DATA-DRIVEN MACHINE LEARNING MODELS FOR COMPLEX BUILDING ENERGY LOAD PREDICTIONS: ANALYSIS AND ASSESSMENT OF CONCEPT DRIFT SCENARIOS

Kai Derzsi[1]*, Florian Stinner[1], Matthias Fenske[1], Thomas Schreiber[1], Dirk Müller[1]

[1] E.ON Energy Research Center, Institute for Energy Efficient Buildings and Indoor Climate, RWTH Aachen University, Germany

*Corresponding Author: kai.derzsi@eonerc.rwth-aachen.de

ABSTRACT

Optimized predictive control of complex building energy systems requires accurate load predictions. Today, increasing sensor installations and obtainable monitoring data enable exploiting continuous data streams and developing data-driven load prediction models. Recent research on load prediction has focused on model development while assuming static relationships. However, real-world data streams are usually evolving over time mainly due to two types of interactions with the environment. First, underlying relationships in energy monitoring data are strongly influenced by user behavior, which is only implicitly represented in timeseries data. Second, system changes lead to concept drifts and deteriorate the prediction performance of data-driven models. To capture the two underlying types of interactions and to increase the prediction accuracy, drift detection methods enable adaptive machine-learning models. In the present study we trained random forest models on monitoring data on two scales of building energy systems representing two types of concept drifts. In particular, we modeled the power consumption of a public building to investigate sudden concept drifts according to user behavior and the heat load of a district heating system to examine gradual concept drifts according to system changes. To analyze passive and active model adaption strategies, we implemented periodical retraining schemes and state-of-the-art concept drift detection methods. Comparing static reference models to the proposed adaptive machine learning models, we demonstrated model deterioration in connection to concept drifts and the potentials for performance enhancements. The results show that model adaption strategies are promising solutions to ensure accurate demand predictions for optimized energy management.

1 INTRODUCTION

The ambitious climate targets set by the Paris Agreement underscore the urgent need for a substantial reduction in CO_2 emissions. Notably, the building sector plays a significant role, accounting for 30 % of global energy consumption and 27 % of global operational CO_2 emissions (Hamilton *et al*., 2022).
Advanced predictive control approaches emerge as promising solutions to contribute to set goals by optimizing and efficiently managing building energy system operations. However, the successful realization of these approaches requires accurate load prediction models. In recent years data-driven load prediction models for individual building energy systems have gained significant research attention. With the increasing installation of low-cost sensors, data-driven models hold the ability to exploit readily available monitoring data (Wang *et al*., 2019), while decreasing modeling cost and increasing prediction accuracy in comparison to physics-based models (Zhang *et al.*, 2021). Data-driven modeling in building energy systems has mainly focused on the development of accurate models assuming stationary concepts. Yet, monitoring data is accessed in the form of continuous and potentially unbound data streams, that are closely associated with non-stationary behavior and dynamically changing concepts (Bayram *et al*., 2022), (Wares *et al*., 2019). This work addresses the challenge of non-stationary data streams in building energy systems, which can occur due to shifts in user behavior, equipment, regulations, or environmental factors, rendering previously trained models obsolete.

https://doi.org/10.52202/077185-0216

In general, data-driven models map a joint data distribution $p_t(X, y)$ of input features X and target features y and implicitly represent a data generating process or an underlying concept at a point in time t (Read and Žliobaitė, 2023). The term concept drift describes the dynamic and unexpected change of a concept over time (Gama *et al*., 2014). Concept drift is commonly defined in terms of Bayesian decision theory as the change of a joint data distribution $p_t(X, y)$ between two points in time t_0 and t_1. This change is either or both affected by differences in the prior probabilities $p_t(y)$ and the conditional probabilities $p_t(X|y)$ of the target feature, while a change in the prior probabilities of the input features $p_t(X)$ does not necessarily constitute a concept drift (Webb *et al*., 2016). Real concept drift is considered as a change that affects the conditional probabilities of the target feature $p_t(X|y)$ with or without an effect on the prior probabilities of the input features $p_t(X)$, whereas virtual drift only affects the prior probabilities of the input features $p_t(X)$ (Webb *et al*., 2016). The most common differentiation of concept drift is characterized by the temporal pattern of change as sudden drift, gradual drift, and recurring drift (Webb *et al*., 2016).

Concept drift in timeseries forecasting is attributed to behavioral changes in populations, aging effects in technical devices, climate fluctuations and technological progress (Ditzler *et al*., 2015)., which are all to be expected in building energy systems. The occurrence of concept drift leads to model performance deterioration (Bayram *et al*., 2022). This circumstance creates a necessity for the utilization of dynamic model adaption to changing concepts in the practical application of load prediction models.

To address concept drift handling in non-stationary environments, common approaches leverage traditional batch learning in single models, allowing the utilization of established machine learning models by retraining new models from scratch and replacing old, expired models (Hoi *et al*., 2021). The model adaption process is typically distinguished between passive- and active approaches. Passive model adaption approaches enforce blind model retraining at a fixed rate. Active model adaption approaches enforce an informed model retraining, that is triggered by explicit concept drift detection methods (Ditzler *et al*., 2015).

Research on concept drift handling in non-stationary environments has been focused on classification tasks and was mainly conducted on artificial data sets (Bayram *et al*., 2022). Only a few recent studies have addressed this challenge for timeseries regression tasks, especially in the context of building energy systems.

Mehmood et al. (2021) evaluated a selection of state-of-the art drift detection methods for active model adaption in regression tasks on two artificial and two real datasets. In artificial data sets the active approaches utilizing the Drift Detection Method (DDM) and Adaptive Windowing (ADWIN) as drift detectors, achieved the best average performances. Nevertheless, in the real datasets model adaption did not yield performance improvements compared to a static model.

Ji et al. (2021) compared passive and active model adaption approaches on a Long-Short Term Memory network (LSTM) for load prediction using real power consumption data from an industrial park located in the Minhang district of Shanghai, China, spanning from June to December 2019. Initial training used data from the first month, with prequential evaluation spanning over the subsequent five months. Passive model adaption involved daily retraining rates with one week of training data size, while the active approach used ADWIN for concept drift detection. Both approaches showed comparable performance, with the active approach offering reduced processing times and memory usage.

Mariano-Hernández et al. (2022) examined passive and active model adaption approaches on a selection of advanced machine learning models for multistep-ahead load prediction. They used real power consumption data from two single buildings at the University of Valladolid in Spain from 2016 to 2019, that reflected sudden user behavioral changes and gradual efficiency improvements. Initial training was carried out on the first three years with evaluation on the last year. The study compared static models with passive model adaption of daily retraining and active model adaption utilizing ADWIN and Kolmogorov-Smirnov Windowing (KSWIN) for concept drift detection. All adapted models outperformed the static model on average performance metrics, with the passive approach showing the best performance. The active approach applying KSWIN achieved comparable results while reducing retraining frequency by more than half.

Three timeseries regression tasks within a single building were conducted, employing various model types, by Toquica et al. (2020). Initial training spanned one month, followed by evaluation of passive

https://doi.org/10.52202/077185-0216

model adaption methods with training data sizes ranging from 7 to 28 days, and active adaption approaches utilizing different concept drift detection methods over one year. The active approach, particularly with ADWIN, demonstrated superior performance, although average model performance exhibited minimal differences of 2 to 5% across approaches. Notably, the passive adaption method benefitted from smaller training data sizes.

In summary, timeseries data in building energy systems often exhibit non-stationary behavior, and leveraging model adaption methods significantly enhances load prediction performance. However, research on non-stationary learning for regression tasks in building energy systems remains limited, mostly focusing on single buildings (Mariano-Hernández *et al.*, 2022), (Toquica *et al*., 2020). While passive and active approaches show no clear performance preference, applying concept drift detectors developed for classification tasks, most notably DDM and ADWIN, effectively reduces retraining numbers (Mehmood *et al.*, 2021), (Ji *et al.*, 2021), (Mariano-Hernández *et al.*, 2022), (Toquica *et al*., 2020). Yet, crucial parameters such as retraining rate and training data size are not extensively discussed within these batch learning frameworks. Explicit concept drift occurrence is often overlooked, with some evaluations seeming to consider incomplete initial training data rather than necessarily actual concept drift instances (Ji *et al.*, 2021), (Toquica *et al*., 2020).

In our study, we analyze two real-world use cases for load prediction demonstrating non-stationary behavior in building energy systems with sudden drift in a single building and gradual drift in a district heating network. We demonstrate static model deterioration and compare passive and active model adaption solutions, focusing on average performance, but also on error distributions, retraining rates, and training data sizes. Additionally, we explicitly discuss drift occurrence and concept drift detection.

The remainder of this paper is organized as follows: In Section 2, we outline the use cases, the data-driven model, and the model adaption approaches. Section 3 presents the results of one year of prequential evaluation simulating the practical application of initially trained models. Finally, Section 4 concludes with a summary.

2 METHODOLOGY

2.1 Use Cases

Two real-world datasets were considered as use cases for data-driven load prediction with a forecast horizon of one hour, each covering a two-year timeframe at an hourly resolution and representing two different scales of building energy systems. The first dataset comprises power consumption monitoring data from a daycare building located in a large city in North Rhine-Westphalia, Germany from January 2019 to December 2020. The second dataset consists of heat consumption monitoring data from a district heating network situated in Munich, Germany spanning over the timeframe from January 2020 to December 2021.

First, the daycare building demonstrates a strong correlation between power consumption and user behavior, influenced by daily and weekly schedules, as well as seasonal variations, particularly during vacation periods. Notably, the dataset captures sudden drifts in 2020 due to behavioral shifts in building usage, triggered by the closure of the daycare starting on March 16th and an emergency operation commencing on May 15th, following government regulations amid the COVID-19 pandemic outbreak (Stinner *et al.*, 2021).

Second, the district heating grid exhibits behavior strongly associated with seasonal changes on a larger scale. Over time, a gradual drift occurs, linked to the increasing heat demand within the city district supplied by the grid. This concept drift is particularly noticeable during the heating season, reflecting the expansion and connection of new buildings to the grid.

2.2 Data-Driven Model

The data-driven models are based on the standard implementation of the Random Forest Regressor (RF) in the python package scikit-learn (Pedregosa et al., 2018). The RF model was chosen due to the demonstrated effectiveness and robustness of tree-based algorithms in terms of hyperparameter tuning for complex timeseries regression tasks (Elsayed *et al*., 2021).

For both use cases, equal input features were selected, to map their relationships with their respective target features of power consumption and heat consumption. These input features included weather data

https://doi.org/10.52202/077185-0216

as outdoor temperature, timestamp data as information on public holidays, information on school holidays, weekday, hour of the day and week of the year and target feature lags of one hour for the daycare building and three hours for the district heating network. Furthermore, the input features hour of the day and week of the year were cyclically encoded. The number of target feature lags was determined through a Partial Auto-Correlation Function (PACF) analysis.

2.3 Model Adaption

In both use cases, initial models were trained on the first year of available data instances. Subsequently, prequential validations were conducted for the remaining second year to simulate the practical deployment of the models in data stream settings. These initial models, without model adaption, served as static references. Furthermore, passive model adaption approaches were evaluated along with active model adaption approaches employing DDM and ADWIN as drift detection methods, within a single model batch learning framework. This framework allows for the utilization of any arbitrary traditional machine learning model, such as RF.

2.3.1 Passive approach

The passive model adaption approach involves periodically retraining a new model from scratch with a specified retraining rate and training data size based on sliding windows. To optimize this process, a grid search was conducted, exploring retraining rates ranging from 1 day to 28 days and training data sizes from 7 days to 182 days. The optimal configuration for load prediction in the daycare building and the district heating network was found to be a retraining rate of 1 day. For the daycare building, the optimal training data size was determined to be 56 days, while for the district heating network, it was 28 days.

In the subsequent evaluation, only these optimal configurations for passive model adaptation are further considered. However, it is noteworthy that across both use cases every combination of retraining rate and training data size for passive model adaptation outperformed static models in terms of average performance metrics during the validation period.

2.3.2 Active approach

The active model adaption approach encompasses informed retraining a new model from scratch upon explicit concept drift detection. Most concept drift detection methods assume that concept drift must arise in conjunction with model deterioration. Thereby, these supervised algorithms rely on the identification of statistical changes in model performance over time (Hu *et al.*, 2020), (Gemaque *et al.*, 2020). DDM and ADWIN are widely regarded as state-of-the-art drift detection methods for learning in non-stationary environments and were further utilized in the active model adaption approach (Wares *et al.*, 2019).

2.3.2.1 *Drift Detection Method*

DDM is founded on the assumption that model error rates remain stable or decrease over time for stationary concepts. Arriving data instances from a data stream are processed sequentially for each timestep i computing the error rate e_i and the standard deviation of the error rate σ_i for a growing landmark window. Minima for error rate $e_{\min}$ and standard deviation $\sigma_{\min}$ are continuously reevaluated and updated with each new data instance. Trigger conditions are established based on selected confidence intervals for the error rate statistics, with warning and drift levels corresponding to confidence levels of 95 % and 99 %, as respectively outlined in Equations (1) and (2).

$$e_i + \sigma_i \geq e_{\min} + 2 * \sigma_{\min} \tag{1}$$

$$e_i + \sigma_i \geq e_{\min} + 3 * \sigma_{\min} \tag{2}$$

Upon detecting a drift, a context window is created, serving as a data buffer representing the current concept. This window includes a minimum number of historic data instances and all data instances that arrived between the warning and drift levels, used as training data for model retraining. (Gama *et al.* 2004)

https://doi.org/10.52202/077185-0216

In both use cases, the minimum size of the context window was set to 7 days. Following drift detection and model retraining, the landmark window was reset. Additionally, the error threshold for error rate calculation was estimated using the Interquartile Range (IQR) Method on the Mean Absolute Error (MAE) of the most recent model.

2.3.2.2 *Adaptive Windowing*

ADWIN operates under the premise that the average model error remains stable over time for stationary concepts. New data instances from a data stream are sequentially processed, utilizing an adaptive window W of length n that extends with each arriving data instance until drift detection is triggered. This adaptive window is then partitioned into two sufficiently large sub-windows, W_0 and W_1, each with lengths n_0 and n_1, respectively. The expected values of the errors $\hat{\mu}_{W_0}$ and $\hat{\mu}_{W_1}$ are calculated and their difference compared against a threshold ϵ_{cut} given in Equations (3).

$$\left|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}\right| \geq \epsilon_{\text{cut}} \tag{3}$$

The threshold ϵ_{cut} is computed according to Equations (4) to (6) using the harmonic mean of the sub-window lengths m, the observed variance of performance metric values within the adaptive window ${\sigma_W}^2$, and the confidence value δ.

$$m = \frac{1}{\frac{1}{n_0} + \frac{1}{n_1}} \tag{4}$$

$$\delta' = \frac{\delta}{n} \tag{5}$$

$$\epsilon_{\text{cut}} = \sqrt{\frac{2}{m} * {\sigma_W}^2 * \ln\left(\frac{2}{\delta'}\right)} + \frac{2}{3 * m} * \ln\left(\frac{2}{\delta'}\right) \tag{6}$$

Drift detection is triggered when the difference between the expected error values of the two sub-windows exceeds the threshold, at which point the adaptive window W is replaced by and cut to the most recent sub-window W_1 representing the new concept. The process of sub-window generation is repeated for all possible combinations of sub-windows data instance by data instance. The training data for model retraining is determined by the most recent sub-window at drift detection. (Bifet and Gavaldà, 2007)

The minimum window size was fixed to 7 days. Furthermore, grid search was performed for both use cases to determine the confidence value δ, with values ranging from 0.7 to 0.0001, and set to the optimal value of 0.001 in terms of average performance over the validation timeframe.

3 RESULTS AND DISCUSSION

For the evaluation of model adaption approaches, prequential validation was conducted over the most recent years of monitoring data for both the daycare building and the district heating network, respectively. Performance was accessed using Maximum Error (E_{max}) and average performance metrics including MAE, Coefficient of Determination (R^2), and Coefficient of the Variation of the Root Mean Square Error (CVRMSE) and absolute error distributions. Additionally, the number of retrained models (n_{models}) and the average training data size in terms of the timeframe ($t_{training}$) in days were tracked over the validation timeframe. Finally, exemplary occurrences of detected drifts are thoroughly discussed in detail.

3.1 Performance Evaluation

The results of various model adaption approaches are summarized in Table 1 for power load prediction in the daycare building and in Table 2 for heat load prediction in the district heating network.

The average performance metrics for power load prediction in the daycare building do not clearly indicate an advantage of the model adaption approaches over the static model. Although maximum

https://doi.org/10.52202/077185-0216

observed errors may tend to be higher and the MAE of the active model adaption approach applying DDM is slightly elevated, all model adaption approaches demonstrate improved ability to explain the variability of the energy systems behaviour, as evidenced by higher R^2 values. Notably, only the passive model adaptation approach shows significant improvement in prediction performance, reducing the static model's CVRMSE by approximately 24 %.

Table 1: Number of retrained models, average training data size, maximum error and average performance metrics of the load prediction for the daycare building for one year of model validation

Model Adaption	n_{models}	$t_{training}$	E_{max}	MAE	R^2	CVRMSE
Static	1	-	5,835 kW	0,479 kW	0,873	0,409
Passive	365	56 d	6,078 kW	0,355 kW	0,926	0,312
DDM	10	7,17 d	7,094 kW	0,503 kW	0,877	0,402
ADWIN	27	10,38 d	6,28 kW	0,436 kW	0,895	0,371

In contrast, average performance metrics for predicting heat load in the district heating network clearly demonstrate that all model adaption approaches lead to performance improvements. This underscores a more notable impact of the observed gradual drift, arising from the expansion of the city district, on the static model's performance.

Table 2: Number of retrained models, average training data timeframe, maximum error and average performance metrics of the load prediction for the district heating network for one year of model validation

Model Adaption	n_{models}	$t_{training}$	E_{max}	MAE	R^2	CVRMSE
Static	1	-	1.411 kW	124 kW	0,830	0,233
Passive	365	28 d	1.353 kW	72 kW	0,923	0,159
DDM	16	7,50 d	1.149 kW	97 kW	0,869	0,204
ADWIN	34	7,75 d	1.305 kW	90 kW	0,897	0,181

Overall, the comparison between the two use cases reveals that load prediction in the district heating network performed better, showcasing lower CVRMSE values.
In both cases, the passive model adaption approach outperformed other models, followed by the active model adaption approach using ADWIN for concept drift detection, although the results were indistinct for the active model adaption approach utilizing DDM.
Active model adaption substantially reduced retraining rates compared to passive model adaption over the one-year validation period, with DDM triggering retraining 53 to 63 % less frequently than ADWIN. The training data sizes determined by concept drift detection algorithms for model retraining are generally smaller than those selected by grid search for passive approaches and appear to approach the set minimum data instance buffers of 7 days. Notably, the average training data size of ADWIN reflects a trend towards shorter training timeframes for the district heating network compared to power load prediction in the daycare building. This observation is consistent with the results of the performed grid search for the passive model adaption approach, suggesting effective data usage.
Figure 1 displays absolute error statistics for load predictions over the one-year validation period for both use cases over the model adaption approaches.
The results confirm the observed behavior, highlighting the strong performance of the passive model adaption approach. Additionally, the greater benefit of model adaption approaches in predicting the heat load for the district heating network, where the static model exhibits a significantly wider range of absolute errors, is further emphasized.

https://doi.org/10.52202/077185-0216

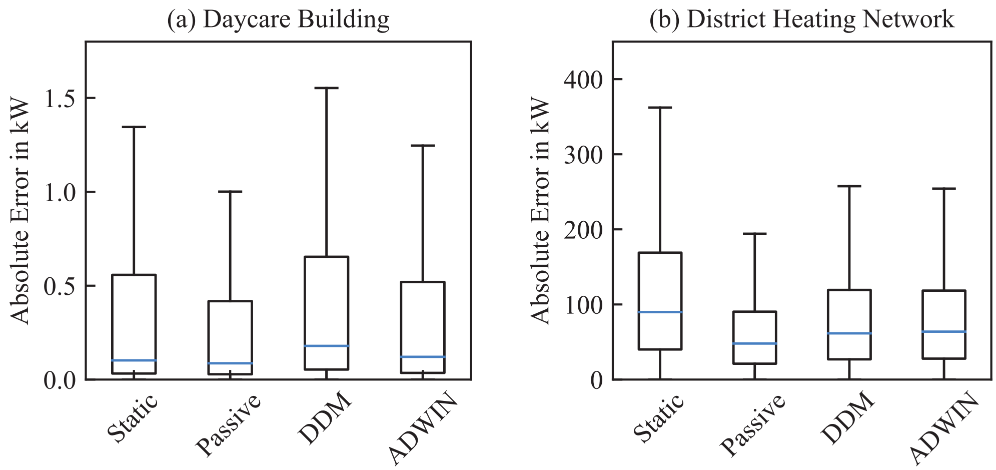

Figure 1: Absolute error statistics of the load predictions for one year validation timeframes over the model adaption method

3.2 Concept Drift Detection

In Figure 2, the monitoring datasets spanning the entire two-year period for both the power consumption of the daycare building and the heat consumption of the district heating network are depicted.

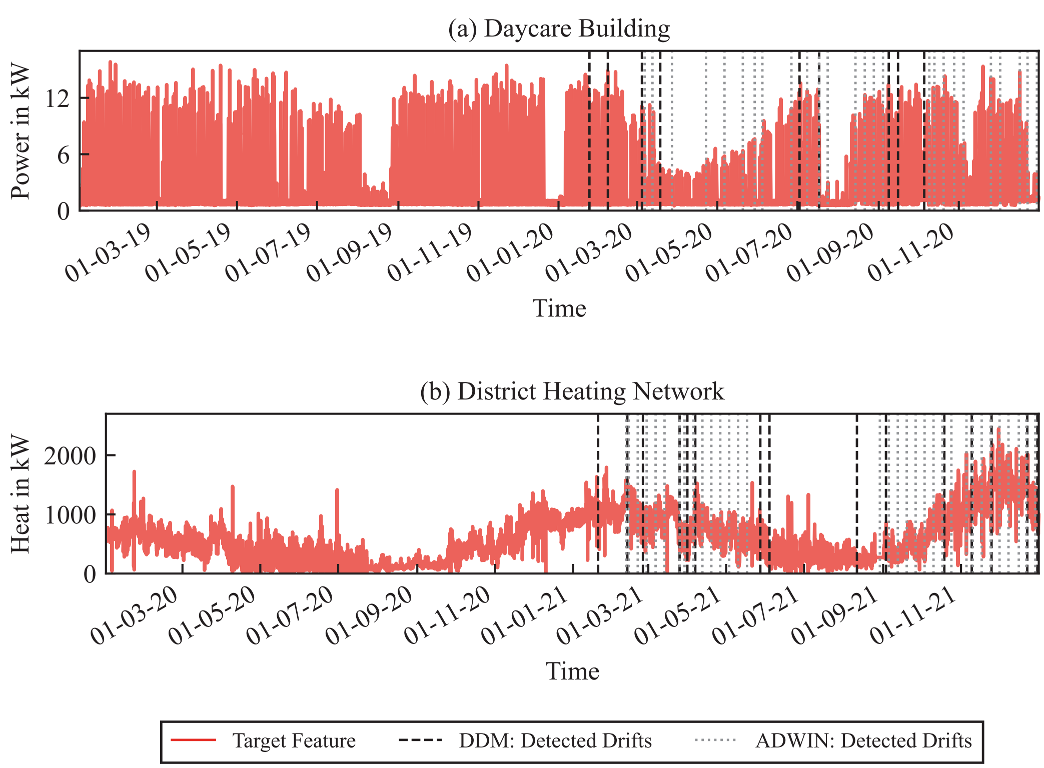

Figure 2: Overview of the monitoring data of the target feature over the first years of initial model training and the second years of model validation with occurrence drift detections for DDM and ADWIN in the active model adaption approaches

https://doi.org/10.52202/077185-0216

The first year was dedicated to initial model training, while the second year was utilized for prequential validation, simulating practical model deployment.

The occurrences of concept drift detections for both DDM and ADWIN are highlighted. The highlights visualize the variations for the retraining rates of the active model adaption approaches applying DDM and ADWIN. The higher rates for both DDM and ADWIN in the district heating network compared to the daycare building can be attributed to differences in observed change characteristics, specifically gradual drift versus sudden drift.

In the power consumption of the daycare building, both drift detection methods trigger retraining following the daycare closure on March 16th, 2020, during the COVID-19 pandemic outbreak as sudden drift. Subsequently, new drift occurrences are signaled continuously as the behavior appears to gradually change from spring to summer and especially ADWIN not fully recovering the change in the rest of the year missing discarded information on past concepts.

Throughout the 2021 heating seasons in the district heating network, characterized by increased heat demand compared to the previous year, both drift detection methods consistently trigger model retraining. However, during the summer period, both concept drift detection methods maintain a stable model.

3.2.1 Sudden Drift Occurrence

Exemplary data from the validation timeframe of the daycare building spanning the timeframe from March 9th to March 22nd, 2020, is presented in Figure 3. The figure displays monitoring data, alongside power load predictions for both the static model and the active model adaption approach using DDM.

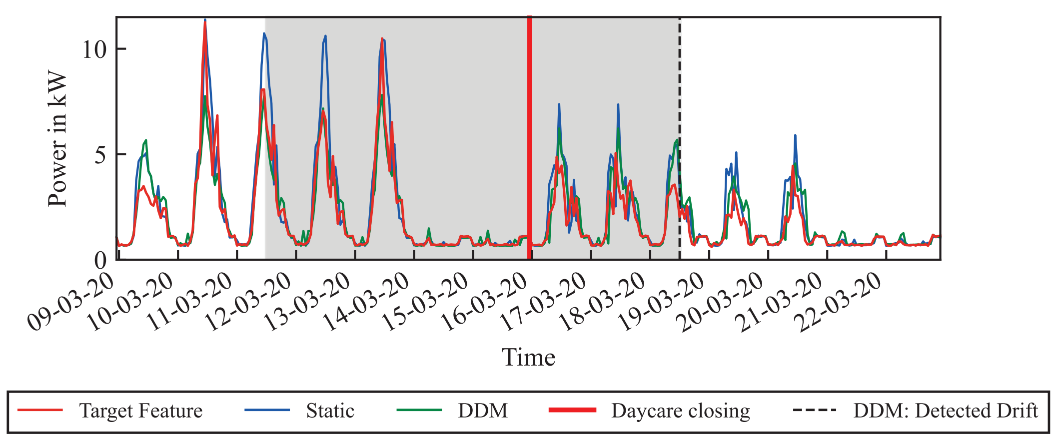

Figure 3: Two weeks of power load predictions and drift detection with highlighted training data for a daycare building with closure starting on March 16th, 2020, during the COVID-19 pandemic outbreak within the validation timeframe

The closure of the daycare on March 16th during the COVID-19 pandemic outbreak and the detected concept drift on March 19th is marked. The highlighted training data also indicates the warning level, signalled by deviations in prediction performance in the preceding week and triggered retraining during the first week of the closure once the drift level was reached. Following retraining and model replacement, the adaptive model demonstrates improved estimation of daily load peaks compared to the static model. Additionally, the adaptive approach may provide valuable information, as behavioural changes were correctly identified. It is worth noting that also the static model shows adaptability within certain boundaries, potentially attributed to the utilization of lag features by the random forest regressor, while the new behaviour remains within the previously observed physical system's bounds.

https://doi.org/10.52202/077185-0216

3.2.2 Gradual Drift Occurrence

Figure 4 presents two weeks of exemplary data from the validation timeframe of the district heating network during the heating season from November 29th to December 12th, 2021. The showcased data includes the monitoring data of the heat consumption in the city district and the load predictions of both the static model and the active model adaption approach using ADWIN, with highlighted drift detection occurrences. During the first week of the observed timeframe, neither ADWIN nor the static model accurately match the heat demand level of the district heating network. ADWIN triggers model retraining, which visibly improves performance. However, another concept drift is detected during the second week, likely due to errors in the initial operation of the new model, highlighting instability for the selected minimum window size. Hence, high model retraining rates were observed. After the second model retraining, the model appears to accurately follow the measured heat load.

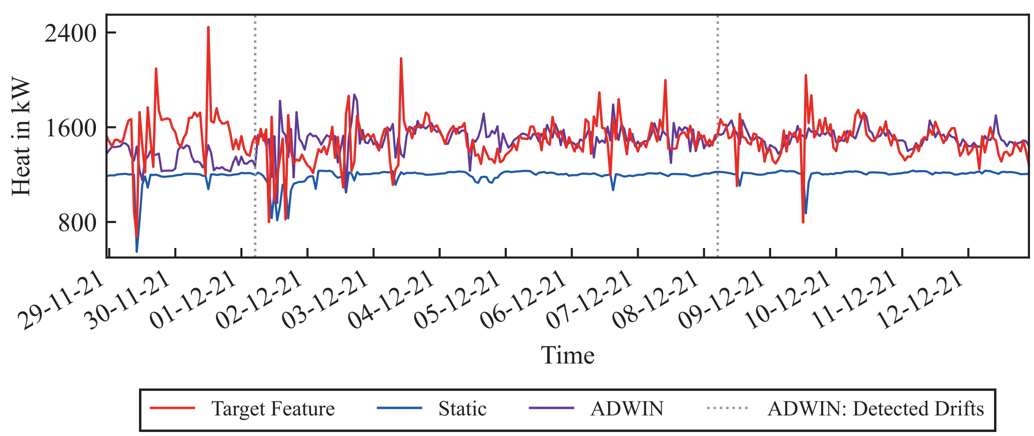

Figure 4: Two weeks of heat load prediction and drift detections for a district heating grid with increased heat demand during the winter within the validation timeframe

In contrast, the static model is deteriorated and demonstrates the inability to make meaningful predictions. The static model lacks information on the extended system bounds due to the addition of connected buildings to the district heating network and increased demand during the heating season.

4 CONCLUSION

Accurate load prediction models for the building sector form the basis of predictive optimized energy management. However, building energy systems are characterized by dynamic behavior and require model adaption in practical application. The active model adaption approach for data-driven models in single model batch learning is underexplored in research on load prediction and concept drift detection occurrences are usually not explicitly discussed. In our study, we evaluated both passive and active model adaption approaches using state-of-the-art concept drift detection methods, namely DDM and ADWIN, coupled with random forest regressors. Our analysis focused on load prediction tasks within two real-world building energy systems of distinct scales, a daycare building and a district heating network, each characterized by unique change characteristics. Both, DDM and ADWIN have been proven valuable in efficiently automating the model retraining rates and training data selections, reducing the retraining cost compared to passive model adaption while achieving similar performance. The district heating network exhibited gradual changes that led to altered system bounds, rendering static models ineffective, whereas model adaption approaches demonstrated significant performance improvements. Conversely, the static model adequately predicted power consumption in the daycare building on average, since the behavioral changes were sudden and temporary, while remaining within previous system bounds. This highlights the remaining challenges in single model-based batch learning model adaption approaches of recovering previous information for recurring concepts and retrieving

https://doi.org/10.52202/077185-0216

global concepts. To address these challenges, our future work will involve expanding our study on active model adaption approaches to include additional concept drift detection methods and to investigate ensemble models.

NOMENCLATURE

e	error rate
m	harmonic mean
n	number of a count
p	probability
X	input features
y	target features
ϵ_{cut}	threshold ADWIN
δ	confidence value
σ	standard deviation
$\hat{\mu}$	expected value

Subscript

i	timestep
t	time
W	window

REFERENCES

Bayram, F., Ahmed, B. S., Kassler, A., 2022, From concept drift to model degradation: An overview on performance-aware drift detectors, *Knowledge-Based Systems*, vol. 245: p. 108632. https://doi.org/10.1016/j.knosys.2022.108632

Bifet, A., Gavaldà, R., 2007, Learning from Time-Changing Data with Adaptive Windowing, *Proceedings of the 2007 SIAM International Conference on Data Mining,* SIAM: p. 443–448. https://doi.org/10.1137/1.9781611972771.42

Ditzler, G., Roveri, M., Alippi, C., Polikar, R., 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, vol. 10, no. 4: p. 12–25. https://doi.org/10.1109/MCI.2015.2471196

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., Schmidt-Thieme, L., 2021, Do We Really Need Deep Learning Models for Time Series Forecasting?, *arXiv*, http://arxiv.org/abs/2101.02118

Gama, J., Medas, P., Castillo, G., Rodrigues, P., 2004, Learning with Drift Detection, *In:* Bazzan A., L., C., S. Labidi, S., *Advances in Artificial Intelligence – SBIA 2004*, Springer, Berlin Heidelberg: p. 286-295. https://doi.org/10.1007/978-3-540-28645-5_29

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A., 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, vol. 46, no. 4: p. 1–37. https://doi.org/10.1145/2523813

Gemaque, R. N., Costa, A. F. J., Giusti, R., Dos Santos, E. M., 2020, An overview of unsupervised drift detection methods, *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 6: p. 1381. https://doi.org/10.1002/widm.1381

Hamilton, I., et al., 2022, Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector, United Nations Environment Programme (2022), Nairobi

Hoi, S. C. H., Sahoo, D., Lu, J., Zhao, P., 2021, Online learning: A comprehensive survey, *Neurocomputing*, vol. 459: p. 249–289. https://doi.org/10.1016/j.neucom.2021.04.112

Hu, H., Kantardzic, M., Sethi, T. S., 2020, No Free Lunch Theorem for concept drift detection in streaming data classification: A review, *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 2: p. 1327. https://doi.org/10.1002/widm.1327

Ji, Y., Geng, G., Jiang, Q., 2021, Enhancing Model Adaptability Using Concept Drift Detection for Short-Term Load Forecast, *2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, IEEE: p. 464–469. https://doi.org/10.1109/ICPSAsia52756.2021.9621522

https://doi.org/10.52202/077185-0216

Mariano-Hernández, D., Hernández-Callejo, L., Solís, M., Zorita-Lamadrid, A., Duque-Pérez, O., Gonzalez-Morales, L., García, F. S., Jaramillo-Duque, A., Ospino-Castro, A., Alonso-Gómez, V., Bello, H. J., 2022, Analysis of the Integration of Drift Detection Methods in Learning Algorithms for Electrical Consumption Forecasting in Smart Buildings, *Sustainability*, vol. 14, no. 10: p. 5857. https://doi.org/10.3390/su14105857

Mehmood, H., Kostakos, P., Cortes, M., Anagnostopoulos, T., Pirttikangas, S., Gilman, E., 2021, Concept Drift Adaptation Techniques in Distributed Environment for Real-World Data Streams, *Smart Cities*, vol. 4, no. 1: p. 349–371. https://doi.org/10.3390/smartcities4010021

Read, J., Žliobaitė, I., 2023, Learning from Data Streams: An Overview and Update, *arXiv*, http://arxiv.org/abs/2212.14720

Stinner, F., Kümpel, A., Müller, D., 2021, Impact of COVID-19 countermeasures on the energy and water consumption and CO 2 concentration of municipal buildings, *Journal of Physics: Conference Series*, vol. 2042, no. 1: p. 12129. https://doi.org/10.1088/1742-6596/2042/1/012129

Toquica, D., Agbossou, K., Malhamé, R., Henao, N., Kelouwani, S., Cardenas, A., 2020, Adaptive Machine Learning for Automated Modeling of Residential Prosumer Agents, *Energies*, vol. 13, no. 9: p. 2250. https://doi.org/10.3390/en13092250

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2018, Scikit-learn: Machine Learning in Python, *arXiv*, http://arxiv.org/abs/1201.0490

Wang, Y., Chen, Q., Hong, T., Kang, C., 2019, Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges, *IEEE Transactions on Smart Grid*, vol. 10, no. 3: p. 3125–3148. https://doi.org/10.1109/TSG.2018.2818167

Wares, S., Isaacs, J., Elyan, E., 2019, Data stream mining: Methods and challenges for handling concept drift, *SN Applied Sciences*, vol. 1, no. 11: p. 1412. https://doi.org/10.1007/s42452-019-1433-0

Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., Petitjean, F., 2016, Characterizing concept drift, *Data Mining and Knowledge Discovery*, vol. 30, no. 4: p. 964–994. https://doi.org/10.1007/s10618-015-0448-4

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., Livingood, W., 2021, A review of machine learning in building load prediction, *Applied Energy*, vol. 285: p. 116452. https://doi.org/10.1016/j.apenergy.2021.116452

ACKNOWLEDGEMENT

We gratefully acknowledge the financial support provided by the BMWK (Federal Ministry for Economic Affairs and Climate Action), promotional reference 03SBE0006A.

https://doi.org/10.52202/077185-0216