Applied Energy 368 (2024) 123364

Contents lists available at ScienceDirect

Applied Energy

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

The impact of heat pumps on day-ahead energy community load forecasting

Leo Semmelmann [a,*], Matthias Hertel [a], Kevin J. Kircher [b], Ralf Mikut [a], Veit Hagenmeyer [a], Christof Weinhardt [a]

[a] Karlsruhe Institute of Technology, Karlsruhe, Germany
[b] Purdue University, West Lafayette, USA

ARTICLE INFO

Keywords:
Aggregation levels
CEEMDAN
Energy community
Heat pumps
Load forecasting
Transformer

ABSTRACT

The rapid ramp-up of heat pump installations in modern power systems constitutes an outstanding challenge for energy community and distribution grid operators. Accurate load forecasts can help community and grid operators to reduce electricity demand peaks by managing flexible devices. This paper shows that installed heat pumps change the load patterns, autocorrelation and peak loads of energy communities, as well as the most suitable forecasting methods. Based on a case study with real-world household and heat pump loads from Hamelin, Germany, we show significant improvements in forecasting quality by employing Transformer models. We publish our underlying data set, feature engineered data, forecasting results, best-performing methods, and benchmarking pipeline open-source, to contribute to the advancement of load forecasting in energy communities with heat pumps.

1. Introduction

Many European countries plan to install hundreds of thousands of heat pumps annually over the coming decades [1]. This leads to additional loads and burdens for distribution grids, for instance, through the overloading of transformers and power lines [2,3]. To postpone heat pump-induced grid reinforcement measures, alternatives such as Demand Side Management or Battery Energy Storage Systems (BESS) can be used [4,5]. One widely discussed concept for managing low voltage nodes are so-called energy communities, which combine tens to hundreds of households in a neighborhood to manage electricity needs [6] collectively. Operators of energy communities have to plan supply and demand under grid constraints to minimize purchase costs for the community members. A critical aspect of managing energy communities and distribution girds is scheduling flexibility measures. This requires an accurate forecast of upcoming and day-ahead loads [7].

Although several studies discuss different methods to forecast day-ahead loads in energy communities and distribution grids, most focus on traditional load patterns, mainly dominated by conventional household appliances [7,8]. These traditional load patterns will change in many countries by transforming the heating sector towards heat pumps [9]. This development has a severe impact on the operators of energy communities. Previous studies have not addressed two main questions: First, it is unclear if the same forecasting methods perform well for traditional household loads and heat pump loads. Second, the potential impact of the aggregation level on energy community load forecasts has not been investigated: it is unclear if operators of energy communities should directly forecast the whole load of the energy community, consisting of heat pump and traditional household loads, or if separate forecasts for the household and heat pump loads should be conducted and then aggregated. Several past studies underline that a higher aggregation level improves the quality of the forecast [10,11]. However, it has not been investigated if this holds true for forecasting different types of loads that follow distinct distributions.

In summary, this paper addresses the following research questions:

- Do the same methods perform well for forecasting traditional household loads and heat pump loads?
- Does the aggregation level of energy community loads – in particular, the decision between directly forecasting the whole energy community load vs. forecasting heat pump loads and household loads separately – have an impact on the forecasting quality?
- How are the presented forecasting methods and aggregation strategies performing in an actual battery-based peak shaving use case of energy community operators?

We answer these questions by suggesting a state-of-the-art methodology, including feature engineering, feature selection, sophisticated Bayesian hyperparameter optimization, sliding window forecasting,

* Corresponding author.
E-mail addresses: leo.semmelmann@kit.edu (L. Semmelmann), matthias.hertel@kit.edu (M. Hertel), kkirche@purdue.edu (K.J. Kircher), ralf.mikut@kit.edu (R. Mikut), veit.hagenmeyer@kit.edu (V. Hagenmeyer), weinhardt@kit.edu (C. Weinhardt).

https://doi.org/10.1016/j.apenergy.2024.123364
Received 17 December 2023; Received in revised form 6 March 2024; Accepted 30 April 2024
Available online 18 May 2024

Nomenclature

Δt	Time resolution [h]
D	Amount of days
E^{tot}	Maximum energy capacity [kWh]
P^{BESS}	Battery charging or discharging power [kW]
P^{Comb}	Aggregated load (including heat pump and household load) [kW]
P^{HH}	Household load [kW]
P^{HP}	Heat pump load [kW]
$P^{predict}$	Predicted load [kW]
P^{real}	Observed load [kW]
SOC	State of Charge [%]
T	Amount of time steps
t	Time index

and detailed benchmarking. The presented methodology is applied to a recent dataset of heat pump and household loads from an energy community in Hamelin, Germany [12]. We publish our preprocessing approach, the feature-engineered data, our results and the best-performing methods open-source.

The present paper is structured as follows. In Section 2, state-of-the-art related work is presented. Section 3 covers our methodology, focusing on the investigated models. Section 4 depicts the researched case study with energy community load data from Hamelin, Germany [12]. Section 5 presents our results in light of the previously introduced research questions. In Section 6, the results are discussed in detail. Finally, Section 7 presents the conclusion.

2. Related work

A wide range of studies discuss potential methods for load forecasting [7,8,13]. The overarching goal of these methods is to forecast upcoming loads based on previous observations. The time horizon of the load forecast can range from the next minutes to the next day, up to several days, months, and years [8,14,15]. Also, the time resolution of the underlying data can range from a few minutes to a single hour, multiple hours, and whole days. All these factors play a role in the resulting quality of the forecast and the selection of the best-performing methods [8]. Our study mainly focuses on a day-ahead forecast of hourly loads, which is especially relevant for operational aspects in energy communities such as energy trading or scheduling flexibilities, as applied in several studies [16–19].

A broad spectrum of possible methods for day-ahead load forecasting tasks is discussed in the literature. The first advances in the field were made through statistical models such as the Autoregressive Integrated Moving Average (ARIMA) method [20]. The Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) [21] is an extension of the method. A different approach for load forecasting is using tree-based methods such as random forests [22] or XGBoost [21]. Tree-based methods use decision trees at their core to split the input data to make predictions over upcoming loads. Advantages of tree-based methods, for instance, XGBoost, are a high computational efficiency, good performance, and easy handling of multivariate data [23]. For multivariate load forecasting, further input features like temperature measurements can be used, which can also be an important factor of electrical load forecasts [24]. Over recent years, also neural networks have been increasingly used for load forecasting tasks, such as Long Short-Term Memory neural networks [25] or Transformers [26,27]. The same methods are analogously commonly used for heat load forecasts [28,29].

Most papers on load forecasting strictly differentiate between forecasts for traditional household loads and heat loads, which are either based on district heating systems [28,30], radiators installed at single-family houses [31] or individual heat pumps [32,33]. However, the effective and reliable management of distribution grids and energy communities of the future requires consideration of heat pump-induced loads, which will lead to significant additional loads [34]. This leads to several practical considerations. First, it is unclear if the same methods that perform well for the forecasting of common household loads also perform well for the task of heat pump load forecasting. Through an increased share of heat pumps and thereby, a change of load structures, the recommended forecasting methods might also change. Second, whether the aggregation type impacts the load forecasting quality has not been investigated. Although many studies have shown that the higher the aggregation level, the better the forecasting quality due to stochastic smoothing [10,11], it is unclear if this holds true for aggregating household and heat pump loads. Hence, we investigate if household and heat pump loads should be directly aggregated and then forecasted or if the different loads should be individually forecasted, and then the forecasts should be aggregated. This also has practical implications for the management of the energy community: If individually-aggregated forecasts perform better than a directly-aggregated forecast, it might be worthwhile to advocate for data sharing of heat pump loads of households [35]. Fourth, most forecasting studies are decoupled from the actual use case in energy communities and distribution grids. The quality of the presented forecast methodologies is solely measured in terms of metrics such as the Mean Absolute Percentage Error (MAPE) or Root Mean Squared Error (RMSE) [36], without a thorough discussion of the metrics and its applicability for distribution grid-related tasks, such as peak shaving at transformers [37].

The present paper fills these research gaps with a state-of-the-art methodology that considers the latest developments in load forecasting research [38]. We also aim to address some common pitfalls in forecasting and machine-learning-based science itself. Recent studies found that many machine learning results are not reproducible due to a lack of transparency [39,40], which is further aggravated through scarce open-source datasets [41]. We address that by making our underlying data, results, and evaluation methodology open-source, thereby enabling researchers to easily build upon our results and benchmark their results against this study. In addition, for the sake of reproducible results [40], we avoid the use of complex hybrid models and instead focus on base models. However, we encourage researchers to use our open-source data set, results, and benchmarking methodology to show possible advances of sophisticated hybrid models over the presented models.

3. Methodology

Our methodology follows the latest advances and common practices in load forecasting literature [13,38]. In the first step, additional input features are engineered to enrich the dataset. Subsequently, the feature set is reduced through a feature selection technique. Then, several forecasting models are presented. In the next step, we introduce another recent advance in load forecasting, the decomposition technique Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [42]. We conclude our methodology by introducing the two investigated aggregation levels, the underlying metrics, and our investigated peak shaving application of the presented forecasting strategies.

3.1. Feature engineering

An integral step in load forecasting is to create additional features that might help to capture additional patterns underlying the data [43]. We create the following features, in addition to given load and perfect foresight weather data, based on previous studies:

Type-of-day features: Several load forecasting related studies create additional features for type-of-day variables [44]. These binary features indicate if the given observation lies on a weekday, weekend, or holiday. We create the corresponding binary type-of-day features based on the timestamps of given observations.

Cyclical calendric features: Features such as the hour or month have a cyclical character, which might not be captured by representing them with their actual values [45]. For instance, hour *0* and hour *24* would be interpreted as far away through a regression model, although they are the same value. This misinterpretation can be avoided by applying a *sine* and *cosine* transformation to the *day* and *month* observations, as described in [45]. We also create features for a twofold and fourfold *sine* and *cosine* transformation for possible consideration of patterns that occur with a higher frequency.

Rolling average of apparent temperature: It has been shown that rolling averages of the observed temperature are important input features for load forecasts due to the thermal inertia of buildings [24, 46]. Hence, we create a rolling average for the apparent outdoor temperatures' last 24 and 48 h. We selected these intervals based on a pre-evaluation of correlations between temperatures and loads. We created the rolling temperature features based on the apparent temperature instead of the actual temperature due to the higher correlation between loads and apparent temperature observations. The actual temperature is the objective air temperature measurement, while the apparent temperature includes factors that affect the perception of temperature, for instance, humidity, wind speed and solar radiation.

Average load at same time step: For a better capture of the medium-term effects on (heat) load, [29] suggest creating an additional feature with the average load at the same hour over the last week. Considering the load of the previous seven days goes beyond including lagged variables, which are only provided for two days in this study for computational reasons.

Past loads: Previous studies showed that past loads are amongst the most important load forecasting features [38]. How these previous loads are given as input features to the model depends on the type of the model. While the following neural network-based methods, such as LSTMs, can handle whole feature vectors as input [47], classical machine learning methods, such as Random Forests, require a tabular representation of the data [23]. This means a one-dimensional feature vector is used to predict one target value. Past load features are included as lagged to accommodate the tabular representation. On day d, the feature vector includes 48 past loads x_t based on the first timestep $t_0(d)$: $\{x_{t_0(d)-1}, x_{t_0(d)-2}, \dots, x_{t_0(d)-48}\}$. Given the hourly time resolution, we consider 48 past loads, which equal two days, based on literature and initial experiments [48,49]. We take lagged features in relation to the first timestep of the respective day, $t_0(d)$, to ensure that the classical machine learning methods are working with the same input features as the neural network-based methods, which receive the two day-before past loads as an input vector.

3.2. Feature selection

We select the most relevant features through a filter and an embedded feature selection method [50] for computational efficiency and a reduction of potential overfitting. We separately conduct the feature selection process for the household-only, heat pump-only and aggregated energy community datasets, to ensure a fair comparability of the aggregation levels, which is described later in further detail. First, we filter out irrelevant features with a Pearson correlation lower than 0.1 [51]. Then, the Random Forest algorithm is used as an additional embedded method to rank the potential features based on their predictive power [50]. We only consider the ten features with the highest Random Forest feature ranking [51]. The Random Forest method itself is explained in detail in the following "Models" subsection. Thereby, we combine the advantages of the correlation-based filter method (quickly reducing the search space) and the random forest-based embedded method (identifying features with high predictive power through a forecasting model) [50,52]. The final feature set includes only features that pass both feature selection methods. The feature selection process is applied before enriching the dataset with the lagged features, to ensure comparability between neural network-based and classical machine learning methods.

We note that we include the cyclical hourly features, based on the previously described *sine* and *cosine* transformation, independently from the feature selection results, to maintain a relationship between past loads and the current observation for the classical machine learning methods. Since the past loads are based on the first daily time step $t_0(d)$ for the classical machine learning methods, the timesteps of the observations are essential to capture the relationship with past loads.

3.3. Models

In the following section, we introduce the investigated models in our study. We selected the underlying models based on a thorough analysis of benchmarking studies, identifying the most common and latest methods used for load forecasting tasks [29,38]. We note that we excluded hybrid models for the sake of reproducibility of our results and that there are several further potential candidate models whose evaluation would go beyond the scope of this study.

3.3.1. Random forests

Random forests are a machine learning method that combines an ensemble of decision tree predictors with random sampling [53]. In the first step, random samples are drawn from the underlying dataset used to build decision trees. The splitting of these trees is based on a random subset of features from which the best split is used. Possible splitting decisions are evaluated according to decision tree algorithms such as the Classification and Regression Tree (CART) method. Finally, an ensemble of a large number of trees is created, which is then used to make its prediction as the average of the included trees. Random forests have been applied in several studies for day-ahead load forecasting tasks due to their high computational efficiency, rather low overfitting, and good quality of forecasts [54,55].

3.3.2. XGBoost

The XGBoost algorithm, introduced by [23], is a highly efficient machine learning algorithm applied in various forecasting tasks. Comparable to the previously introduced Random Forest method, XGBoost utilizes an ensemble of CART models. During the model's training, the loss function's gradient is constantly calculated. At the same time, new tree learners are added iteratively to the model to reduce the error of the model. The optimization function of the model includes a regularization term, which helps the model to prevent overfitting. Additional measures to prevent overfitting are the "shrinkage method" which reduces the influence of individual trees in the model, and column subsampling, which also increases the computational speed of the model. In general, the high computational efficiency and strong prediction accuracy make XGBoost a popular model for load forecasting studies [56].

3.3.3. LSTMs

A highly popular method for time series forecasting problems are Long Short-Term Memory networks (LSTMs) [47,57]. LSTMs are a special form of Recurrent Neural Networks, that use gate units and memory cells to "forget" irrelevant information over long-term patterns but remember important information. The ability to recognize patterns and to capture long-term dependencies makes LSTMs a popular choice for time series forecasting problems, amongst other load forecasting tasks. The original study from Hochreiter and Schmidhuber [57] only contains one input layer, one hidden layer (which includes the memory cells and gate units and can be called "LSTM layer") and one output layer. Based on recent studies that apply LSTM networks for load forecasting, we include an additional hidden layer [25] and the option to include a dropout layer to prevent from overfitting [58].

3.3.4. Transformers

A novel neural network architecture was introduced by [59], which is increasingly used for natural language processing and computer vision tasks. Recently, the Transformer architecture was successfully applied to short-term load forecasting problems, due to its good performance in handling long-term patterns [42,60]. In a more recent study, the applicability of Transformer models for long-term time series forecasting was debated, given that simple linear models were outperforming them on several datasets [61]. However, with the right training strategy and enough training data, Transformers outperform linear models and other baselines for short-term and long-term load forecasting [27,62].

The standard Transformers are based on an encoder–decoder structure, although other variants exist [63]. The load, calendar and weather features for the past time steps are fed into the encoder, and the calendar and weather features for the time steps to predict are fed into the decoder. The encoder layer consists of a stack of identical layers which in turn include two sublayers: a multi-head self-attention mechanism and a fully connected feed-forward network. The multi-head attention layer allows the model to access information from various representation subspaces at varying positions. The decoder contains, in addition to the multi-head self-attention layers, multi-head cross-attention layers accessing the output of the encoder. Overall, the Transformer architecture incorporates attention mechanisms at three different points: self-attention in the encoder, self-attention in the decoder and cross-attention that allows the decoder to access the output of the encoder. A final linear layer transforms the decoder output into the predicted load values.

3.4. Bayesian hyperparameter selection

An essential part of setting up machine learning models is the selection of the right model parameters, so-called hyperparameter tuning. A novel method for the optimal selection of hyperparameters is based on a Bayesian Optimization model [64], which especially comes with the benefit of high computational efficiency and fast convergence times. The model utilizes a Gaussian Process probabilistic model to map hyperparameters to an underlying optimization function, which aims to minimize the forecasting error of the model. The model uses an acquisition function to determine new hyperparameters, as a trade-off between exploration of new areas in the space of possible hyperparameters and the exploitation of existing well-performing observations. Our Bayesian hyperparameter model is initialized with 10 randomly drawn hyperparameter sets, the κ value of the model is set at 3, determining the trade-off between exploration and exploitation. We run the Bayesian model for 100 iterations. For every iteration, we run the target model with the respective hyperparameters given by the Bayesian model, which is then updated with the Root Mean Squared Error (RMSE) achieved by the target model. The data used for the hyperparameter tuning is not included in the test dataset, which is later used for the evaluation. We note that we discretize the hyperparameter search space, which is commonly done, but comes with some drawbacks, since the parameter spaces of categorical variables, such as the activation function of neural network methods, are disjoint [65,66]. We accept this drawback for the sake of the computational efficiency of the method and given the fairness since all models are using the same approach.

3.5. CEEMDAN decomposition

An increasing number of load forecasting studies is applying decomposition techniques to improve the model performance [42,67,68]. Decomposition techniques decompose a given signal – such as a time series of loads – into subcomponents for a better understanding of underlying patterns and trends. One recent advance is the so-called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method [69]. The method first adds white noise to the target signal. Then, the signal is decomposed into different Intrinsic Mode Functions (IMFs) and the respective residue is calculated. The process is repeated and the IMFs are re-calculated until the residue cannot be decomposed anymore. For a thorough description of the method we refer to [42,69]. Aggregating the resulting IMFs and residue yields the underlying signal.

The CEEMDAN algorithm has several advantages over alternative decomposition methods like the Empirical Mode Decomposition: it exhibits an improved handling of the mode mixing problem (having similar oscillations in different modes), it is more robust to noise, as well as being non-stationary [69]. We first decompose the target load time series into IMFs with the CEEMDAN method on a monthly rolling basis [70]. Then, we train a dedicated model for every IMF. Finally, we aggregate the forecasts of the forecasted IMFs to get the resulting forecast for the target load. We compare the CEEMDAN method extension with the respective base models, taking over the same Bayesian-optimized hyperparameters from the base model.

3.6. Aggregation levels

Our study aims at forecasting loads at a low-voltage transformer (which would be for our German case between the 400 V low-voltage level and the 20 kV medium-voltage level [71]), to which multiple households of an energy community are connected. Given a dataset of multiple individual household loads, we retrieve the aggregated energy community household load P_t^{HH} by aggregating the individual loads of all N households for each time step t:

$$P_t^{HH} = \sum_{i=1}^{N} P_t^{i,HH} \qquad \forall t \in T \tag{1}$$

The same procedure is repeated for the energy community heat pump load P_t^{HP}:

$$P_t^{HP} = \sum_{i=1}^{N} P_t^{i,HP} \qquad \forall t \in T \tag{2}$$

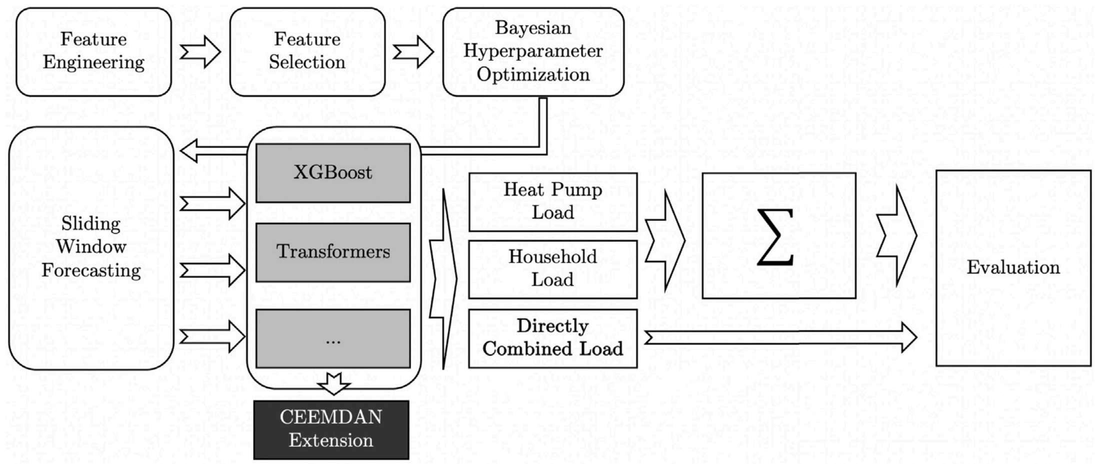
The final transformer power P_t^{Comb} consists of the sum of the household and heat pump load:

$$P_t^{Comb} = P_t^{HH} + P_t^{HP} \qquad \forall t \in T \tag{3}$$

From the perspective of the energy community or distribution grid operator, it remains unclear if household and heat pump loads should be:

- **Separate:** equaling to an individual forecast of the aggregated heat pump and household loads, and then summing the forecasts up to get the whole transformer load. This approach is reasoned by the different underlying distributions of heat pump and household load data (as shown later in the paper), which might make it reasonable to train distinct models for more accurate forecasting results [72].
- **Combined:** equaling to aggregating first all heat pump and household loads, and then predicting the whole transformer load. This approach is reasoned by the frequently observed pattern that the higher the aggregation level, the better the forecasting results [11].

Analyzing the effects of the aggregation level on forecasting quality has practical implications: predicting household and heat pump loads individually requires the operator to be able to separately access them, which might be challenging given the currently low level of observability in distribution grids [73]. Our study investigates if having additional, separate heat pump load data leads to a respective improvement in forecasting accuracy, justifying additional efforts for data retrieval.



(a)

Fig. 1. Forecasting methodology.

We illustrate the overall methodology in Fig. 1. To enable a fair comparison of methods, we conduct a separate feature engineering, feature selection and Bayesian hyperparameter optimization for every dataset (Households (HH), Heat Pumps (HP) and Combined Load ($Comb$)). Thereafter, we conduct a sliding window forecast over one year for the presented forecasting models, either as a standalone model or combined with the CEEMDAN decomposition. Based on the sliding window forecast, we train the models on every first day of the investigated months based on the data from the past year. Then, the day-ahead loads of every day in the investigated month are forecasted. The process is repeated for every month in the test set. Then, the forecasting results are evaluated based on the metrics depicted in the following.

3.7. Metrics

We evaluate the forecast quality by widely proliferated evaluation metrics, such as the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R^2 [36]. The presented metrics combine advantages like intuitive interpretability (MAPE, MAE, R^2), share of explained variance through the forecasting model (R^2), appropriate consideration of large errors (RMSE) and reduced sensitivity to outliers (MAE) [36,74,75].

The MAPE is calculated as the mean of the percent deviation from predicted loads $P_t^{predict}$ from real observations P_t^{real} over all timesteps T, multiplied by 100:

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{P_t^{predict} - P_t^{real}}{P_t^{real}}\right| \times 100 \tag{4}$$

The RMSE is calculated as the root of the mean squared deviation between $P_t^{predict}$ and P_t^{real}:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}|P_t^{predict} - P_t^{real}|^2} \tag{5}$$

The MAE is the mean of the absolute errors:

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|P_t^{predict} - P_t^{real}| \tag{6}$$

R^2 is calculated by dividing the squared error between $P_t^{predict}$ and P_t^{real} and the squared error between the average load P^{mean} and actual values P_t^{real}:

$$\text{R}^2 = 1 - \frac{\sum_{t=1}^{T}(P_t^{predict} - P_t^{real})^2}{\sum_{t=1}^{T}(P^{mean} - P_t^{real})^2} \tag{7}$$

Higher values of R^2 indicate a higher forecast quality; values can range from $-\infty$ to 1. When a model yields negative R^2 values, it indicates that its predictions for the target variable are less accurate than simply using the mean as a forecast [75].

3.8. Applicability

Most forecasting literature solely focuses on comparing and improving methods based on widely proliferated metrics, such as the MAPE or RMSE [36]. However, a critical evaluation of how well the presented methods perform in actual use cases is often missing. Hence, our study compares the performance of the presented forecasting methods for the whole energy community load in an actual use case: reducing the peak aggregated energy community load by scheduling day-ahead charging and discharging of a BESS based on the day-ahead load forecast [37]. Reducing the peak load of the energy community is important to save the underlying distribution grid from degradation, to avoid costly reinforcement measures and to reduce possible peak power grid charges [76].

We formulate the underlying optimization based on [37], simultaneously targeting peak shaving and load smoothing. For that, we minimize the squared sum of the forecasted load $P(t)^{predict}$ and the BESS power $P(t)^{BESS}$ multiplied by the time resolution Δt, over the forecasting horizon N, as depicted in Eq. (8). The BESS operations are accommodated with a few constraints: the maximum power P^{max} shall never be exceeded (Constraint (8b)). The State of Charge SOC at time t is defined by the previous charging operations $P(t)^{\text{BESS}}$ multiplied with the time resolution Δt and divided by the maximum BESS capacity E^{tot} (Constraint (8c)). Furthermore, the SOC has to be kept within 0 and 1 (Constraint (8d)). For our study, we simply set the maximum BESS capacity E^{tot} at the peak load of the previous year $P^{max,y-1}$ times the time resolution Δt, which is in our case one hour. The maximum BESS charging and discharging power P^{max} is set at half the capacity E^{tot}, divided by the time resolution Δt. Finally, in Constraint (8e), we limit the amount of allowed full cycles to one equivalent full cycle per day d (equaling to one full charging and one full discharging cycle). We note that investigating different BESS sizes and peak shaving strategies might bring additional insights, but this would go beyond the scope of this study.

$$\min \quad \left(\sum_{t=1}^{T}\Delta t[P(t)^{\text{predict}} + P(t)^{\text{BESS}}]^2\right) \tag{8a}$$

$$\text{s.t.} \quad P^{\max} \geq |P_t^{\text{BESS}}|, \quad \forall t \in T, \tag{8b}$$

$$SOC_t = SOC_{t-1} + \frac{\Delta t \cdot P(t)^{\text{BESS}}}{E^{\text{tot}}}, \quad \forall t \in T, \tag{8c}$$

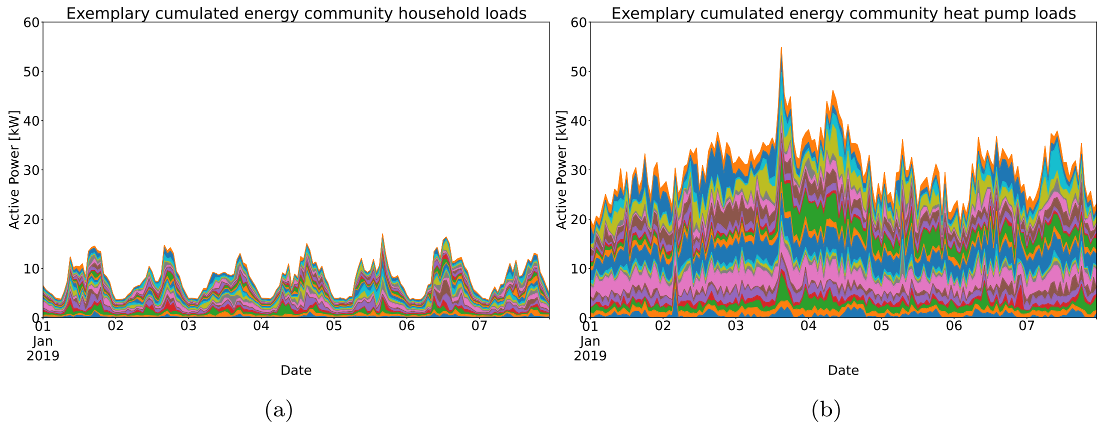


Fig. 2. Exemplary load profiles of energy community in Hamelin, Germany.

$$0 \le SOC_t \le 1, \quad \forall t \in T, \tag{8d}$$

$$\sum_{t=1}^{T} |P(t)^{\text{BESS}}| \cdot \Delta t \le E^{\text{tot}} \cdot \Delta t \cdot D \cdot 2 \tag{8e}$$

For every forecasting method, the previously depicted peak shaving and smoothing optimization is conducted using the Mixed-Integer Linear Programming solver Gurobi [77]. Then, the suggested BESS charging operations P^{BESS} are applied to the actual observed loads P^{real}. Subsequently, based on the respective forecasting methods, we can compare the achieved peak reductions through a day-ahead scheduling of BESS operations. Thereby, we can evaluate the actual applicability of the presented methods for an energy community peak shaving task.

4. Case study

In this section, the underlying dataset and the results of our hyperparameter tuning process are presented.

4.1. Data

The previously presented methodology is applied to a high-quality dataset of household loads in an energy community in Hamelin, Germany [12]. Initially, the dataset includes active and reactive power, voltage and current measurements of 38 households equipped with water-to-water heat pumps and an additional heating rod as backup heater. The dataset includes separate measurements for the households and heat pumps in 10 s to 60 min resolution from mid-2018 to the end of 2020. The heat pumps from the dataset are both responsible for covering heating and hot water demand. For our study, we use the hourly resolution of the active power and 21 out of the 38 households that do not have missing data. We see the agglomeration of the 21 households as an exemplary, small energy community, which can be found in a comparable size in existing studies [78,79]. We note that we are solely focusing on forecasting the aggregated active power. However, phase imbalance or voltage issues might arise through the installation of heat pumps [80] in single- or three-phase configurations, and these are interesting directions for future work.

We depict an exemplary weekly load profile of the household and heat pump loads in Fig. 2. We can observe that the underlying household loads follow a completely different pattern than the heat pump loads. While the household loads follow a daily pattern, with load peaks in the morning and evening, and load valleys in the night, the heat pump loads are rather on a constant high level over days, mainly caused by low temperatures. Furthermore, the heat pumps are for the exemplary illustrated winter weeks up to 8 times higher than the household loads, underlining the additional stress caused by heat pump installations on distribution grids [2]. Overall, through the installation of heat pumps the peak load in our dataset is raised from 20.1 kW to 80.1 kW, which represents a fourfold increase.

In Fig. 3, the yearly energy community load is illustrated, showing again the heavy impact of heat pump installations on the load curve, with distinct new peak loads during winter months. Also, the autocorrelation profile depicted in Fig. 4 shows the differences between heat pump and common household loads. While household loads are strongly correlated with the same daily hours, heat pump loads strongly correlate with loads in the previous hours. Potential autocorrelation patterns of future energy communities, including both household and heat pump loads, rather resemble the heat pump load autocorrelation structure.

The distinct profile of heat pump loads and their difference to the regular household loads underlines our question if separately forecasting aggregated household and heat pump loads before summing them might yield an advantage over directly forecasting the whole energy community loads, due to the different distributions and properties of the load curves.

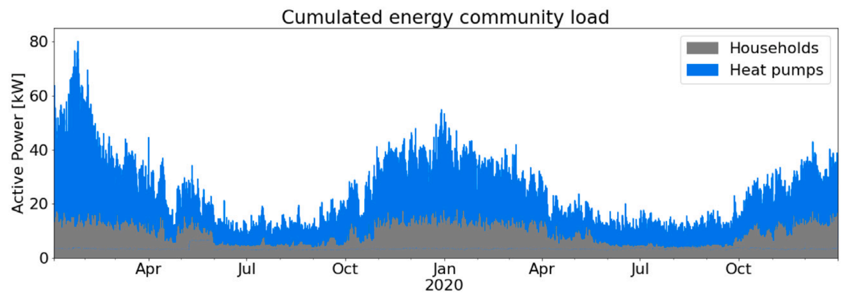
Of the presented dataset, we use observations between the beginning of 2019 till the end of 2020 for our study. The data of 2019 is used for hyperparameter selection, with the first 6 months being used for training and the last six months for testing. We split our data half-half during the hyperparameter tuning process, since we want to cover different seasonalities. Then, the data in 2020 is used for the actual benchmarking of the methods, with the previously determined parameters. We use a sliding window forecast, that trains the models at the beginning of each month based on the last twelve months. For forecasting the day-ahead hourly loads, features from the past two days are used, as previously described.

In the following, we present the results of the feature selection process and the subsequent hyperparameter selection, as well as the particular structure of the utilized models.

4.2. Feature selection

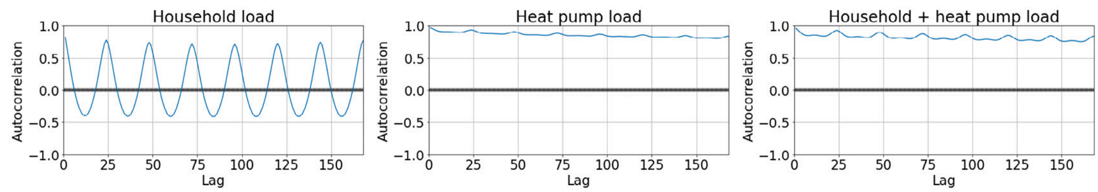
After conducting our feature selection process for the household, heat pump and combined dataset separately, we obtain the resulting feature sets in Table 1. We conduct separate feature selection processes to ensure an unbiased evaluation process in light of the comparison between individually aggregated or directly combined energy community load forecasting.

We find differences in relevant features between the different types of loads. For instance, the cosine of the hour would usually not be included in the heat pump load dataset, which is instead more focused on temperature features such as the apparent temperature or the probability of precipitation, which would not be included in the household and directly combined model. We note that we include the sine and cosine of the hour nonetheless (marked in brackets in Table 1), to maintain the temporal relationship between current and past loads for the tree-based models, as explained in Section 3.1.



(a)

Fig. 3. Yearly aggregated energy community load.



(a)

Fig. 4. Autocorrelation of household and heat pump loads.

Table 1
Resulting feature sets for different load models. Features that were excluded by the feature selection process, but were kept in the dataset to maintain temporal relationships, are marked in brackets (further details in the text).

	HH	HP	Comb
Apparent temperature		✓	✓
Apparent temperature: Rolling average 24 h	✓	✓	✓
Apparent temperature: Rolling average 48 h	✓	✓	✓
Past loads (48 h)	✓	✓	✓
Probability of precipitation		✓	
Relative humidity	✓	✓	✓
Temperature		✓	✓
Wind speed	✓	✓	✓
Cosine of hour	✓	(✓)	✓
Sine of hour	(✓)	(✓)	(✓)
Average load at same hour last week	✓	✓	✓
Dropped features (e.g. wind direction, ...)			

4.3. Hyperparameters and model structure

Based on the previously described Bayesian hyperparameter optimization and selected feature sets, we investigate for each model optimal parameters.

We use the first six months of 2019 for training during the hyperparameter selection process, and evaluate based on the last six months. Through our Bayesian hyperparameter optimization, we obtain different hyperparameters for heat pump, household and directly combined loads, as depicted in Table 2. The XGBoost parameter search space is based on [29,81], the Random Forest (RF) search space is based on [29,82], the LSTM parameters are based on [35,83], and the Transformer search space is based on [27].

The LSTM neural network is built with one bi-directional LSTM layer, two dense layers, from which the second dense layer has half the neurons of the first ones, and one dropout layer, before one final dense layer with neurons in the amount of the prediction horizon (in our case 24 h) [35].

5. Results

In this section, we present the results of our study. First, reached metrics for household and heat pump load forecasting are presented. Second, the results of different aggregation strategies are depicted. Finally, we show the results for the application of the forecasted loads on the peak shaving case study.

5.1. Forecasting results

Table 3 displays the results for the forecasting of household loads. Depending on the metrics the best results are achieved by the Transformer (RMSE, R^2) and Random Forest (MAPE, MAE) model. Although the Transformer model is amongst the best models, we do not see a remarkably better performance of it over tree-based models such as the Random Forest. In addition, the CEEMDAN extension deteriorates the models rather than improving them.

Table 4 shows the forecasting results for heat pump-only loads, which differ strongly from the household-only results. The Transformer models significantly outperform the alternative models. Based on the MAPE and MAE, the Transformer model yields the best results, while based on the RMSE and R^2 metric, the Transformer-CEEMDAN model yields the best results. The CEEMDAN extension improves the forecast quality for the neural network-based methods while significantly deteriorating the tree-based methods. We note that we have compared the variance of the forecasting results over multiple runs of the underlying methods after obtaining the initial results, to analyze the uncertainty connected with them, as depicted in Appendix. Although we observe a higher standard deviation of the RMSE of the neural network-based methods, the order of the results remains the same.

5.2. Aggregation level

Table 5 presents the results over the aggregation levels and methods. Overall, the results underline the superiority of the Transformer models: the “Transformer-CEEMDAN: combined” model achieves the best result for two of four metrics (RMSE, R^2), the “Transformer: separate” model achieves the best results for three of four metrics (MAPE, MAE, R^2). All Transformer models reach the highest observed R^2 score of 0.9. While the tree-based methods yielded comparable results for the household-only case, they are significantly worse when heat pump loads are added. This has implications for energy community and grid operators: forecasting models that have achieved good results in the

Table 2
Results of the Bayesian hyperparameter tuning process for each model.

Model	Parameter	Values	HH	HP	Comb
XGBoost	max_depth	(3,10)	3	5	6
	subsample	(0.4,1.0)	0.569	0.467	0.711
	min_child_weight	(2,6)	2.172	2.059	4.693
	colsample_bytree	(0.6,1.0)	0.633	0.887	0.744
	n_estimators	(10,200)	29	21	19
RF	max_depth	(1,500)	326	32	368
	learning_rate	(0.01,0.2)	0.035	0.079	0.01
	min_samples_split	(2,10)	2	6	10
	min_samples_leaf	(1,10)	5	4	1
	n_estimators	(10,200)	20	176	162
LSTM	batch_size	[32,64,128,256,512,1024]	256	128	64
	lstm_neurons	[32,64,128,256]	256	256	128
	lstm_first_layer	[32,64,128,256]	256	256	32
	dropout	(0.3,0.7)	0.3	0.7	0.3
	activation_function	[tanh, relu, sigmoid]	sigmoid	sigmoid	relu
	optimizer	[adam, adagrad, rmsprop]	sigmoid	sigmoid	sigmoid
	learning_rate	(0.0001,1)	0.0001	0.0001	0.0001
Transformer	num_layers	(1,6)	1	1	1
	d_model	[16,32,64,128,256,512,1024]	128	1024	128
	num_heads	[1,2,4,8]	8	1	2
	batch_size	[16,32,64,128,256,512]	16	64	16
	learning_rate	(0.0001,0.1)	0.0008	0.0001	0.0014

Table 3
Household load forecasting results for evaluation data set (2020).

	MAPE	MAE	RMSE	R^2
Random Forest	**12.84**	**959.35**	1362.61	**0.74**
Random Forest CEEMDAN	15.51	1093.85	1481.39	0.69
XGB	13.11	969.79	1362.03	**0.74**
XGB CEEMDAN	14.53	1036.19	1421.26	0.71
LSTM	14.08	1041.63	1447.21	0.70
LSTM CEEMDAN	17.12	1161.97	1522.47	0.67
Transformer	13.19	965.39	**1352.30**	**0.74**
Transformer CEEMDAN	15.60	1049.55	1373.21	0.73

Table 4
Heat pump load forecasting results for evaluation data set (2020).

	MAPE	MAE	RMSE	R^2
Random Forest	52.20	2054.40	2756.48	0.88
Random Forest CEEMDAN	97.54	2985.40	3897.60	0.76
XGB	49.17	2064.44	2817.45	0.88
XGB CEEMDAN	70.06	2617.10	3458.96	0.81
LSTM	70.51	2240.22	2908.37	0.87
LSTM CEEMDAN	49.32	1860.33	2482.39	0.90
Transformer	**33.03**	**1602.49**	2280.83	**0.92**
Transformer CEEMDAN	47.93	1690.77	**2226.14**	**0.92**

Table 5
Forecasting results for the whole energy community, including household and heat pump loads, for evaluation data set (2020).

	MAPE	MAE	RMSE	R^2
Random Forest: Separate	16.79	2377.94	3137.00	0.87
Random Forest: Combined	17.24	2466.57	3269.30	0.86
Random Forest CEEMDAN: Separate	25.40	3274.45	4247.91	0.77
Random Forest CEEMDAN: Combined	22.94	3277.46	4478.73	0.74
XGB: Separate	16.48	2375.88	3168.89	0.87
XGB: Combined	16.58	2432.93	3256.33	0.86
XGB CEEMDAN: Separate	20.80	2913.93	3827.78	0.81
XGB CEEMDAN: Combined	21.31	3056.01	4097.77	0.78
LSTM: Separate	18.76	2525.47	3265.96	0.86
LSTM: Combined	17.11	2509.65	3330.41	0.86
LSTM CEEMDAN: Separate	16.47	2314.98	3029.79	0.88
LSTM CEEMDAN: Combined	17.66	2426.39	3167.11	0.87
Transformer: Separate	**13.43**	**2014.11**	2743.59	**0.90**
Transformer: Combined	13.76	2062.61	2776.97	**0.90**
Transformer CEEMDAN: Separate	16.16	2150.27	2754.22	**0.90**
Transformer CEEMDAN: Combined	14.95	2089.65	**2736.40**	**0.90**

past might not be the most suitable ones in a future with significant heat pump loads.

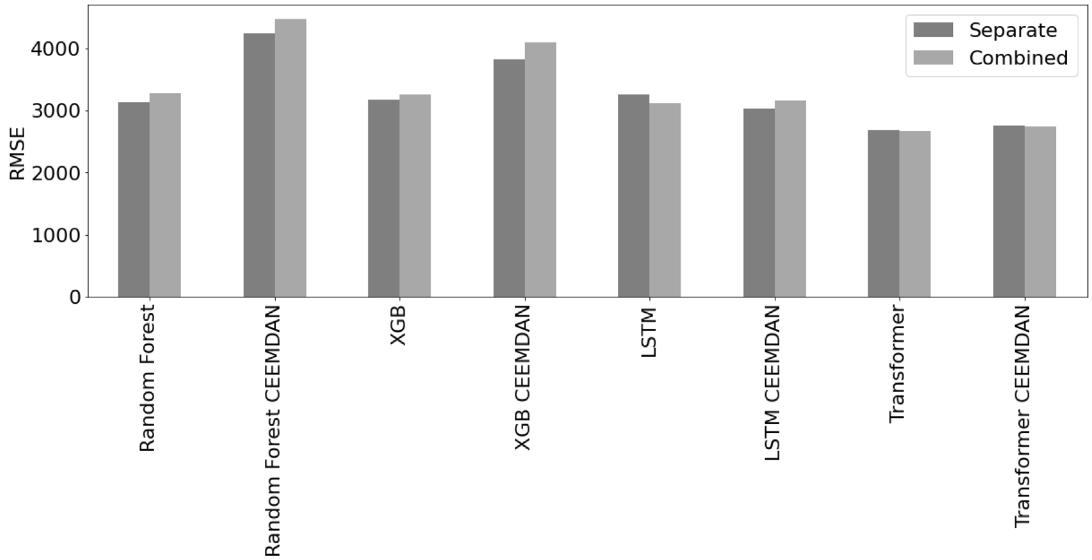
We compare the effects of the aggregation level in Fig. 5. Although separately forecasting heat pump and household loads and then aggregating the forecast brings improvements for some methods, especially the best-performing methods only exhibit negligible performance differences, or even perform better when using the directly combined aggregation level (Transformer-CEEMDAN).

5.3. Applicability

In the following, we investigate the applicability of the presented forecasts for day-ahead scheduling of a BESS sized at the hourly peak energy consumption of the year before (80.1 kWh). The overall yearly peak reduction, based on the optimization model detailed in Section 3.8, is presented in Fig. 6. We can again observe a strong performance of the Transformer-based methods, yielding solid peak reductions. The highest peak reduction of 5.7 kW is achieved with the Transformer model and the "combined" aggregation level, which represents 57% of the theoretical optimal peak reduction based on a perfect foresight forecast.

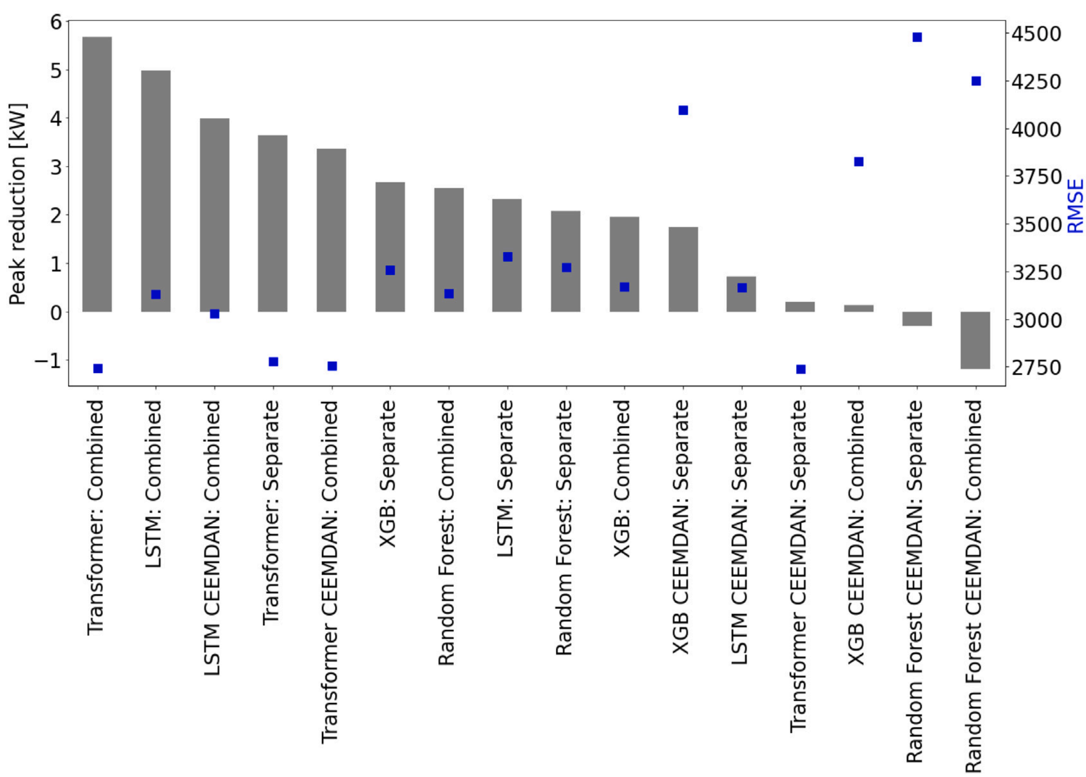
On the other hand, models with weak forecasting performance, such as the Random Forest CEEMDAN model, yield negligible or even negative peak reductions through the day-ahead scheduling, underlining the importance of high-quality forecasts. We compare peak shaving on the day with the highest peak load with the method that yields the best peak reduction results (Transformer Combined) and the worst (Random Forest CEEMDAN Combined) by considering the load curves depicted in Fig. 7. The Random Forest CEEMDAN method forecasts the highest peak in the afternoon, while the actual peak takes place in the morning. In addition, the load is consistently underestimated. Consequently, the peak load is not sufficiently reduced. Although the Transformer model also forecasts an afternoon peak, the overall load level and timing of peaks is forecasted better. Consequently, the BESS operation reduces the peak load level. We also observe that the three methods achieving the highest peak reduction follow the "combined" aggregation level, thereby delivering another indication that obtaining separate heat pump measurements does not necessarily lead to an operationally relevant improvement of the forecasting quality.

Although we see a tendency that models with good overall forecasting quality yield a reasonable decision basis for day-ahead BESS peak



(a)

Fig. 5. RMSE per method and aggregation level.



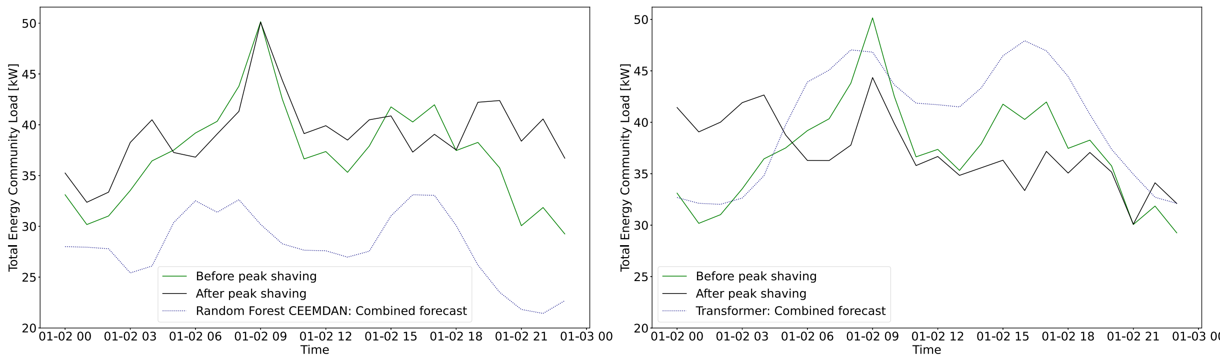
(a)

Fig. 6. Achieved peak reduction based on day-ahead scheduling of BESS based on forecasts. RMSE displayed in blue squares.

shaving scheduling, the "LSTM-CEEMDAN: Separate" or "Transformer-CEEMDAN: Separate" forecast with individually predicted household and heat pump loads raise awareness for potential problems when relying too much on forecast results. Even though both methods exhibit a solid forecasting performance, they lead to only limited peak reductions since they fail to correctly forecast the peak shape on the day with the worst peak load.

6. Discussion

The results of our study have implications for energy community and grid researchers and operators. We show that through the installation of heat pumps in energy communities, the autocorrelation patterns and peak load magnitudes significantly change. Based on that, the choice of adequate forecasting methods should be reviewed



(a) (b)

Fig. 7. Peak shaving results on the day with the highest peak load, based on day-ahead scheduling with the Random Forest CEEMDAN and Transformer model.

and re-evaluated. While for traditional energy communities, tree-based forecasting models, such as XGBoost or Random Forests, are delivering reasonable forecasting results, it is not the case anymore when heat pumps are installed. Then, in our case study, Transformer-based models are significantly outperforming all other investigated models.

Thereby, we also contribute to the general discussion about Transformer models in load forecasting: [61] find that linear models outperform multivariate Transformer models for long-term forecasting. However, with a global training strategy, Transformer models outperform linear models and other baselines for short-term and long-term forecasting [27,62,63]. On very aggregated load time series, Transformer models can also outperform several baselines significantly [60]. In our experiments with traditional household loads, we see no large difference in forecasting quality between the Transformer model and tree-based models. However, when investigating heat pump loads and total energy community loads including heat pumps, the Transformer models considerably outperform the other models.

We can see a comparable pattern for the CEEMDAN technique, which decomposes the load time series in different Intrinsic Mode Functions that are separately forecasted and later aggregated. While the method does not improve the traditional energy community forecast, it constitutes one of the best methods for forecasting loads of heat pumps and energy communities with them.

We can transfer these results also to our application case, in which peak loads are reduced through an external BESS, based on the forecast of the depicted methods. The highest peak reductions are consistently reached through the Transformer method, which also achieved good forecasting results. However, we note that we only analyzed a limited case study with a given storage size and limited load data. Hence, future research should also critically evaluate the applicability of load forecasts on energy community- and grid operation-related, actual tasks. Although the forecasting methods that achieve good forecasting metrics also tend to show good results in the actual peak reduction task, we can also see discrepancies between forecast quality and actual effectivity, for instance through failing predicting the peak load. Since most forecasting literature is focused on evaluating common metrics like the RMSE, without considering an actual use case, we call for a more task-centric forecast evaluation and the consideration of alternative metrics that might be more aligned with the task at hand.

Our result that separately predicting aggregated household and heat pump loads and then summing the forecasts up does not bring a meaningful advantage over forecasting directly the load of the whole energy community indicates that efforts to gather separate heat pump data might not be worthwhile. Instead, energy community and grid operators should focus on gathering solid load measurements at the transformer level, which can be used for forecasting models and operational decisions built upon them. We note that our study is focused on the low-voltage level. It has to be investigated if our findings hold true for the medium- and high-voltage levels.

Our study is based on the energy community household and heat pump load data from [12]. The water-to-water heat pumps from the dataset are operated based on desired household temperature levels, neglecting potential price-based demand response signals [84]. Through an increasing level of households with dynamic household prices, the load forecasting uncertainty could also rise, which should be considered in future studies [84]. In addition, new heat pump technologies, tariff structures and regulation can lead to concept drifts that make an adjustment of forecasting models necessary [85]. For instance, the German government has announced a new set of rules for controllable consumer devices – which will be implemented from 1st January 2024 on (§14a EnWG) [86] – allowing grid operators to reduce the electricity consumption of heat pumps and electric vehicle chargers down to 4.2 kW during overloading events. Applying these new rules could lead to considerable changes of heat pump load profiles and an increase in forecasting uncertainty. Also, we note that the most proliferated type of heat pumps in Germany are air-to-water ones [1], which can exhibit slightly different load profiles than the water-to-water heat pumps from in the underlying data set. However, due to comparable heat pump coefficients of performance, we argue that the heating demand of energy communities with water-to-water and air-to-water heat pumps – and respective forecasting outcomes – should be comparable [87].

Overall, we understand our work as a first step towards the discussion of forecasting energy community loads with heat pumps, which will get increasingly important over coming years, given the increasing number of heat pump installations [1]. Because of the observed changes in load curves, novel forecasting methods should be discussed and applied. However, we note that we have only explored a limited amount of models, given the high number of novel models published in recent years [13]. Hence, we publish large parts of our study open-source, including the feature engineering and selection steps, the underlying final data set, the evaluation pipeline, the peak shaving application, the best-performing methods and our resulting forecast. Thereby, we simplify benchmarking novel methods against our results and contribute to open-source load forecasting research[1].

We note that our study has a couple of limitations. The peak shaving application of our forecasted loads is based on a retrospective simulation, which might neglect practical factors. We aim to empirically validate our results in further studies. Also, we focus on forecasting aggregated energy community loads. In further studies, it might be interesting to analyze the effects of forecasting individual household

[1] Our best performing forecasting models, all our feature-engineered and preprocessed data, our benchmarking pipeline and all our results are published open-source at https://github.com/leloq/load-forecasting-with-heatpumps. We encourage fellow researchers to benchmark novel forecasting methods against our results.

loads before aggregation. Furthermore, the underlying dataset includes perfect foresight weather data, which might lead to slightly overestimating the forecasting quality. We argue that this does not interfere with the general direction of our results, given the overall good level of weather data forecasts and that all models are based on the same data, hence a fair comparison is given.

7. Conclusion

Our study investigates the impact of the installation of heat pumps in energy communities on day-ahead load forecasting with a state-of-the-art forecasting pipeline. The installation of heat pumps leads to remarkable changes in autocorrelation patterns and peak loads of the energy community. This has implications for the overall load forecasting process. In particular, we find that:

- The best-performing forecasting methods change after the installation of heat pumps. While for traditional energy communities, also tree-based models such as Random Forests or XGBoost deliver a reasonable forecasting quality, after installing heat pumps, Transformer-based methods outperform them significantly.
- The day-ahead energy community load forecasting quality cannot be notably increased by obtaining separate measurements of heat pump loads, which would constitute an additional effort for energy community or distribution grid operators.
- Transformer-based models are also delivering the best performance in a real-world peak reduction BESS use case for the investigated energy community with heat pumps. However, we see a discrepancy between forecasting metrics and actual results in the application task for some models.

Our findings have practical implications for operators of distribution grids and energy communities, making a re-evaluation of applied forecasting methods necessary and advocating against potentially expensive efforts to obtain separate heat pump load measurements.

We note that our study has some limitations: it is limited to a selected energy community, the practical application is based on a retrospective simulation and has not been empirically validated, the underlying data only includes water-to-water heat pumps, the forecast is based on perfect foresight weather data and only a selection of forecasting methods were applied.

Hence, we encourage researchers to use our dataset, results and evaluation pipeline, which we publish open-source, to benchmark novel methods against them to advance accurate forecasting techniques for loads of energy communities with heat pumps and to apply our methodology on alternative datasets. We also call for a more task-centric evaluation of forecasting methods, which might include the introduction of novel metrics that are more aligned with the application area of the produced forecasts. Also, future studies should investigate the impact of heat pump installations and aggregation levels on the forecast quality of the medium- and high-voltage grid, and empirically validate our results for energy communities and underlying low-voltage grids. Further studies should also consider the effect of severe weather events on the forecasting quality of energy communities with heat pumps.

CRediT authorship contribution statement

Leo Semmelmann: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matthias Hertel:** Visualization, Software, Methodology, Investigation, Conceptualization, Writing – original draft, Writing – review & editing. **Kevin J. Kircher:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Ralf Mikut:** Writing – review & editing, Supervision. **Veit Hagenmeyer:** Writing – review & editing. **Christof Weinhardt:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

See Table 6.

Table 6
Descriptive statistics of five runs per method, showing higher standard deviation (std) of neural network-based methods. Focusing on non-CEEMDAN methods, since the CEEMDAN method exhibits multiple runs internally through several IMFs.

	Mean	std	min	max
LSTM HH	1452.56	6.33	1447.21	1461.48
LSTM HP	2989.20	90.29	2908.37	3083.38
LSTM Comb	3096.27	58.89	3014.16	3152.67
XGB HH	1362.03	0.00	1362.03	1362.03
XGB HP	2817.45	0.00	2817.45	2817.45
XGB Comb	3256.33	0.00	3256.33	3256.33
Random Forest HH	1368.29	6.93	1362.01	1374.83
Random Forest HP	2761.41	4.09	2756.48	2765.77
Random Forest Comb	3256.02	10.81	3242.94	3269.30
Transformer HH	1324.57	13.72	1310.44	1340.74
Transformer HP	2319.38	27.72	2277.90	2335.79
Transformer Comb	2727.31	41.40	2688.11	2764.41

References

[1] Bundesverband Wärmepumpe (BWP). Wärmepumpenabsatz 2022: Wachstum von 53 prozent gegenüber dem vorjahr. 2023, URL: https://www.waermepumpe.de/presse/pressemitteilungen. [Accessed 02 April 2023].
[2] Protopapadaki C, Saelens D. Heat pump and PV impact on residential low-voltage distribution grids as a function of building and district properties. Appl Energy 2017;192:268–81.
[3] Çakmak HK, Hagenmeyer V. Using open data for modeling and simulation of the all electrical society in eASiMOV. In: 2022 open source modelling and simulation of energy systems. OSMSES, IEEE; 2022, p. 1–6.
[4] Logenthiran T, Srinivasan D, Shun TZ. Demand side management in smart grid using heuristic optimization. IEEE Trans Smart Grid 2012;3(3):1244–52.
[5] Stecca M, Elizondo LR, Soeiro TB, Bauer P, Palensky P. A comprehensive review of the integration of battery energy storage systems into distribution networks. IEEE Open J Ind Electron Soc 2020;1:46–65.
[6] Barone G, Buonomano A, Forzano C, Palombo A, Russo G. The role of energy communities in electricity grid balancing: A flexible tool for smart grid power distribution optimization. Renew Sustain Energy Rev 2023;187:113742.
[7] Coignard J, Janvier M, Debusschere V, Moreau G, Chollet S, Caire R. Evaluating forecasting methods in the context of local energy communities. Int J Electr Power Energy Syst 2021;131:106956.
[8] Kuster C, Rezgui Y, Mourshed M. Electrical load forecasting models: A critical systematic review. Sustainable Cities Soc 2017;35:257–70.
[9] Love J, Smith AZ, Watson S, Oikonomou E, Summerfield A, Gleeson C, Biddulph P, Chiu LF, Wingfield J, Martin C, et al. The addition of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial. Appl Energy 2017;204:332–42.
[10] Shaqour A, Ono T, Hagishima A, Farzaneh H. Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building. Energy AI 2022;8:100141.
[11] Sevlian R, Rajagopal R. A scaling law for short term load forecasting on varying levels of aggregation. Int J Electr Power Energy Syst 2018;98:350–61.
[12] Schlemminger M, Ohrdes T, Schneider E, Knoop M. Dataset on electrical single-family house and heat pump load profiles in Germany. Sci Data 2022;9(1):56.
[13] Wang H, Alattas KA, Mohammadzadeh A, Sabzalian MH, Aly AA, Mosavi A. Comprehensive review of load forecasting with emphasis on intelligent computing approaches. Energy Rep 2022;8:13189–98.
[14] Khuntia SR, Rueda JL, van Der Meijden MA. Forecasting the load of electrical power systems in mid-and long-term horizons: a review. IET Gener Transm Distrib 2016;10(16):3971–7.
[15] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods: Tools for decision making. European J Oper Res 2009;199(3):902–7.
[16] Xu X, Jin X, Jia H, Yu X, Li K. Hierarchical management for integrated community energy systems. Appl Energy 2015;160:231–43.

[17] Javadi MS, Gough M, Nezhad AE, Santos SF, Shafie-khah M, Catalão JP. Pool trading model within a local energy community considering flexible loads, photovoltaic generation and energy storage systems. Sustainable Cities Soc 2022;79:103747.

[18] Anees A, Dillon T, Wallis S, Chen Y-PP. Optimization of day-ahead and real-time prices for smart home community. Int J Electr Power Energy Syst 2021;124:106403.

[19] Parra D, Norman SA, Walker GS, Gillott M. Optimum community energy storage for renewable energy and demand load management. Appl Energy 2017;200:358–69.

[20] Hagan MT, Behr SM. The time series approach to short term load forecasting. IEEE Trans Power Syst 1987;2(3):785–91.

[21] Tarsitano A, Amerise IL. Short-term load forecasting using a two-stage SARIMAX model. Energy 2017;133:108–14.

[22] Dudek G. Short-term load forecasting using random forests. In: Intelligent systems' 2014: proceedings of the 7th iEEE international conference intelligent systems iS'2014, September 24-26, 2014, Warsaw, Poland, volume 2: tools, architectures, systems, applications. Springer; 2015, p. 821–8.

[23] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 785–94.

[24] Semmelmann L, Resch O, Henni S, Weinhardt C. Privacy-preserving peak time forecasting with learning to rank XGBoost and extensive feature engineering. IET Smart Grid 2023.

[25] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 2017;10(1):841–51.

[26] Gao J, Chen Y, Hu W, Zhang D. An adaptive deep-learning load forecasting framework by integrating transformer and domain knowledge. Adv Appl Energy 2023;10:100142.

[27] Hertel M, Beichter M, Heidrich B, et al. Transformer training strategies for forecasting multiple load time series. Energy Inform 2023;6(Suppl 1):20. http://dx.doi.org/10.1186/s42162-023-00278-z.

[28] Gong M, Zhao Y, Sun J, Han C, Sun G, Yan B. Load forecasting of district heating system based on Informer. Energy 2022;253:124179.

[29] Chung WH, Gu YH, Yoo SJ. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. Energy 2022;246:123350.

[30] Powell KM, Sriprasad A, Cole WJ, Edgar TF. Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy 2014;74:877–85.

[31] Bacher P, Madsen H, Nielsen HA, Perers B. Short-term heat load forecasting for single family houses. Energy Build 2013;65:101–12.

[32] Xu C, Chen H, Xun W, Zhou Z, Liu T, Zeng Y, Ahmad T. Modal decomposition based ensemble learning for ground source heat pump systems load forecasting. Energy Build 2019;194:62–74.

[33] Song Y, Peskova M, Rolando D, Zucker G, Madani H. Estimating electric power consumption of in-situ residential heat pump systems: A data-driven approach. Appl Energy 2023;352:121971.

[34] Fischer D, Madani H. On heat pumps in smart grids: A review. Renew Sustain Energy Rev 2017;70:342–57.

[35] Semmelmann L, Henni S, Weinhardt C. Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data. Energy Inform 2022;5(1):1–21.

[36] vom Scheidt F, Medinová H, Ludwig N, Richter B, Staudt P, Weinhardt C. Data analytics in the electricity sector–A quantitative and qualitative literature review. Energy AI 2020;1:100009.

[37] Reihani E, Motalleb M, Ghorbani R, Saoud LS. Load peak shaving and power smoothing of a distribution grid with high renewable energy penetration. Renew Energy 2016;86:1372–9.

[38] Hou H, Liu C, Wang Q, Wu X, Tang J, Shi Y, Xie C. Review of load forecasting based on artificial intelligence methodologies, models, and challenges. Electr Power Syst Res 2022;210:108067.

[39] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns 2023;4(9).

[40] Pfenninger S, DeCarolis J, Hirth L, Quoilin S, Staffell I. The importance of open data and software: Is energy research lagging behind? Energy Policy 2017;101:211–5.

[41] Gilbert C, Browell J, Stephen B. Probabilistic load forecasting for the low voltage network: forecast fusion and daily peaks. Sustain Energy Grids Netw 2023;34:100998.

[42] Ran P, Dong K, Liu X, Wang J. Short-term load forecasting based on CEEMDAN and Transformer. Electr Power Syst Res 2023;214:108885.

[43] Zhu J, Dong H, Zheng W, Li S, Huang Y, Xi L. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. Appl Energy 2022;321:119269.

[44] Ziel F. Modeling public holidays in load forecasting: a German case study. J Mod Power Syst Clean Energy 2018;6(2):191–207.

[45] Haben S, Arora S, Giasemidis G, Voss M, Greetham DV. Review of low voltage load forecasting: Methods, applications, and recommendations. Appl Energy 2021;304:117798.

[46] Wang P, Liu B, Hong T. Electric load forecasting with recency effect: A big data approach. Int J Forecast 2016;32(3):585–97.

[47] Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Comput 2000;12(10):2451–71.

[48] Fan S, Chen L. Short-term load forecasting based on an adaptive hybrid method. IEEE Trans Power Syst 2006;21(1):392–401.

[49] Liang R-H, Cheng C-C. Short-term load forecasting by a neuro-fuzzy based approach. Int J Electr Power Energy Syst 2002;24(2):103–11.

[50] Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Front Bioinform 2022;2:927312.

[51] Koprinska I, Rana M, Agelidis VG. Correlation and instance based feature selection for electricity load forecasting. Knowl-Based Syst 2015;82:29–40.

[52] Hu Z, Bao Y, Xiong T, Chiong R. Hybrid filter–wrapper feature selection for short-term load forecasting. Eng Appl Artif Intell 2015;40:17–27.

[53] Breiman L. Random forests. Mach Learn 2001;45:5–32.

[54] Lahouar A, Slama JBH. Day-ahead load forecast using random forest and expert input selection. Energy Convers Manage 2015;103:1040–51.

[55] Fan G-F, Zhang L-Z, Yu M, Hong W-C, Dong S-Q. Applications of random forest in multivariable response surface for short-term load forecasting. Int J Electr Power Energy Syst 2022;139:108073.

[56] Abbasi RA, Javaid N, Ghuman MNJ, Khan ZA, Ur Rehman S, Amanullah. Short term load forecasting using XGBoost. In: Web, artificial intelligence and network applications: proceedings of the workshops of the 33rd international conference on advanced information networking and applications (WAINA-2019) 33. Springer; 2019, p. 1120–31.

[57] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.

[58] Lin J, Ma J, Zhu J, Cui Y. Short-term load forecasting based on LSTM networks considering attention mechanism. Int J Electr Power Energy Syst 2022;137:107818.

[59] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;30.

[60] Hertel M, Ott S, Schäfer B, Mikut R, Hagenmeyer V, Neumann O. Evaluation of Transformer architectures for electrical load time-series forecasting. In: Proceedings 32. Workshop computational intelligence. Vol. 1, 2022, p. 93.

[61] Zeng A, Chen M, Zhang L, Xu Q. Are Transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence. Vol. 37, 2023, p. 11121–8.

[62] Emami P, Sahu A, Graf P. BuildingsBench: A large-scale dataset of 900k buildings and benchmark for short-term load forecasting. Adv Neural Inf Process Syst (NeurIPS) 2023.

[63] Nie Y, H. Nguyen N, Sinthong P, Kalagnanam J. A time series is worth 64 words: Long-term forecasting with Transformers. In: International conference on learning representations. 2023.

[64] Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. J Electron Sci Technol 2019;17(1):26–40.

[65] Wistuba M, Schilling N, Schmidt-Thieme L. Hyperparameter search space pruning–a new component for sequential model-based hyperparameter optimization. In: Machine learning and knowledge discovery in databases: European conference, ECML pKDD 2015, Porto, Portugal, September 7-11, 2015, proceedings, part II 15. Springer; 2015, p. 104–19.

[66] Lévesque J-C, Gagné C, Sabourin R. Bayesian hyperparameter optimization for ensemble learning. 2016, arXiv preprint arXiv:1605.06394.

[67] Li K, Huang W, Hu G, Li J. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. Energy Build 2023;279:112666.

[68] Karijadi I, Chou S-Y. A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction. Energy Build 2022;259:111908.

[69] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing. ICASSP, IEEE; 2011, p. 4144–7.

[70] Duan Y, Zhang J, Wang X. Henry hub monthly natural gas price forecasting using CEEMDAN-bagging-HHO-SVR. Front Energy Res 2023;11:1323073.

[71] Bayer B, Matschoss P, Thomas H, Marian A. The German experience with integrating photovoltaic systems into the low-voltage grids. Renew Energy 2018;119:129–41.

[72] Kell A, McGough AS, Forshaw M. Segmenting residential smart meter data for short-term load forecasting. In: Proceedings of the ninth international conference on future energy systems. 2018, p. 91–6.

[73] Bhela S, Kekatos V, Veeramachaneni S. Enhancing observability in distribution grids using smart meter data. IEEE Trans Smart Grid 2017;9(6):5953–61.

[74] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30(1):79–82.

[75] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623.

[76] Uddin M, Romlie MF, Abdullah MF, Abd Halim S, Kwang TC, et al. A review on peak load shaving strategies. Renew Sustain Energy Rev 2018;82:3323–32.

[77] Gurobi Optimization, LLC. Gurobi optimization, LLC. 2023, URL: https://www.gurobi.com. Gurobi Optimizer version 9.5.
[78] Van Der Stelt S, AlSkaif T, Van Sark W. Techno-economic analysis of household and community energy storage for residential prosumers with smart appliances. Appl Energy 2018;209:266–76.
[79] Dong S, Kremers E, Brucoli M, Brown S, Rothman R. Impact of household heterogeneity on community energy storage in the UK. Energy Rep 2020;6:117–23.
[80] Navarro-Espinosa A, Mancarella P. Probabilistic modeling and assessment of the impact of electric heat pumps on low voltage distribution networks. Appl Energy 2014;127:249–66.
[81] Wang Z, Hong T, Piette MA. Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 2020;263:114683.
[82] Walther J, Spanier D, Panten N, Abele E. Very short-term load forecasting on factory level–a machine learning approach. Procedia CIRP 2019;80:705–10.
[83] He Y, Tsang KF. Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. Energy Rep 2021;7:6473–88.
[84] Klaiber S, Bretschneider P, Waczowicz S, Mikut R, Konotop I, Westermann D. A contribution to the load forecast of price elastic consumption behaviour. In: 2015 IEEE eindhoven powerTech. IEEE; 2015, p. 1–6.
[85] Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. ACM Comput Surv (CSUR) 2014;46(4):1–37.
[86] Bundesnetzagentur. Bundesnetzagentur sets rules for the integration of controllable consumer devices. 2023, URL: https://www.bundesnetzagentur.de/SharedDocs/Pressemitteilungen/EN/2023/20231127_14a.html. [Accessed 14 December 2023].
[87] Çakır U, Çomaklı K, Çomaklı O, Karslı S. An experimental exergetic comparison of four different heat pump systems working at same conditions: As air to air, air to water, water to water and water to air. Energy 2013;58:210–9.