Preface

University	Aalborg University
Study Program	Urban, Energy and Environmental Planning
Specialisation	Sustainable Energy Planning and Management
Semester	4th Master Semester
Company	EMD International A/S
Thesis Period	February 2023 - June 2023
Semester Theme	Master's Thesis
Submission Format	Research Article
Student	Klaas Mielck
Supervisors University	Meng Yuan
Supervisors Company	Anders N. Andersen

Summary

The optimal operation of district heating systems (DHS) requires heat load forecasts. Data-driven models are the state-of-the-art approach for heat load forecasting. They use known data to learn the relations between the heat load and influential variables. However, concept drift may occur over time, e.g., changes in user behaviour due to an energy crisis or district heating grid expansion due to increasing demand, which changes the relationships between the heat load and the influential variables. Such changes can affect the accuracy of forecasting models. This could further negatively influence the optimal operation of DHS. Previous research on heat load forecasting in DHS has not analysed different strategies to handle concept drift. This study addresses the research gap by analysing the research questions “How can different learning strategies be applied to data-driven heat load forecasting models to handle concept drift in DHS?”. The article focuses on grid expansions, using data from a Danish DHS.

First, concept drifts are synthetically inserted through different incremental grid expansion magnitudes. Second, different learning strategies, i.e., offline with retraining and online, were examined using linear regression. The models were trained with historical weather forecasts, time features and observed recent heat load values to forecast hourly heat load values for a horizon of one week. The strategies to handle concept drift were evaluated in a data stream environment, where new data arrived sequentially in the temporal order. The performance was measured by the root mean squared error (RMSE), mean absolute percentage errors (MAPE) and the mean error (ME).

The findings indicate that daily retraining and online learning can sufficiently handle concept drift, as they were not sensitive to different magnitudes of concept drift insertion. In contrast, the baseline strategy of offline linear regression without retraining is sensitive to concept drift. However, including observed recent heat load features could improve the robustness of the baseline strategy. Nevertheless, this study recommends frequent retraining of offline models, e.g., daily and online learning to be robust to concept drift. The findings provide guidance to forecast model developers and DHS operators that need to handle declining heat load forecast accuracy due to changes in the DHS.

Contents

Heat Load Forecasting:
Handling Concept Drifts in District Heating Systems

Klaas Mielck (Aalborg University)

May 31, 2023

Abstract

Heat load forecasting is essential for optimising the operation of district heating systems (DHS). Data-driven models are the state-of-the-art approach for heat load forecasting. However, concept drift may occur over time, e.g., changes in user behaviour due to an energy crisis or district heating grid expansion due to increasing demand, which changes the relationships between the heat load and the influential variables. Such changes can affect the accuracy of forecasting models, which could further negatively impact the optimal operation of DHS. This article fills the research gap by investigating the research questions "How can different learning strategies be applied to data-driven heat load forecasting models to handle concept drift in DHS?". The special focus is on grid expansions, using a Danish DHS as a case study. First, concept drifts are synthetically inserted through different incremental grid expansion magnitudes. Second, different learning strategies, i.e., offline with retraining and online, were analysed using linear regression to predict hourly heat load values for a forecast horizon of one week. The results indicate that daily retraining and online learning can sufficiently handle concept drift, as they were robust to different magnitudes of concept drift insertion. In contrast, the baseline strategy of offline linear regression without retraining is sensitive to concept drift. However, including observed recent heat load features could enhance the robustness of the baseline strategy. The findings of this study could guide forecast model developers and DHS operators in securing stable heat load forecast performance when DHS change.

Abbreviations

DHS	District Heating System(s)
CHP	Combined Heat and Power
MAE	Mean Absolute Error
DMSFS	Direct Multi-Step Forecasting Strategy
LR	Linear Regression
RSS	Residual Sum of Squares
SGD	Stochastic Gradient Descent
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error
ME	Mean Error

1 Introduction

The optimal operation of district heating systems requires heat load forecasts (Finkenrath et al., 2022). Data-driven forecasting approaches are state-of-the-art (Ntakolia et al., 2022). They use known data to define a mathematical system description by learning the underlying relations between the heat load and the influential variables (Fumo, 2014).

However, grid expansions, building renovations, or changes in consumer behaviour can lead to changes in the aggregated heat load in reality. For example, the energy crisis in 2022 has demonstrated that consumer behaviour can change significantly and impact the aggregated heat load of DHS. Consequently, the accuracy of data-driven forecast models can decay when they do not incorporate the changes. Inaccuracy can cause non-optimal system operations such as increasing heat generation from peak load boilers (Finkenrath et al., 2022). The problem, when the relation between the input and target variables changes over time, is referred to as *concept drift* (Gama et al., 2014). The general assumption is that the concept drift happens unexpectedly and unpredictably (Gama et al., 2014) in contrast to, for example, trend.

To ensure accurate forecasts even though DHS change and the corresponding forecasting models underlie concept drift, the research question of this study is "How can different learning strategies be applied to data-driven heat load forecasting models to handle concept drift in district heating systems?". Theoretically, the ways in which concept drifts can occur in heat load forecasting in DHS are inexhaustible. Thus, this study focuses on concept drift scenarios that could likely happen in practice - district heating grid expansions. Hourly linear regression models for a forecast horizon of one week are developed. The capability of different offline and online learning strategies to adapt these models to concept drift in DHS is analysed.

1.1 Previous Research

Various studies have analysed different data-driven methods to forecast heat loads in DHS, such as regression (Fang and Lahdelma, 2016; Idowu et al., 2016; Kurek et al., 2021; Rusovs et al., 2021), tree-based (Idowu et al., 2016), ensemble (Finkenrath et al., 2022; Geysen et al., 2018; Wei et al., 2021; Xue et al., 2019), support-vector-machines (Idowu et al., 2016; Wei et al., 2021; Xue et al., 2019) or neural network-based methods (Finkenrath et al., 2022; Idowu et al., 2016; Kurek et al., 2021; Wei et al., 2021; Xue et al., 2019).

However, most studies assume that future data looks similar to past data. The assumption is that a static environment where the relationship between input variables and the heat load as the target variable does not change over time. Consequently, these studies' results do not provide information on how heat forecasting models can obtain their performance under concept drift. This knowledge is of particular importance in practice where DHS change over time.

The literature on heat load forecasting can be grouped into *offline* and *online learning*. In offline learning, models can only be used for forecasting when the training with the entire dataset is completed (Gama et al., 2014). The models are static and not updated. Updating offline models would require retraining with a new batch of data from scratch. Then, the new model replaces the old one. This approach is also referred to as *batch learning* (Read and Zliobaite, 2023).

In contrast to offline learning, online learning processes data sequentially (Gama et al., 2014) and updates are carried out at every point in time with a single, most recent training pair (Read and Zliobaite, 2023). Online learning refers to a subgroup of incremental learning defined as updating the current model using the most recent data in general (Gama et al., 2014).

Only a few studies address the problem of concept drifts in the context of heat load forecasting in DHS and analyse these different learning approaches. Suryanarayana et al. (2018) evaluated their forecasting methods by retraining the models daily to recalibrate them with the most recent data and, therefore, to simulate real-life conditions. The authors emphasise that the simplicity of their linear models allows for quick retraining. Thus, retraining every day in real-time would be a feasible option. However, whether periodic retraining (e.g., every day) is preferable over other approaches to cope with concept drifts is unclear and not further investigated in their study.

Similarly, Potočnik et al. (2021) recommend regular model updates (e.g. monthly) for an efficient forecasting operation in practice. For more dynamic systems, adaptive forecasting methods may be considered. However, these recommendations were not explicitly analysed in their study.

In contrast to offline learning approaches and the corresponding retraining strategies, Provatas et al. (2014) relate their work to the master thesis by Provatas (2014). They address the non-stationary nature of heat load in DHS through online ensemble bagging of Fast Incremental Model Trees with Drift Detection (Ikonomovska et al., 2011), which can detect and adapt to concept drifts. The authors state they are first applying online machine learning to heat load forecasting. They conclude that the algorithm has a strong and robust predictive ability and is efficient in processing heat load stream data and their non-stationary behaviour. However, this study has a few limitations: First, Provatas et al. (2014)'s study lacks an offline benchmark model. Thus, it is unclear if their proposed online model actually performs better than offline approaches such as retraining. Second, their study does not focus on concept drifts, and it is unclear if their used dataset explicitly consists of concept drifts, which offline models could struggle with. Third, the used dataset does not include summer months, where the heat load usually shows different patterns than in winter. The follow-up study focuses on implementing heat load forecasting in an operational environment rather than the previously stated limitations (Johansson et al., 2017).

The study by Grosswindhager et al., 2011 has similar limitations. The authors stress the non-stationary characteristics of heat loads in DHS. They applied Seasonal Autoregressive Integrated Moving Average (SARIMA) combined with Kalman Recursion to update knowledge of the heat load forecasting models each time a new observation comes in. However, the study does not compare offline and online approaches. Thus, it is unclear whether an online approach is actually better than an offline approach to cope with the problem of non-stationarity or concept drift.

Jan (2021) addresses the problem of concept drifts regarding heat load forecasting. A probabilistic forecasting algorithm for one year forecast horizon that adapts to concept drifts is proposed. The author concludes that the algorithm can generate reliable medium- to long-term forecasts under concept drift. Nevertheless, whether the proposed approach is more advisable than simple retraining is unclear. Also, the study's forecast horizon is about forecasting a one-year horizon. Shorter forecast horizons are not investigated, which is essential to optimise the operation in DHS.

In conclusion, most previous studies on heat load forecasting assume a static environment. They ignore the potential impact of concept drift in heat load forecasting models caused by, for example, grid expansions, heat saving measures or changes in consumer behaviour. Even though a few studies have addressed the problem of concept drifts in DHS, none have systematically analysed different strategies to handle them. The analysis and comparison of offline and online learning approaches are lacking.

1.2 Contribution

There are several contributions this study makes: To the best of the author's knowledge, this research systematically compares offline and online learning approaches to handle concept drifts in DHS for the first time. Accordingly, this study proposes a data stream environment which regards the fact that data streams evolve and algorithms must react to the change. A strength of this research is the synthetic insertion of concept drift by simulating different grid expansions. Thus, the different learning approaches are analysed in scenarios where the presence of concept drift is actually known.

In reality, DHS underlies change over time. Therefore, concept drift is inevitable. Consequently, comparing different approaches to handle concept drift is vital to ensure that the forecast quality after deployment is similar to the quality in the research environment. In other words, this study contributes to securing a stable forecast performance over time. It provides valuable and novel information for developers and users that work with DHS and need solutions to handle concept drift in corresponding data-driven forecasting models.

2 Data

This study used three types of datasets. First, a heat load time series of the DHS in Ringkøbing, Denmark, as the forecasting target. Second, heat demand time series from smart meters installed in residential buildings in Aalborg, Denmark, to simulate grid expansions and corresponding concept drifts. Third, historical weather forecast data as input data to train the heat load forecast models. The following explains these three datasets.

2.1 Aggregated Heat Load Data from Danish District Heating System

Aggregated heat load data was obtained from the district heating system in Ringkøbing, Denmark. The town is located on the west coast of Denmark and has around 10000 citizens.

An orange circle marks the location on Figure 1.

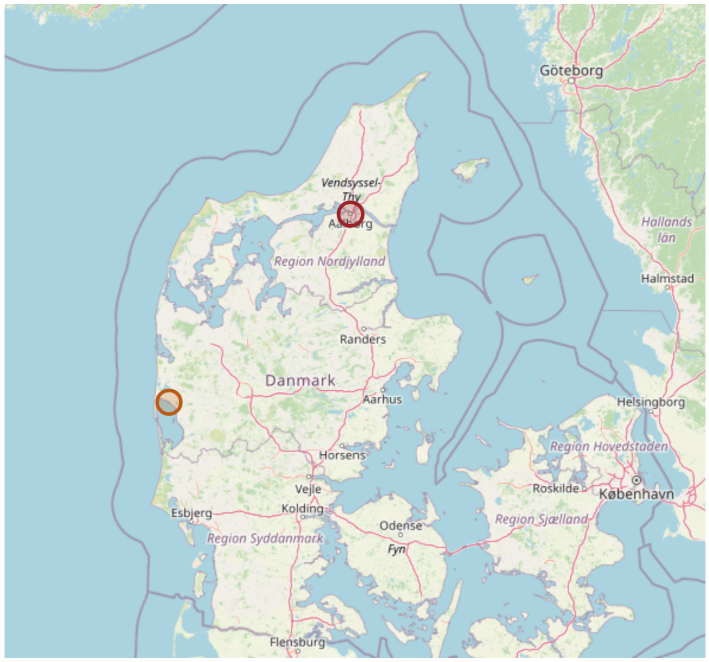

Figure 1: Locations of Aalborg (red marker) and Ringkøbing (orange marker) (Map and data from OpenStreetMap Contributors (2023))

Various heat production units supply approximately 4700 consumers (Ringkøbing Fjernvarme, 2023b). Table 2 shows the heat production units and their capacities. Note that the table indicates the approximate maximal capacity of the energy units. The actual capacities may vary. For example, the heat pump capacity varies due to changing COP values. Figure 2 shows the time series of the aggregated heat load. The mean heat load of the time series is approximately 13 MW. A grid map is available at www.rfv.dk/vores-ledningsnet (Ringkøbing Fjernvarme, 2023a). Live operation data is shown at www.energyweb.dk/rfvv/ (EMD International A/S, 2023).

Unit	Max. Capacity
Solar thermal collector	22 MW (30000 m^2)
Gas combined heat and power (CHP) engine	Heat: 10 MW; Electricity: 9 MW
Heat pump	4 MW
Electric boiler	12 MW
Natural gas boilers	40 MW
Thermal storages	400 MWh

Table 2: Energy units of Ringkøbing's DHS

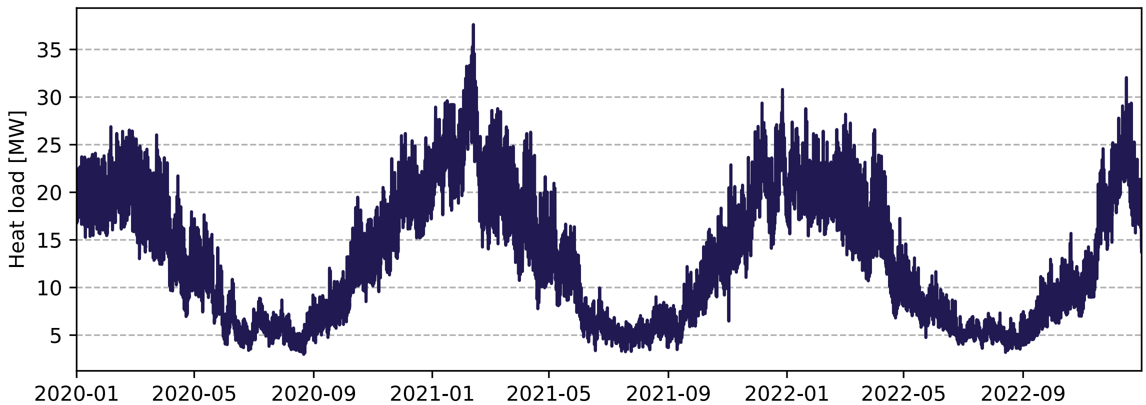

Figure 2: Time series of the aggregated heat load of Ringkøbing‘s DHS

Figure 3 visualises the weekly profiles of the aggregated heat load. The data are grouped and averaged into different weekdays and heating seasons according to the book by Frederiksen and Werner (2013). The figure shows a strong seasonal and daily variation as well as a minor load drop on weekends. The base load and the daily variation of the summer months are significantly smaller than in the winter months.

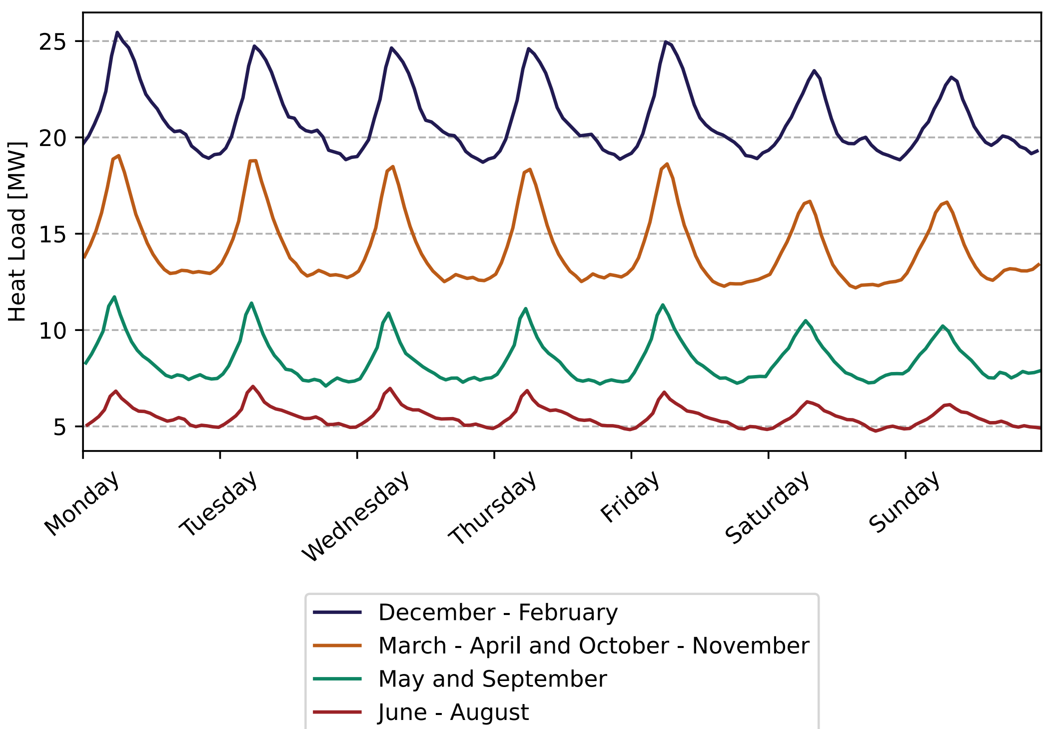

Figure 3: Weekly profiles of heat load data of Ringkøbing‘s DHS grouped into weekdays and heating seasons

The raw time series was cleaned as follows. First, data points were marked as outliers if they were greater than the rolling median plus 4 times the rolling standard deviation of a centred window size of 73 hours, or if they were lower than the rolling median minus 2.5 times the rolling standard deviation of a centred window size of 73 hours.

Second, all gaps were imputed with a rolling median of a centred window size of 7 hours. If the gaps were too large to calculate the rolling median (minimum 3 data points in the rolling window), they were filled with values from the closest time stamp of the same hour and weekday.

2.2 Smart Meter Data from Residential Buildings

To simulate the different grid expansion scenarios (Section 3.1), three years of hourly data from smart meters installed in residential buildings in Aalborg, Denmark (Red circle in Figure 1), were used. The data were screened, interpolated, imputed and published by Schaffer et al. (2022).

The data embraces measurements of heat energy, volume flow at supply, supply and return temperature. From the heat energy, hourly time steps of the heat demand were obtained and used in this study. Based on the Danish *Building and Housing Register (BBR)* (Danish Property Assessment Agency, 2023), the authors classified the buildings into 2460 single-family houses, 474 terraced houses, 88 apartments, 8 non-residential buildings and 97 unclear buildings. This study did not distinguish between these types. The grid expansions were modelled by randomly selecting buildings from the underlying distribution. The data were deemed appropriate to model grid expansions since they are from the same country as the data from the DHS in Ringkøbing. Due to the relatively small distance between the locations (Figure 1), the datasets were assumed to reveal similar demand patterns due to similar climate conditions and social behaviour.

Figure 4 shows the mean heat demand of the 3127 different smart heat meters. Figure 5 illustrates the corresponding weekly profiles grouped and averaged into weekdays and heating seasons similar to Figure 3. The average heat load of the winter seasons decrease slightly from 2019 to the end of the winter of 2020. Furthermore, the data reveal morning peaks similar to the aggregated heat load of the DHS in Ringkøbing. As opposed to this, the smart meter data from residential buildings in Aalborg show small evening peaks.

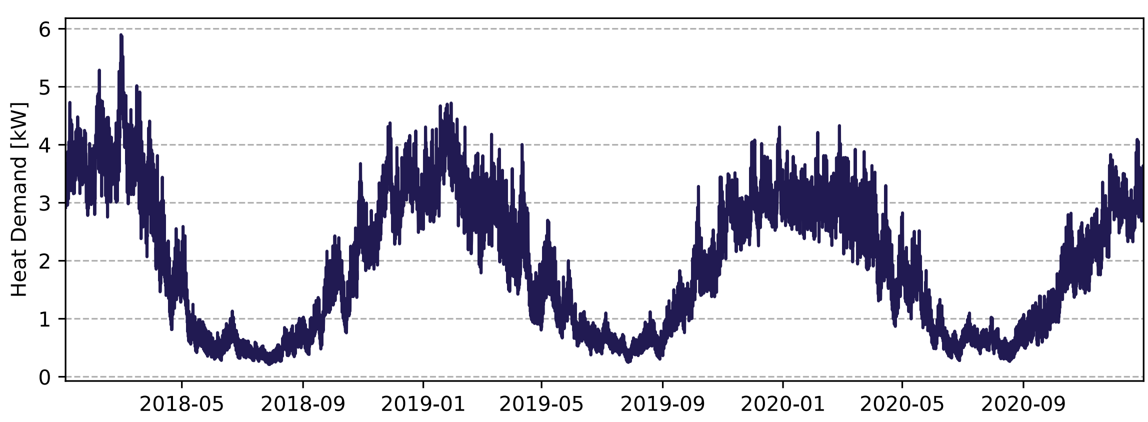

Figure 4: Mean heat demand time series of 3127 smart meters installed in residential buildings in Aalborg, Denmark

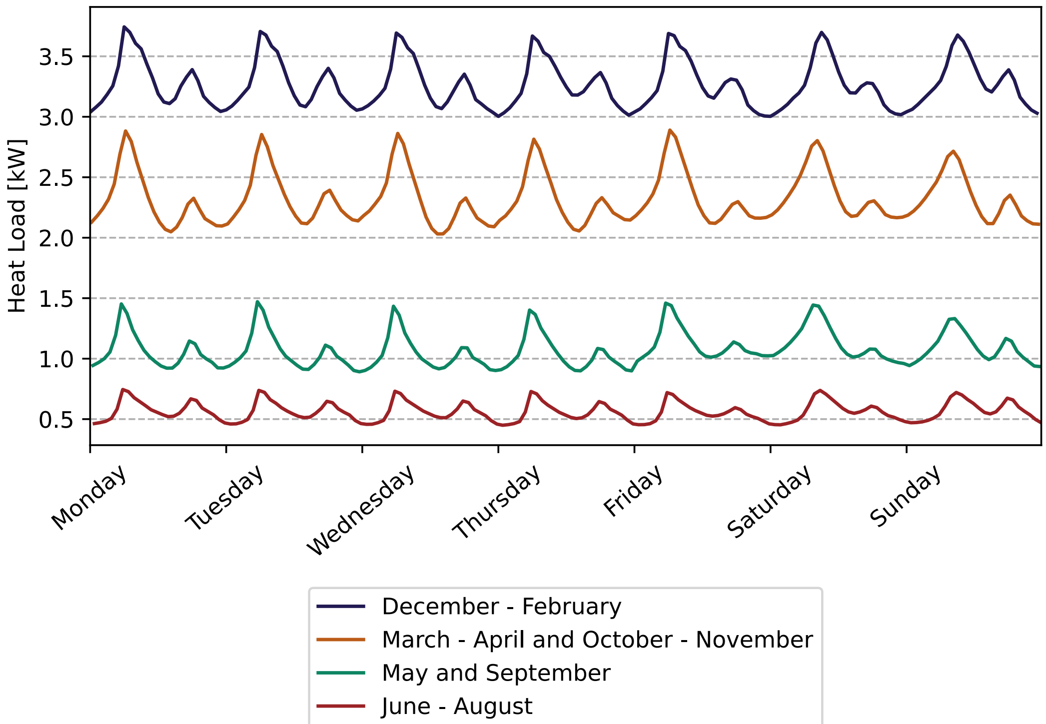

Figure 5: Weekly profiles of the mean heat demand of 3127 smart meters installed in residential buildings in Aalborg, Denmark

2.3 Weather Data

Weather data were used as input for the forecast models. They were obtained from the provider OpenWeather (OpenWeather, 2023). The data provide historical 16-day weather forecasts with hourly steps. Thus, the data include the weather forecasts errors of specific forecast steps. Consequently, the models were exposed to weather input uncertainty during the training stage, which could enhance the model's robustness (Wang et al., 2020). Moreover, the model evaluation was closer to a production environment where models are fed with real weather forecasts with uncertainty.

This study used the outdoor temperature and wind speed data for a one-week forecast horizon from the location of Ringkøbing, Denmark. The data ranged from 2020 to 2022 and aligned with the time range of the aggregated heat load data (Section 2.1).

Only a few implausibly high wind speed values were identified as outliers. These values were dropped and subsequently linearly interpolated. Figure 6 shows the outdoor temperature and wind speed time series by taking the example of the forecast step zero where the calculation time of the weather forecasts equals the forecast step time.

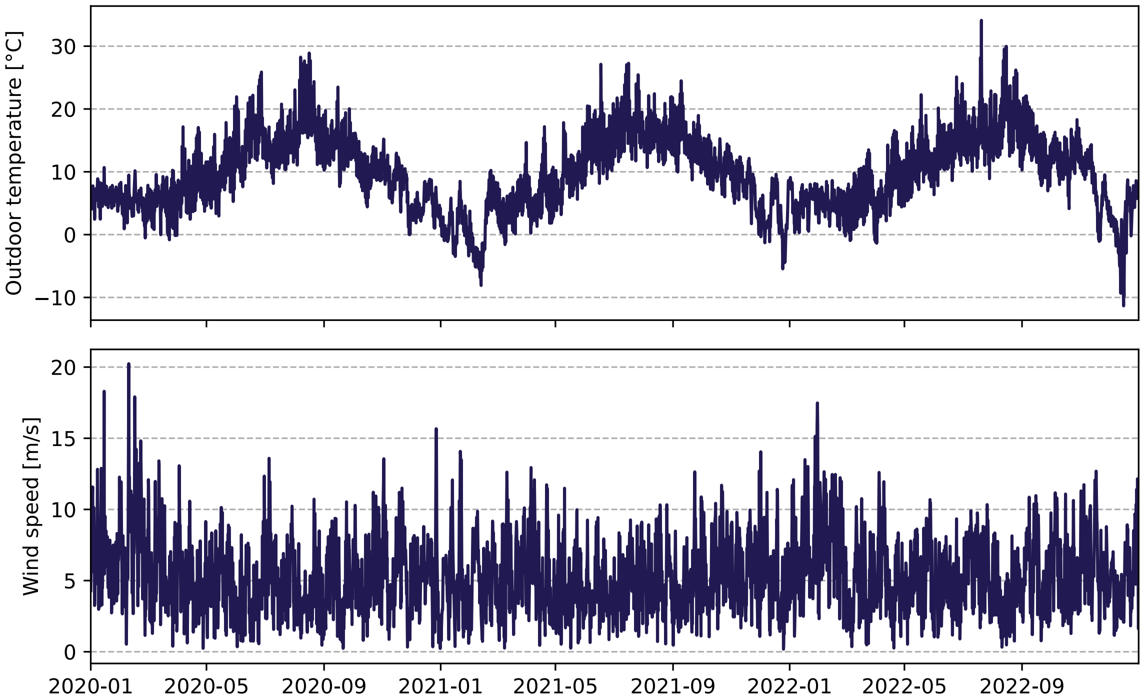

Figure 6: Outdoor temperature and wind speed data from Ringkøbing obtained from the forecast step zero of historical weather forecast data from OpenWeather (OpenWeather, 2023)

Figure 7 shows that the weather forecast errors increase over time as expected. These errors were incorporated into the heat load forecasting models. The undulate curve form is related to the fact that the weather forecast models were recalculated four times per day (OpenWeather, 2023). The forecast errors were calculated by computing the mean absolute error (MAE) as follows (scikit-learn, 2022):

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

where

$\hat{y}_i$ is the predicted value of the i-th instance;
y_i is the corresponding true value (forecast step zero);
n is the total number of instances.

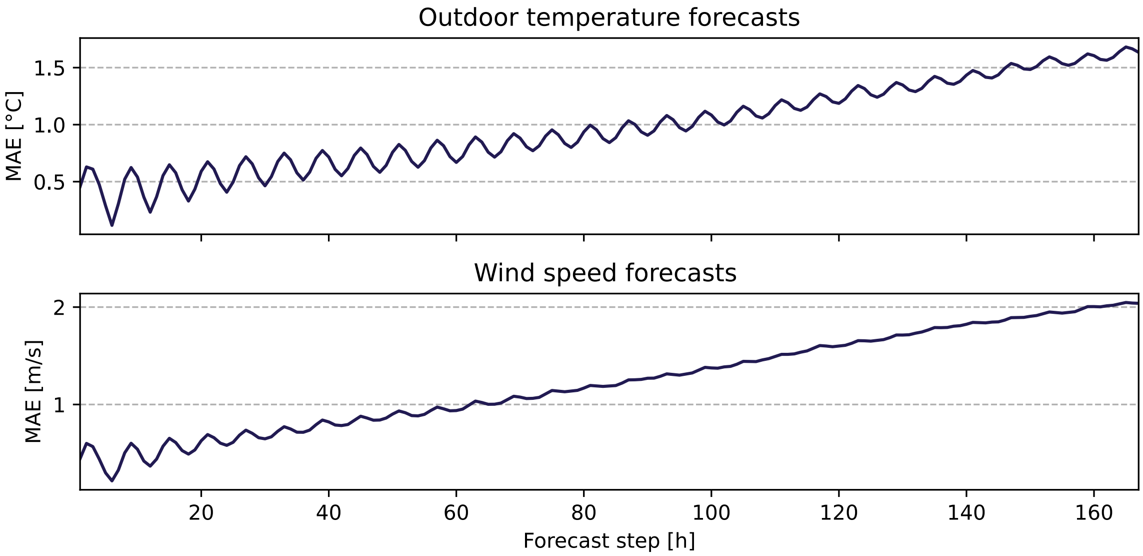

Figure 7: Forecasts errors of historical outdoor temperature and wind speed forecast data from Ringkøbing obtained from OpenWeather (OpenWeather, 2023)

3 Methodology

Figure 8 provides an overview of the methodology applied to investigate the research question "How can different learning strategies be applied to data-driven heat load forecasting models to handle concept drift in district heating systems?". The first part of this section is about how concept drifts were synthetically inserted by simulating different grid expansion magnitudes. Then, this methodology section describes the forecasting framework, including the forecast horizon, the corresponding multi-step forecasting strategy, the input variables to predict heat loads, and the data stream environment. Next, the learning strategies to handle concept drift are explained. They are divided into offline and online linear regression models. The offline models were retrained with several different data batch sizes and retraining triggers. Finally, the models were evaluated based on their forecasting performances measured by the root mean squared error (RMSE), the mean absolute percentage error (MAPE) and the mean error (ME). The following explains the methods shown in Figure 8 in detail.

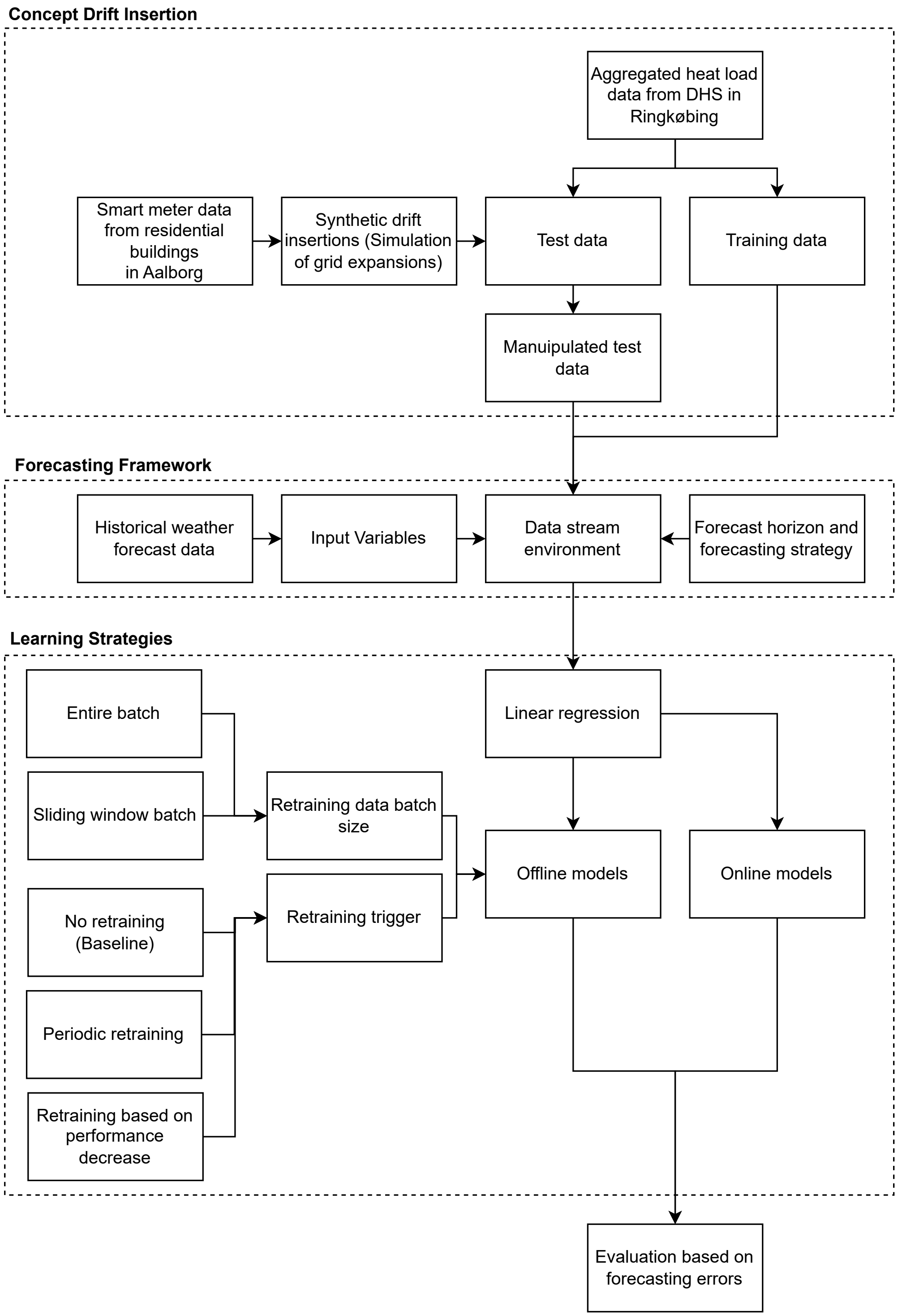

Figure 8: Overview of methodology

3.1 Concept Drift Insertion

Concept drifts were synthetically inserted by using the example of grid expansions. Grid expansions would inevitably change the relationship between the input variables and the aggregated heat load of the district heating system. For example, the aggregated heat load would be higher at the same outdoor temperature than before the grid expansion. Moreover, the statistical properties of the aggregated heat load time series would change as newly connected buildings come with specific load profiles that would influence, for example, the mean and the standard deviation. Note that the simulated grid expansions only increased the aggregated heat load. Simultaneous load decrease due to, e.g., heating saving measures was not simulated.

The heat load time series of the DHS was divided into 75% training data and 25% test data, corresponding to approximately three years of training data and one year of test data. The concept drift insertions were solely applied to the test data. This study analysed increases of 10%, 20%, and 30% of the initial mean heat load due to residential buildings added incrementally over a one-year horizon. In addition, a baseline 0% (no synthetic insertion) was examined.

The grid expansions were simulated by adding real smart meter heat load data from residential buildings to the existing heat load time series. The data were grouped into the week of the year, the weekday, and the hour of the day. Data in these groups were subsequently averaged, merged with the time series of the district heating system based on the respective group indexes, and finally added. This way of merging the data ensured that yearly, weekly, and daily patterns were considered when the recorded year differed between the time series. In addition, yearly differences were averaged out. 0.4 kW constant heat loss per connected building was assumed since it corresponds to approximately 20% of the mean of the residential heat load time series data.

The specific time series of the smart meter dataset (described in Section 2.2), the number of added buildings per step and the incremental step were selected randomly but under the condition that the total addition over one year respectively corresponded to approx. 10%, 20% and 30% of the initial mean of test data time series. The rationale behind randomly selecting these parameters was to introduce an incremental concept drift and insert some variability at the same time.

Figure 9 demonstrates the train-test split and an example of the 30% increase (drift insertion). The corrupted heat load of the test data (with drift insertion) is shown to be higher than the heat load of the uncorrupted test data (light-blue time series in the background). Accordingly, the mean and standard deviation of the heat load was increased. Statistically speaking, nonstationarity was introduced to the test data. Table 3 shows the mean values and standard deviations of the test set's heat loads before (0%) and after the drift insertions. For instance, the mean heat load values of the test set increased from 12.1 MW to 14 MW when 30% of the initial mean heat load, which is 12.1 MW, was incrementally added over one year. Correspondingly, the mean heat load of the uncorrupted test data in Figure 9 is 12.1 MW and the mean heat load of the corrupted test data in Figure 9 is 14.0 MW.

It is important to emphasise that the aggregated heat load data probably included information on changes in DHS before the synthetic concept drift insertion. However, the synthetic corruption ensured that the different approaches to handling concept drift were tested on known system changes. Also, note that 2022 was affected by the energy crisis, which is further discussed in Section 5.

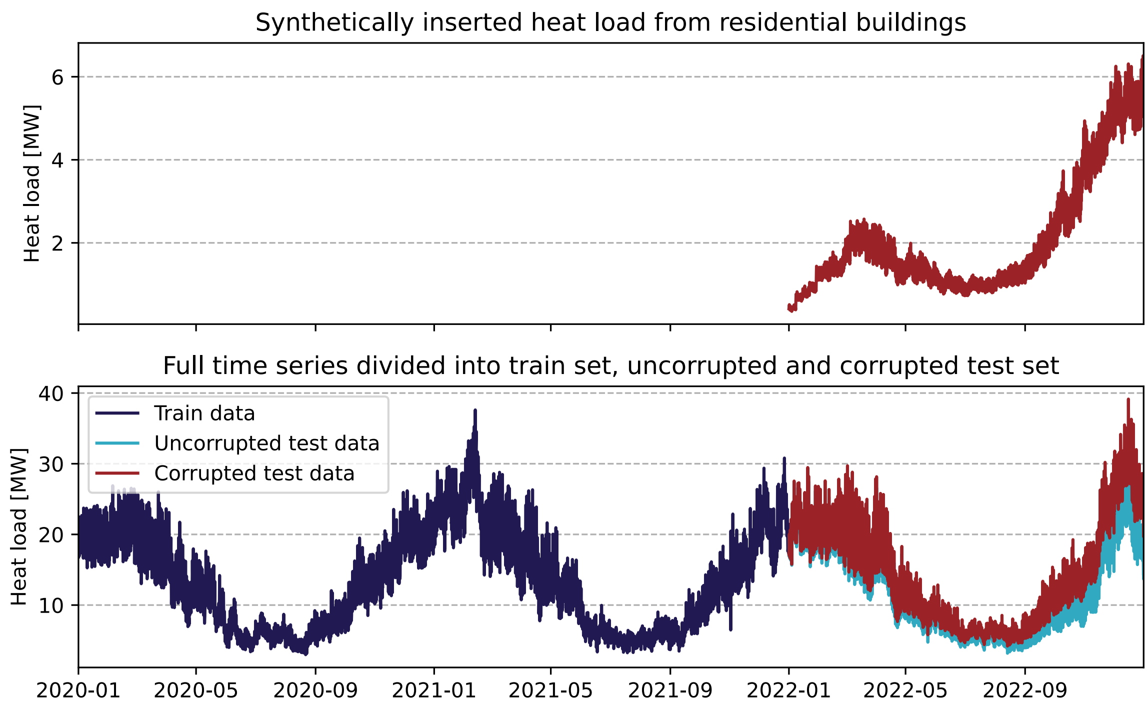

Figure 9: Train-test split and an example of test set's heat load values before and after the 30% incremental drift insertion

	Drift insertion			
	0 %	10 %	20 %	30 %
Mean [MW]	12.1	12.7	13.4	14.0
STD [MW]	6.5	6.7	7.0	7.3

Table 3: Yearly mean and standard deviation (STD) of the heat load of the test sets before (0%) and after (10%, 20%, 30%) drift insertions

3.2 Forecasting Framework

3.2.1 Forecast Horizon and Direct Multi-step Forecasting Strategy

In the Danish context, heat load forecasts relevant to the day-ahead electricity market must have at least a horizon of 36 hours since the day-ahead market closes at 12 pm and bids for the next day 00:00 - 24:00 must be made (Dahl et al., 2018). However, one could argue that the longer the forecast horizon, the longer the horizon the unit scheduling can be optimised. Especially when thermal storages are part of the DHS and electricity price forecasts are available for a horizon longer than 36 hours, a heat load forecast horizon greater than 36 hours might be beneficial. Since the DHS of the case area in Ringkøbing contains thermal storages, and electricity price forecasts are available for one week ahead, the heat load forecast horizon was chosen to be one week (168 hours).

Accordingly, the forecasting task of this study is framed as multi-step time series forecasting. The problem that, for example, the heat load of the previous hour is not available when forecasting the second time step ahead was solved through a direct multi-step forecasting strategy (DMSFS). The DMSFS forecasts each forecast step h independently from the others (Ben Taieb et al., 2012). H individual models f_h are learnt (one for each forecast step). Figure 10 shows an example of the DMSFS. Note that the forecast step 0 is considered future.

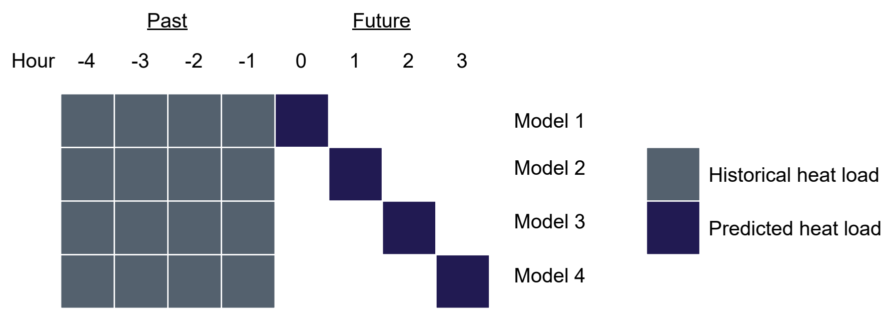

Figure 10: Example of the Direct Multi-Step Forecasting Strategy (Mielck, 2023)

The DMSFS has already been applied by Xue et al. (2019) in the context of heat load forecasting. However, the DMSFS comes with drawbacks, such as higher computational costs than a recursive strategy and disregarding dependencies between the predicted heat load values (Ben Taieb et al., 2012; Xue et al., 2019). On the other hand, the advantage of the DMSFS of being immune to the accumulation of errors (Xue et al., 2019) was regarded as superior due to the comparatively long forecast horizon of one week. Additionally, the DMSFS allowed the training of individual models for each forecast step. Therefore, the models were exposed to respective weather forecast uncertainty during the training stage. In addition, the DMSFS can provide more insight into how the forecast is changing over the forecast horizon since the forecast steps are modelled individually. Taking this together, the DMSFS was considered beneficial for the research purposes of this study.

3.2.2 Input Variables

The following describes the variables (features) used as input to the forecasting models. They are categorised into *weather*, *time* and *observed recent heat load* features.

Weather features Historical forecast data of the outdoor temperature and wind speed were used as input variables. In addition, the mean outdoor temperature and the mean wind speed of the last 24 hours were considered to incorporate an inert reaction of heat load to the weather.

Due to data availability, solar radiation data were not utilised. Including solar radiation would have had probably no significant effect since several previous studies have shown that solar radiation is of minor importance in forecasting the heat load (Fang and Lahdelma, 2016; Liu et al., 2020; Mielck, 2023; Potočnik et al., 2021; Wojdyga, 2008).

Time features The hour of the day was included to capture the daily seasonality. Since the aggregated heat load data (Figure 3 in Section 2.1) revealed noticeable differences between weekdays and weekends, an input variable that flagged a day as either a weekday (0) or a weekend (1) was included to incorporate the weekly seasonality.

Observed recent heat load features Furthermore, features referring to observed recent heat load values were used. The purpose of these features was to include information on the previous heat load (past behaviour) and their relation to the future heat loads, also known as autocorrelation. Figure 11 shows the autocorrelation and the partial autocorrelation diagram. In contrast to autocorrelation, partial autocorrelation describes the correlation that remains after removing the impact of any correlations due to the terms at shorter lags (Metcalfe and Cowpertwait, 2009).

According to the study by Xue et al. (2019), a threshold of ± 0.15 regarding the partial autocorrelation was set (red dashed line in Figure 11) to determine the number of lagged features. Consequently, the previous 25 heat load values were used as lagged features. Additionally, the mean heat load of the past 24 hours was included. While the lagged values were used to capture the recent hourly information of the heat load, the mean heat load incorporated a cumulative influence of the heat load over the past 24 hours on the forecasts.

The lagged heat load features of the models were highly correlated among themselves and induced the problem of multicollinearity. However, the multicollinearity may affect the coefficients of the linear regression models, which are later described in Section 3.3.1, but it does not influence the predictions (Neter et al., 1996, as cited in Frost, 2019).

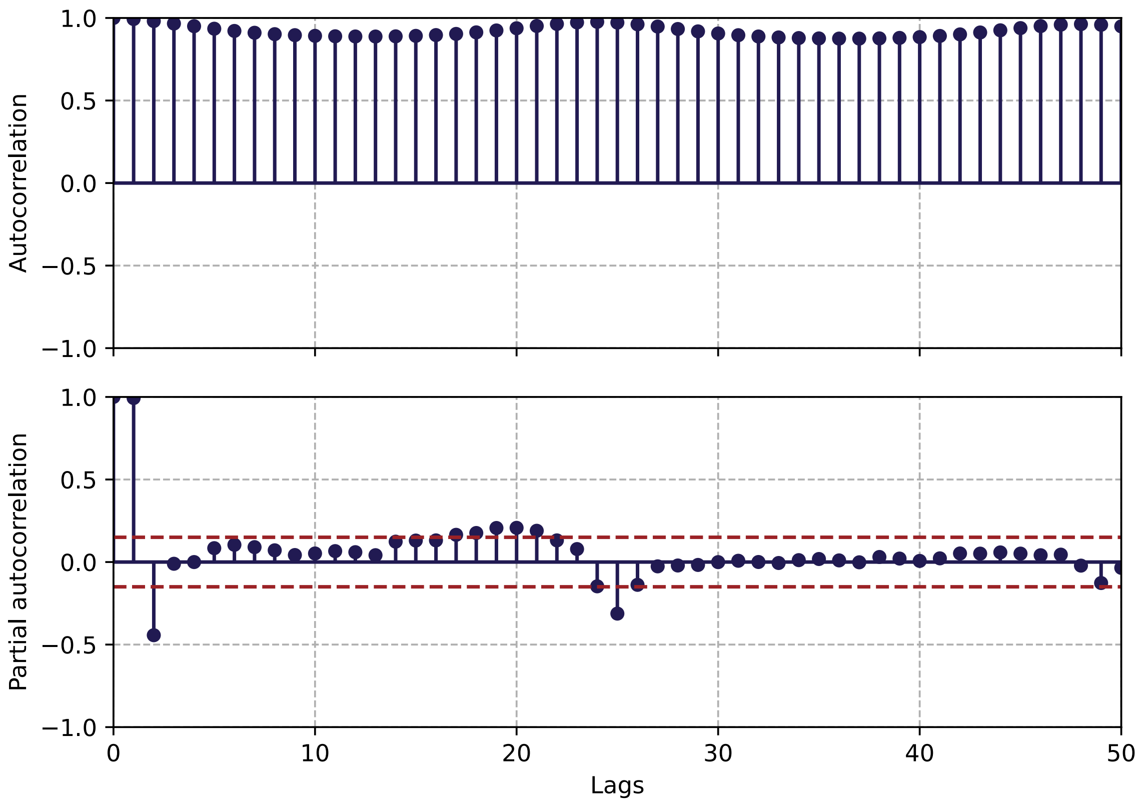

Figure 11: Autocorrelation and partial autocorrelation of aggregated heat load data of the DHS in Ringkøbing

Table 4 shows an overview of the features. In addition, polynomial features (higher-order terms) of the degree of 4 and interaction features were created from the outdoor temperature and the hour of the day to capture non-linear relations between them and the heat load. Moreover, all input variables were standardised to obtain a zero mean and unit variance. More precisely, running means and running variances were used to standardise the incoming data sequentially concerning the data stream environment explained in the following.

Weather	**Time**	**Observed recent heat load features**
Outdoor temperature	Hour of the day	Heat load values of past 1 to 25h
Mean outdoor temperature of past 24h	Weekday or weekend	Mean heat load of past 24h
Wind speed		
Mean wind speed of past 24h		

Table 4: Input variables grouped into *weather*, *time* and *observed recent heat load* features

3.2.3 Data Stream Environment

In contrast to the vast majority of previous studies (Section 1.1), which assume a static environment, this study analysed heat load forecasting approaches in an environment where data arrived continuously in the form of data streams. Bifet et al. (2018) define data streams as an algorithmic abstraction of a sequence of instances. The instances arrive one by one in a temporal order. The forecasting model interacts with these data streams in real-time.

The data stream environment is much closer to reality, where heat load and weather data arrive in real-time. The stream environment regards the fact that data streams evolve over time, and algorithms must react to the change. This study assumed that new instances, such as observed recent heat load values and new weather forecasts, arrive in hourly time steps.

Given this data stream environment, different learning strategies to handle concept drift were applied. They are grouped into offline and online learning and described in the following.

3.3 Learning Strategies to Handle Concept Drift

3.3.1 Linear Regression

Linear regression (LR) models were chosen to forecast the heat loads based on the input variables. They are easy to formulate, and their training is computationally cheap and quick (Suryanarayana et al., 2018). In addition, they come with high explainability and interpretability.

LR models were trained for both offline and online learning strategies. However, the way the optimal coefficients of the models were determined differed. If the entire training batch is available, as in offline learning, there is a mathematical *closed-form* solution that provides the results for the coefficients directly (Géron, 2019). In contrast, in online learning, the coefficients were updated sequentially through the Stochastic Gradient Descent (SGD) algorithm, which is further explained in Section 3.3.5.

LR models were trained to forecast each forecast step independently according to the DMSFS (Section 3.2.1). Formally, a generic linear regression model is defined as (Hastie et al., 2009):

$$\hat{Y} = \beta_0 + \sum_{j}^{p} X_j \beta_j \tag{2}$$

where

$\hat{Y}$ is the predicted value;
β_0 is a constant;
p is the number of input variables (features);
β_j is a coefficient referring to the input variable X_j.

To train the linear regression model, the coefficients β were determined by minimising the residual sum of squares (RSS) (Hastie et al., 2009):

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - x_i^T \beta)^2 \tag{3}$$

where

β denotes the coefficents;
N is the number of instances;
y_i is the target variable of the i-th instance;
x_i is the features vector of the i-th instance.

3.3.2 Offline Learning

In offline learning, the models are not updated constantly. Instead, the models must be retrained with a new batch of data. A new model replaces the old model. This raises two questions: When is a good time to retrain the model (Retraining trigger)? What data should be used to retrain the model (Retraining batch size)? Different retraining triggers and retraining data batch sizes were analysed to answer these questions. These are explained in the following. Algorithm 1 demonstrates the pseudo-code to simulate the corresponding stream data environment.

Algorithm 1 Offline Leaning in Data Stream Environment

1. Given a offline model f_h for the h-th forecast step pre-trained on the tabular training set D_{Train}.
2. For each feature vector x_i and target variable y_i of the i-th instance (timestamp) of the test set D_{Test}:
   (a) Predict the feature vector x_i and store the resulting prediction $\hat{y}_i$
   (b) Check if retraining should be triggered based on periodic intervals or drift detection
   (c) If retraining is triggered:
      i. Retrain model $\widetilde{f_h}$ with new training set $\widetilde{D}_{Train}$ of a certain batch size
      ii. Replace old model f_h with the new model $\widetilde{f_h}$

3.3.3 Retraining Triggering

One option to cope with concept drift is retraining offline models with a new batch of data from scratch and replacing the old model completely. The strategies to trigger this retraining are explained in the following.

No Retraining (Baseline) First, the models were not retrained to create a baseline that can be compared to other strategies. The offline models were trained once with the initial training set and not updated throughout the data stream simulation (Algorithm 1).

Periodic Periodic retraining strategy refers to a blind concept drift adaption (Gama et al., 2014) since the offline models were adapted through retraining without any explicit detection of change. More precisely, the models were retrained periodically, and the following two retraining intervals were tested: Daily according to the suggestion by Suryanarayana et al. (2018) and monthly according to the suggestion by Potočnik et al. (2021). However, determining periodic intervals depends on the specific use case, the type as well as the rate of the change (concept drift). Thus, the two values proposed in the previous studies were used to evaluate the impact of periodic retraining strategies in comparison to other approaches.

Sliding Mean Squared Error In contrast to a blind adaption strategy, such as periodic retraining, an informed strategy is reactive and depends on whether a drift detector has been flagged (Gama et al., 2014). This study used a sliding mean squared error (MSE) to detect concept drift.

The MSE was calculated over a sliding window size of one month. When the MSE of the following month was 25% higher than the previous MSE, retraining was triggered. The MSE was calculated as follows (scikit-learn, 2022):

$$MSE(Q, \hat{Q}) = \frac{1}{n} \sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2 \tag{4}$$

where

$\hat{Q}_i$ is the predicted heat load value of the i-th instance of the set of the sliding window D_{SW};
Q_i is the corresponding true value;
n is the total number of instances in the sliding window set D_{SW}.

This drift detection technique is similar to the one used by Jan (2021), where the total percentage error over a six-month sliding window was calculated. The acceptable error threshold was set to 5%. This study used the MSE as it penalises high forecast errors higher than the percentage error since the errors are squared. Accordingly, the threshold was set higher (to 25%) in comparison to 5%. Also, the sliding window size was smaller, as this study refers to a forecast horizon of one week instead of long-term one-year-ahead forecasts. Nevertheless, selecting the optimal parameter for a sliding performance drift detector is challenging as the type, magnitude and speed of the change are generally not known in advance and depend on the use case. Therefore, the selected parameters represent one possibility of a sliding performance drift detector to compare this approach to other strategies handling concept drift.

3.3.4 Retraining Data Batch Size

When retraining was triggered, the following training data batch sizes were evaluated:

Entire batch The new model was retrained with the entire batch data (Old training data and the new data up to the moment of retraining). This approach requires that all the data are stored to retrain the models.

Sliding window batch In contrast to retraining the model with the entire batch size, models were also retraining with a fixed window size of the latest samples. This method requires storing the data of the corresponding window size.

This sliding window batch considers the most recent data more strongly than using the entire batch of data for retraining since old data that are not part of the sliding window are excluded. On the other hand, the models forget old data to some extent that could still contain valuable information to predict heat load. Hence, there is a trade-off. Retraining with sliding window batches could be considered a way of emulating incremental learning with new data batches, which is referred to as *batch-incremental learning* (Read & Zliobaite, 2023). To a certain degree, the parameters retraining triggers and retraining data batch sizes can be regarded as a learning rate that determines the speed of adaption. For example, the more often the retraining and the smaller the batch sizes, the faster the model adapts to the most recent data and the faster the model forgets old data. In this sense, there is a similarity to online learning explained Section 3.3.5.

Sliding window sizes of three and six months were used in this study. The rationale behind choosing these window sizes is the following: The window sizes are relatively small compared to the initial length of the training data, which was approximately two years (Section 3.1). If the models are retrained with relatively small window sizes, they can focus more on recent data, which is likely to be more relevant for forecasting current heat load values if concept drift has happened recently. At the same time, the model can respectively retain information from older data from the latest three or six months. Therefore, the intention was to balance the previously mentioned trade-off. However, similar to the parameters of the other retraining triggers and retraining batch sizes, it is challenging to choose the best parameter for the sliding window batch size since the type, extent and rate of the change are usually unpredictable and not known in advance. As a result, the parameters chosen in this study represent just potential sliding batch sizes, which can be used for comparison against the other methods to handle concept drift.

3.3.5 Online Learning

In contrast to offline learning, online learning is a branch of machine learning where the learnt model is updated through sequential steps. Online learning does not require retraining from scratch to cope with concept drifts as the models are updated with the latest data that contain information about the changed concept. Due to the algorithmic nature of online learning algorithms, the data stream environment was slightly different to the offline learning data stream environment (Algorithm 2). The models were pre-trained on the training data set sequentially. Moreover, no retraining had to be triggered. Instead, predictions were generated and stored. Then, the model was updated with the feature vector and the corresponding true target value. The target is only revealed to the model after a certain delay corresponding to the forecast step since not all true values of the entire forecast horizon are available each hour. For example, when the heat load 168 hours ahead is predicted, it takes 168 hours until the corresponding true value is available to update the model.

Algorithm 2 Online Leaning in Data Stream Environment

1. Given an online model f_h for the h-th forecast step
2. For each feature vector x_i and target variable y_i of the i-th instance (timestamp) of the training set D_{Train}:
   (a) Update model f_h with x_i and y_i
3. For each feature vector x_i and target variable y_i of the i-th instance (timestamp) of the test set D_{Test}:
   (a) Predict the feature vector x_i and store the resulting prediction $\hat{y}_i$
   (b) Update model f_h with x_i and y_i

In offline linear regression (Section 3.3.2), the coefficients of the linear model to minimise the RSS can be found directly through a mathematical equation (Géron, 2019). This requires access to the entire dataset. In contrast, online linear regression updates the coefficients sequentially as new data arrives. Accordingly, the models adapt to change blindly and constantly.

Stochastic Gradient Descent (SGD) (Robbins, 1951, as cited in River-Contributors, 2023) was applied to minimise the RSS in order to find the optimal coefficients of the online linear regression model. The SGD algorithm allowed the iterative minimisation of the squared residuals and was, therefore, suitable to update the coefficients of the linear model sequentially in an online manner. However, due to its stochastic nature, the SGD algorithm only finds solutions that are close to the mathematical optimum (Géron, 2019).

The learning rate of the SGD algorithm describes the step size at each iteration while moving towards a minimum of a loss function (Géron, 2019). In terms of online learning, the learning rate can be regarded as the speed of adaption (Géron, 2019). However, there is a trade-off. With a low learning rate, the model will have more inertia - it learns more slowly (Géron, 2019). At the same time, the model is less sensitive to noise and nonrepresentative data points (outliers) (Géron, 2019). This study determined the learning rate based on the training dataset without explicit drift insertion. Thus, leaking information from the test was avoided. Learning rates of 0.01, 0.001, and 0.0001 were evaluated. The learning rate of 0.001 performed best based on the root mean squared error (RMSE), which is later explained in Equation (5) in Section 3.4, and was used in this study.

3.3.6 Overview of Learning Strategies to Handle Concept Drift

Table 5 lists all previously described strategies to handle concept drift. The table states the strategy names which are referred to in the results (Section 4) and discussion section (Section 5).

Model	Retraining Trigger	Retraining Batch Size	Name
Offline LR	None	None	Baseline
Offline LR	Daily	Sliding 3 months	Offline-1
Offline LR	Daily	Sliding half a year	Offline-2
Offline LR	Daily	Entire batch	Offline-3
Offline LR	Half a year	Sliding 3 months	Offline-4
Offline LR	Half a year	Sliding half a year	Offline-5
Offline LR	Half a year	Entire batch	Offline-6
Offline LR	Sliding MSE (20% decrease, window size of one month)	Sliding 3 months	Offline-7
Offline LR	Sliding MSE (20% decrease, window size of one month)	Sliding half a year	Offline-8
Offline LR	Sliding MSE (20% decrease, window size of one month)	Entire batch	Offline-9
Online LR	-	-	Online

Table 5: Overview of analysed learning strategies to handle concept drift

3.4 Evaluation

According to Section 3.1, the aggregated heat load time series of the DHS in Ringkøbing was divided into two years of training data (2020 and 2021) and one year of test data (2022). Subsequently, the test data was manipulated by inserting concept drift. Then, the models were tested on this corrupted test data, and their forecast performances were evaluated.

According to the DMSFS (Section 3.2.1), each forecast step was modelled by an independent model resulting in performance metrics for each model. To evaluate the strategies' capabilities to handle concept drift, the average root mean squared error (RMSE), the mean absolute percentage error (MAPE), (scikit-learn, 2022) and the mean error (ME) were calculated by averaging the individual metrics of each model of the entire forecast horizon. The variables of the following metric equations are denoted as follows:

$\hat{Q}_i$ is the predicted heat load value of the i-th instance of the test set D_{Test};
Q_i is the corresponding true value;
n is the total number of instances in the test set;
ϵ is an arbitrarily small, strictly positive number to avoid undefined results when Q_i is zero.

Root Mean Squared Error (RMSE) The RMSE describes the forecast accuracy - the lower the RMSE, the better the forecast accuracy. It penalises over- and underpredictions equally, as the errors are squared, leading to only positive values. At the same time, the squaring penalises outliers strongly. The square root retains the original unit, which is the heat load measured in megawatts (MW) in this study. The RMSE is specific to the case area. It can support the DHS operator in estamating the specific impact of the forecast errors, for example, if there is a linear relationship between the forecast errors and the operation costs. Moreover, the RMSE can be related to the heat load of a specific DHS to assess the significance of the forecast errors.

$$RMSE(Q, \hat{Q}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2} \qquad (5)$$

Mean Absolute Percentage Error (MAPE) In contrast to the RMSE, the MAPE relates the absolute forecast errors to the true values in fact. Therefore, the MAPE provides a general picture of the forecast accuracy, which can be theoretically compared to other case studies. However, a clear comparison is hardly possible, as the accuracy depends on several factors, such as the used training and test data and its quality. In this study, the MAPE must be interpreted carefully, as the true heat load values Q_i of the denominator increased depending on the drift insertion. Consequently, the MAPE values can decrease even though the absolute errors increase, which is further discussed in Section 5.3.

$$MAPE(Q, \hat{Q}) = \frac{1}{n} \sum_{i=1}^{n} \frac{|Q_i - \hat{Q}_i|}{max(\epsilon, |Q_i|)} \tag{6}$$

Mean Error (ME) The ME indicates systematic errors (biases). A positive ME reveals systematic underprediction, whereas a negative ME indicates systematic overprediction. For an unbiased regression model, the ME equals zero (Frost, 2019). Consequently, the closer the ME is to zero, the less biased the model is.

$$ME(Q, \hat{Q}) = \frac{1}{n} \sum_{i=1}^{n} (Q_i - \hat{Q}_i) \tag{7}$$

4 Results

All previously described methods were implemented in Python. The open-source libraries *NumPy* (Harris et al., 2020), *Pandas* (McKinney, 2010) and *Statsmodels* (Seabold and Perktold, 2010) were used for general data handling and exploration. *FeatureEngine* (Galli, 2021) was partly used for feature engineering. The offline regression was implemented through *Scikit-learn* (Pedregosa et al., 2011), whereas the online regression was implemented through the online machine learning package *River* (Montiel et al., 2021).

This section is structured as follows: First, the results of retraining offline models are described. Then, this offline learning approach is compared to online learning. The results are summarised at the end of this section.

4.1 Results of Retraining Offline Models

This section describes the results of the different retraining strategies to train the offline linear regression models and replace them with new models. This study investigated periodic retraining and retraining based on a decrease in the MSE as described in Section 3.3.3. Furthermore, this study analysed retraining these new models with different data batches as explained in Section 3.3.4. Table 5 in Section 3.3.6 provides an overview of the strategies and their names, which are referred to in the following figures.

Figure 12 shows the results of retraining offline models based on the three drift insertions (10%, 20% and 30%) and the original test data without any manipulation (0%). The evaluation is based on RMSE and compared to the baseline, where no retraining was conducted. The RMSE refers to the average of the entire forecast horizon (one week of hourly forecasts).

Figure 12 shows that retraining does not improve the RMSE significantly. The RMSE differences to the baseline are not greater than approximately 1 MW. Some retraining strategies perform even worse than the baseline. In relation to the yearly mean heat load values of the test data sets, which are shown in Table 3 in Section 3.1 range between 12.1 MW and 14.0 MW, the changes in the RMSE are considered minor.

Moreover, the Figure 12 indicates that the slopes of the lines do not significantly change under different drift insertions. Since the slope can be interpreted as the sensitivity to concept drift or the capability to handle concept drift, the results show that all retraining strategies are not very sensitive to the concept drift insertion. However, the curves of *Offline-1* and *Offline-2* are least sensitive to the magnitude of concept drift.

The best performance under drift insertion that is greater than 10% shows the *Offline-1* strategy where offline models were retrained daily with a sliding data batch of the past three months. The corresponding RMSE is approximately 1.4 MW. However, the RMSE difference between the baseline and the *Offline-1* scenario is only around 0.4 MW under 30% drift insertion. This difference is not significant in relation to the mean of the corresponding test data set (14.0 MW, as shown in Table 3).

Also, the RMSE of the strategy *Offline-7* did even decrease with increasing concept drift. However, the RMSE was approximately 0.7 MW higher than the baseline, indicating that this strategy does not perform well in general, as the RMSE under no synthetic concept drift insertion is around 2 MW.

More generally, the relatively high RMSE errors of some retraining strategies under no synthetic concept drift insertion, such as *Offline-4* and *Offline-7*, illustrate that there is also a risk of applying a retraining strategy that can lead to worse performances than applying no retraining strategy (the baseline).

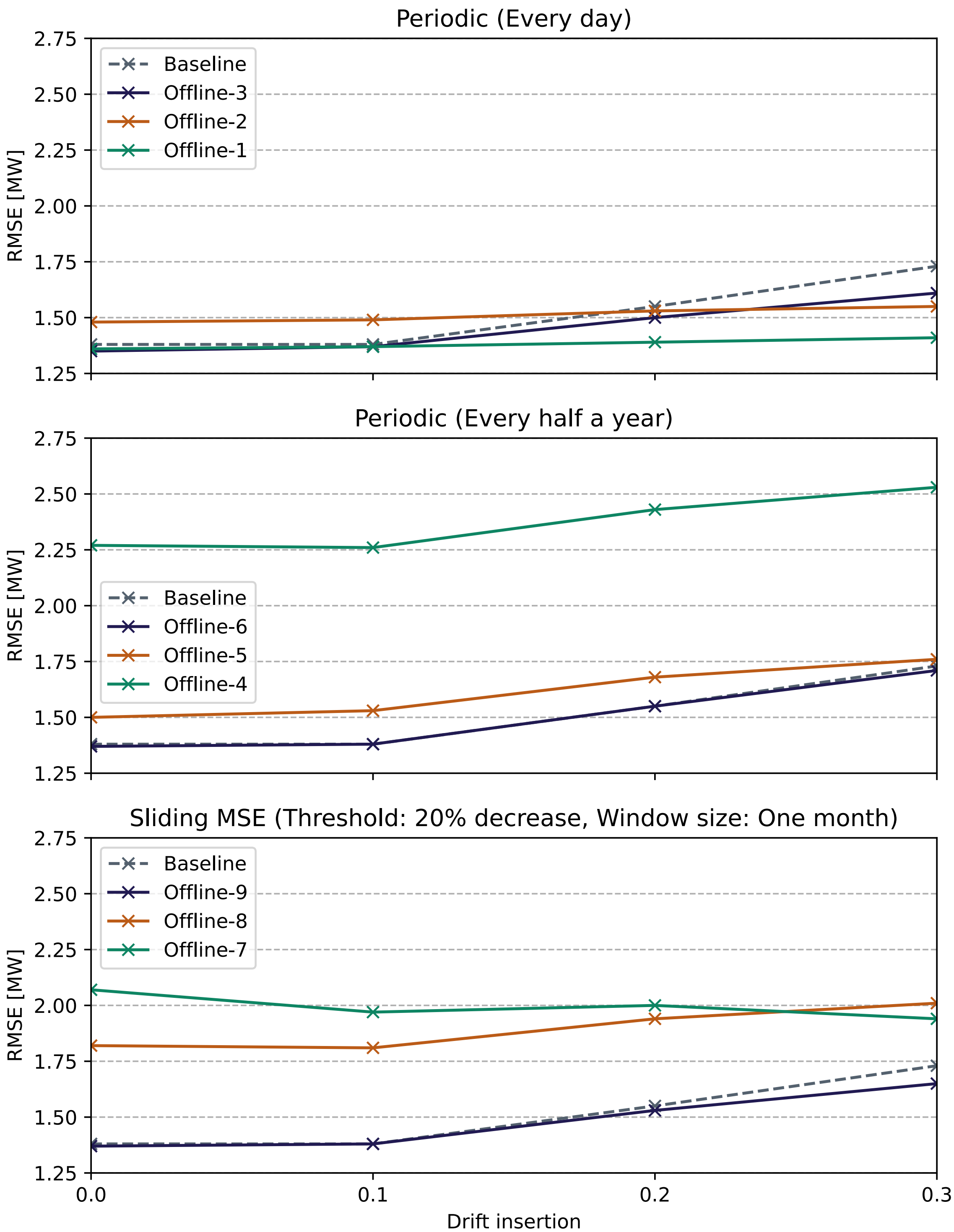

Figure 12: Average RMSE (corresponding to the entire forecast horizon) of the retraining strategies divided into different retraining triggers (subplots) and retraining batch sizes (lines) for different drift insertions

4.2 Comparison of Offline and Online Learning

Can online learning handle concept drift in DHS better than the retraining of offline models? Figure 13 compares offline regression with online regression. The offline regression is divided into the baseline (no retraining) and the previously mentioned retraining strategy *Offline-1* that performed best of all retraining strategies. The figure demonstrates that offline learning performed better than online learning over all drift insertions, even without retraining (*Baseline*). As previously mentioned, the strategy *Offline-1* indicate an RMSE of around 1.4 MW that increases very slightly with increasing concept drift insertion. In contrast, the online linear regression models show an RMSE of approximately 1.75 MW without any significant difference between different drift insertions. The baseline performance starts to increase after a drift insertion of 10% from 1.4 MW to approximately 1.72 MW. Considering these values again in relation to the mean values of the test data sets (Table 3), RMSE differences in the order of 0.4 MW cannot be considered significant.

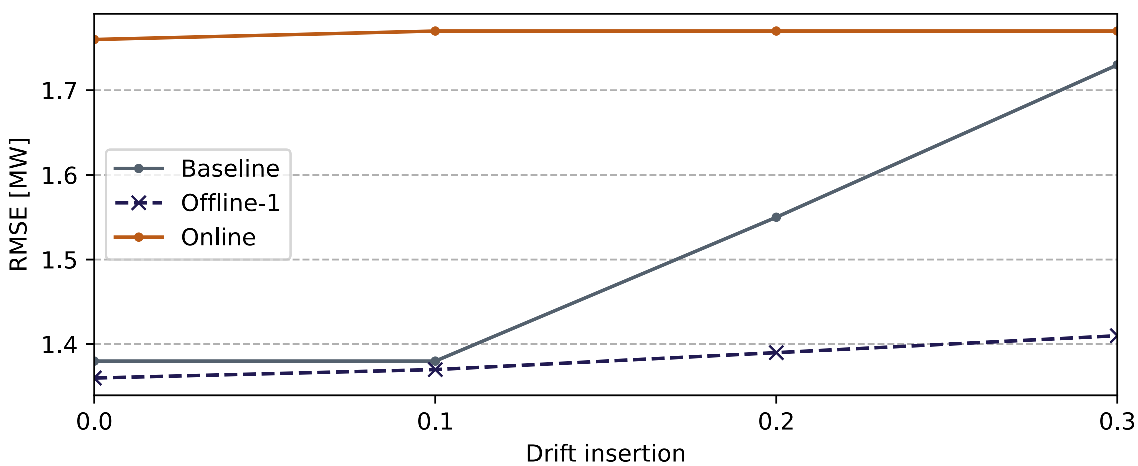

Figure 13: Average RMSE (corresponding to the entire forecast horizon) of the baseline strategy (*Baseline*), online linear regression (*Online*) and the best-performed retraining strategy (*Offline-1*) for different drift insertions

Viewing the results with the MAPE provides a different perspective on the results. Figure 14 indicates that the MAPE decreases with increasing drift insertion. The relation between the absolute errors and the mean values becomes slightly smaller with regard to the equation of the MAPE, as shown in Equation (6) in Section 3.4. Even though the absolute errors increase according to the RMSE shown in Figure 13, the relative error decreases slightly. Only the baseline shows an opposite trend after a drift insertion of 10% when the MAPE starts to increase.

As explained in Section 3.4, it is important to emphasise that the MAPE is sensitive to the way of the concept drift insertion. More precisely, the MAPE is sensitive to the true heat load values to which the absolute forecast errors are related. Depending on the drift insertion, these true heat load values increase or decrease and consequently decrease or increase the final MAPE. This explains why MAPE values decline in Figure 14. Consequently, the MAPE has to be interpreted carefully, which is further discussed in Section 5.3.

The slopes of the lines in Figure 14 illustrate that the strategies *Offline-1* and *Online* can properly handle the concept drift insertion. In contrast, the MAPE of the baseline increases after 10% drift insertion. Consequently, one could argue that the strategies *Offline-1* and *Online* can handle concept drift better than the baseline. However, the different capabilities do not result in significantly better absolute forecast errors in light of the RMSE and the mean values of the test sets.

Moreover, Figure 14 shows that the MAPE difference between *Offline-1* and *Baseline* is constantly around 2.5%. However, the difference originates from the general algorithmic performance rather than from the capabilities to handle concept drift since the slopes of the respective line are similar. The MAPE difference of 2.5 % was already present under 0% drift insertion and corresponded only to 0.4 MW in terms of the RMSE, as shown previously in Figure 13.

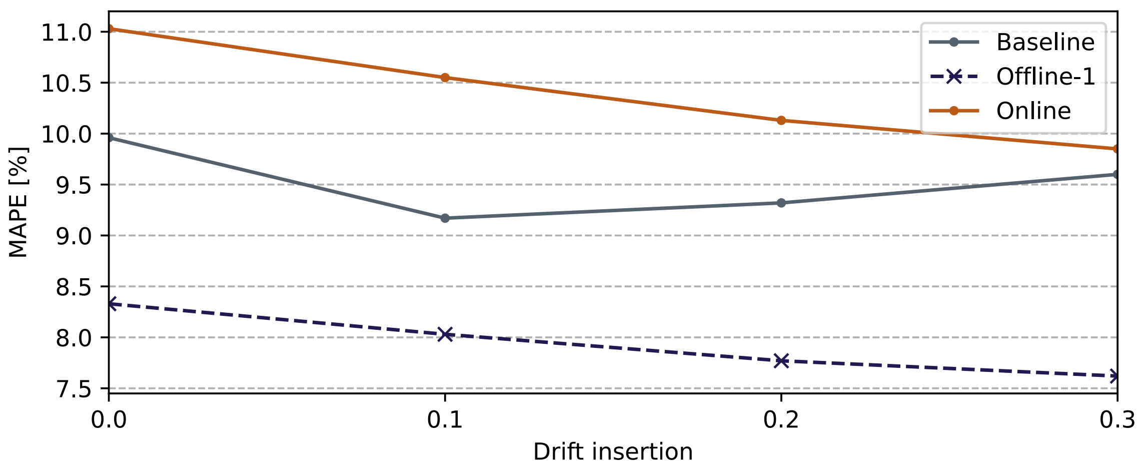

Figure 14: Average MAPE (corresponding to the entire forecast horizon) of the baseline strategy (*Baseline*), online linear regression (*Online*) and the best-performed retraining strategy (*Offline-1*) for different drift insertions

Figure 15 provides an additional angle on the results by showing the ME over the different forecast steps. In contrast to RMSE and the MAPE, the ME did not calculate the square and, therefore, not the absolute values of the errors, as explained in Section 3.4. Therefore, the ME can indicate forecast biases. A positive ME refers to systematic underprediction, whereas a negative ME reveals systematic overprediction.

Figure 15 shows that the baseline overpredicted without any synthetic drift insertion. The ME decreases from forecast step zero to the last forecast step by approximately 0.5 MW. This fact is probably related to the effect of the energy crisis on the test data and is further discussed in Section 5.3. The overprediction of the baseline is compensated under 10% drift insertion. Under 20% and 30% drift insertion, the bias of the baseline increased and led to underpredictions. The edge case of 30% drift insertion and the last forecast step indicates a ME of around 0.6 MW. In relation to the mean heat load of 14 MW, it is, however, not a significant underprediction.

Additionally, Figure 15 reveals that the *Online* strategy is robust. It does not indicate any systematic error of different forecast steps or drift insertions. In contrast, the strategy *Offline-*

1 started to overpredict slightly after a forecast step of approximately 4 days. The maximum ME of the *Offline-1* is approximately 0.25 MW, and there are no clear differences between the drift insertion magnitudes.

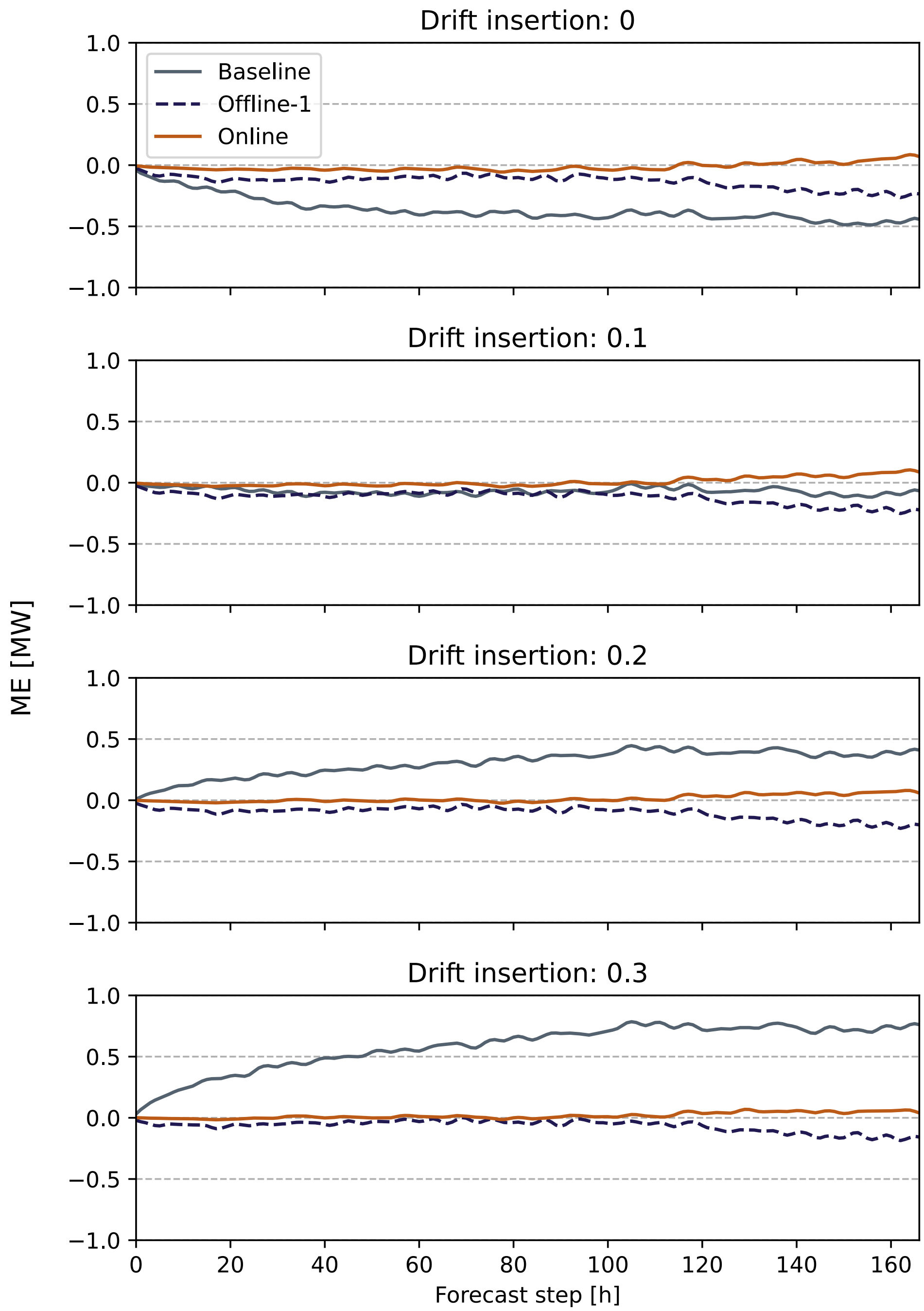

Figure 15: ME of of the baseline strategy (*Baseline*), online linear regression (*Online*), and the best-performed retraining strategy (*Offline-1*) of different forecast steps and drift insertions

The bias of the baseline increased with an increasing forecast step. This is plausible, as the importance of observed recent heat load features decreases over the forecast horizon (Mielck, 2023). Since these features incorporate recent information on the heat load, they incorporate information on the change at the same time. However, this information becomes less with an increasing forecast horizon leading to an increasing bias of the baseline strategy.

To further explain the role of the observed recent heat load features, Figure 16 shows the ME of the strategies *Baseline*, *Offline-1* and *Online* with and without observed recent heat load features. The figure explains that the observed recent heat load features can help to handle concept drift to some extent without any specific learning strategy (baseline). The slope of the lines indicates the robustness under concept drift and, therefore, the capability to handle concept drift. The slope of the baseline strategy *with* observed recent heat load features is flatter than *without* observed recent heat load features. Under 30% concept drift insertion, the ME difference between them is around 0.7 MW. However, in light of the mean heat load of approx. 14 MW, the difference is minor.

Furthermore, the strategies *Offline-1* and *Online* show ME of approximately zero. Therefore, they indicate no significant systematic error. Thus, they are capable of coping with concept drift without indicating over- or underpredicting.

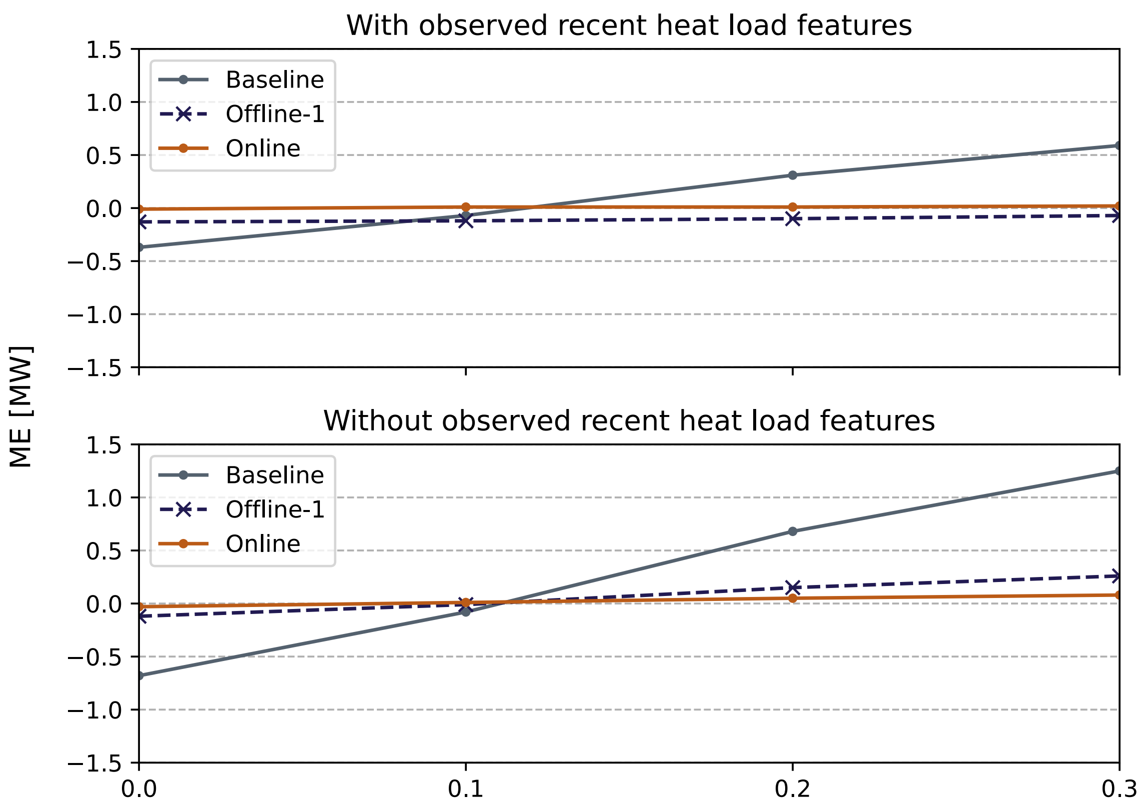

Figure 16: Average ME (corresponding to the entire forecast horizon) of the baseline strategy (*Baseline*), online linear regression (*Online*), and the best-performed retraining strategy (*Offline-1*) over different drift insertion divided into models trained *with* and *without* observed recent heat load features

4.3 Summary of Results

In summary, the interpretation of the results depends on the perspective. On the one hand, the strategies differ in terms of their general capability to handle concept drift based on their sensitivity to concept drift insertions. Daily retraining and online learning could sufficiently handle concept drift, as they were robust to different magnitudes of concept drift. The baseline strategy of offline linear regression without any retraining was more sensitive to concept drift. However, including observed recent heat load features mitigated the sensitivity of the baseline strategy.

On the other hand, when the RMSE and ME are put into the perspective of the mean of the aggregated heat load of the case area, the different capabilities to handle concept drift did not result in significantly different absolute errors, especially when observed recent heat load features are included.

5 Discussion

5.1 Practical Implications

DHS play an important role in future energy systems, as they can integrate renewable energy sources, couple the heat and electricity sector and utilise excess heat (Lund et al., 2014). Accordingly, they allow the use of local resources and have a high potential to decrease the dependency on energy imports. Especially the energy crisis in 2022 has increased the demand for district heating dramatically. Consequent changes, such as grid expansion, can lead to concept drift in the corresponding forecast models. The findings of this study provide novel information for model developers and practitioners to secure stable heat load forecasting performances when changes like this happen. This study provides the following guidance.

Observed recent heat load features can neglect the need for retraining under minor concept drift, but monitoring remains essential

The results revealed that forecast errors measured by the RMSE do not significantly decrease if no retraining or online learning strategy is applied and observed recent heat load features are incorporated. Consequently, offline models can be deployed and operated without retraining if the DHS does not change majorly. This, however, does not neglect the need for monitoring the models so that retraining can be induced if a significant performance decrease is observed.

Generally, the findings underline the importance of observed recent heat load features in practice. When offline modelling without retraining is applied, model developers should ensure that models incorporate autocorrelation to condition forecasts on the latest measured heat load values. This is important to be less sensitive to concept drift. At the same time, observed recent heat load features can lead to better forecast accuracy in general.

This study recommends frequent retraining of offline models or online learning to be robust to concept drift

The findings point towards the need for retraining offline models or online learning if the concept drift is greater than the magnitudes analysed in this study. This study recommends frequent retraining, e.g., daily or online learning, as both were robust to the magnitude of drift insertion. However, the specific periods and data batch sizes of retraining may be adjusted with regard to the case area and the expected change, as the results showed that there is a risk of worsening the performance if the wrong retraining parameters are chosen. Generally, the more often the retraining, the better. LR models may allow this as they are computationally cheap (Suryanarayana et al., 2018). A high degree of retraining automatisation, such as automated data cleaning, is advisable to ensure low retraining costs in general.

Linear regression can enhance trust in adaptive systems

An additional advantage of using linear LR models is their simplicity. They are easy to interpret in contrast to complex *black box* models. Relying on hardly interpretable models is not popular among experts (Gama et al., 2014; Machlev et al., 2022). This is especially the case in the district heating industry, which requires a high level of reliability. Hence, the high degree of explainability of LR models can improve usability and trust in adaptive learning systems in practice.

Databases must be easily accessible to allow frequent retraining or online learning

Lastly, the findings underline that SCADA systems and the underlying databases must be easily accessible and connectable to forecasting models, e.g. through a REST API. They are the foundation for retraining models with the most recent data, providing observed recent heat load values as features or applying online learning. This complements the demand for high-quality databases for heat load forecasting in general, according to Zhao et al. (2022).

5.2 Contextualisation of Results Within Previous Research

This section aims to integrate the findings into the literature that was introduced in Section 1.1. To summarise, not many studies have investigated the problem of concept drift in terms of heat load forecasting. The vast majority focused on a static environment with no change. The few studies that acknowledge the problem of concept drift applied online learning as a solution and did not thoroughly investigate retraining. In addition, the simple question of whether retraining or online learning is even required, if input variables that incorporate change are included, is overlooked.

The study by Suryanarayana et al. (2018) state that daily retraining could be a feasible option to cope with concept drift. This study confirms this. The findings of this study indicate that daily retraining was not sensitive to the magnitude of synthetically inserted concept drift. In addition, this study complements Suryanarayana et al. (2018)'s research by showing that online learning through linear regression can effectively handle concept drift as well, even though the general forecast accuracy measured by the RMSE and MAPE was slightly worse.

Potočnik et al. (2021) suggest retraining, for example, every half a year. On the contrary, the findings of this study indicate that retraining more often (daily) is better than retraining less often (retraining every half a year). At first view, this might seem obvious. However, in other concept drift scenarios, for example, if the drift happens very slowly, daily retraining might be too often. It can lead to a waste of resources, such as human resources, if the retraining requires manual interaction to, for example, clean the training data. In addition, computational resources may be wasted if many models with many data or very complex models are trained. However, computational resources may be a minor criterion in terms of linear regression as it is comparatively computationally cheap (Suryanarayana et al., 2018).

The research by Provatas et al. (2014), Provatas (2014), Grosswindhager et al. (2011) and Jan (2021) conclude that their online approaches work well to forecast heat loads. On the one hand, this study confirms that online learning can handle concept drift in district heating systems properly. On the other hand, this study compared offline and online approaches. The findings indicate that an online approach may be unnecessary when observed recent heat load features are included, and concept drift is minor. Then, the absolute error difference to the online strategy of this study was not significantly large. In addition, the findings of this study show that retraining, especially frequent retraining, can be an alternative to online learning and can result in slightly better forecast accuracy.

Finally, this study point towards the importance of observed recent heat load features if offline learning without retraining is applied, as it mitigates the sensitivity to concept drift. This converges with the results of the study by Mielck (2023), which analysed the feature importance in a static environment. This study complements these findings by indicating that observed recent heat load features can be even more important when heat load forecasting models face the problem of concept drift.

5.3 Limitations & Future Directions

Data

This study used two datasets to model the heat load (Section 2). First, a heat load time series from the DHS in Ringkøbing, Denmark, as the forecasting target. These data ranged from 2020 to 2022. Second, heat load time series from smart meters to simulate a grid expansion and a corresponding concept drift. These data ranged from 2018 to 2020. Therefore, the first mentioned dataset entailed the energy crisis in 2022. Additionally, both datasets involve the effects of the COVID-19 pandemic.

Hence, the models might have been trained on comparatively high heat loads (training data from 2020 to 2021), assuming that the COVID-19 pandemic led to higher aggregated heat loads. Then, the models were tested on comparatively low aggregated heat loads (test data from 2022). Consequently, the low simulated grid expansions might have compensated for the declining heat demand due to the energy crises. The models were already trained on higher heat loads. This has probably resulted in the better capability to handle the inserted concept drifts and explains the low effect of the 10% drift insertions on the baseline strategy (see e.g. Figure 13 or Figure 14 in Section 4.2) and the overpredictions of the baseline under no synthetic concept drift insertion (see Figure 15 and Figure 16 in Section 4.2).

However, the data (Figure 2 and Figure 4 in Section 2) indicate no major deviation between different years. Especially the drift insertions that correspond to adding 20% or 30% of the mean heat load over one-year (Section 3.1) have likely exceeded the effect of less heat load in the test data due to the energy crisis. Consequently, the baseline strategy proved to be affected by these drift insertions (e.g. Figure 13).

More time or computational capacity could have made cross-validation feasible, referring to training and testing the models with different data subsets (folds) and averaging the results subsequently (Géron, 2019). This could have increased the generalisability and mitigated the bias from the single train-test split. Correspondingly, the effect of the energy crisis on the test set could have been averaged out to some extent.

Concept Drift Insertion

Moreover, the findings of this study are limited to the simulated incremental grid expansions to introduce concept drift, which does not fully capture the complexities of the real world.

Besides *incremental* concept drifts, they may also happen *suddenly/abruptly* or *recurrent* (Gama et al., 2014). In reality, mixtures of many types can be observed (Gama et al., 2014). For example, grid expansions might happen suddenly if two individual grids are merged. This change might come along with, for example, heat-saving measures. Consequently, the analysed strategies to handle concept drift might have indicated different results. Further research could explore how different concept drift types, such as sudden concept drift, impact the performance of heat load forecasting models.

Direct Multi-step Forecasting Strategy

Furthermore, there are limitations regarding the DMSFS. The results show that offline models can cope with the inserted drift since the observed recent heat load features can capture change to some extent. In simple words, when predictions depend on previous heat loads, and the models have learnt that dependency, the models adjust the prediction when the previous heat loads have changed due to concept drift.

However, the strength of learning individual models for each forecast step (DMSFS) has a drawback in terms of concept drift handling. The importance of features to forecast different horizons varies (Liu et al., 2020; Mielck, 2023). The importance of observed recent heat load features decreases over the forecast horizon, whereas the importance of, e.g., the outdoor temperature, increases (Mielck, 2023). Hence, the capability to handle concept drift can decrease over the forecast horizon due to declining autocorrelation and therefore declining importance of observed recent heat load features. This effect was visible in Figure 15in Section 4.2.

Time series models that train one model and forecast recursively, such as the SARIMAX model, e.g., applied in the study by Grosswindhager et al. (2011), would not face this problem. A recursive forecasting strategy would only train one model that recursively generates predictions for the entire forecast horizon (Ben Taieb et al., 2012). However, a recursive multi-step forecasting strategy would not be able to learn the varying importance of features at different forecast steps. It could perform worse in the context of no or little concept drift. Additionally, recursive forecasting propagates forecast errors (Ben Taieb et al., 2012). The longer the forecast horizon, the more problematic this can be.

Strategies to Handle Concept Drift

Another important limitation is that this study did not compare all possible algorithms and strategies to handle concept drift in DHS. Instead, it focused on comparatively simple and explainable linear regression models to test online learning as well as different strategies to retrain offline models. Other forecasting algorithms, input variables, drift detectors or hyperparameters, such as retraining periods, retraining batch sizes or online learning rates, might change the results. However, the findings provide a general understanding of how different strategies handle concept drift and can guide decisions in practice.

Further research could use this study's methodological framework to analyse other approaches, such as recursive forecasting strategies, drift detectors or neural network- or ensemble-based algorithms. In addition, methods that separately model the level and the remainder (residuals) of a heat load time series could be explored, as they have been successfully applied in electricity load forecasting (Heidrich et al., 2022).

Evaluation

The study evaluated the capability of different learning approaches to handle concept drift by measuring the RMSE. The RMSE is symmetric and penalises overestimations and underestimations equally. However, from a security of supply point of view, there is a considerable difference between the over- and underestimation (Mbiydzenyuy et al., 2021). Overestimation of the demand might cause some amount of energy waste. In contrast, underestimation can make the district heating network fail at its primary function, for instance, to guarantee that heat is available on demand (Mbiydzenyuy et al., 2021). For example, an unscheduled peak load plant might have to ramp up, usually entailing high operation costs.

In contrast to the RMSE, the ME, shown in Equation (7) in Section 3.4, indicates over- and underestimation. However, the ME indicates systematic errors rather than accuracy since positive and negative errors might compensate for themselves even though they might be high. Since it is quite likely that concept drift leads to systematic errors, the ME can be an appropriate metric to access the sensitivity of strategies to concept drift, but it should be applied in combination with other metrics, such as the RMSE or MAPE, focussing on the general accuracy .

Accordingly, this study evaluates the strategies to handle concept drift through the MAPE. On the one hand, the MAPE allows an intuitive interpretation and gives a direct idea of the average percentage error. On the other hand, the MAPE is biased towards underestimation and, therefore, towards heat load increases such as grid expansions. This explains why the MAPE values decrease (Figure 14) even though the RMSE values are stable or increase (Figure 13). For example, if district heating grid reductions had been simulated instead of grid expansions, the true heat load values Q_i would have decreased. As Q_i is in the denominator of the MAPE, as shown in Equation (6) in Section 3.4, the MAPE values would have been greater even though the absolute errors might have remained constant. Moreover, viewing the MAPE from the previously discussed security of supply point of view, the MAPE has to be interpreted very carefully as it does the opposite of penalising underestimation higher than overestimations.

This work may be extended by investigating the impact of concept drift and the corresponding forecast errors in the context of a holistic energy system modelling approach. Then, this research could, for example, quantify the different impacts of over- and underprediction caused by concept drift.

Lastly, other criteria besides pure forecast performance could have been used. Time and memory are additional resource dimensions of a data stream learning process (Bifet et al., 2018). For example, offline learning and corresponding retraining with new data batches from scratch generally take more training time and memory usage since old data must be stored to retrain the model from the ground up. In contrast, online learning is comparatively fast and memory efficient. Nevertheless, in practice, these advantages of online learning might be of minor importance in forecasting the aggregated heat load in DHS. The necessary data is often stored in any case for several other analytical purposes. The computational speed of retraining offline models is probably also feasible as significant changes in the aggregated heat load, such as grid expansions, do not take place within seconds. In addition, simple models such as linear regression are quick and computationally cheap (Suryanarayana et al., 2018). Time and memory could be more important in terms of, for instance, IoT applications.

6 Conclusion

Heat load forecasting is crucial to optimise the operation of DHS. However, the changing relationship between input and output variables over time, known as concept drift, can negatively affect the forecast accuracy, which can further influence the optimal operation.

This article addressed this problem and provides guidance to practitioners in securing stable heat load forecasts under concept drift. It fills the research gap by investigating the research question "How can different learning strategies be applied to data-driven heat load forecasting models to handle concept drift in district heating systems?". The focus was on grid expansions to synthetically inserted incremental concept drifts. Offline and online learning strategies were analysed in a data stream environment using linear regression to predict hourly heat load values for a one-week horizon. The strategies were evaluated on the performances measured by RMSE, MAPE and ME.

The results indicate that daily retraining and online learning can sufficiently handle concept drift, as they were not sensitive to different magnitudes of incremental concept drift insertion. In contrast, the baseline strategy was sensitive to concept drift. Including observed recent heat load features could mitigate that sensitivity.

Taking together, this offers a novel perspective on handling concept drift in DHS: Including observed recent heat load features can mitigate the requirement for retraining models or online learning. The shorter the forecast horizon and the lower the concept drift, the less the necessity for strategies to handle concept drift. Putting the results into the perspective of the case area and the corresponding aggregated heat load, retraining or online might not be necessary, as the baseline did not significantly perform worse than the other strategies in terms of the absolute errors.

However, daily retraining with data from the most recent three months and online learning could sufficiently handle concept drift, as it was not sensitive to the magnitude of the synthetic concept drift insertion. Therefore, this study generally recommends frequent, e.g. daily, retraining and online learning since the concept drift might be stronger in practice, especially when change happens over several years. The specific intervals and data batch sizes of retraining may be adjusted with regard to the expected change and the degree of the retraining automatisation in practice. Due to low computational costs and high explainability, linear regression models can be a good choice in the district heating industry.

Future research could investigate additional drift detectors, forecasting algorithms and strategies, or types of concept drift. Furthermore, future studies may extend this study by analysing the impact of concept drift through a holistic energy system modelling approach.

Acknowledgement

The present article's author would like to thank Meng Yuan (Aalborg University), Anders N. Andersen (EMD International A/S), Lasse Svenningsen (EMD International A/S) and Jonathan Refsgaard (EMD International A/S) for the supervision and discussions throughout this study.

References

Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, *39*(8), 7067–7083. https://doi.org/10.1016/j.eswa.2012.01.039

Bifet, A., Gavaldà, R., Holmes, G., & Pfahringer, B. (2018). *Machine Learning for Data Streams: With Practical Examples in MOA*. The MIT Press. https://doi.org/10.7551/mitpress/10654.001.0001

Dahl, M., Brun, A., Kirsebom, O., & Andresen, G. (2018). Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data. *Energies*, *11*(7), 1678. https://doi.org/10.3390/en11071678

Danish Property Assessment Agency. (2023). Bygnings- og Boligregistret (BBR). Retrieved March 15, 2023, from https://bbr.dk/forside

EMD International A/S. (2023). Ringkøbing District Heating - Live Operation Data. Retrieved March 15, 2023, from https://www.energyweb.dk/rfvv/

Fang, T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied Energy*, *179*, 544–552. https://doi.org/10.1016/j.apenergy.2016.06.133

Finkenrath, M., Faber, T., Behrens, F., & Leiprecht, S. (2022). Holistic modelling and optimisation of thermal load forecasting, heat generation and plant dispatch for a district heating network. *Energy*, *250*, 123666. https://doi.org/10.1016/j.energy.2022.123666

Frederiksen, S., & Werner, S. (2013). *District heating and cooling*. Studentlitteratur AB.

Frost, J. (2019). *Regression analysis: An intuitive guide for using and interpreting linear models*. Statisics By Jim Publishing.

Fumo, N. (2014). A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews*, *31*, 53–60. https://doi.org/10.1016/j.rser.2013.11.040

Galli, S. (2021). Feature-engine: A Python package for feature engineering for machine learning. *Journal of Open Source Software*, *6*(65), 3642. https://doi.org/10.21105/joss.03642

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37. https://doi.org/10.1145/2523813

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Geysen, D., De Somer, O., Johansson, C., Brage, J., & Vanhoudt, D. (2018). Operational thermal load forecasting in district heating networks using machine learning and expert advice. *Energy and Buildings*, *162*, 144–153. https://doi.org/10.1016/j.enbuild.2017.12.042

Grosswindhager, S., Voigt, A., & Kozek, M. (2011). Online Short-Term Forecast of System Heat Load in District Heating Networks. *ISF 2011 - Prague PROCEEDINGS - 31st International Symposium on Forecasting*. Retrieved April 7, 2023, from http://hdl.handle.net/20.500.12708/66419

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Heidrich, B., Ludwig, N., Turowski, M., Mikut, R., & Hagenmeyer, V. (2022). Adaptively coping with concept drifts in energy time series forecasting using profiles. *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, 459–470. https://doi.org/10.1145/3538637.3539759

Idowu, S., Saguna, S., Åhlund, C., & Schelén, O. (2016). Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings*, *133*, 478–488. https://doi.org/10.1016/j.enbuild.2016.09.068

Ikonomovska, E., Gama, J., & Džeroski, S. (2011). Learning model trees from evolving data streams. *Data Mining and Knowledge Discovery*, *23*, 128–168. https://doi.org/10.1007/s10618-010-0201-y

Jan, J. (2021). *Adaptive Algorithm for Forecasting of Medium-term District Heating Demand* (Master's thesis). KTH, School of Electrical Engineering and Computer Science (EECS). Stockholm, Sweden. Retrieved April 7, 2023, from https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1544226&dswid=-3172

Johansson, C., Bergkvist, M., Geysen, D., Somer, O. D., Lavesson, N., & Vanhoudt, D. (2017). Operational Demand Forecasting In District Heating Systems Using Ensembles Of Online Machine Learning Algorithms. *Energy Procedia*, *116*, 208–216. https://doi.org/10.1016/j.egypro.2017.05.068

Kurek, T., Bielecki, A., Świrski, K., Wojdan, K., Guzek, M., Białek, J., Brzozowski, R., & Serafin, R. (2021). Heat demand forecasting algorithm for a Warsaw district heating network. *Energy*, *217*, 119347. https://doi.org/10.1016/j.energy.2020.119347

Liu, Y., Hu, X., Luo, X., Zhou, Y., Wang, D., & Farah, S. (2020). Identifying the most significant input parameters for predicting district heating load using an association rule algorithm. *Journal of Cleaner Production*, *275*, 122984. https://doi.org/10.1016/j.jclepro.2020.122984

Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th Generation District Heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. *Energy*, *68*, 1–11. https://doi.org/10.1016/j.energy.2014.02.089

Machlev, R., Heistrene, L., Perl, M., Levy, K., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, *9*, 100169. https://doi.org/10.1016/j.egyai.2022.100169

Mbiydzenyuy, G., Nowaczyk, S., Knutsson, H., Vanhoudt, D., Brage, J., & Calikus, E. (2021). Opportunities for Machine Learning in District Heating. *Applied Sciences*, *11*(13), 6112. https://doi.org/10.3390/app11136112

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceeding of the 9th Python in Science Conference (SCIPY 2010)*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Metcalfe, A. V., & Cowpertwait, P. S. (2009). *Introductory Time Series with R*. Springer New York. https://doi.org/10.1007/978-0-387-88698-5

Mielck, K. (2023). *Permutation-based Feature Importance Analysis for Medium-Term Heat Load Forecasting in District Heating Systems Through Machine Learning* (Student report). Aalborg University. Aalborg.

Montiel, J., Halford, M., Mastelini, S. M., Bolmier, G., Sourty, R., Vaysse, R., Zouitine, A., Gomes, H. M., Read, J., Abdessalem, T., & Bifet, A. (2021). River: Machine learning for streaming data in Python. *The Journal of Machine Learning Research*, *22*(1), 4945–4952. https://doi.org/10.5555/3546258.3546368

Neter, J., Kutner, M., Wasserman, W., & Nachtsheim, C. (1996). *Applied linear statistical models* (4th ed). McGraw-Hill.

Ntakolia, C., Anagnostis, A., Moustakidis, S., & Karcanias, N. (2022). Machine learning applied on the district heating and cooling sector: A review. *Energy Systems*, *13*(1), 1–30. https://doi.org/10.1007/s12667-020-00405-9

OpenStreetMap Contributors. (2023). OpenStreetMap. Retrieved May 16, 2023, from https://www.openstreetmap.org

OpenWeather. (2023). History Forecast Bulk. Retrieved March 15, 2023, from https://openweathermap.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Potočnik, P., Škerl, P., & Govekar, E. (2021). Machine-learning-based multi-step heat demand forecasting in a district heating system. *Energy and Buildings*, *233*, 110673. https://doi.org/10.1016/j.enbuild.2020.110673

Provatas, S. (2014). *An Online Machine Learning Algorithm for Heat Load Forecasting in District Heating Systems* (Master's thesis). Blekinge Institute of Technology. Retrieved December 18, 2022, from https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A830782&dswid=-3808

Provatas, S., Lavesson, N., & Johansson, C. (2014). An Online Machine Learning Algorithm For Heat Load Forecasting In District Heating Systems. *The 14th International Symposium on District Heating and Cooling, Stockholm, Sweden.*

Read, J., & Zliobaite, I. (2023). Learning from Data Streams: An Overview and Update. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4326595
Preprint

Ringkøbing Fjernvarme. (2023a). Ledningsnet. Retrieved April 6, 2023, from https://www.rfv.dk/vores-ledningsnet/

Ringkøbing Fjernvarme. (2023b). Om Os - Vores Historie. Retrieved February 20, 2023, from https://www.rfv.dk/om-os/vores-historie/

River-Contributors. (2023). River API Documentation - SGD. Retrieved May 3, 2023, from https://riverml.xyz/0.15.0/api/optim/SGD/

Robbins, H. E. (1951). A Stochastic Approximation Method. *Annals of Mathematical Statistics*, *22*, 400–407. https://doi.org/10.1214/AOMS/1177729586

Rusovs, D., Jakovleva, L., Zentins, V., & Baltputnis, K. (2021). Heat Load Numerical Prediction for District Heating System Operational Control. *Latvian Journal of Physics and Technical Sciences*, *58*(3), 121–136. https://doi.org/10.2478/lpts-2021-0021

Schaffer, M., Tvedebrink, T., & Marszal-Pomianowska, A. (2022). Three years of hourly data from 3021 smart heat meters installed in Danish residential buildings. *Scientific Data*, *9*(1), 420. https://doi.org/10.1038/s41597-022-01502-3

scikit-learn. (2022). Scikit-learn User Guide. Retrieved December 18, 2022, from https://scikit-learn.org/stable/user_guide.html

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceeding of the 9th Python in Science Conference (SCIPY 2010)*, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Suryanarayana, G., Lago, J., Geysen, D., Aleksiejuk, P., & Johansson, C. (2018). Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods. *Energy*, *157*, 141–149. https://doi.org/10.1016/j.energy.2018.05.111

Wang, Z., Hong, T., & Piette, M. A. (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, *263*, 114683. https://doi.org/10.1016/j.apenergy.2020.114683

Wei, Z., Zhang, T., Yue, B., Ding, Y., Xiao, R., Wang, R., & Zhai, X. (2021). Prediction of residential district heating load based on machine learning: A case study. *Energy*, *231*, 120950. https://doi.org/10.1016/j.energy.2021.120950

Wojdyga, K. (2008). An influence of weather conditions on heat demand in district heating systems. *Energy and Buildings*, *40*(11), 2009–2014. https://doi.org/10.1016/j.enbuild.2008.05.008

Xue, P., Jiang, Y., Zhou, Z., Chen, X., Fang, X., & Liu, J. (2019). Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. *Energy*, *188*, 116085. https://doi.org/10.1016/j.energy.2019.116085

Zhao, B., Jin, Y., Li, W., & Zheng, H. (2022). Analysis on the Technical Situation and Applied Difficulties of District Heating Load Forecasting. *Thermal Engineering*, *69*(6), 464–472. https://doi.org/10.1134/S0040601522060088