Adaptively coping with concept drifts in energy time series forecasting using profiles

Benedikt Heidrich
Institute for Automation and Applied Informatics
Karlsruhe Institute of Technology
Germany
benedikt.heidrich@kit.edu

Nicole Ludwig
Cluster of Excellence Machine Learning
University of Tübingen
Germany
nicole.ludwig@uni-tuebingen.de

Marian Turowski
Institute for Automation and Applied Informatics
Karlsruhe Institute of Technology
Germany
marian.turowski@kit.edu

Ralf Mikut
Institute for Automation and Applied Informatics
Karlsruhe Institute of Technology
Germany
ralf.mikut@kit.edu

Veit Hagenmeyer
Institute for Automation and Applied Informatics
Karlsruhe Institute of Technology
Germany
veit.hagenmeyer@kit.edu

ABSTRACT

Accurate electrical load forecasts are necessary to stabilize the electricity grid, e.g., by optimally operating energy storage systems or using demand-side management. However, an implicit assumption of most load forecasting methods is that future data looks similar to past data. Unfortunately, this assumption often does not hold; for example, recruiting new staff or a pandemic can lead to demand changes resulting in so-called concept drifts in the underlying data. Most methods for coping with such concept drifts rely on computationally expensive retraining. We propose a new method for coping with concept drifts based on profiles and a linear regression model that avoids expensive retraining. Compared to a simple baseline and five state-of-the-art benchmark models on two different data sets, our method has lower computational costs and higher forecast accuracy, making it especially interesting for smart grid applications.

CCS CONCEPTS

• **Applied computing → Forecasting**; • **Computing methodologies → Machine learning**.

KEYWORDS

energy time series, concept drift, profile, forecasting

ACM Reference Format:
Benedikt Heidrich, Nicole Ludwig, Marian Turowski, Ralf Mikut, and Veit Hagenmeyer. 2022. Adaptively coping with concept drifts in energy time series forecasting using profiles. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22), June 28–July 1, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3538637.3539759

e-Energy '22, June 28–July 1, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-9397-3/22/06.
https://doi.org/10.1145/3538637.3539759

1 INTRODUCTION

As the share of renewable energy sources increases, the energy supply becomes more volatile. This volatility increases the difficulty for grid operators to balance supply and load. To ease any balancing task, grid operators require not only accurate supply but also accurate load forecasts. Therefore multiple short-term load forecasting (STLF) methods exists such as statistical [7], neural network-based [12, 13, 15], and hybrid [9] forecasting methods.

All these forecasting methods assume that the electrical load in the future is similar to that in the past. However, in reality, recruiting new employees, a pandemic, renovating a building, or installing more photovoltaic panels changes the load and its underlying concepts. The resulting changes are known as concept drifts and can reduce the forecast accuracy of STLF.

Typically, four concept drifts are distinguished [23]; sudden, gradual, incremental, and recurring concept drifts. A sudden concept drift describes an abrupt change in a time series, while a gradual concept drift refers to a transition phase, where an old and a new concept are alternately active with an increasing activity of the new concept over time. An incremental concept drift occurs when concepts are consecutively replaced by slightly different concepts. Due to their similar behavior in recorded time series, we refer to gradual and incremental concept drifts as incremental concept drifts in the following. Lastly, a concept drift is recurring when an old concept reappears in the time series. Since all these concept drifts can occur in electrical load time series, efficiently coping with concept drifts is essential in STLF.

Multiple approaches to cope with concept drifts in general time series exist such as online learning [10, 14], ensemble methods [1, 22], or drift detection methods [2, 16] for triggering retraining. However, only very few paper explicitly consider concept drifts in load forecasting. Online learning methods for load time series include Online Support Vector Regression [19], Random Functional Link Networks [18], or Online Adaptive Recurrent Neural Networks (OARNN) [6]. With regard to ensemble methods, Jagait et al. [11] combine the OARNN with a moving ARIMA to address concept drifts. Lastly, Vrablecová et al. [20] propose a detection-based method where they retrain the linear regression model when the threshold-based detection recognizes a concept drift. However, all proposed methods rely on retraining or model selection methods that tend to be computationally expensive. In smart meter environments, such methods may not be suitable if computing resources are limited.

As [20], we assume that concept drifts only influence the level of load time series. Thus, we propose a new method for coping with concept drifts that models the level and the remainder of a time series separately. We describe the level part with profiles and the remainder –for simplicity– with a linear regression model, which we then add up for the final forecast. This separation allows us to only update the profile instead of the whole model. Given an appropriate updating strategy, the proposed method's ability to cope with concept drifts then only depends on updating the profiles and does not need expensive retraining. The low computational cost and high forecast accuracy make the proposed method especially interesting for smart grid applications. Altogether, the present paper contributes a new method to cope with concept drift using profiles that does not need a concept drift detection method and expensive retraining.

The remainder of the paper is structured as follows. In Section 2, we introduce our method for coping with concept drifts comprising a profile and a linear regression. In Section 3, we describe the experimental setting applied in the evaluation. Next, Section 4 evaluates our method using different profiles, and Section 5 compares it with existing methods. In Section 6, we discuss the results and our method. Finally, in Section 7, we wrap up the paper and describe potential further research.

2 PROFILES FOR CONCEPT DRIFTS

In this section, we introduce our method for coping with concept drifts. For our method, we assume like [20] that concept drifts in load forecasting mainly influence the level of a time series. Based on this assumption, we propose to model the level and the remainder of a time series separately and to add them later on (see Figure 1). For the level of a time series, we use adaptive profiles similar to [9] to consider their periodic nature. For the remainder of the time series, we use a linear regression that is trained to forecast the difference between the profile and the actual consumption. Finally, the prediction of the remainder and the profile are added to form the prediction. In the following, we introduce both different profile calculations and the linear regression in detail.

2.1 Profile

To model the level of a time series, we use profiles representing the typical load of consumers. For the calculation of these profiles, different methods exist. Each profile calculation method considered in the present paper distinguishes between weekdays and weekends or holidays and uses past data. However, the past data used and the weight given, differ between the profile calculation methods (see Figure 2). In the following, we introduce four exemplary profile calculation methods which we later use in our experiments, by first explaining each method before describing its benefits and drawbacks.

Static Profile. The simplest profile calculation method is the Static Profile. It is determined on the training set and remains unchanged throughout the test period. We calculate the Static Profile using

$$p_t = \sum_{i=0}^{T_0} \frac{1}{n_{t,[0,T_0]}} \begin{cases} l_i, & (h_t = h_i) \wedge (t \bmod S = i \bmod S) \\ 0, & \text{else}, \end{cases} \tag{1}$$

where t is the current time, T_0 is the end of the training set, S is the seasonal length (for profile calculation, we use 24), $t \bmod S$ is the position of t in the seasonality (in our case the hour of the day), l_t is the load at t, h_t is a binary variable, which is one if t is a holiday or weekend else h_t is zero, and $n_{t,\chi}$ is the number of elements with $(h_t = h_i) \wedge (t \bmod S = i \bmod S)$ in the interval χ which is used to calculate the profile. Although we do not expect the Static Profile to adapt to concept drifts, it serves as a baseline and might be beneficial whenever a time series returns to the original concept after a drift.

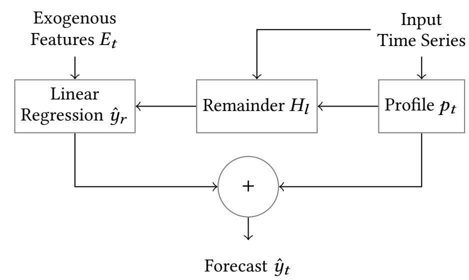


Figure 1: The proposed method for coping with concept drifts in short-term load forecasting sums a profile and a linear regression to obtain a prediction.

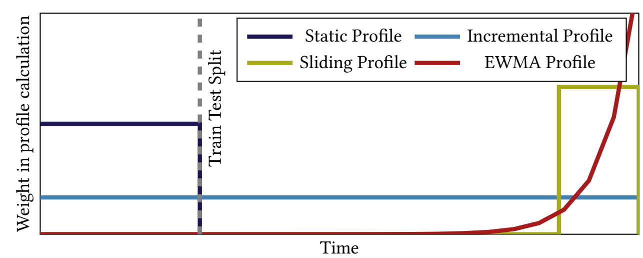


Figure 2: The four considered profile calculation methods use and weight past data differently.

Incremental Profile. To overcome the missing adaption to concept drifts in the Static Profile, the Incremental Profile adds all data available at each new time step t to the profile calculation. It is thus calculated as

$$p_t = \sum_{i=0}^{t-1} \frac{1}{n_{t,[0,t-1]}} \begin{cases} l_i, & (h_t = h_i) \wedge (t \bmod S = i \bmod S) \\ 0, & \text{else}, \end{cases} \tag{2}$$

where t, S, $t \bmod S$, l_t, h_t, and $n_{t,\chi}$ are defined as for the Static Profile. Since all historical data available at time t are used for the calculation, the Incremental Profile slightly adapts to concept drifts. However, although the data belonging to the old concept are outdated, they are not discarded and thus still influence the profile. For this reason, the Incremental Profile changes slowly and never fully adapts to new concepts.

Sliding Profile. To forget outdated data and only focus on the most current data, we apply a sliding window approach. This results in the Sliding Profile, which is defined as

$$p_t = \sum_{i=t-W}^{t-1} \frac{1}{n_{t,[t-W,t-1]}} \begin{cases} l_i, & (h_t = h_i) \wedge (t \bmod S = i \bmod S) \\ 0, & \text{else}, \end{cases} \tag{3}$$

where t, S, $t \bmod S$, l_t, h_t, and $n_{t,\chi}$ are defined as before and W is the window length. The window length has to be set by the user (for the content of this paper, we set $W = 28$ days). The Sliding Profile requires saving all data used for calculating the profile at a certain window. Otherwise, it would not be possible to remove the oldest sample from the sliding window as it moves further. Although the Sliding Profile only focuses on the most current data, there is also a lag of W time steps until it completely adapts to a concept drift.

Exponential Weighted Moving Average (EWMA) Profile. To reduce the adaption time, we can weight current values higher than past values, resulting in an Exponential Weighted Moving Average (EWMA) Profile. It consists of a subprofile $p_{t,0}$ for workdays and a subprofile $p_{t,1}$ for holidays and weekends, resulting in

$$p_t = \begin{cases} p_{t,1}, & h_t = 1 \\ p_{t,0}, & else, \end{cases} \tag{4}$$

where t and h_t are defined as before. Both subprofiles are calculated based on the previously defined h_{t-S}, which is 1 if $t - S$ is a holiday or weekend. As a result, both subprofiles are defined as

$$p_{t,h'} = \begin{cases} (1-\alpha) \cdot p_{t-S,h'} + \alpha \cdot l_{t-S}, & h_{t-S} = h' \\ p_{t-S,h'}, & else, \end{cases} \tag{5}$$

where t, S, l_t, and h_t are defined as before, $h' \in 0, 1$, $p_{0,h'} = 0$, and α is the smoothing factor of the Exponential Weighted Moving Average. The smoothing factor has to be set by the user (for the content of this paper, we set $\alpha = 0.3$).

Due to its exponential weighted average, the EWMA Profile responds more quickly to changes than the Sliding Profile. Moreover, it does not require saving the data of a particular window. Nevertheless, the EWMA Profile is more vulnerable to anomalies as they can strongly influence the profile as the most recent value.

2.2 Linear Regression

We choose a linear regression to model the remainder of a time series. As inputs, it takes the historical remainder time series and exogenous features for the values to be forecast. We obtain the historical remainder time series by subtracting the selected profile from the input time series. The exogenous features comprise calendar information, i.e. the hour of the day, month of the year, and a flag indicating whether the value to be forecast is a weekend, holiday, or neither. Given these inputs, we train the linear regression to predict future values of the remainder time series. Formally, the linear regression is defined as

$$\hat{y}_r = c + \sum_j \beta_j \cdot E_j + \sum_l \gamma_l \cdot H_l, \tag{6}$$

where c is a constant, E_j are all exogenous variables such as calendar and weather information, and H_l are the historical remainder input values. The resulting prediction $\hat{y}_r$ is then added to the selected profile to obtain the final forecast $\hat{y}_t$.

3 EXPERIMENTAL SETTING

This section describes how we evaluate our method for coping with concept drifts. We first introduce the data sets before explaining how the forecast models are trained and retrained. Finally, we describe the applied evaluation metrics and the used hard- and software.

3.1 Data Sets

We use two real-world data sets to evaluate our method. Both data sets consist of real-world load time series. However, they differ in the observed concept drifts; while we insert synthetic concept drifts into the first, the second already contains recorded concept drifts.

3.1.1 *Data with Synthetic Concept Drifts.*

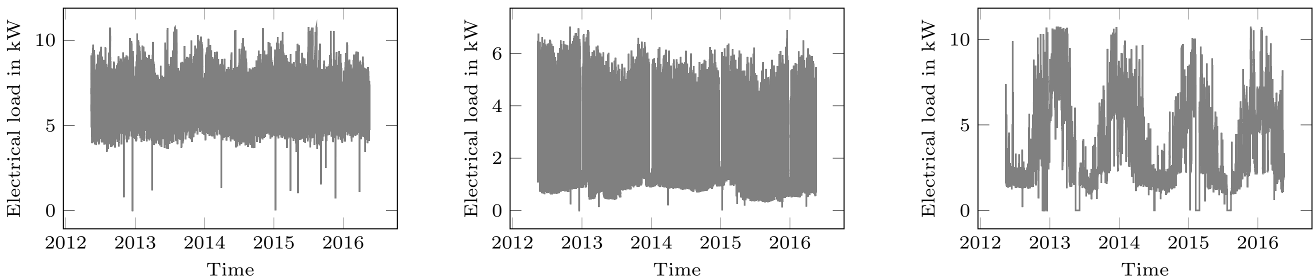
Sufficiently evaluating concept drift adaption methods benefits from information on the concept drifts' characteristics, such as their intensity, type, and position. However, this information is typically unavailable in real-world datasets containing recorded concept drifts. Thus, we artificially insert concept drifts, which allows us to control their characteristics. In the following, we first briefly describe the data set in which we insert the synthetic concept drifts and then explain how we insert the synthetic concept drifts.

Data Set Description. The basic real-world time series data, in which we insert synthetic concept drifts, comprises the electrical load of three consumers on a university campus, namely U1, U2, and U3 [9, 21]. The time series are recorded from January 1, 2006 to May 18, 2016 with a resolution of 15 minutes. We filter out measurements over the 99% quantile and below zero, likely to be errors, and linearly interpolate these values as well as single missing values. For our experiments, we work on an hourly resolution, and thus aggregate the data. The resulting time series are visualized in Figure 3. In addition to the load measurements, we use temperature and humidity as recorded by a nearby weather station of the German Meteorological Service [5].

Concept Drift Insertion. In the pre-processed time series, we insert six concept drifts with different characteristics (see Table 1). To create two types of synthetic concept drifts, i.e. incremental and sudden, we add or subtract either a linear function for incremental concept drifts or a step function for sudden concept drifts. The linear function increases in 999 steps from 0 to *intensity*. The step function increases in one step from 0 to *intensity*. To analyze whether the intensity of the drifts makes a difference, we use two different intensities. Lastly, to avoid distorting the results through artifacts in the raw data, we vary their positions. The three considered positions are July (Position 1), August (Position 2), and October (Position 3). For recurring synthetic concept drifts, the old concept is restored in December (Position 1), January (Position 2), or March (Position 3).

3.1.2 *Real Data with Concept Drifts.*

Apart from the synthetic concepts drifts, we also evaluate our method on a data set that already contains recorded concept drifts. The selected data set is the open



(a) Consumer U1 has low seasonal and low daily variations.

(b) Consumer U2 has low seasonal and strong daily variations.

(c) Consumer U3 has strong seasonal and low daily variations.

Figure 3: The three consumers from an university campus, that are used to insert synthetic concept drifts, show different seasonal and daily variations.

Table 1: We consider six different concept drifts that vary in their type, intensity, and recurrence.

Name	Type	Intensity in kW	Recurrence
Sudden 1	Sudden	2	Yes
Sudden 2	Sudden	4	Yes
Sudden 3	Sudden	2	No
Incremental 1	Incremental	2	Yes
Incremental 2	Incremental	4	Yes
Incremental 3	Incremental	2	No

UCI Electricity Load Dataset[1] [4]. This data set comprises the electrical load of several consumers in Portugal with a quarter-hourly resolution from the beginning of 2011 until the end of 2014. From this data set, we select the electrical load of three consumers (see Figure 4). Two of the consumers, namely C118 and C188, contain a sudden concept drift, and one consumer C157 contains an incremental concept drift. For the evaluation, we again aggregate the chosen electrical load time series to an hourly resolution.

3.2 Training and Retraining

Given these two different data sets, we evaluate if using profiles improves the forecast accuracy. We train our methods on a train set and evaluate them on a separate test set. Additionally, some of the methods apply a retraining during the evaluation, thus repeating the training procedure with additional past data from the test set.

Training. The train and test set used for the evaluation is different for the two data sets. For consumers U1, U2, and U3, we use about three years of data (May 14, 2012 until May 17, 2015) for training and one year for testing (May 18, 2015 until May 18, 2016); for consumers C118, C157, and C188, about one year (Jan 1, 2012 until Jan 4, 2013) for training and about two years for testing (Jan 5, 2013 until Dec 31, 2014). We train our method on this historical electrical load time series together with calendar information, and, if available, weather information.

Retraining. We evaluate the forecasts of each model mini batch by mini batch. Each mini batch comprises one day of the test set (24 samples). For each sample, the considered forecasting model performs a 24-hours forecast. Based on the forecasts for all samples from a mini batch, we make the decision whether it is necessary to retrain the forecasting model or not. For this decision, we apply one of the following three retraining strategies: **None** never triggers a retraining, **Periodic** triggers a retraining after every 30 days, and **Detection** triggers a retraining if ADWIN [2] as concept drift detection method detects a concept drift.

[1] https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

3.3 Metrics

To evaluate if using profiles improves the forecast of time series with concept drifts, we use one metric for forecast accuracy and one for computational cost. We introduce both in the following.

Forecast Accuracy. For assessing the quality of the forecasts, we measure the forecast accuracy with the Mean Absolute Scaled Error (MASE). We choose the MASE as it enables us to compare the forecast accuracy of different load time series. It is calculated as

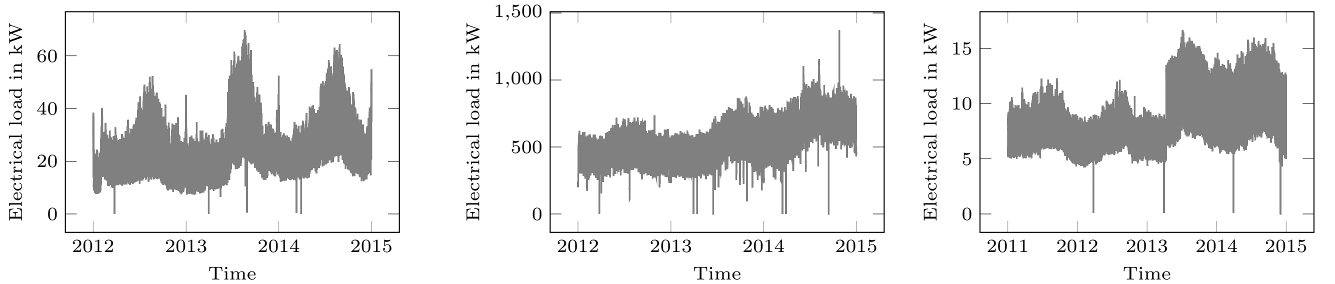
$$\text{MASE} = \frac{\frac{1}{n}\sum_{t=1}^{n} \mid y_t - \hat{y}_t \mid}{\frac{1}{n}\sum_{t=1}^{n} \mid y_t - y_{t-S} \mid}, \quad (7)$$

where n is the size of the dataset, $\hat{y}_t$ is the forecast, y_t is the true value, and S is the length of the seasonality. For our evaluation, we choose $S = 168$, resulting in a scaling by the persistence forecast with a lag of one week y_{t-S}. The division by the persistence forecast makes the metric invariant of the data’s scale. To comprehensively evaluate the forecast accuracy, we calculate the MASE in two different ways. First, we calculate the MASE over the whole test data to assess the overall forecast accuracy of the forecasting method. Second, we calculate a sliding MASE with a window of 24 days that slides over the data to gain insights into forecast accuracy changes over time.

Computational Cost. Besides the forecast accuracy, we are also interested in how computationally expensive the methods for coping with concept drifts are. We, therefore, consider the training and retraining time of the introduced methods and take their sum in seconds as a second metric in our evaluation.

3.4 Hardware and Software

For a better comparability of the results, we apply the same hardware throughout our evaluation, namely an off-the-shelve computer



(a) Consumer C118 contains a sudden concept drift.

(b) Consumer C157 contains an incremental concept drift.

(c) Consumer C188 contains a sudden concept drift.

Figure 4: We select three clients with different concept drifts from the UCI Electricity Load Dataset.

with a 4 core i7 CPU and 16GB of RAM. Furthermore, all evaluated forecasting models are implemented in Python. More specifically, the used Linear Regression model is implemented with SKLearn [17] and the neural network-based models with Keras [3]. To automate the evaluation using these implementations, we employ pyWATTS [8].

4 EVALUATING PROFILES TO COPE WITH CONCEPT DRIFTS

To evaluate whether profiles improve how the forecasting methods cope with concept drifts, we perform three experiments on the consumers U1, U2, and U3 with synthetic concept drifts. In the first experiment, we use the different previously introduced profiles in our method and evaluate them based on the resulting forecast accuracy. In the second experiment, we examine how retraining strategies influence the forecast accuracy of our method. In the third experiment, we examine how our method behaves over time considering the sliding forecast accuracy.

4.1 Evaluating Different Profiles

To evaluate the different previously introduced profiles, we employ them together with the linear regression model as in **??** and compare the forecast accuracy.

Table 2 shows the MASE of our method with different profiles over the whole test time series. Using the EWMA profile achieves the best MASE for consumers U1 and U2 while using the Sliding profile obtains the best MASE for consumer U3. For all consumers, using the Static and Incremental profile performs worst. Additionally, the MASE of our method with the EMWA profile has the smallest fluctuation across different concept drifts, while the MASE of our method with the Static and Incremental profile have the highest fluctuations. Furthermore, we observe that the position of the concept drift does not have an influence on the accuracy.

4.2 Evaluating Retraining Strategies

To examine how retraining strategies influence the forecast accuracy of our method, we compare it with the three previously introduced retraining strategies regarding the resulting forecast accuracy.

Table 3 reports the results of comparing the three training strategies for consumer U1 and the concept drifts at Position 1 (for the other positions and consumers, see Tables 5 to 7 in the Appendix). The results show that retraining does not noticeably improve the forecast accuracy of our method with the best performing profile EWMA; in all cases, our method with the profile EWMA and without retraining shows at least the same forecast accuracy.

4.3 Evaluating Profiles over Time

To examine our method's behavior over time, we make use of the sliding MASE. For this purpose, we exemplarily consider two sliding MASE, one for the Sudden 1 and one for the Incremental 1 concept drift for consumer U1.

Based on the results in Figure 5, we make three observations. First, our method with the Static, and Incremental profiles are strongly affected by the concept drift in August and their MASE decreases when the old concept recurs. Second, for both concept drifts, our method with the EWMA and Sliding profiles have the lowest MASE. For the sudden concept drift, an exception is a short period after the second concept drift, where the MASE is comparatively higher. For the same short period and the incremental concept drift, the MASE of our method with the EWMA and Sliding profiles is on par with the MASE of the our method with the other profiles. Third, for the sudden concept drift, the MASE of our method with the EWMA and Sliding profiles show two peaks when the concept drifts occur (see Figure 5a). However, such peaks are not observable for the incremental concept drift (see Figure 5b).

5 BENCHMARKING PROFILES TO COPE WITH CONCEPT DRIFTS

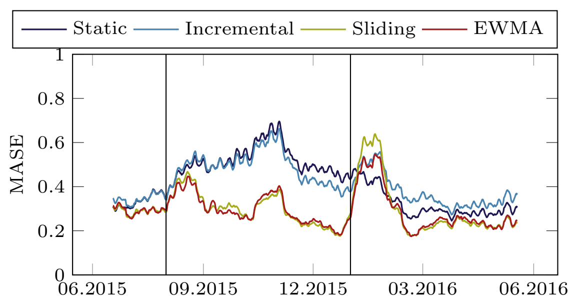
Having established our method with the best performing profiles for coping with concept drifts, namely the EMWA and Sliding profile, we compare our method with these profiles to six benchmark models. We analyze the performance on both electrical load time series with synthetic and recorded concept drifts. In the following, we first introduce the selected benchmark models, before reporting the results.

Table 2: The MASE of our method with different profiles for the three consumers with six different synthetic concept drifts at the three different positions. The best performing profile for each consumer and position is highlighted in bold.

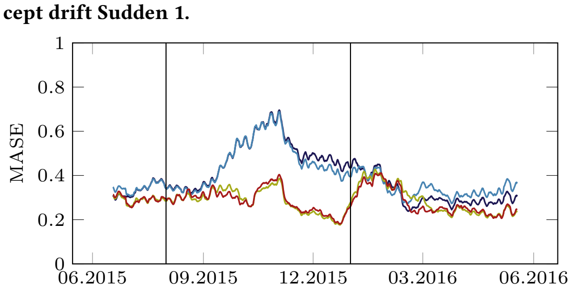
		Position 1						Position 2						Position 3					
Consumer	Profile	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3
U1	Static	0.86	1.25	2.09	0.88	1.35	2.12	0.85	1.24	1.81	0.87	1.31	1.77	0.90	1.33	2.41	0.91	1.36	2.33
	Incremental	0.88	1.34	1.82	0.89	1.40	1.89	0.87	1.30	1.71	0.89	1.34	1.71	0.94	1.46	1.99	0.93	1.46	1.98
	Sliding	0.61	0.63	0.63	**0.63**	0.67	**0.63**	0.63	0.66	0.65	0.64	0.68	0.65	0.63	0.67	0.65	0.64	0.67	0.65
	EWMA	**0.60**	**0.59**	**0.61**	**0.63**	**0.63**	**0.63**	**0.61**	**0.60**	**0.62**	**0.63**	**0.62**	**0.63**	**0.61**	**0.60**	**0.62**	**0.63**	**0.62**	**0.64**
U2	Static	1.18	1.76	3.34	1.17	1.79	3.27	1.17	1.70	2.80	1.25	1.84	2.77	1.19	1.84	4.05	1.21	1.84	3.60
	Incremental	1.47	2.21	3.54	1.44	2.19	3.55	1.42	2.05	3.16	1.51	2.19	3.19	1.49	2.35	4.13	1.50	2.30	3.78
	Sliding	0.73	0.79	0.76	0.74	0.80	0.76	0.72	0.78	0.74	0.72	0.79	0.76	0.73	0.80	0.78	0.74	0.79	0.76
	EWMA	**0.65**	**0.65**	**0.65**	**0.66**	**0.66**	**0.66**	**0.64**	**0.64**	**0.65**	**0.65**	**0.65**	**0.66**	**0.65**	**0.64**	**0.67**	**0.66**	**0.66**	**0.67**
U3	Static	**0.45**	0.47	0.54	**0.45**	0.48	0.53	**0.44**	0.46	0.51	**0.46**	0.48	0.52	**0.45**	0.47	0.57	**0.46**	0.49	0.56
	Incremental	**0.45**	0.46	0.49	**0.45**	0.47	0.49	0.45	0.45	0.49	**0.46**	0.47	0.50	**0.45**	0.46	0.51	**0.46**	0.47	0.51
	Sliding	0.46	**0.44**	**0.45**	0.46	**0.45**	**0.45**	0.45	**0.43**	**0.45**	**0.46**	**0.45**	**0.46**	0.46	**0.44**	**0.46**	**0.46**	**0.46**	**0.46**
	EWMA	0.48	0.47	0.48	0.49	0.48	0.48	0.48	0.47	0.48	0.49	0.48	0.49	0.48	0.47	0.49	0.49	0.48	0.49

Table 3: The MASE of our method with the different profiles and retraining strategies for consumer U1 with a synthetic concept drift at Position 1. The best retraining strategy for each profile and consumer is highlighted in bold.

Consumer	Profile	Retraining Strategy	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3
U1	Static	None	0.86	1.25	2.09	0.88	1.35	2.12
		Periodic	**0.64**	**0.62**	**0.70**	**0.67**	**0.65**	**0.70**
		Detection	0.68	0.64	0.74	0.71	0.75	0.83
	Incremental	None	0.88	1.34	1.82	0.89	1.40	1.89
		Periodic	**0.65**	**0.64**	**0.71**	**0.69**	**0.67**	**0.72**
		Detection	0.70	0.66	0.74	0.74	0.77	0.84
	Sliding	None	**0.61**	0.63	**0.63**	**0.63**	0.67	**0.63**
		Periodic	0.62	0.61	0.65	0.64	**0.62**	**0.63**
		Detection	**0.61**	**0.60**	0.65	0.64	0.65	0.65
	EWMA	None	**0.60**	**0.59**	**0.63**	**0.63**	**0.61**	**0.63**
		Periodic	0.62	0.60	0.65	0.64	0.65	0.64
		Detection	0.61	**0.59**	0.64	**0.63**	0.63	0.64



(a) The MASE of our method with different profiles on the concept drift Sudden 1.



(b) The MASE of our method with different profiles on the concept drift Incremental 1.

Figure 5: The MASE of our method with different profiles and synthetic concept drifts at Position 1 on consumer U1. The vertical lines indicate the start of a concept drift.

5.1 Benchmark Models

For the comparison, we select six benchmark models, including the previously introduced linear regression model without a profile, two deep learning models and three online learning models. We introduce the latter two groups in the following.

5.1.1 Deep Learning Models. As state-of-the-art deep learning models we select the Recurrent Inception Network (RIN) [12] and the Profile Neural Network (PNN) [9]. The RIN combines LSTMs with an 1-D convolution inception module comprising parallel convolution layers with different kernel sizes. The PNN aims to combine advantages from statistical and deep learning forecasting by splitting the forecasting task in three modules, which respectively represent the trend component, the standard load profile and the colorful noise of a time series. Their outputs are finally weighted and aggregated into a final prediction. For the RIN, we use our own implementation[2]; for the PNN, we use the pyWATTS implementation.

5.1.2 Online Learning Models. As online learning models, we select the Online-Sequential Extreme Learning Machine (OS-ELM)[14], the Online Adaptive Recurrent Neural Network (OARNN) [6], and

[2]https://github.com/KIT-IAI/Coping-with-Concept-Drift-using-Profiles

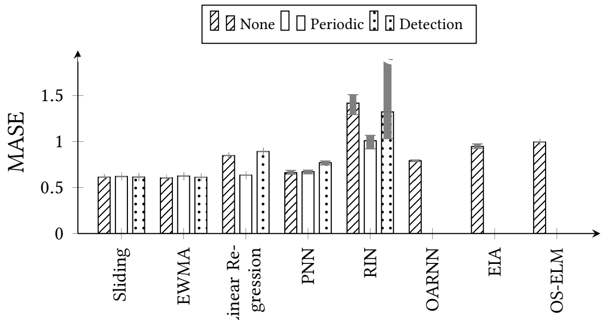


Figure 6: The MASE of our method with the EWMA and Sliding profiles and the benchmarks on consumer U1 with the Sudden 1 concept drift at Position 1 for the three retraining strategies. The bars indicate the average MASE. The errors bars show the worst and the best achieved MASE of the corresponding model.

the Error Intersection Approach (EIA) [1]. The OS-ELM [14] is commonly used for concept drifts. It is a three-layered neural network architecture with random weights between the input and the hidden layer. To learn the weights between the hidden and output layer, the Least Square Method is used instead of the backpropagation algorithm. The OARNN [6] has been designed to cope with concept drift in load time series data. It consists of a simple RNN embedded in an online learning framework, which updates the RNN and the normalization parameters after each time step. Additionally, the OARNN performs a Bayesian optimization for the RNN's hyperparameters if the forecasting error is too high. The final online learning approach, EIA [1], is an easy to implement ensemble approach and consists of a simple (persistence forecast) and a complex model (MLP). The EIA measures the errors of both models and if the error curves intersect, the forecast model is changed[3]. For the EIA and the OARNN, we also use our own implementation[4], while for the OS-ELM, we use an existing implementation[5].

5.2 Benchmark with Synthetic Concept Drifts

To compare our method with the EWMA and Sliding profiles to the benchmarks, we use two different evaluations on the consumers with synthetic concept drifts. First, we measure the forecast accuracy. Afterward, we evaluate the computational costs.

Forecast Accuracy. Considering the forecast accuracy of the different models, we exemplarily report the results for the concept drift Sudden 1 and Consumer 1 at Position 1. For the other concept drifts, consumers, and positions, the results are similar (for the other consumers, see Figure 8, and for the concept drift Incremental 1, see Figure 9 in the Appendix). In Figure 6, we make two observations. First, our method with the EWMA and Sliding profiles outperforms the other models regardless of the applied retraining strategy. Second, if profiles are used, as in our method and the PNN, the effect of retraining is low. In contrast, for the models not using profiles, retraining improves the forecast accuracy.

[3]Note that the authors state that this model is only suitable for sudden concept drifts.
[4]https://github.com/KIT-IAI/Coping-with-Concept-Drift-using-Profiles
[5]https://github.com/leferrad/pyoselm

Table 4: The computational cost for the Sudden 1, Sudden 2, Incremental 1, and Incremental 2 concept drifts at Position 1 for consumer U1.

Model	Retraining Strategy	Sudden 1	Sudden 2	Incremental 1	Incremental 2
	None	0.65	0.65	0.65	0.65
Sliding	Periodic	3.29	3.26	3.30	3.29
	Detection	1.08	1.11	0.88	1.06
	None	0.67	0.67	0.67	0.67
EWMA	Periodic	3.26	3.37	3.35	3.32
	Detection	1.08	1.07	0.87	0.90
	None	0.67	0.67	0.67	0.67
Linear Regression	Periodic	3.36	3.30	3.24	3.28
	Detection	1.07	1.28	1.13	1.30
	None	104.80	104.80	104.80	104.80
PNN	Periodic	113.69	112.69	112.83	112.42
	Detection	106.55	106.33	106.61	105.95
	None	517.26	517.26	517.26	517.26
RIN	Periodic	541.01	541.63	540.94	541.40
	Detection	531.43	530.70	526.68	528.29
EIA		75.78	75.78	75.78	75.78
OARNN		307.73	309.08	310.90	310.21
OS - ELM		50.98	50.81	51.33	51.29

Computational Cost. To compare the computational cost of our method for coping with concept drifts with the benchmarks, we sum the training and all retraining times for the Sudden 1, Sudden 2, Incremental 1, and Incremental 2 concept drifts at Position 1 for consumer U1. Table 4 shows the resulting computational costs. In this table, we make two observations. First, we observe that the benchmarks have clearly higher computational costs than our method for coping with concept drift. Second, the type or intensity of concept drifts does not influence the computational costs.

5.3 Benchmark with Recorded Concept Drifts

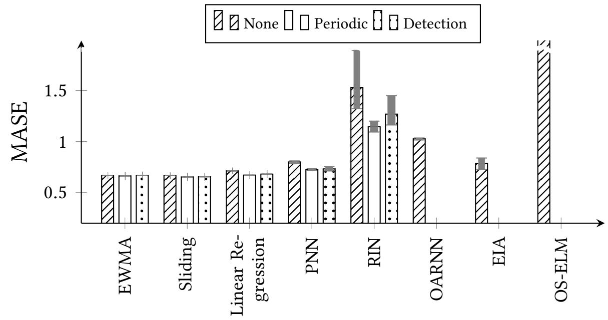
To compare our method with the EWMA and Sliding profiles to the benchmarks, we also use recorded concept drifts as found in the three consumers C118, C157, and C188.

Based on Figure 7, we make three main observations. First, our method with the EWMA and Sliding profiles always outperforms the benchmarks regardless of the retraining strategy. Only for consumer C118, the performance of the linear regression is similar to our method with the EWMA and Sliding profiles. Second, contrary to the benchmarks, our method with the EWMA and Sliding profiles does not benefit from any retraining. Third, the MASE of our method with the EWMA and Sliding profiles appears stable for all consumers and varies less than the other benchmarks.

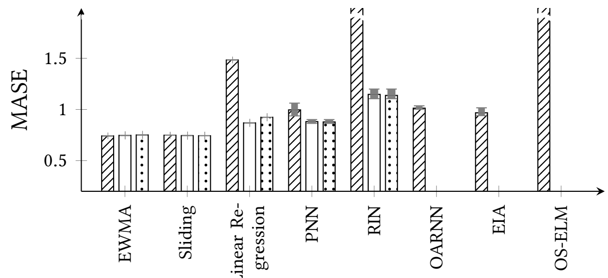
6 DISCUSSION

Based on the previous section results, we discuss our method using profiles in the following regarding how well the method can cope with concept drifts, the robustness of results concerning different concept drifts, and some artifacts in the sliding MASE results.

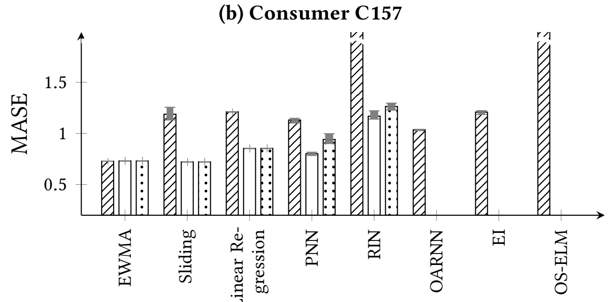
Profiles Can Cope with Concept Drifts. In load time series forecasting, the results indicate that retraining strategies do not improve



(a) Consumer C118



(b) Consumer C157



(c) Consumer C188

Figure 7: The MASE of the proposed method and benchmarks on the three consumers with the recorded incremental and sudden concept drifts. The bars indicate the average MASE. The errors bars show the worst and the best achieved MASE of the corresponding model.

the forecast accuracy of our method with the EWMA and Sliding profiles. Consequently, it is sufficient to use profiles without expensive retraining strategies for coping with concept drifts. Additionally, using profiles leads to good forecasting results compared to state-of-the-art methods even if we select a simple model, i.e. a linear regression, for the remainder. As a result, using profiles helps to cope with concept drift without the need to apply a detection method and retraining. However, it is essential to use an appropriate profile calculation method, as the methods' adaption speed differs. For example, the EWMA profile quickly adapts to new concepts. Nevertheless, responding too quickly to newly received data, that, for example, contain anomalies, can reduce the forecast accuracy. To handle such effects, future work may mix different profiles for improving the performance. Besides the appropriate selection of the profile calculation method, it is also necessary for the Sliding and EWMA profiles to select values for the parameters window length and smoothing factor α. However, without further tuning of these parameters, our method with these profiles provide a good forecasting accuracy.

Robust Results. Additionally, we observe that the type and intensity of the concept drifts do not influence our method with the Sliding and EWMA profiles as much as our method with the Static, and Incremental profiles or a linear regression without a profile. Therefore, we assume that our method with the EWMA and Sliding profiles provides more robust predictions across various concept drifts than the other methods. This observation might be essential when deploying a forecasting system in environments with volatile consumption patterns, where the position, intensity, and type of concept drifts is unknown.

Peaks in the Sliding Forecast Accuracy. Moreover, we observe peaks in the Sliding MASE of our method with the EWMA and Sliding profiles. The reason for these peaks is the adaption time of the EWMA and Sliding profiles. During an incremental concept drift, the time series changes slowly, leading to more time for our method to adapt to the concept drift and thus to avoid peaks.
For recurring concept drifts, the MASE of our method with the Static and Incremental profile does not show such peaks. The reason is that these profiles reflect the old concept. Hence, it may be helpful for recurring concept drifts to restore the profiles of the past concept to reduce the adaption time.

7 CONCLUSION

The present paper proposes a new computationally efficient method for coping with concept drifts based on profiles that does not need a concept drift detection method and expensive retraining. The new method assumes that concept drifts mainly affect the level of a time series and that the time series can be decomposed into a level part and a remainder. Based on this decomposition, we introduce profiles to describe the level part and a linear regression model to describe the remainder. We then add the profile to the regression output for the final forecast.
We show on two different real-world datasets, one with synthetic and one with recorded concept drifts, that our method leads to more accurate forecasts than five state-of-the-art benchmark models. More specifically, we evaluate our method with four different profiles and find that an exponential weighted moving average profile performs best in most cases. At the same time, computationally expensive retraining is not necessary. The new method is more accurate and computationally cheaper than all benchmark methods, making it especially interesting for smart grid applications.
Future work will focus on more complex concept drifts, and extend the forecasting method for the remainder.

ACKNOWLEDGMENTS

This project is funded by the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI, the Helmholtz Association under the Program "Energy System Design", and the German Research Foundation (DFG) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645.

REFERENCES

[1] Lucas Baier, Marcel Hofmann, Niklas Kühl, Marisa Mohr, and Gerhard Satzger. 2020. Handling Concept Drifts in Regression Problems–the Error Intersection Approach. (2020). arXiv:2004.00438

[2] Albert Bifet and Ricard Gavalda. 2007. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM)*, David Skillicorn, Bing Liu, Chid Apte, and Srinivasan Parthasarathy (Eds.). Society for Industrial and Applied Mathematics, 443–448. https://doi.org/10.1137/1.9781611972771.42

[3] François Chollet et al. 2015. Keras. https://keras.io.

[4] Dheeru Dua and Casey Graff. 2019. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml.

[5] DWD Climate Data Center (CDC). 2018. Historical Hourly Station Observations of 2m Air Temperature and Humidity for Germany, Version V006. https://cdc.dwd.de/portal/

[6] Mohammad Navid Fekri, Harsh Patel, Katarina Grolinger, and Vinay Sharma. 2021. Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. *Applied Energy* 282 (2021). https://doi.org/10.1016/j.apenergy.2020.116177

[7] Stephen Haben, Georgios Giasemidis, Florian Ziel, and Siddharth Arora. 2019. Short term load forecasting and the effect of temperature at the low voltage level. *International Journal of Forecasting* 35, 4 (2019), 1469–1484. https://doi.org/10.1016/j.ijforecast.2018.10.007

[8] Benedikt Heidrich, Andreas Bartschat, Marian Turowski, Oliver Neumann, Kaleb Phipps, Stefan Meisenbacher, Kai Schmieder, Nicole Ludwig, Ralf Mikut, and Veit Hagenmeyer. 2021. pyWATTS: Python Workflow Automation Tool for Time Series. *arXiv preprint arXiv:2106.10157* (2021).

[9] Benedikt Heidrich, Marian Turowski, Nicole Ludwig, Ralf Mikut, and Veit Hagenmeyer. 2020. Forecasting Energy Time Series with Profile Neural Networks. In *The Eleventh ACM International Conference on Future Energy Systems (e-Energy'20)*. ACM, 220–230. https://doi.org/10.1145/3396851.3397683

[10] Elena Ikonomovska, João Gama, and Sašo Džeroski. 2015. Online tree-based ensembles and option trees for regression on evolving data streams. *Neurocomputing* 150 (2015), 458–470. https://doi.org/10.1016/j.neucom.2014.04.076

[11] Rashpinder Kaur Jagait, Mohammad Navid Fekri, Katarina Grolinger, and Syed Mir. 2021. Load Forecasting Under Concept Drift: Online Ensemble Learning With Recurrent Neural Network and ARIMA. *IEEE Access* 9 (2021), 98992–99008. https://doi.org/10.1109/ACCESS.2021.3095420

[12] Junhong Kim, Jihoon Moon, Eenjun Hwang, and Pilsung Kang. 2019. Recurrent inception convolution neural network for multi short-term load forecasting. *Energy and Buildings* 194 (2019), 328–341. https://doi.org/10.1016/j.enbuild.2019.04.034

[13] Ming Lei, Liyang Tang, Mingxing Li, Zhenyu Ye, and Liwei Pan. 2019. Forecasting Short-Term Residential Electricity Consumption Using a Deep Fusion Model. In *Proceedings of 2018 Chinese Intelligent Systems Conference*. Springer, Singapore, 359–371. https://doi.org/10.1007/978-981-13-2291-4_36

[14] Nan-ying Liang, Guang-bin Huang, P. Saratchandran, and N. Sundararajan. 2006. A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Transactions on Neural Networks* 17, 6 (2006), 1411–1423. https://doi.org/10.1109/TNN.2006.880583

[15] Daniel L. Marino, Kasun Amarasinghe, and Milos Manic. 2016. Building Energy Load Forecasting using Deep Neural Networks. In *42nd Annual Conference of the IEEE Industrial Electronics Society (IECON 2016)*. IEEE, 7046–7051. https://doi.org/10.1109/IECON.2016.7793413

[16] E. S. Page. 1954. Continuous Inspection Schemes. *Biometrika* 41, 1/2 (1954), 100–115. https://doi.org/10.2307/2333009

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

[18] Xueheng Qiu, Ponnuthurai Nagaratnam Suganthan, and Gehan A.J. Amaratunga. 2018. Ensemble incremental learning Random Vector Functional Link network for short-term electric load forecasting. *Knowledge-Based Systems* 145 (2018), 182–196. https://doi.org/10.1016/j.knosys.2018.01.015

[19] Petra Vrablecová, Anna Bou Ezzeddine, Viera Rozinajová, Slavomír Šárik, and Arun Kumar Sangaiah. 2018. Smart grid load forecasting using online support vector regression. *Computers & Electrical Engineering* 65 (2018), 102–117. https://doi.org/10.1016/j.compeleceng.2017.07.006

[20] Petra Vrablecová, Viera Rozinajová, and Anna Bou Ezzeddine. 2017. Incremental Adaptive Time Series Prediction for Power Demand Forecasting. In *Data Mining and Big Data*, Ying Tan, Hideyuki Takagi, and Yuhui Shi (Eds.). Springer International Publishing, Cham, 83–92. https://doi.org/10.1007/978-3-319-61845-6_9

[21] Long Wang, Yong Ding, Till Riedel, Andrei Miclaus, and Michael Beigl. 2017. Data analysis on building load profiles: A stepping stone to future campus. In *2017 International Smart Cities Conference (ISC2)*. 1–4. https://doi.org/10.1109/ISC2.2017.8090823

[22] Jianhua Xiao, Zhu Xiao, Dong Wang, Jing Bai, Vincent Havyarimana, and Fanzi Zeng. 2019. Short-term traffic volume prediction by ensemble learning in concept drifting environments. *Knowledge-Based Systems* 164 (2019), 213–225. https://doi.org/10.1016/j.knosys.2018.10.037

[23] Indrė Žliobaitė. 2010. Learning under concept drift: an overview. (2010). arXiv:1010.4784

A ADDITIONAL RESULTS

Table 5: The MASE of our method with the different profiles and retraining strategies over the whole test time series for the consumer U2 and U3 with a synthetic concept drift at Position 1. The best performing retraining strategy for each profile and consumer is highlighted in bold.

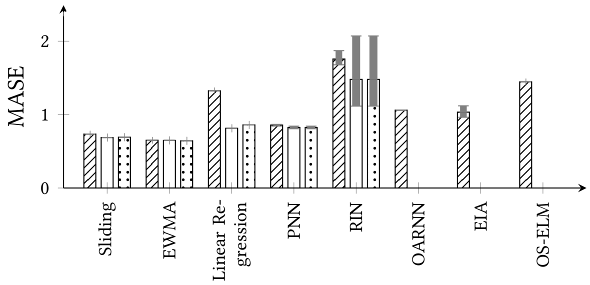
Consumer	Profile	Retraining Strategy	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3
U2	Static	None	1.18	1.76	3.34	1.17	1.79	3.27
		Periodic	0.73	0.74	0.85	0.77	0.82	0.92
		Detection	**0.61**	**0.59**	**0.63**	**0.64**	**0.63**	**0.64**
	Incremental	None	1.47	2.21	3.54	1.44	2.19	3.55
		Periodic	**0.73**	**0.71**	**0.81**	**0.79**	**0.77**	**0.87**
		Detection	0.78	0.78	0.90	0.84	0.91	1.03
	Sliding	None	0.73	0.79	**0.76**	0.74	0.80	0.76
		Periodic	**0.69**	**0.69**	**0.76**	**0.70**	**0.67**	**0.71**
		Detection	**0.69**	0.70	0.77	0.71	0.71	0.75
	EWMA	None	0.65	0.65	**0.65**	0.66	0.66	0.66
		Periodic	0.65	0.64	0.69	**0.65**	**0.64**	**0.65**
		Detection	**0.64**	**0.63**	0.68	0.66	0.65	0.66
U3	Static	None	**0.45**	0.47	0.54	**0.45**	0.48	0.53
		Periodic	0.46	**0.45**	**0.47**	0.46	**0.45**	**0.47**
		Detection	0.46	0.47	0.50	0.46	0.47	0.50
	Incremental	None	**0.45**	0.46	0.49	**0.45**	0.47	0.49
		Periodic	0.46	**0.45**	**0.46**	0.47	**0.46**	**0.46**
		Detection	0.46	0.46	0.49	0.46	0.47	0.48
	Sliding	None	**0.46**	**0.44**	**0.45**	**0.46**	**0.45**	**0.45**
		Periodic	**0.46**	**0.44**	**0.45**	**0.46**	**0.45**	**0.45**
		Detection	0.47	0.45	0.46	0.48	0.47	0.46
	EWMA	None	**0.48**	**0.47**	**0.48**	**0.49**	**0.48**	**0.48**
		Periodic	0.49	0.48	0.49	0.50	0.49	0.49
		Detection	0.49	0.48	0.49	0.50	0.49	0.49

Table 6: The MASE of our method with the different profiles and retraining strategies for each consumer with a synthetic concept drift at Position 2. The best performing retraining strategy for each profile and consumer is highlighted in bold.

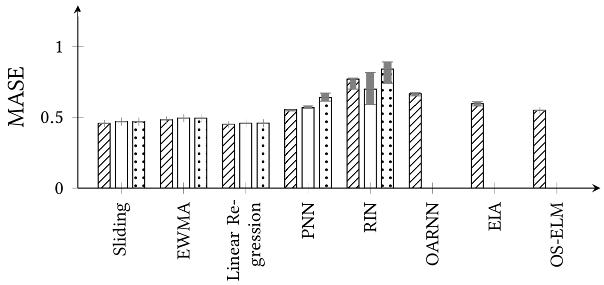
Consumer	Profile	Retraining Strategy	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3
U1	Static	None	0.85	1.24	1.81	0.87	1.31	1.77
		Periodic	**0.62**	**0.60**	**0.66**	**0.65**	**0.64**	**0.70**
		Detection	0.66	0.66	0.75	0.70	0.74	0.82
	Incremental	None	0.87	1.30	1.71	0.89	1.34	1.71
		Periodic	**0.64**	**0.61**	**0.68**	**0.66**	**0.66**	**0.71**
		Detection	0.75	0.68	0.76	0.71	0.77	0.84
	Sliding	None	0.63	0.66	0.65	0.64	0.68	0.65
		Periodic	**0.61**	**0.59**	**0.64**	**0.63**	**0.61**	**0.64**
		Detection	0.62	0.61	0.67	**0.63**	0.63	0.66
	EWMA	None	**0.61**	0.60	**0.62**	**0.63**	0.62	**0.63**
		Periodic	**0.61**	**0.59**	0.64	**0.63**	**0.61**	0.64
		Detection	**0.61**	0.60	0.64	0.64	**0.61**	0.64
U2	Static	None	1.17	1.70	2.80	1.25	1.84	2.77
		Periodic	**0.68**	**0.67**	**0.78**	**0.74**	**0.72**	**0.83**
		Detection	0.74	0.74	0.86	0.79	0.91	1.26
	Incremental	None	1.42	2.05	3.16	1.51	2.19	3.19
		Periodic	**0.70**	**0.68**	**0.79**	**0.76**	**0.74**	**0.85**
		Detection	0.78	0.80	0.95	0.89	1.01	1.25
	Sliding	None	0.72	0.78	0.74	0.72	0.79	0.76
		Periodic	**0.66**	**0.65**	**0.73**	**0.66**	**0.63**	0.69
		Detection	0.67	0.67	0.75	0.68	0.64	**0.68**
	EWMA	None	0.64	0.64	**0.65**	0.65	0.65	0.66
		Periodic	**0.63**	0.62	0.66	**0.64**	**0.60**	**0.64**
		Detection	**0.63**	**0.61**	0.66	**0.64**	0.61	**0.64**
U3	Static	None	**0.44**	0.46	0.51	**0.46**	0.48	0.52
		Periodic	0.45	**0.44**	**0.46**	0.47	**0.45**	**0.47**
		Detection	0.45	0.45	0.48	**0.46**	0.46	0.49
	Incremental	None	**0.45**	0.45	0.49	**0.46**	0.47	0.50
		Periodic	0.46	**0.44**	**0.46**	0.47	**0.46**	**0.47**
		Detection	0.46	0.45	0.48	0.48	**0.46**	0.48
	Sliding	None	**0.45**	**0.43**	**0.45**	**0.46**	**0.45**	**0.46**
		Periodic	0.46	0.45	0.46	0.48	0.47	0.48
		Detection	0.46	0.44	0.46	0.48	0.46	0.48
	EWMA	None	**0.48**	**0.47**	**0.48**	**0.49**	**0.48**	**0.49**
		Periodic	0.49	0.48	0.49	0.50	0.49	0.50
		Detection	0.49	**0.47**	0.49	0.50	**0.48**	0.50

Table 7: The MASE of our method with the different profiles and retraining strategies for each consumer with a synthetic concept drift at Position 3. The best performing retraining strategy for each profile and consumer is highlighted in bold.

Consumer	Profile	Retraining Strategy	Sudden 1	Sudden 2	Sudden 3	Incremental 1	Incremental 2	Incremental 3
U1	Static	None	0.90	1.33	2.41	0.91	1.36	2.33
		Periodic	0.65	0.66	**0.75**	**0.68**	**0.68**	**0.74**
		Detection	**0.64**	**0.65**	0.79	0.74	0.91	1.41
	Incremental	None	0.94	1.46	1.99	0.93	1.46	1.98
		Periodic	0.66	**0.67**	**0.75**	**0.70**	**0.70**	**0.76**
		Detection	**0.65**	**0.67**	0.78	0.75	0.77	0.83
	Sliding	None	0.63	0.67	**0.64**	0.67	0.65	**0.65**
		Periodic	**0.62**	0.62	0.68	0.65	**0.63**	0.66
		Detection	**0.62**	**0.61**	0.67	**0.64**	0.64	0.66
	EWMA	None	**0.61**	**0.60**	**0.62**	**0.63**	**0.62**	**0.64**
		Periodic	0.65	0.66	0.75	0.68	0.68	0.74
		Detection	0.62	0.61	0.65	0.64	**0.62**	0.65
U2	Static	None	1.19	1.84	4.05	1.21	1.84	3.60
		Periodic	**0.71**	**0.73**	**0.88**	**0.76**	**0.73**	**0.82**
		Detection	0.77	0.76	0.94	0.79	0.83	0.93
	Incremental	None	1.49	2.35	4.13	1.50	2.30	3.78
		Periodic	**0.73**	**0.74**	**0.89**	**0.78**	**0.75**	**0.85**
		Detection	0.80	0.80	1.01	0.85	0.92	1.03
	Sliding	None	0.73	0.80	**0.78**	0.74	0.79	0.76
		Periodic	0.69	**0.70**	0.80	**0.70**	**0.66**	**0.71**
		Detection	**0.68**	0.71	0.82	0.74	0.71	0.76
	EWMA	None	**0.65**	0.64	**0.67**	0.66	0.66	0.67
		Periodic	**0.65**	0.65	0.71	**0.65**	**0.62**	**0.66**
		Detection	**0.65**	**0.63**	0.69	0.66	0.63	**0.66**
U3	Static	None	**0.45**	0.47	0.57	**0.46**	0.49	0.56
		Periodic	0.46	**0.45**	**0.48**	0.47	**0.46**	**0.48**
		Detection	0.47	0.48	0.53	0.47	0.49	0.52
	Incremental	None	**0.45**	0.46	0.51	**0.46**	0.47	0.51
		Periodic	0.46	**0.45**	**0.47**	0.47	**0.46**	**0.47**
		Detection	0.46	0.47	0.51	**0.46**	0.48	0.51
	Sliding	None	**0.46**	**0.44**	**0.46**	**0.46**	**0.46**	**0.46**
		Periodic	0.47	0.46	0.48	0.48	0.47	0.48
		Detection	0.47	0.46	0.48	0.48	0.47	0.47
	EWMA	None	**0.48**	**0.47**	**0.49**	**0.49**	**0.48**	**0.49**
		Periodic	0.50	0.49	0.50	0.50	0.50	0.50
		Detection	0.50	0.49	0.51	0.50	0.50	0.50

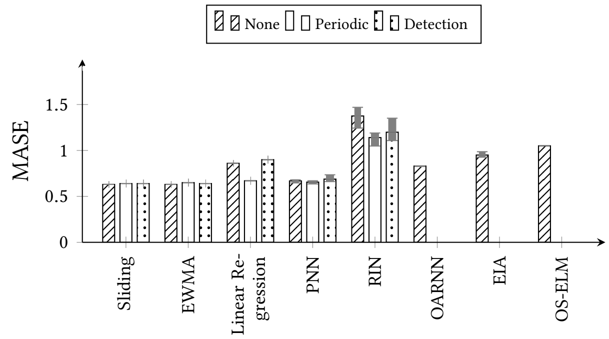


(a) Consumer U2

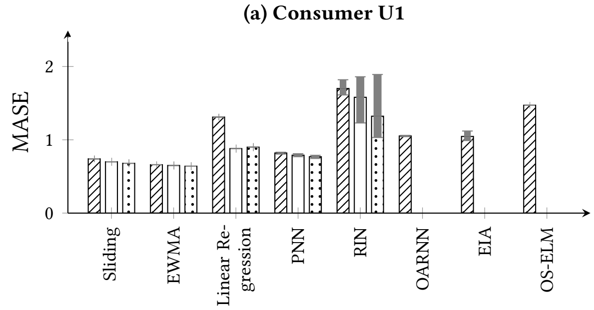


(b) Consumer U3

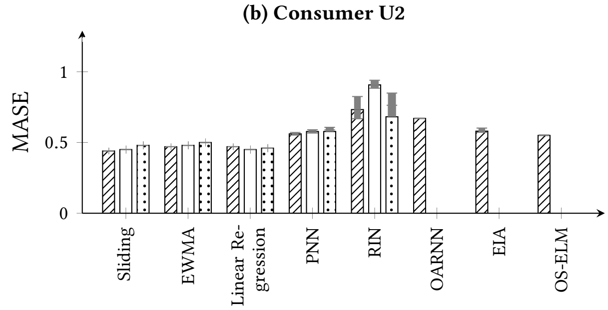
Figure 8: The MASE of our method with the EWMA and Sliding profiles and the benchmarks on the the consumers U2 and U3 with the Sudden 1 concept drift at Position 1 for the three retraining strategies. The bars indicate the average MASE. The errors bars show the worst and the best achieved MASE of the corresponding model.



(a) Consumer U1



(b) Consumer U2



(c) Consumer U3

Figure 9: The MASE of our method with the EWMA and Sliding profiles and the benchmarks on the three consumers with the Incremental 1 concept drift at Position 1 for the three retraining strategies. The bars indicate the average MASE. The errors bars show the worst and the best achieved MASE of the corresponding model.