mathbf{E}[Y_{\ell,t+k}|Y = y_1, \ldots, y_t]$ is the unknown conditional mean of the future values of the most granular observed series, i.e. the reconciled forecasts. The $\epsilon(k)$ represents the error between the base forecasts and their expected value, the coherency error $\hat{Y} - \tilde{Y}$. The error $\epsilon(k)$ is assumed to have zero mean and covariance matrix, Σ. Hence, the generalised least squares estimation of $\beta(k)$ in Eq. (9). If Σ is assumed to be known and the base forecasts are unbiased, the reconciled forecasts can be estimated by

$$\tilde{y} = S(S^{\mathrm{T}}\Sigma^{-1}S)^{-1}S^{\mathrm{T}}\Sigma^{-1}\hat{y}, \tag{10}$$

where the matrix $\boldsymbol{G} = (S^{\mathrm{T}}\Sigma^{-1}S)^{-1}S^{\mathrm{T}}\Sigma^{-1}$. The issue with this approach is that the covariance of the coherency errors Σ is not identifiable, as shown by Wickramasuriya et al. [13]. Consequently, multiple authors have suggested possible alternatives. Hyndman et al. [11] argued that Σ would be difficult to estimate and replaced it with an identity matrix, thus placing equal weights on all base forecasts. Hyndman et al. [45] proposed estimating Σ by using weighted least squares where the variance of the one-step ahead base forecast is used in place of Σ. Athanasopoulos et al. [44] proposed three different structures of the estimator based on the in-sample base forecast errors. However, all of the estimators disregard variance between groups and levels in the hierarchy. Wickramasuriya et al. [13] introduce what they call the minimum trace (MinT) reconciliation, which uses the full variance–covariance matrix of the base forecast errors. In Nystrup et al. [29], it is also proposed to use the full variance–covariance matrix to maximise the accuracy potential of the reconciliation process. While other methods have since been proposed, e.g. the Combined Conditional Coherent forecasts (CCC) by Hollyman et al. [32], the MinT has seemingly become somewhat of a go-to method for forecast reconciliation. Hence, the full covariance of the base forecast errors will be used for the reconciliation process in this study.

Operational forecasts often need to be re-estimated multiple times per day according to the operation it is needed for. Therefore, the reconciled forecast also needs to be re-estimated with the same frequency as the base forecast. This requires fast computations such that the forecasts are available when they are needed. For example, temperature optimisation in district heating networks is usually done on an hourly resolution based on new heat load observations for the past hour. The data needs to be sent between servers and needs to be quality-checked before it is used. Input variables also need to be available as soon as possible, but they also take time. Then the base forecast needs to be updated and used e.g., for recalculation of the optimal set point of the temperature optimisation. Hence, the computation of the reconciliation forecasts needs to be fast.

To ensure rapid computations Bergsteinsson et al. [12] propose using the exponential smoothing method to estimate the covariance,

$$\hat{\Sigma}_{\mathrm{t}} = \lambda\hat{\Sigma}_{\mathrm{t}-1} + (1-\lambda)e_{\mathrm{t}}e_t^{\mathrm{T}}. \tag{11}$$

where e_{t} is the newest base forecast error at time t, λ is the forgetting factor and $\hat{\Sigma}_{\mathrm{t}-1}$ is the previously estimated covariance matrix. Hence, updating the estimator with new information is quick, and only the previous estimator and current error need to be stored. This method also makes the estimator adaptive due to the forgetting factor where past information is exponentially down-weighted, i.e. newer observations have more influence on the estimation. This is useful, for instance, when the system is non-stationary, and forecasts need to adapt quickly to changes that occur. We therefore propose to estimate the covariance estimator using the exponential smoothing method, as the forecasts in this paper are used operationally and are updated on an hourly basis to handle the non-stationary heat load.

The accuracy improvements achieved using the proposed reconciliation process will be demonstrated using the Relative Root Mean Squared Error (RRMSE),

$$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\mathrm{RMSE}^{\mathrm{base}}} - 1, \tag{12}$$

where $\mathrm{RMSE}^{\mathrm{base}}$ is the Root Mean Squared Error (RMSE) of the base forecasts, and RMSE is of the reconciled forecast. The result demonstrates either improvement or decline in performance compared to the base forecast, where negative values correspond to improvements in accuracy over the base forecast. The RRMSE is frequently used to compare forecasts between different methods due to its interpretability of the relative measure [51].

4. Results

This section presents the results of applying the reconciliation process to the heat load forecasts for the spatial hierarchy. Section 4.1

Table 1
Estimates of the offline coefficients for the total heat load forecast at *Brønderslev Forsyning*.

a_T	a_W	a_G	λ
0.937	0.816	0.980	0.992

presents an example of the *simple* base forecast model and demonstrates its performance. Section 4.2 shows the optimisation of the forgetting factor for the covariance estimator. The forgetting factor must be optimised over an in-sample period to achieve optimal improvement in accuracy over the out-of-sample period. In Section 4.3 the improvements for the case study *Brønderslev Forsyning* are presented. Furthermore, in Section 4.3.1, the difference in accuracy for the reconciliation process when using the one-step prediction errors versus using the corresponding prediction horizon for each horizon for the estimation of the covariance matrix is investigated. Finally, Section 4.4 presents the improvements in *Fjernvarme Fyn* and demonstrates the difference in accuracy when using different hierarchy structures.

4.1. Base forecast

The model identification and validation will follow the steps described in Bacher et al. [38], find the optimal model by extending the model by investigating if there are any missing dynamics left by residual analysis of the one-step error as well as investigating if the errors and other inputs are correlated using the cross-correlation function (CCF) for each modelling step. The CCF is used to identify any remaining dynamics that an input variable can explain or if different transformations for the current inputs can be used to improve the model. The models are thereby constructed by using the forward selection principle, i.e. adding new inputs sequentially and examining which input had the lowest error score during the scoring-period period. The same approach will be applied here to find the optimal models, and using the Root Mean Squared Error (RMSE) will be used to compare the performance of models. However, this process will not be shown in this work as explaining the process is tedious. Similar models have been proposed. For instance, Bergsteinsson et al. [12] uses the same method to establish a forecasting model for each temporal level.

An example of a model created for this study is the total forecast in *Brønderslev Forsyning*. The training period is from 2020-02-01 to 2020-05-01, with the first month used as a "burn-in" period, i.e., discarded when calculating the error score. The scoring period is then 2020-03-01 to 2020-05-01. The final model is,

$$\begin{aligned}\hat{Y}_{t+k|t} = \theta_{0,k} + \theta_{1,k} Y_t + \theta_{2,k} H_{a_T}(q) T^{\text{a,NWP}}_{t+k|t} + \\ \theta_{3,k} H_{a_W}(q) W^{\text{a,NWP}}_{t+k|t} + \theta_{4,k} H_{a_G}(q) G^{\text{a,NWP}}_{t+k|t} + \\ \theta_{5,k} T^{\text{a,NWP}}_{t+k|t} + \theta_{6,k} W^{\text{a,NWP}}_{t+k|t} + \theta_{5,k} G^{\text{a,NWP}}_{t+k|t},\end{aligned} \tag{13}$$

where the filters are of the from

$$H_a(q) = \frac{1-a}{1-aq^{-1}}, \tag{14}$$

with the backward shift operator ($[q^{-1}]$). The offline coefficients (i.e. the constants for the model which are estimated in the in-sample period) of the model are: the time constants for ambient air temperature ($[a_T]$), wind speed ($[a_W]$), global radiation ($[a_G]$) and the forgetting factor ($[\lambda]$). The off-line coefficients are estimated during the in-sample period from 01/02/2020 to 01/05/ 2020, removing the first month in the scoring period, i.e. the first month is treated as a burn-in period, while the other errors are used to calculate the RMSE, i.e. the scoring period. The forecast horizons used for the estimation of the offline coefficients are $k = \{3, 6, 12, 18, 24\}$. The offline coefficient estimates are presented in Table 1. The θ coefficients are estimated recursively as new observations become available, see Bacher et al. [37] for further details.

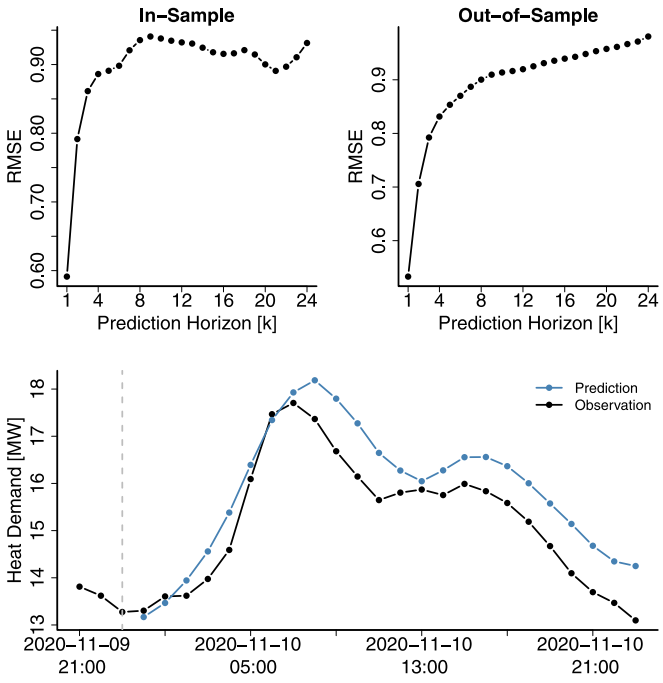


Fig. 5. The top plots shows the RMSE for the in-sample period on the left and the out-of-sample period on the right for each horizon. The bottom plot shows one forecasting realisation for the total heat load forecast from one to 24 step-ahead created at 2020-11-09 23:00, as highlighted with a vertical dashed grey line.

The RMSE for each horizon is shown in Fig. 5. The top left and right plots show the performance of the model for the in-sample and out-of-sample periods. The estimation of the offline coefficients shows that they are valid estimates, as the RMSE for the out-of-sample period is similar to that for the in-sample period and shows no significant increase in error. The bottom graph in Fig. 5 shows a realisation of the model prediction to demonstrate the performance of the model prediction. The grey line in the plot indicates when the prediction was made, at time t = 2020-11-09 23:00 for the next 24 h.

4.2. Optimisation of hyperparameters

The forgetting factor λ is used to update the empirical covariance estimator and needs to be optimised, as discussed in Section 3.3, to achieve the optimal accuracy improvements. The forgetting factor is determined by minimising the RMSE of the reconciliation forecasts for the total and the areas in the hierarchy during the training period, between 2020-10-31 to 2021-03-01. It is estimated either as one optimal forgetting factor for all horizons by minimising the RMSE for all horizons or as a single optimal forgetting factor for each horizon by minimising the RMSE for each horizon. The value of the forgetting factor may vary for the different horizons, as shown in Nielsen and Madsen [26]. It is therefore important to investigate whether the optimal forgetting factor is different for each horizon or whether a global forgetting factor is sufficient. In this study, horizons from one to 24 h ahead are investigated. Therefore, the covariance estimator will have either one forgetting factor or 24 forgetting factors.

The result of the investigation of the forgetting factor for the *Brønderslev Forsyning* utility is shown in Fig. 6. First, two cases are examined: (1) The *operational* base forecast is used for every level in the hierarchy (case one). The results for case one are shown in the left plots, (2) The *operational* base forecast is used for three areas only, and the total forecast uses the *simple* base forecast from the forecast model in Eq. (13) (case two). The results for case two are shown in the right plots. The top plots illustrate the performance of the reconciliation

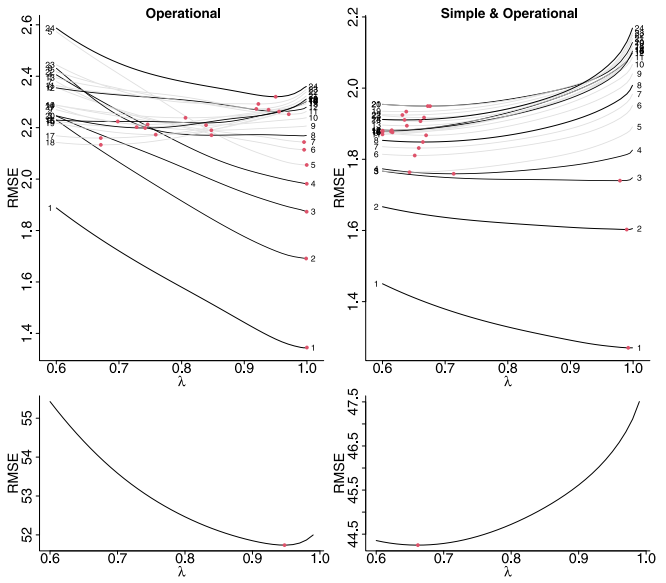


Fig. 6. Figure shows RMSE of the prediction errors versus the forgetting factor for each horizon in the top plots while the bottom plots show it for all horizons for Brønderslev. The shorter the prediction horizon, the longer memory is needed; as the horizon increases, the memory decreases. The red dots show the optimal λ, i.e. lowest RMSE for each horizon. Some of the horizons are coloured grey to reduce too many black lines in the same space.

Table 2
The optimal forgetting factors for each horizon and the optimal forgetting factors for all horizons for both *operationa* and *simple* base forecast are shown in the table with the corresponding RMSE value. The values in the parenthesis are the result of using a global forgetting factor. Notice that the forgetting factor and RMSE for the *simple* base forecast are significantly lower than for the *operational*.

Horizon	*Operational*		*Operational* & *Simple*	
[k]	Forgetting factor [λ]	RMSE	Forgetting factor [λ]	RMSE
1	0.9999 (0.9475)	1.345 (1.38)	0.9922 (0.6617)	1.27 (1.403)
2	0.9987 (0.9475)	1.691 (1.723)	0.9898 (0.6617)	1.603 (1.646)
3	0.9995 (0.9475)	1.873 (1.908)	0.979 (0.6617)	1.74 (1.754)
4	0.9997 (0.9475)	1.982 (2.008)	0.7136 (0.6617)	1.759 (1.762)
5	0.9996 (0.9475)	2.055 (2.08)	0.6431 (0.6617)	1.764 (1.764)
6	0.9955 (0.9475)	2.114 (2.131)	0.6515 (0.6617)	1.811 (1.81)
7	0.9953 (0.9475)	2.144 (2.158)	0.6579 (0.6617)	1.832 (1.831)
8	0.9999 (0.9475)	2.17 (2.172)	0.6645 (0.6617)	1.849 (1.848)
9	0.8475 (0.9475)	2.19 (2.199)	0.6696 (0.6617)	1.867 (1.867)
10	0.839 (0.9475)	2.21 (2.226)	0.616 (0.6617)	1.876 (1.877)
11	0.971 (0.9475)	2.252 (2.254)	0.6 (0.6617)	1.87 (1.876)
12	0.9999 (0.9475)	2.28 (2.263)	0.6148 (0.6617)	1.882 (1.885)
13	0.9999 (0.9475)	2.293 (2.27)	0.6389 (0.6617)	1.894 (1.895)
14	0.8062 (0.9475)	2.238 (2.262)	0.6 (0.6617)	1.88 (1.886)
15	0.7282 (0.9475)	2.202 (2.264)	0.6 (0.6617)	1.878 (1.883)
16	0.8363 (0.9475)	2.218 (2.258)	0.6 (0.6617)	1.878 (1.884)
17	0.671 (0.9475)	2.16 (2.257)	0.6136 (0.6617)	1.878 (1.881)
18	0.671 (0.9475)	2.133 (2.25)	0.6607 (0.6617)	1.907 (1.907)
19	0.7587 (0.9475)	2.173 (2.249)	0.638 (0.6617)	1.934 (1.934)
20	0.7418 (0.9475)	2.199 (2.256)	0.6723 (0.6617)	1.949 (1.949)
21	0.7458 (0.9475)	2.213 (2.27)	0.6759 (0.6617)	1.949 (1.949)
22	0.9192 (0.9475)	2.273 (2.279)	0.6318 (0.6617)	1.924 (1.925)
23	0.8679 (0.9475)	2.293 (2.295)	0.6663 (0.6617)	1.917 (1.917)
24	0.9501 (0.9475)	2.32 (2.32)	0.6352 (0.6617)	1.911 (1.911)

forecasts using a covariance estimator that is estimated for each horizon with a forgetting factor that was optimised for each horizon, while the bottom plots illustrate the performance using a single global forgetting factor for all horizons for the covariance estimation. To investigate the effect of the forgetting factor on the reconciliation accuracy performance, the profile of the RMSE is calculated for different forgetting factors and visualised in Fig. 6. The forgetting factor on the interval $[0.6, 0.99]$ with a step of 0.01 was calculated. Also, two additional forgetting factors were added for the profile calculation, 0.995 and 0.999. The optimal value was then optimised using `nlminb()` in R [52], shown as a red dot for each horizon.

The results show that for the short horizons, the forgetting factors are found to be optimal with a high forgetting factor where the RMSE is minimised, as can be seen in the upper plots in Fig. 6. For instance, the first seven horizons for the first case and the first three for the second case, where the *simple* base forecasts are used for the total. At higher horizons, the forgetting factor is optimal for both cases with a rather small memory; however, there is a rather flat curve, i.e. the optimum is not very well defined, especially for the second case. For example, a forgetting factor of 0.8 translates to effective memory of $N_{\text{eff}} = 5$. This is quite a small number of effective observations that can be used to estimate the empirical covariance matrix, which could be prone to very large prediction errors.

When comparing the optimal factors between the two cases, it can be seen that the RMSE is significantly lower when using the author's base forecast for the total load. This can be seen in more detail in Table 2, where the optimal forgetting factor and corresponding RMSE are shown for both optimisations, a single forgetting factor for each horizon and one for all horizons which are shown inside the parenthesis. So, these results suggest that finding the optimal forgetting factor for each horizon should lead to greater accuracy improvements than using a single forgetting factor. At the very least, use different forgetting factors for the lower and higher horizons, as they tend to cluster together, and the RMSE curves are quite flat around their optimal points.

4.3. Accuracy improvements in brønderslev

Using the forgetting factor found in Section 4.2 to adaptively update the covariance estimator for the reconciliation process to produce reconciled forecasts as new information becomes available. The RMSE of the base and the reconciliation forecasts for the two cases are calculated for the total area level and the RRMSE to demonstrate the accuracy performance for each horizon in percentage. The results for the *Brønderslev Forsyning* demo case are shown in Fig. 7. The accuracy improvements by using forgetting factors optimised for each horizon or a single forgetting factor for all horizons are also shown in the plots. The two upper plots show the result for the in-sample period (2020-10-31 to 2021-03-01), while the two lower plots show the result for the out-of-sample period (2021-03-01 to 2022-05-01). The lower plots for in-sample and out-of-sample plots show the improvement in accuracy between the base forecast and the reconciled forecast using the RRMSE score. Fig. A.17 shows the same plots for the bottom three areas.

Case two, where the *simple* base forecast was used for the total area in the hierarchy, shows significant improvements compared to case one, where the *operational* base forecast was used at all levels. Case one shows a lower accuracy for the reconciliation forecast compared to the base forecast for almost all areas and horizons, the longer horizon however demonstrates a slight improvement in accuracy, as can be seen in Fig. 7 and Fig. A.17. We can also see that using the optimal forgetting factor for each horizon results in greater improvements for the first four steps ahead than the single forgetting factor. Using a single forgetting factor for all horizons usually results in a worse accuracy than the base forecast for the one-step horizon. After the first horizon, the improvements in accuracy are similar to the other horizons. This is consistent with the result of optimising the forgetting factor in Fig. 6. For example, in case two, the forgetting factor for the first three horizons was high, and the single forgetting factor was low and similar to the forgetting factor for the other horizons.

RRMSE is between −0.02 and −0.2 which is an accuracy improvement for the state-of-the-art forecast of 2%–20% (Fig. 9, bottom row), further measured in terms of RMSE the relative error is improved from around 0.9 to 0.7 (Fig. 9, third row). S These results become even more apparent when calculating the cumulative sum of squared errors for the base forecast and the reconciled forecast with the optimal forgetting factor for each horizon in the total heat load. This is shown in Fig. 8 for

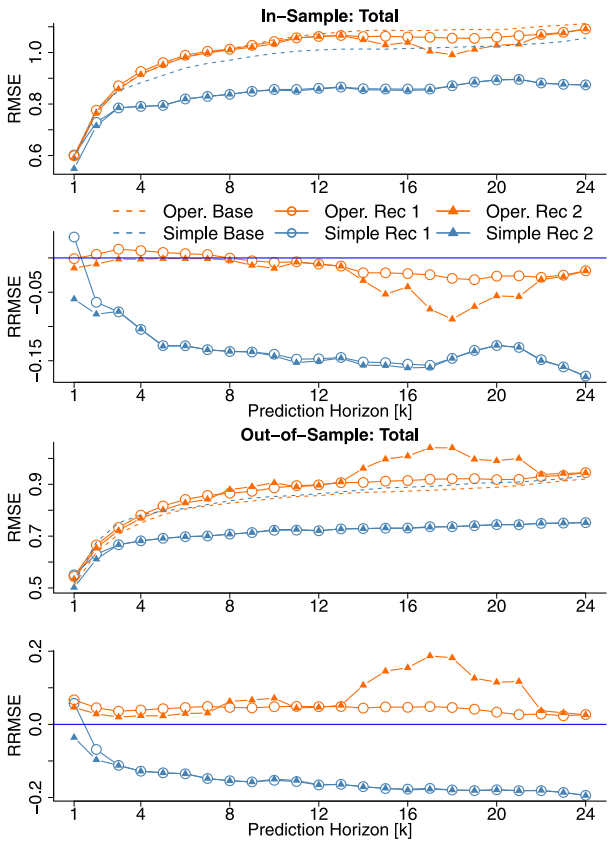


Fig. 7. The upper plots in the in-sample (2020-10-31 to 2021-03-01) or out-of-sample (2021-03-01 to 2022-05-01) period illustrate the RMSE of the total forecast error for the two cases (1) operational for all levels, (2) *simple* at the top and *operational* at the bottom. Case 1 is shown in orange, and case 2 is in black. The base forecast errors for both cases are marked with a dashed line. Also, the use of either one forgetting factor for each horizon with a line and dot and a unique forgetting factor for each horizon with a line and a triangle.

one step and 24 steps ahead for both the in-sample and out-of-sample forecasts. This shows a significant improvement in the accuracy using the proposed method. It can also be observed that the gains in accuracy are most significant in the colder periods. In contrast, in the warm period (summer), the error slope is similar (flat) when comparing the base and reconciled errors.

The correlation matrices shown in Fig. 9 are calculated from the covariance estimator at the time shown in the plots from the 24-ahead forecast errors for both cases, using only the *operational* forecast (plots on the left) and the *operational* and *simple* forecasts (the plots on the right) as base forecast. The upper and lower plots show instances from the summer and winter periods. Only using the *operational* forecast results in a higher correlation, while using the author forecast results in a lower correlation. This could indicate that using forecasts produced with the same model leads to too similar errors and therefore cannot share any useful information in the reconciliation process. Similar results are found and discussed in Nystrup et al. [31].

Fig. 10 shows an example of the realisation of the prediction for both the base and reconciled forecast of the one to 24 steps ahead. Each plot shows a different area, e.g. the top left plot shows the total load. The top plot shows both the base forecast of *operational* and *simple* base forecast and the corresponding reconciled forecast with a unique forgetting factor for each horizon to estimate the covariance matrix. From these plots, it is difficult to tell which forecast performs best; however, the base forecasts for the total heat load are quite different.

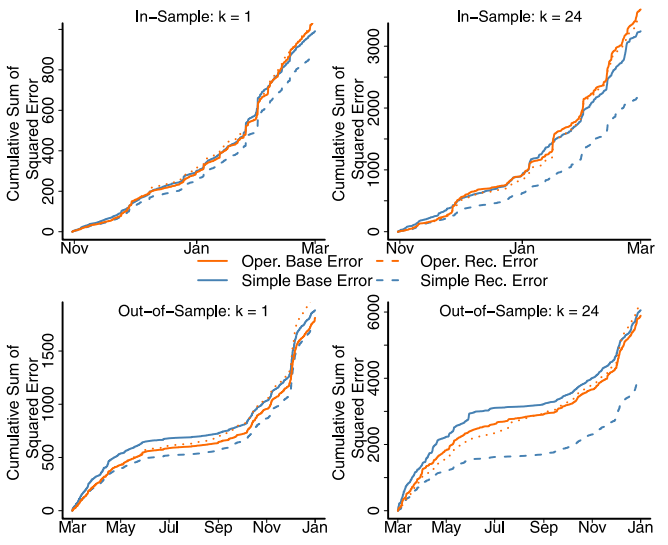


Fig. 8. Figure shows the cumulative sum of squared error for the base and reconciled forecast of the total heat load where the reconciled forecast estimated using optimal forgetting for each horizon using the authors base forecast of the total aggregation.

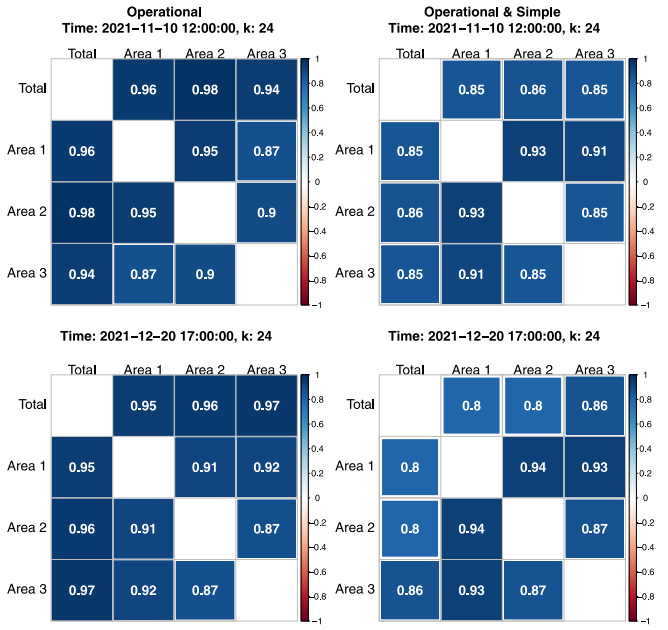


Fig. 9. Figure illustrates the correlation between the areas when using only *operational* base forecast in the left plots and using *operational* and *simple* forecast as the base in the right plots. The plots show correlation instances at periods in summer in the top plots and lower plots in winter.

4.3.1. *One-step ahead empirical covariance matrix*

In the literature, one-step errors are usually used to tune models, i.e. to estimate the model's coefficients. These estimates are then used for multi-step predictions. However, this can lead to sub-optimal results because the correlation between the output and input variables can change depending on the horizon. The same is true for the reconciliation forecast, as the empirical covariance matrix is estimated using forecast errors. The accuracy improvements from the reconciled process result from the base forecast errors from the recursively estimated empirical covariance matrix. Therefore, having a covariance estimator for each horizon might be beneficial. We used an empirical covariance matrix for each horizon calculated from the corresponding k-step prediction error in the previous section. However, we would like to

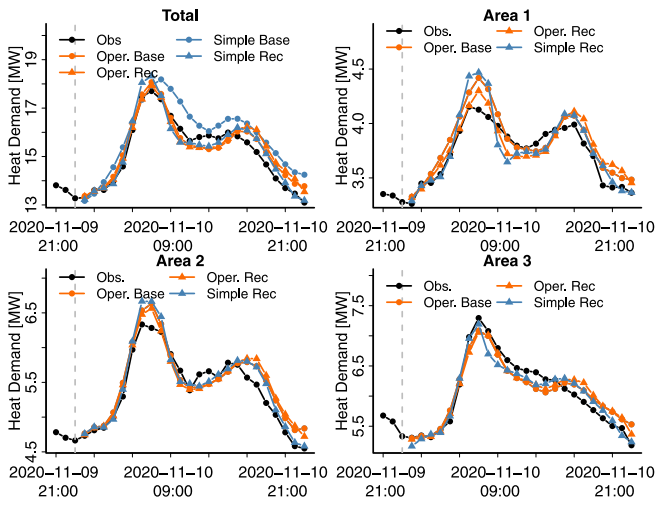


Fig. 10. Realisations of base and reconciled forecast for each area in the *Brønderslev Forsyning* hierarchy.

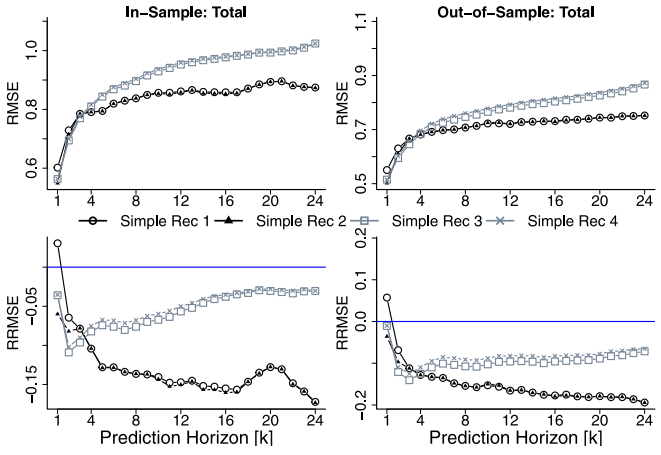


Fig. 11. The figure shows the difference in accuracy between using errors from all prediction horizons and using errors only one step ahead when estimating the empirical covariance matrix. *Simple Rec 1* and *Simple Rec 2* show the results of using errors from each prediction horizon when a single forgetting factor is optimised for all horizons or each horizon has a unique forgetting factor. *Simple Rec 3* and *Simple Rec 4* show the result of optimising the single forgetting factor for all horizons or each horizon has a unique forgetting factor using only the error from the one step ahead to estimate the covariance matrix for all horizons. The base forecasts from case two were used here.

investigate whether there is a significant difference in accuracy when we use either the one-step error for all horizons or the k-step error for the covariance estimator. Here we will only use case 2 with the *simple* total forecast and the *operation* forecast for the bottom areas.

The results are shown in Fig. 11, which shows the accuracy improvements when using the one-step or multi-step prediction errors to estimate the covariance matrix. The top plots show the RMSE score, and the bottom plot shows the RRMSE, while the left and right plots show the in-sample and out-of-sample results. For simplicity, only the results for the total area are shown. *Simple Rec 1* is the result of using errors from all horizons and only one forgetting factor, while *Simple Rec 2* uses errors from all horizons and has an optimal forgetting factor for each horizon. *Simple Rec 3* and *Simple Rec 4* show the same, but only using the base error from the one-step prediction with one forgetting factor or optimal for each horizon. The improvements using only the errors from the one-step forecast are significantly lower than those using multi-step prediction errors. You can see the optimised forgetting factors for each horizon and the forgetting factor for all horizons in Table A.3.

4.4. Accuracy improvements in FYN and hierarchy structure investigation

The current hierarchy at Fjernevarme Fyn is the total and the 12 different areas however due to the high number of areas and how they are linked together as seen in Fig. 2. They are quite grouped together, with some areas located far from the central point. Hence, it gives the opportunity to create more aggregation levels to extend the hierarchy structure and hopefully further enhance the accuracy improvements. An hourly average of the heat load for each weekday for the four seasons is shown in Fig. A.16. The four heating seasons in Denmark are listed below,

1. ***Winter*** *(December, January, February, March)* during cold periods when the heating demand is high due to the high consumption of space heating to keep the indoor climate comfortable.
2. *Spring (April, May)* during the transition period from cold to warm with the influence of solar irradiance of warming houses, thus lowering the amount of space heating needed.
3. *Summer (June, July, August, September)* when space heating is usually not needed in Denmark, only domestic hot water is needed (e.g. hot tap water and showering).
4. *Fall (October, November)* during the transition period when space heating is required again due to lowering ambient air temperatures.

Comparing the areas' heating dynamics together makes it evident that each area has unique heat dynamics; however, there are some possibilities to group some of them depending on their behaviour from Fig. A.16 when comparing their shape. For instance, Areas 6, 8 and 12 exhibit similar shapes. This analysis can give an idea of which areas should be aggregated together. Based on these results, new aggregate levels have been added. Three new aggregations have been added, which aggregated four areas together. Aggregation 1 is the aggregation of { Area 1, Area 2, Area 3, Area 4 }, Aggregation 2 is {Area 5, Area 9, Area 10, Area 11 }, and Aggregation 3 is {Area 6, Area 7, Area 8, Area 12} based on from the result in Fig. A.16 and the layout in Fig. 2.

The base forecast for the new aggregation level is then created (*simple*) while still using the *operational* base forecast for the total and the 12 areas. The forecasting model used for the new aggregation level is similar to the model for the total for *Brønderslev Forsyning* as shown in Eq. (13). The prediction horizon will also be the same, one- to 24-steps ahead. We will use both hierarchies, a total of 12 areas and a total with three aggregations of the bottom 12 areas, to compare the results using a simple hierarchy and a hierarchy with an additional aggregation level. Thus, investigate when more information is added to the hierarchy and how it will affect the accuracy improvements of the reconciled forecasts. The hyperparameters are optimised as done in Section 4.2 during the in-sample period (2019-10-31 to 2020-04-01). The result can be found in Table A.4 for both hierarchy structures.

The accuracy improvements for the total aggregation are shown in Fig. 12 where it shows that the reconciled forecast for both cases improves the accuracy in almost all cases, especially in the longer horizons as was the case for *Brønderslev Forsyning* study. The reconciled forecasts with *Oper. Rec 1* and *Oper. Agg Rec 1* shows the reconciled forecast where the same forgetting factor is optimised for all horizons while *Oper. Rec 2* and *Oper. Agg Rec 2* shows the result for optimising the forgetting factor for each horizon. *Oper. Rec* is where the simple hierarchy is used and *Oper. Agg Rec* is when the additional aggregation level has been added to the hierarchy. In the top plots, the RMSE of the forecasts are shown, and in the bottom, the accuracy improvements as the RRMSE. The left and right plots show the result in the in-sample (2019-10-31 to 2020-04-01) and out-of-sample (2020-04-01 to 2021-12-01) periods.

The results in the plots in Fig. 12 demonstrate that the reconciled forecast for both hierarchy structure cases improves the accuracy of the base forecasts. The first three horizons needed to have their own forgetting factor, the same as we saw for the *Brønderslev Forsyning*

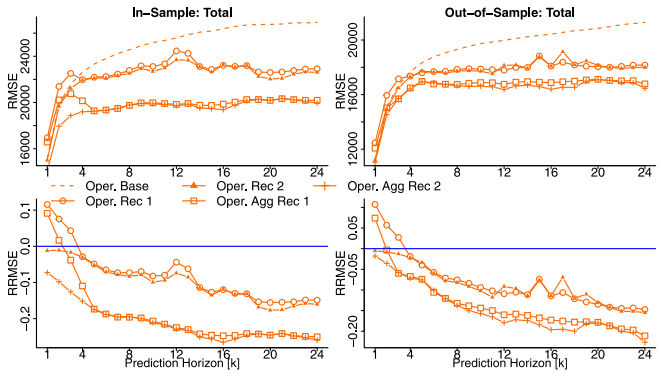


Fig. 12. The plots show the accuracy improvements of the base forecast in *Fjernvarme Fyn*, where two different hierarchies are investigated. The top plots show the RMSE, and the bottom plots show the improvements using the RRMSE, while the left and right plots show the result in the in-sample and out-of-sample periods.

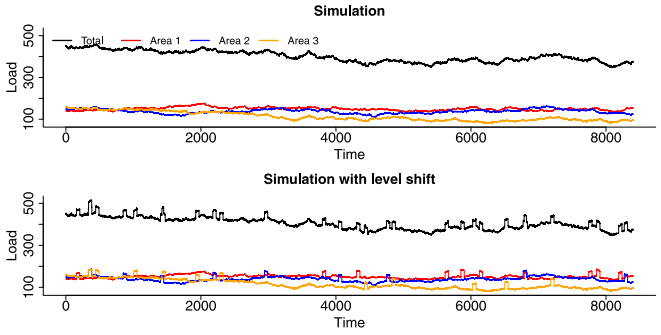


Fig. 13. Simulated heat load measurements. Top: raw simulated measurements from AR(1) models. Bottom: simulated measurements with levels shifts.

case. The longer horizon otherwise improves the accuracy for both setups. Adding an additional aggregation level to the hierarchy structure demonstrates significant accuracy improvement, almost double the improvement. Again, an improvement of the state-of-the-art operation forecast in the range of 2% to 25%.

5. Simulation study

To show the effect of the forgetting factors in a controlled environment, a simulation study is performed. The setup for the simulation study will consist of three areas and a total of one year of hourly measurements each. The simulated load from the three areas are produced by simple Auto-Regressive (AR) models of lag one with different offsets.

$$Y_{\text{Area1,t}} = 0.7Y_{\text{Area1,t-1}} + 140 + \epsilon_{1,t}, \tag{15}$$

$$Y_{\text{Area2,t}} = 0.8Y_{\text{Area2,t-1}} + 150 + \epsilon_{2,t}, \tag{16}$$

$$Y_{\text{Area3,t}} = 0.9Y_{\text{Area3,t-1}} + 158 + \epsilon_{3,t}, \tag{17}$$

Where total is the aggregation of the areas and $\epsilon_{i,t}$ are white noise

$$Y_{\text{Total,t}} = Y_{\text{Area1,t}} + Y_{\text{Area2,t}} + Y_{\text{Area3,t}}. \tag{18}$$

Level shifts are added to the simulated heat load at random intervals to introduce some additional complexity and simulate spikes in heat load. These are added at randomly sampled time points. At each sampled time point, the load is increased for 48 h by a random amount sampled from the distribution $N(30, 2)$ and rounded to the nearest integer. The resulting simulated heat load is visualised in Fig. 13 where the raw simulation is shown in the top plot while the lower plot shows the simulation with level shifts.

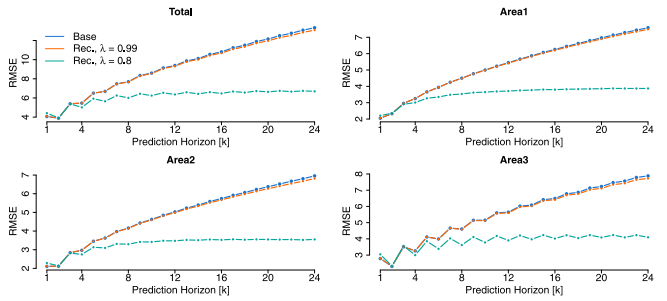


Fig. 14. Accuracy of the base and reconciled forecasts with the two forgetting factors being investigated for the three areas and the total.

To reconcile the spatial hierarchy, base forecasts are, of course, needed. These are again created using the R package onlineforecast, where the model coefficients are estimated using the Ordinary Least Squares method. The models that are constructed are AR(1) processes for both the areas and the total, where the coefficients are re-estimated every time a new observation becomes available. A unique model is created for each prediction horizon. Hence,

$$Y_{\text{Area1,t}} = \phi_{1,t,k} Y_{\text{Area1,t-1}} + \epsilon_{1,t}, \tag{19}$$

$$Y_{\text{Area2,t}} = \phi_{2,t,k} Y_{\text{Area2,t-1}} + \epsilon_{2,t}, \tag{20}$$

$$Y_{\text{Area3,t}} = \phi_{3,t,k} Y_{\text{Area3,t-1}} + \epsilon_{3,t}, \tag{21}$$

$$Y_{\text{Total,t}} = \phi_{4,t,k} Y_{\text{Total,t-1}} + \epsilon_{4,t}. \tag{22}$$

Where the subscript t is the time and k is the prediction horizon. The $\phi_{\text{i,t,k}}$ is the AR(1) coefficient of the model and $\epsilon_{\text{i,t}}$ is the error term.

When reconciling this simulated hierarchy, an investigation is performed into the memory used for updating the empirical covariance matrix when new information is available. This is done by running the reconciliation process twice, once with a short memory and once with a longer memory. The short memory will have the forgetting factor set as $\lambda = 0.8$; thus, the effective memory is the past five time steps, i.e. $N_{\text{eff}} = 5$. The long memory will have the forgetting factor as $\lambda = 0.99$, i.e. $N_{\text{eff}} = 100$. The RMSE for each horizon for the base forecast and the two scenarios of memory for the reconciled forecasts are shown in Fig. 14 for each area. The long memory demonstrates slight improvements compared to the base forecast, while the short memory improves it significantly, except for the first three horizons, where it has lower accuracy.

Fig. 14 shows the accuracy on the one-hour prediction horizon. The long memory with $\lambda = 0.99$ follows the base forecast closely and no clear improvement can be seen. The shorter memory shows a great increase in accuracy across all areas and the total. However mostly for prediction horizons larger than 4. This result indicates that a short memory might be preferable; however this should be investigated further. Hence the one-step and 12-step prediction horizon for the total is examined in Fig. 15. Since the areas and the total all seem to follow the same pattern, only the total will be examined here. Similar plots for the areas can be seen in the appendix. The top rows of Fig. 15 show a segment of the data, the base forecast, and the two reconciled forecasts for prediction horizons $k = 1$ and $k = 12$ respectively. The bottom rows of these figures show the corresponding cumulative sum of squared errors for each forecast. It is clear from Fig. 15 that on the one-hour prediction horizon, the short memory is very volatile, resulting in lower accuracy than the base forecast. However, the longer memory is also unable to improve the base forecast significantly. For the longer 12-hour prediction horizon the shorter memory does significantly outperform the base forecast, as was also seen in Fig. 14. Here again, the longer memory either follows the performance of the base forecast or improves it slightly. These plots also demonstrate that the significant accuracy improvements come from the short memory reconciliation

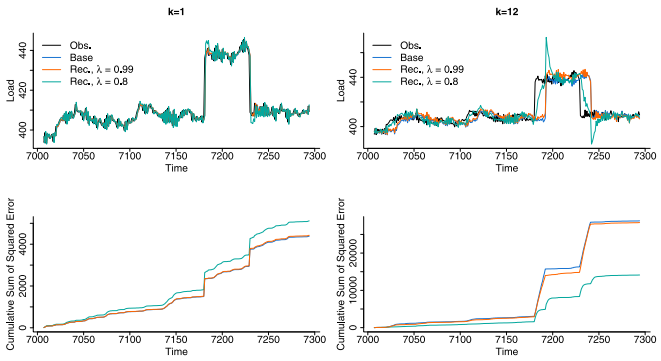


Fig. 15. Top row: Segment of the simulated observations along with base forecasts and reconciled forecasts with memories $\lambda = 0.8$ and $\lambda = 0.99$ for the one-hour prediction horizon for the total. Bottom row: cumulative sum of squared errors for the three forecasts for the total.

better being able to adapt to sudden changes in the data. This is most pronounced at the level shift around 7175 h, which originates in Area 3. The reconciled forecast with the short memory begins to adapt to the level shift shortly after it has occurred, while the longer memory follows the base forecast more closely and is thus slower to adapt. Hence highlighting the power of being able to react quickly to changes in the data. The short memory reconciled forecasts do however show a significant overshoot when changes in the base forecast happen quickly, as seen when the level shift occurs. This is likely due to the weights used in the reconciliation process being estimated based on base forecasts from before the change. It is likely that this could be addressed by improving the base forecasts or tweaking the memory. Improving base forecast to react to changes in data in a more timely manner would result in weights not being so far out of scale. Alternatively, the memory should be tweaked to be a bit longer, thus likely trading off some performance in exchange for a more robust reconciled forecast.

It is clear from this simulation study that choosing a fitting memory for the proposed reconciliation process is vital to ensuring good performance. Too long memory can result in reconciled forecasts not being reactive enough, while too short memory can result in overreacting to changes in the base forecast. As such, it is important to choose the robust forgetting factor based on the data behaviour to ensure both optimal performance and reliability.

6. Discussion

This paper proposes to improve state-of-the-art operational forecast accuracy by using spatial hierarchies with an adaptive covariance matrix in the reconciliation process to produce coherent forecasts for district heating systems. The coherency is introduced by linear constraints defined by the hierarchy, where the lower levels are forced to be aggregated coherently to the upper level. The independent and not necessarily coherent base forecasts are first created and then projected onto the coherent subspace using a projection matrix. The projection matrix is created from the summation matrix of the hierarchy, and the covariance estimator is estimated from the errors of the in-sample base forecasts errors. Due to the non-stationary nature of the heat load, the covariance matrix must be able to adapt as the heat load changes over time. Therefore, adaptive and recursive covariance estimation is performed using exponential smoothing as proposed by Bergsteinsson et al. [12]. Additional forecasts created by the authors were needed to investigate whether the *operational* base forecast was too similar as the accuracy improvements were almost none as discussed in Section 4.3. Also, it was investigated if adding a new level of aggregation to the hierarchy would enhance the accuracy improvements even further, as shown in Section 4.4. A linear regression model was proposed to forecast the additional heat load forecasts as demonstrated in Section 4.1. The two-stage forecasting framework was used, where the input variables are first transformed, and then the coefficients are estimated using RLS with exponential forgetting. This framework is ideal for the inherent non-linearity and non-stationary heat load and for producing robust online forecasts with high accuracy.

We initially defined the spatial hierarchy of district heating with the total as the top level and the individual areas as the lowest levels. This study disregarded heat losses from production to areas for simplicity, as heat losses can be challenging to forecast with high accuracy. The first attempt to make reconciliation forecasts for the *Brønderslev Forsyning* utility showed almost no improvements for the in-sample period and it resulted in worse accuracy than the base forecast in the out-of-sample period. It was found that the similarity of the *operational* base forecasts between levels was too high by investigating the correlation matrix of the errors as shown in Fig. 9. Therefore, no gain in accuracy improvements could be achieved if the correlation between the base forecasts is too high as they did not exchange any useful information, as shown in Nystrup et al. [31]. However, replacing the total forecast with the proposed *simple* base forecast showed high accuracy improvements in both the in-sample and out-of-sample periods, around 15% depending on the horizon. Hence, it is important to not use forecasts generated using the same model and using the same information to gain accuracy improvements when using a covariance matrix estimated from the forecast errors.

We concluded that the forgetting factor needs to be optimised for each horizon, especially for the short horizons, which tend to be very high as demonstrated in Table 2. It was also discovered that the optimum of the forgetting factor for higher horizons is very low and not very well defined. The low forgetting factor uses a small amount of data to estimate the covariance matrix which allows the reconciliation process to react quickly to sudden changes when the base forecast performs poorly. The ability to react quickly to changes, e.g. level shifts, can be useful when sudden and unexpected changes in the heat load occur. This was demonstrated in a simulation study in Section 5 where it was investigated where the accuracy improvements could come from and how the forgetting factor affects the reconciliation process. The base forecast model for forecasting heat load used in this work tends to have a high forgetting factor, which could be the reason why the low forgetting factor leads to such a significant improvement in accuracy as it allows the reconciliation forecasts to adapt quickly when sudden changes occur in the heat load. However, choosing a slightly higher forgetting factor could lead to a more robust estimate and reduce the risk of high prediction errors. It is also argued that a higher forgetting factor enables more robust prediction by reducing overshooting/undershooting of the prediction by having too low a forgetting factor.

The errors used in the covariance estimation were investigated. In literature, the one-step errors are frequently used to estimate coefficients of forecasting models, which are then used for forecasting multiple-step horizons ahead. It could be more beneficial to use a specific forecast model for each k-step ahead forecast and use the appropriate errors to estimate the coefficients. We demonstrated that it is important to use the k-step errors when estimating the covariance estimator to predict the k-step ahead heat load by comparing the accuracy between only using the one-step errors for all horizons when estimating the covariance estimator. The accuracy improvement is almost double when using the corresponding k-step errors for each k-step forecast compared to only using the one-step errors for all k-step forecasts.

Another case study was carried out using the heat load forecast from the *Fjernevarme Fyn* utility to validate the results of the case study of *Brønderslev Forsyning*. It showed similar accuracy improvements and highlighted the importance of finding the optimal forgetting factor for each horizon. We also showed that adding an additional level of aggregation by adding *simple* base forecasts between the bottom and top aggregation levels to the hierarchy significantly improved the accuracy.

Thus, adding more information to the voting process shows that more knowledge is shared between levels about the covariance matrix.

We therefore conclude that spatial hierarchies improve the accuracy of state-of-the-art operational heat load forecasting, as demonstrated in two different case studies with different hierarchy structures. We demonstrated accuracy improvements ranging from 2% to 20% depending on the forecasting horizon and system. Improving operational state-of-the-art forecasts by these amounts is significant for the district heating operator, especially for temperature and production optimisation. This will be important for future energy systems as an improved heat load forecast increases the flexibility potential of district heating by providing more accurate information for decision-making. Furthermore, a spatial coherent forecast for district heating operations is important for temperature, and production optimisation as district heating systems are becoming more decentralised with multiple areas with local heat units or TES systems. We also believe that selecting a robust forgetting factor that reduces the probability of large forecast errors will be important. Too similar forecasts could be problematic in not improving the accuracy, leading to worse accuracy than the base forecast. Hence, it could be essential to have different forecasting models from different methodologies for other areas to achieve higher forecasting accuracy, this should be investigated in more detail in future works. The influence of having a forecast of the heat loss in the network will also need to be analysed. In this work, we discarded the heat loss for simplicity however if a robust and accurate method is found to predict the heat loss, it should have positive effects on the accuracy improvements. There are several possible directions for future research, but a more detailed investigation of where these significant accuracy improvements come from at such a low forgetting factor would be essential before this can be used operationally. Also, different methods for adaptivity when estimating the covariance matrix for the reconciliation than the proposed forgetting factor method in this work could be investigated further, i.e. how to update the covariance matrix when new errors from the base forecast are available.

7. Conclusion

We propose a novel method to increase the accuracy of operational state-of-the-art heat load forecasts by exploiting information between different areas through the spatial hierarchy. Forecasts for different areas in district heating are essential for efficient zonal temperature control, as each area optimising its supply temperature needs the future heat load as input to find optimal set points. The more accurate the future heat load input is, the better the temperature optimisation, resulting in higher savings for the utility. The improved heat load forecasts were computed using a reconciliation process where individual base forecasts are forced to be coherent using the predefined hierarchy structure, and the information is exchanged between them using the proposed covariance estimator. It is proposed to estimate the covariance estimator recursively and make it adaptive using the exponential smoothing formula to handle the non-stationary nature of the heat load. Based on two case studies, it was shown that the proposed method significantly increases the accuracy compared to operational state-of-the-art heat load forecasts for all horizons used in this work, ranging from 2% to 20% improvements. This is highly desired by district heating utilities as future systems will be more decentralised and therefore more local heat load forecasts will be needed. Improving the accuracy of the local heat load forecast will improve the operation of the district heating by optimising the network's temperature levels and production.

It was concluded that the covariance estimator needs to be estimated for each forecast horizon to achieve the highest accuracy improvements for each horizon. Therefore, using information from the same horizon will be more optimal than using only the one-step-ahead prediction errors, as is the tendency to do in forecasting. It was also found that optimising the forgetting factor for each horizon is crucial for covariance estimation to achieve optimal improvement in accuracy. Optimal forgetting factors were found to be quite small compared to what is commonly used in research for forecasts, usually around 0.65 compared to the usual 0.999 for forecasting models. However, a higher forgetting factor of around 0.99 was found for shorter horizons. A low forgetting factor allows the reconciliation process to shift the weights when the errors of the base forecast start to increase, e.g. due to level shifts in the heat load. We argue that it might be reasonable to select a higher forgetting factor than found through optimisation. Higher forgetting will lead to a more robust estimate, as a low forgetting factor might be prone to large prediction errors.

Even more improvements in accuracy were demonstrated by adding an additional level of aggregation in the hierarchy. This provides more information to share in the hierarchy, leading to higher improvements than a hierarchy with fewer levels of aggregation. The definition of the aggregation levels for the spatial hierarchy is trivial compared to a more straightforward hierarchy structure for the temporal hierarchy. An investigation of the optimal hierarchy structure for the spatial hierarchy could therefore be explored in future research.

CRediT authorship contribution statement

Hjörleifur G. Bergsteinsson: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mikkel Lindstrøm Sørensen:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Jan Kloppenborg Møller:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Henrik Madsen:** Funding acquisition, Supervision, Writing – review & editing.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work is funded by Innovation Fund Denmark through the projects Heat 4.0 (8090-00046B), SEM4Cities (0143-0004), TOP-UP (9045-00017 A), Flexible Energy Denmark (8090-00069B), DynFlex (Mission Green Fuels), PTXHeatUtilisation (Energy Cluster Denmark, GreenLab), ARV (EU H2020 101036723), ELEXIA (EU HE-101075656), and the Norwegian FME-ZEN project funded by the ZEN partners and the Research Council of Norway (257660). We thank the energy forecasting provider ENFOR, especially Torben Skov Nielsen, for his support and supplying the forecast and information. We would also like to thank *Fjernvarme Fyn* and *Brønderslev Forsyning* for making their heat load data available.

Appendix

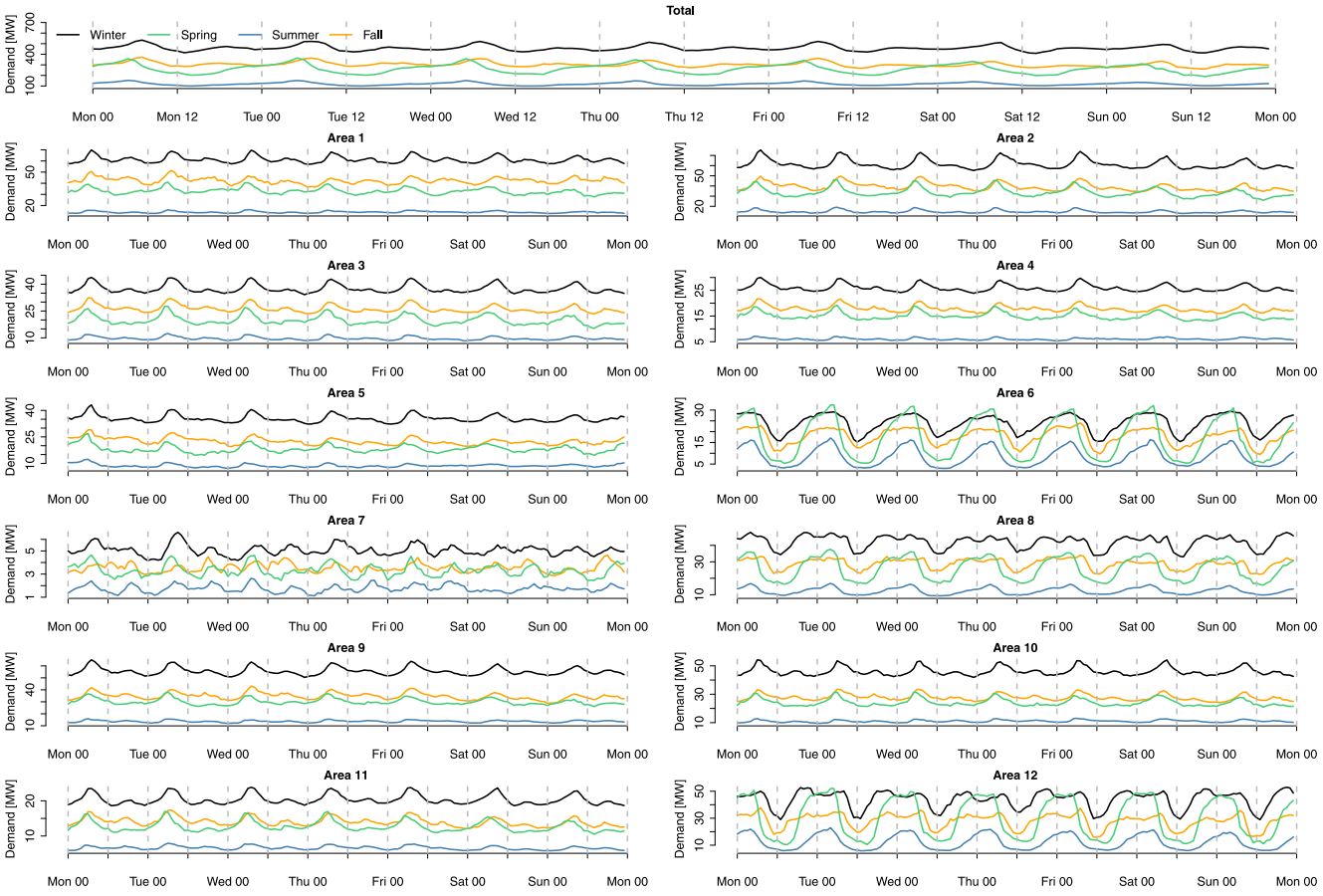


Fig. A.16. The weekday seasonality for the areas at *Fjernvarme Fyn* is visualised here by splitting it up into four seasons; Winter (December, January, February, March), Spring (April, May), Summer (June, July, August, September), and Fall (October, November).

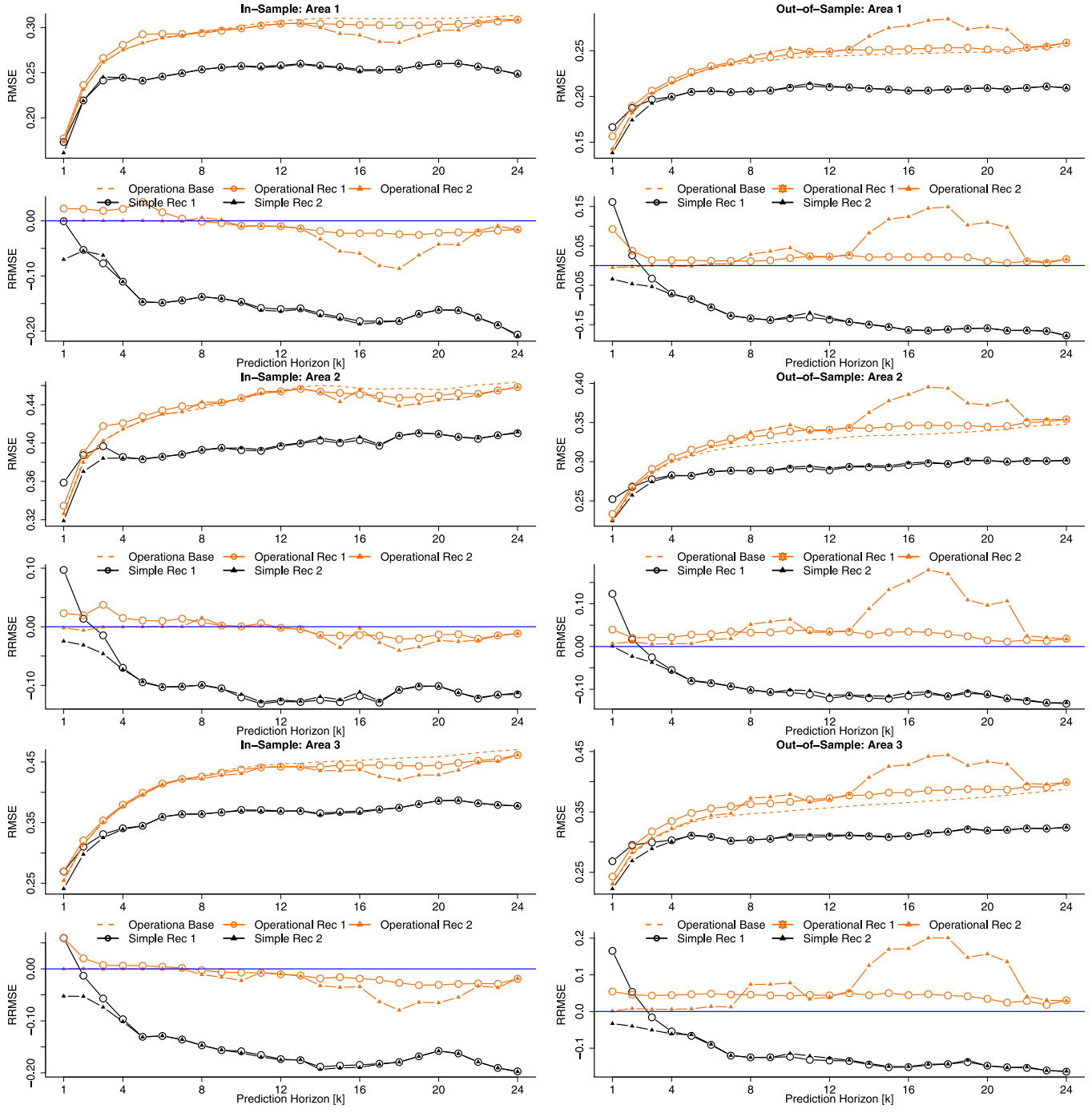


Fig. A.17. Figure illustrates accuracy improvements of the reconciled forecast using aggregation hierarchy for in-sample and out-of-sample for all areas and total.

Table A.3
The optimal forgetting factors for each horizon and the optimal forgetting factors for all horizons for using only one-step ahead errors to estimate the empirical covariance matrix from Section 4.3.1. Notice how the forgetting factor increases again for higher prediction horizons.

Horizon [k]	*Operationa* & *Simple*	
	Forgetting factor [λ]	RMSE
1	0.999	1.2702
2	0.7149	1.5505
3	0.7803	1.7078
4	0.8059	1.7925
5	0.8259	1.8616
6	0.7917	1.9041
7	0.7824	1.9231
8	0.7838	1.9527
9	0.7785	1.9884
10	0.7765	2.0125
11	0.7829	2.0358
12	0.7955	2.0608
13	0.8184	2.0781
14	0.8353	2.0932
15	0.8644	2.1037
16	0.9958	2.1101
17	0.9961	2.1153
18	0.9954	2.1254
19	0.9955	2.1353
20	0.9948	2.1353
21	0.9931	2.1445
22	0.9908	2.1538
23	0.9913	2.1595
24	0.9945	2.1728

Table A.4
The optimal forgetting factors for each horizon and the optimal forgetting factors for all horizons for both *operational* and *simple* base forecast are shown in the table with the corresponding RMSE value for *Fjernvarme Fyn* in Section 4.4. Notice that the forgetting factor and RMSE for the *simple* base forecast are significantly lower than for the *operational*. It is not possible to compare the RMSE between the two case studies in this table as *operational* Agg has another aggregation level, thus a higher RMSE.

Horizon [k]	*Operational*		*Operational* Agg	
	Forgetting factor [λ]	RMSE	Forgetting factor [λ]	RMSE
1	0.9861	38718.26	0.9878	88435.2
2	0.9767	49146.8	0.9829	97964.3
3	0.9649	52476.19	0.9753	99853.03
4	0.8701	54040.31	0.9542	99052.69
5	0.693	54598.05	0.9095	96424.64
6	0.6851	54702.81	0.8797	94701.16
7	0.6844	54994.56	0.8618	93483.41
8	0.6596	55289.8	0.8615	93578.51
9	0.659	55938.88	0.8426	93094.56
10	0.6204	55599.19	0.853	93088.3
11	0.6	55853.17	0.8456	93184.57
12	0.8662	57567.87	0.8553	93069.96
13	0.8475	57528.59	0.8753	94143.76
14	0.7152	56412.53	0.8699	93725.18
15	0.6539	55621.23	0.8729	93569.79
16	0.7652	56567.24	0.8846	94219.94
17	0.6842	56327.91	0.8688	94575.05
18	0.7358	56720.07	0.8723	95835.68
19	0.6376	55265	0.8962	98122.15
20	0.603	55002.29	0.9227	99076.89
21	0.6	55236.32	0.9224	99671.8
22	0.6633	55841.8	0.7555	106038.7
23	0.6431	55896.72	0.9059	97152.2
24	0.6392	55781.87	0.8975	96414.18

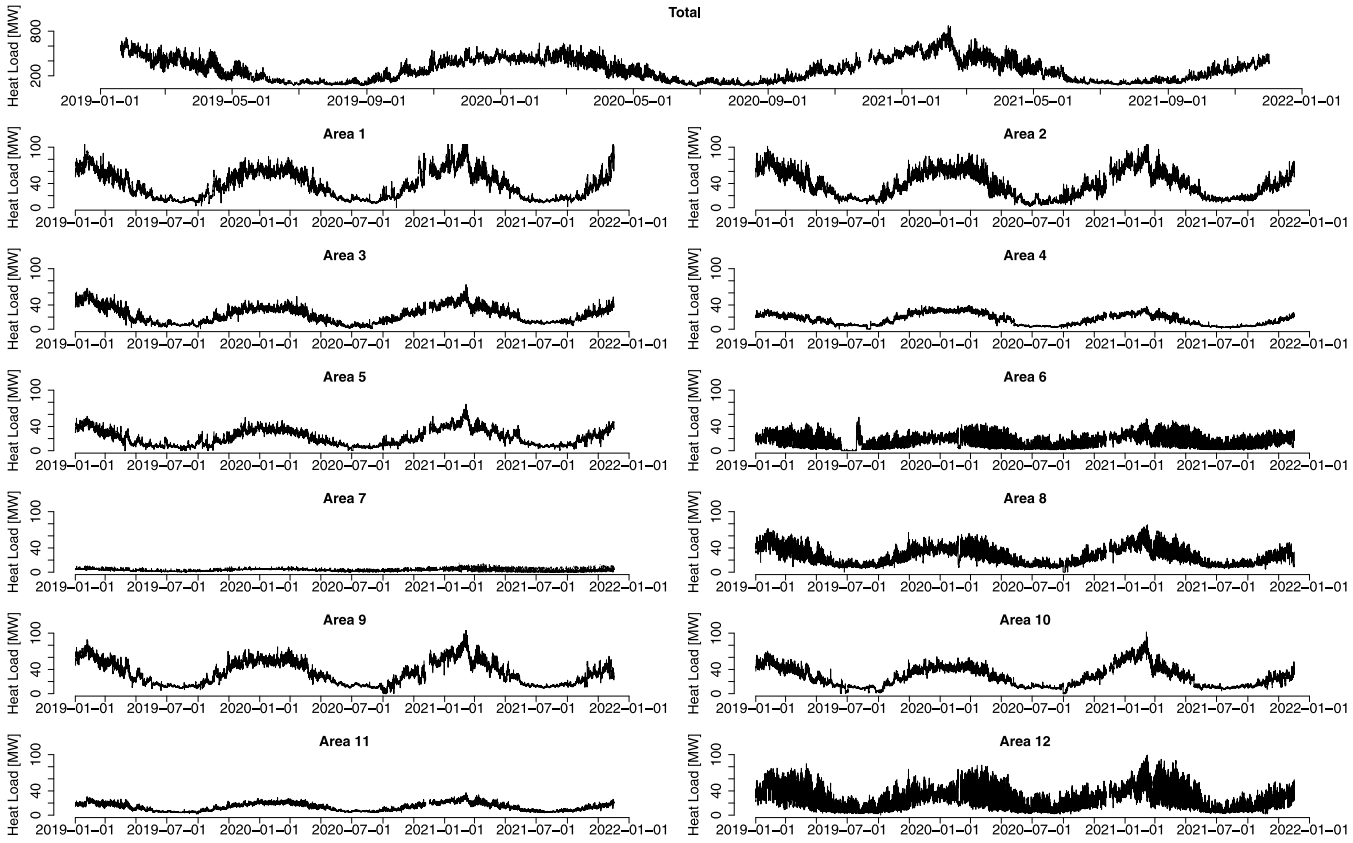


Fig. A.18. Time series plot of the heat load over the two and half year period for each group inside the *Fjernvarme Fyn* system.

References

[1] Vandermeulen A, van der Heijde B, Helsen L. Controlling district heating and cooling networks to unlock flexibility: A review. Energy 2018;151:103–15.

[2] International Energy Agency (IEA). Renewables 2021, analysis and forecasts to 2026. 2021, https://www.iea.org/reports/renewables-2021, EA, Paris, France.

[3] David A, Mathiesen BV, Averfalk H, Werner S, Lund H. Heat roadmap Europe: Large-scale electric heat pumps in district heating systems. Energies 2017;10(4).

[4] Guelpa E, Verda V. Thermal energy storage in district heating and cooling systems: A review. Appl Energy 2019;252:113474.

[5] IRENA. Innovation outlook: thermal energy storage. 2020, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Nov/IRENA_Innovation_Outlook_TES_2020.pdf. International Renewable Energy Agency, Abu Dhabi.

[6] Bloess A, Schill W-P, Zerrahn A. Power-to-heat for renewable energy integration: A review of technologies, modeling approaches, and flexibility potentials. Appl Energy 2018;212:1611–26.

[7] Nielsen TS. Online prediction and control in nonlinear stochastic systems (Ph.D. thesis), Informatics and Mathematical Modelling, Technical University of Denmark, DTU; 2002, URL: http://www2.compute.dtu.dk/pubdb/pubs/792-full.html.

[8] Madsen H, Sejling K, Søgaard HT, Palsson OP. On flow and supply temperature control in district heating systems. Heat Recov Syst CHP 1994;14(6):613–20.

[9] Pinson P, Nielsen TS, Nielsen HA, Poulsen NK, Madsen H. Temperature prediction at critical points in district heating systems. European J Oper Res 2009;194(1):163–76.

[10] Bergsteinsson H, Ben Amer S, Nielsen P, Madsen H. Digitalization of district heating. Technical University of Denmark; 2021.

[11] Hyndman R, Ahmed R, Athanasopoulos G, Shang H. Optimal combination forecasts for hierarchical time series. Comput Statist Data Anal 2011;55(9):2579–89.

[12] Bergsteinsson HG, Møller JK, Nystrup P, Pálsson ÓP, Guericke D, Madsen H. Heat load forecasting using adaptive temporal hierarchies. Appl Energy 2021;292:116872.

[13] Wickramasuriya SL, Athanasopoulos G, Hyndman RJ. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. J Amer Statist Assoc 2019;114(526):804–19.

[14] 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.

[15] Blanco I, Guericke D, Andersen A, Madsen H. Operational planning and bidding for district heating systems with uncertain renewable energy production. Energies 2018;11(12).

[16] Verrilli F, Srinivasan S, Gambino G, Canelli M, Himanka M, Del Vecchio C, et al. Model predictive control-based optimal operations of district heating system with thermal energy storage and flexible loads. IEEE Trans Autom Sci Eng 2016;14(2):547–57.

[17] Nielsen TS, Madsen H. Control of supply temperature in district heating systems. In: Proceedings of the 8th international symposium on district heating and cooling. 2002.

[18] Quaggiotto D, Vivian J, Zarrella A. Management of a district heating network using model predictive control with and without thermal storage. Opt Eng 2021;22(3):1897–919.

[19] Gadd H, Werner S. Daily heat load variations in Swedish district heating systems. Appl Energy 2013;106:47–55.

[20] Nielsen H, Madsen H. Modelling the heat consumption in district heating systems using a grey-box approach. Energy Build 2006;38(1):63–71.

[21] Dotzauer E. Simple model for prediction of loads in district - heating systems. Appl Energy 2002;73(3–4):277–84.

[22] Dahl M, Brun A, Andresen G. Using ensemble weather predictions in district heating operation and load forecasting. Appl Energy 2017;193:455–65.

[23] Grosswindhagera S, Voigtb A, Kozeka M. Online short-term forecast of system heat load in district heating networks. Tsp 2011;1(2).

[24] Dahl M, Brun A, Kirsebom OS, Andresen GB. Improving short-term heat load forecasts with calendar and holiday data. Energies 2018;11(7):1678.

[25] Idowu S, Saguna S, Åhlund C, Schelén O. Applied machine learning: Forecasting heat load in district heating system. Energy Build 2016;133:478–88.

[26] Nielsen H, Madsen H. Predicting the heat consumption in district heating systems using meteorological forecasts. Informatics and Mathematical Modelling, Technical University of Denmark; 2000.

[27] Kato K, Sakawa M, Ishimaru K, Ushiro S, Shibano T. Heat load prediction through recurrent neural network in district heating and cooling systems. In: 2008 IEEE international conference on systems, man and cybernetics. IEEE; 2008, p. 1401–6.

[28] Song J, Zhang L, Xue G, Ma Y, Gao S, Jiang Q. Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model. Energy Build 2021;243:110998.

[29] Nystrup P, Lindström E, Pinson P, Madsen H. Temporal hierarchies with autocorrelation for load forecasting. European J Oper Res 2020;280(3):876–88.

[30] Van Erven T, Cugliari J. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In: Antoniadis A, Poggi J-M, Brossat X, editors. Modeling and stochastic learning for forecasting in high dimensions. Cham: Springer International Publishing; 2015, p. 297–317.
[31] Nystrup P, Lindström E, Mø ller JK, Madsen H. Dimensionality reduction in forecasting with temporal hierarchies. Int J Forecast 2021;37(3):1127–46.
[32] Hollyman R, Petropoulos F, Tipping ME. Understanding forecast reconciliation. European J Oper Res 2021;294(1):149–60.
[33] Sørensen ML, Nystrup P, Bjerregård MB, Møller JK, Bacher P, Madsen H. Recent developments in multivariate wind and solar power forecasting. Wiley Interdiscip Rev Energy Environ 2022;e465.
[34] Jeon J, Panagiotelis A, Petropoulos F. Probabilistic forecast reconciliation with applications to wind power and electric load. European J Oper Res 2019;279(2):364–79.
[35] Arvastson L. Stochastic modeling and operational optimization in district heating systems (Ph.D. thesis), Doctoral theses in mathematical sciences, Centre for Mathematical Sciences, Lund University; 2001.
[36] Ljung L, Söderström T. Theory and practice of recursive identification. The MIT press series in signal processing, optimization, and control, vol. 4, MIT Press; 1983.
[37] Bacher P, Madsen H, Nielsen HA. Online short-term solar power forecasting. Solar Energy 2009;83(10):1772–83.
[38] Bacher P, Madsen H, Nielsen H, Perers B. Short-term heat load forecasting for single family houses. Energy Build 2013;65:101–12.
[39] Rasmussen LB, Bacher P, Madsen H, Nielsen HA, Heerup C, Green T. Load forecasting of supermarket refrigeration. Appl Energy 2016;163:32–40.
[40] Jónsson T, Pinson P, Nielsen HA, Madsen H, Nielsen TS. Forecasting electricity spot prices accounting for wind power predictions. IEEE Trans Sustain Energy 2012;4(1):210–8.
[41] Bacher P, Bergsteinsson HG, Frölke L, Sørensen ML, Lemos-Vinasco J, Liisberg J, et al. Onlineforecast: An R package for adaptive and recursive forecasting. 2021, arXiv preprint arXiv:2109.12915.
[42] Bergsteinsson HG, Møller JK, Thilker CA, Guericke D, Heller A, Nielsen TS, Madsen H. Data-driven methods for efficient operation of district heating systems. In: Handbook of low temperature district heating. Springer; 2022, p. 129–63.
[43] Nielsen T, Madsen H, Nielsen H, Pinson P, Kariniotakis G, Siebert N, et al. Short-term wind power forecasting using advanced statistical methods. In: European wind energy conference and exhibition 2006, EWEC 2006, Vol. 3. 2006, p. 2482–90.
[44] Athanasopoulos G, Hyndman R, Kourentzes N, Petropoulos F. Forecasting with temporal hierarchies. European J Oper Res 2017;262(1):60–74.
[45] Hyndman RJ, Lee AJ, Wang E. Fast computation of reconciled forecasts for hierarchical and grouped time series. Comput Statist Data Anal 2016;97:16–32.
[46] Kourentzes N, Athanasopoulos G. Cross-temporal coherent forecasts for Australian tourism. Ann Tour Res 2019;75:393–409.
[47] Yang D, Quan H, Disfani VR, Liu L. Reconciling solar forecasts: Geographical hierarchy. Sol Energy 2017;146:276–86.
[48] Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Sequential reconciliation. Sol Energy 2019;179:391–7.
[49] Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Taieb SB, et al. Forecasting: theory and practice. Int J Forecast 2022.
[50] Panagiotelis A, Athanasopoulos G, Gamakumara P, Hyndman RJ. Forecast reconciliation: A geometric view with new insights on bias correction. Int J Forecast 2021;37(1):343–59.
[51] Hyndman R, Koehler A. Another look at measures of forecast accuracy. Int J Forecast 2006;22(4):679–88.
[52] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.