



Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features

Lechen Li^a, Christoph J. Meinrenken^{b,d}, Vijay Modi^{c,d,*}, Patricia J. Culligan^{a,d}

^a Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, USA

^b Earth Institute, Columbia University, New York, USA

^c Department of Mechanical Engineering, Columbia University, New York, USA

^d Data Science Institute, Columbia University, New York, USA

HIGHLIGHTS

- Forecasted hourly apartment electricity use (59 apartments, 11 floors, 3 seasons).
- Developed novel neural network with selected auto-regressive features.
- Compared accuracy of model against 4 benchmark models.
- Novel neural network improves forecasting accuracy by up to 25%.
- Identified underlying drivers of accuracy and discussed their application.

ARTICLE INFO

Keywords:

Short term load forecasting
Dynamic feature selection
ConvLSTM neural network
“Default” state of model
Time-series electricity forecasting difficulty
analysis

ABSTRACT

Residential electricity load profiles and their diversity have become increasingly important to realize the benefits of Smart or Transactive Energy Networks (TENs). An important element of TENs will be practical, accurate, and implementable residential load forecasting techniques. While there have been many approaches to short-term load forecasting, few have included forecasting for individual households, partly because the high volatility and idiosyncrasies present in individual household load data can pose significant challenges. In this study, we develop a Convolutional Long Short-Term Memory-based neural network with Selected Autoregressive Features (termed a CLSAF model) to improve short-term household electricity load forecasting accuracy by employing three strategies: autoregressive features selection, exogenous features selection, and a “default” state to avoid overfitting at times of high load volatility. We include aggregations of apartments to floor and building level, because utilities may favor transactive approaches that rely on aggregator models, e.g., a cluster of consumers as opposed to an individual. We demonstrate that the CLSAF model, by virtue of its enhanced feature representation and modest computational resources, can accomplish load forecasting in a multi-family residential building across three spatial granularities (individual apartment/household, floor, and building levels), with an accuracy improvement of up to 25% compared to a persistence model. We propose a data screening technique to characterize time-series electricity-load data. This technique is suitable for integration into a TEN ecosystem and allows one to estimate confidence levels of the load forecasts to optimize computational resources and the risks associated with uncertain forecasts.

1. Introduction

1.1. Background and motivation

In recent years, residential electricity load profiles have become

increasingly varied among neighborhoods and homes due to modified work and leisure patterns, increased use of electronics, and more frequent presence of distributed generation (e.g., roof top photovoltaic) and storage (e.g., electric vehicles) [1]. This increases the benefit of and need for electrical networks such as Transactive Energy Networks

* Corresponding author at: Department of Mechanical Engineering, Columbia University, New York, USA.

E-mail address: modi@columbia.edu (V. Modi).

<https://doi.org/10.1016/j.apenergy.2021.116509>

Received 29 August 2020; Received in revised form 12 January 2021; Accepted 13 January 2021

Available online 1 February 2021

0306-2619/© 2021 Elsevier Ltd. All rights reserved.

(TENS) [1], which could transform homes from being a passive load into a smart storage and demand responsive entity for electric grids, thus enabling a dynamic balance of demand and deeper integration of emerging clean electricity generation (Fig. 1). For example, Zheng et al. [2] introduced a model for leveled storage cost, based on storage lifetime and electricity tariffs, and developed a storage dispatch algorithm to optimize the storage size and the grid demand limits. Similarly, as reviewed by Song et al. [3], a host of novel market mechanisms and respective technology solutions are under consideration to improve the resiliency and reduce the carbon intensity of electricity grids. Most of these innovations will either require, or at least benefit from, the ability to forecast short-term electricity consumption patterns at the level of individual actors in a TEN (with “short-term” typically referring to 30 min to one-week time periods, but not longer) [4]. In the case of multi-family residential buildings, which are common in many urban areas around the world, the level of an individual actor could include an individual apartment, a floor, or an entire apartment building, for example.

Implementation of such intelligent and adaptive elements requires advanced techniques for accurate and precise load demand and power generation forecasting [5]. For short-term load forecasting, many approaches have been studied but few have focused on the electricity load of individual households, for two reasons: First, electricity load profiles of individual households can reveal private information that often cannot be published, contributing to a lack of data availability for the residential sector, especially in multi-family residential buildings. Second, forecasting the electricity load of individual households is conventionally considered challenging due to the volatile nature of household load data [6].

1.2. Research context and previous work

A large portion of the existing work on electricity use forecasting has focused on commercial buildings due to the availability of datasets and the often more easily identifiable diurnal use patterns (reviewed in, e.g., Meinrenken and Mehmani [7]). For residential buildings, researchers have developed various statistical models and machine learning algorithms for load prediction. Many of them used datasets containing only one level of spatial aggregation (e.g., the aggregate load profile of an entire building). A few studies have carried out comparative

experiments based on various scenarios to investigate the influence of forecast granularities (e.g., load at level of individual apartments or their aggregates at floor or building level). Some studies have aimed at improving forecasting accuracy by overcoming some common challenges of machine learning algorithms such as overfitting (reviewed by Amasyali et al. [8]).

In the following, we focus on three specific aspects of load forecasting models that are of particular relevance to the methodology in the present study, and how it builds on previous work, namely (i) particular challenges in the residential sector; (ii) spatial granularity; and (iii) feature selection and the role of sparse models. A summary of the forecasting accuracies in prior studies, separated by spatial granularity, is shown in Fig. 2. For a more complete overview of existing forecasting

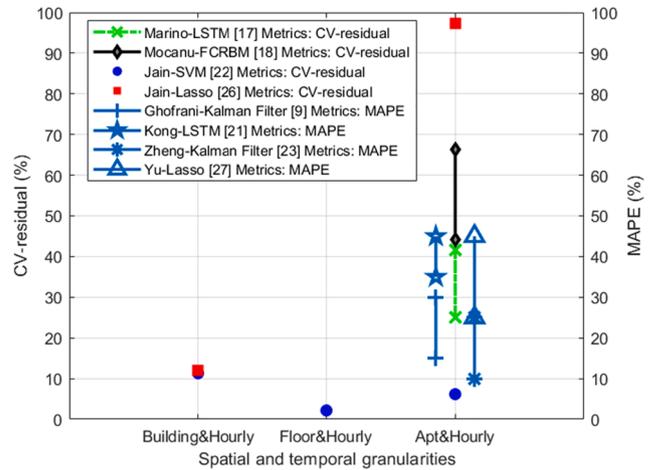


Fig. 2. A summary of forecasting accuracies achieved in previous related studies. CV-residual denotes the Coefficient of Variation, as used, e.g., by Jain et al. [22] and MAPE denotes the Mean Absolute Percent Error, as used, e.g., by Ghofrani et al. [9]. Some of the CV-residual plotted here were originally assessed as root mean squared error (RMSE), as used, e.g., by Marino et al. [17]. In order to make the results more comparable, we divided the RMSE by the respective averages of their corresponding load observations to change the metric from RMSE to CV-residual. The three spatial granularities shown correspond to the ones analyzed in the present study.

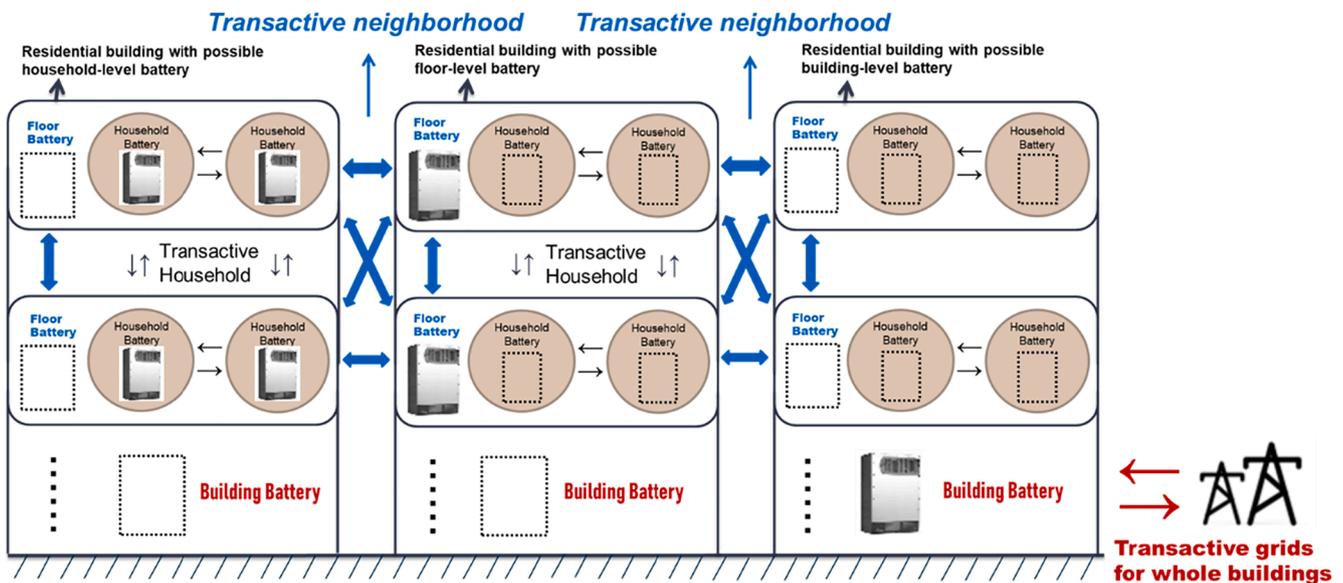


Fig. 1. Illustrative setup of sharing temporarily stored electricity in a Transactive Energy Network (TEN) for a Multifamily Urban Residential Building. Three illustrative multi-family residential buildings show three possible choices of battery systems (one per apartment, one shared per each floor, or one shared per building) which would require load forecasting at the level of apartments, floors, or buildings, respectively.

models, we refer to existing reviews such as the ones cited above.

1.2.1. Load forecasting models for residential electricity use

As noted, few previous studies have focused on the forecasting of electricity load in residential buildings, and only some on individual households. One such study, by Ghofrani et al. [9], forecasted the electricity load of one specific household. A Kalman Filter estimator was applied, and the load was forecasted hourly and sub-hourly as a sum of two separate components: the weather-dependent component and the lifestyle component. The authors used mean absolute percentage error as the accuracy metric and obtained forecasting accuracies between 18% and 30%. Munkhammar et al. [10] employed what is referred to as a “Markov-chain mixture distribution model” to forecast one step ahead (half-hour resolution) residential electricity consumption data from Australia. Iwafune et al. [11] used a multiple linear regression model to predict one-day ahead electricity consumption of a single household with a mean absolute percentage error (MAPE) forecast accuracy of 12.4%.

Previous studies also addressed the problem of identifying an optimal model for residential load forecasting tasks by comparing the accuracies of various machine learning algorithms. For example, Edward et al. [12] implemented seven different models, including multiple linear regression, support vector machine, and deep neural networks, to forecast one-hour ahead electricity loads of a residential building. While the tested models showed reliable forecasts when considering the average coefficient of variation (CV), compared to similar work, their datasets were limited to only three individual households. Besides this, in the work by Iwafune et al. [11] the electricity load data was not measured but simulated based on several energy-usage patterns of some US households, which were recorded in the *Building-America* study [13]. Therefore, the need remains to validate such models on larger datasets of measured electricity consumption.

In order to improve forecasting techniques, some studies have tried hybrid models and demonstrated that these could perform better than a single statistical or machine-learning model. Fan et al. [14] set up a new ensemble model to forecast next-day electricity use and electricity-load peaks by implementing eight base models. They concluded that the ensemble models can achieve higher accuracies compared with single base models. Zhang et al. [15] developed a hybrid approach that used an autoregressive integrated moving average, improved mode decomposition, and a wavelet neural network optimized with a fruit fly optimization approach. Their forecast results illustrated that such model performs better than other single machine learning models in electricity load forecasting accuracy.

In recent years, deep learning models have been shown to offer many advantages, and they often perform better than traditional machine learning. Both Zheng et al. [16] and Marino et al. [17] have succeeded in applying a Long Short-Term Memory (LSTM) Neural Network to short-term load forecasting in residential buildings. They concluded that a LSTM neural network has an advantage in handling data-driven electricity consumption forecasting tasks. Mocanu et al. [18] used a factored conditional restricted Boltzmann machine, a deep learning method, for a residential electricity forecasting task. They obtained notable improvements in comparison with shallow neural networks and support vector machines. Recently, Andriopoulos et al. [19] applied a Convolutional Neural Network (CNN) to a short-term load forecasting task for three individual households. They employed a statistical analysis to convert their original dataset to a format that facilitated leverage of the advantages of the CNN algorithm. They concluded that the proposed CNN can outperform conventional LSTM in cases where the number of data observations are limited (such as the loads in a small energy community) and the load patterns change dynamically.

1.2.2. Forecasting at different spatial and temporal scales

In addition to modeling techniques, the spatial and/or temporal scale of forecasting (or sometimes granularity [20]) is another factor affecting

the forecasting accuracy. Electricity load data in multi-family residential buildings, for example, can be obtained at varying temporal granularity such as 15 mins, 1 h, or 1 day, and at different spatial granularities such as household, floor, or building level. Determining the optimal forecast granularity is an important aspect of improving accuracy of the forecast. Lusi et al. [20] forecasts 30 mins, 60 mins, and 120 mins ahead, using four STLF methods for individual households. They concluded that prediction accuracy is higher for 2-hour ahead forecasts. Kong et al. [21] proposed an LSTM neural network-based approach and conducted load forecasts at both grid sub-station and household level. They compared the forecasting accuracy in two situations, namely, aggregating the forecast for each household or forecasting only the aggregated total substation load. They concluded that aggregating the individual household-load forecasts can improve their prediction accuracy by 1.08% for BPNN and 0.49% for LSTM models.

Determining the optimal spatial and temporal granularities at the same time, Jain et al. [22] applied a support vector regression (SVR) model to make one-step load predictions for a residential building at 10-min, hourly, and daily temporal granularities, as well as household, floor, and building spatial granularities. They found the optimal forecasting granularity to be for one-hour ahead and at floor level. Zheng et al. [23] developed a Kalman filter-based bottom-up method to increase the accuracy of household-load forecasting. They verified the advantages of this approach via granularity analysis at the level of appliances, rooms, and household, and found that the Kalman filter bottom-up method at the appliance level can improve household load forecasting accuracy. Xu et al. [24] applied a probability-based electricity forecasting model for buildings that decomposed the load into a baseline load and an abnormal peak load. They concluded that such a decomposition technique can provide more granular data for forecasting models and hence increase forecasting accuracy.

1.2.3. Feature selection and sparse models

Sparse models and feature selection techniques have been shown to improve electricity load prediction by capturing certain key features [25]. Therefore, these approaches could be utilized to obtain a generalized model by lowering the risk of overfitting, as they can focus on a small amount of core information highly correlated with electricity use.

Regarding the sparse coding techniques used in existing literature, Jain et al. [26] applied a lasso regression model, which is a shrinkage and selection approach to linear regression that approximates sparse coefficients, to forecast energy use in an NYC multifamily residential building. They concluded that the lasso regression model provides competitive performance compared with a support vector machine. Yu et al. [27] studied the use of a sparse coding approach to forecast the electricity load of several individual households. They obtained 10% improvements in the accuracy of forecasting next-day total load and next-week total load (compared with classical methods) when they added sparse code features in ridge regression. Recently, Kaneko et al. [28] proposed a technique to identify annually dominant explanatory variables by devising scenario-dependent models to quantify hourly use. The proposed approach, based on enumerating sparse, partially linear additive models and a linear programming approach, was successful in identifying the key dominant variables. Candanedo et al. [29] presented a data-filtering method by removing non-predictive parameters and unrelated features, to improve the performance of 4 statistical models for the energy use of appliances in a low-energy house. With the method, they concluded that the gradient boosting machines (GBM) outperformed the other 3 used models, which achieved the accuracy of 57% in R^2 . However, their data-filtering method for the GBM and their approach of separating training and testing datasets are not applicable to us, since we need to keep updating the values of our model parameters to handle the changing load profiles of our datasets (introduced in Section 2.1).

Clustering of historical household load curves is another technique that could be combined with sparse coding techniques to better

accomplish electricity load forecasting tasks. Kwac et al. [30] used an adaptive K-means algorithm to cluster historical load curves of households and identified different types of typical household load patterns (without yet using them for forecasting). Benitez et al. [31] analyzed the electricity consumption of residents in Spain by applying a dynamic clustering approach to better evaluate electricity consumption trends of individual households. Chaouch et al. [32] applied some cluster-based features to improve functional time-series forecasting for household electricity loads.

More recently, with respect to forecasting electricity load and other time-series data, some studies have implemented feature selection by integrating advanced deep learning techniques. Hernández et al. developed improved neural networks to forecast the next day's aggregated load at an electrical power station. Seven models based on their designed multilayers of neural networks were proposed, progressively adding input variables after analyzing the influence of climate factors on the aggregated load [33]. Amarasinghe et al. [34] developed a 1-D convolutional neural network (1D-CNN) performing energy load forecasting at individual building level. Their experimental results showed that the CNN outperformed SVR. However, using such an approach here is unlikely to succeed when temporal load profiles have large seasonal volatility and unexpected load changes due to human behavior. Regarding the recently used recurrent neural networks, Wang et al. [35] developed a novel short-term load forecasting method based on the attention mechanism (AM), rolling update (RU) and bi-directional long short-term memory (Bi-LSTM) neural network. When comparing the Bi-LSTM model with AM and RU to a traditional Bi-LSTM model, both the mean absolute percentage error (MAPE) and the root mean square error (RMSE) were shown to decrease in the load forecasting associated with their two data sets. Wilms et al. and Wan et al. [36,37] both employed encoder-decoder layers by using a sequence-to-sequence (Seq2Seq) network and a temporal convolutional network respectively, to yield better hidden representation of features for time-series data forecasting. They concluded that their developed architectures outperform many multivariate regressions techniques.

1.3. Focus of present study and differentiation from previous work

The above-mentioned studies, using sparse coding techniques and advanced neural network techniques, usually automatically obtain the most influential hidden feature representation by using fixed types of features. However, this approach usually does not consider whether these features are always dominant under different situations such as seasonal changes and some idiosyncratic human behaviors. Similarly, although most of the above discussed studies applied past electricity load values as important features for prediction, they did not consider dynamic methods of continuously updating the selection to the most correlated feature types in order to enhance the feature representation before feeding them to the forecasting models. Such a dynamic feature-selection process is needed when electricity consumption of an apartment (as in our dataset described in Section 2.1) could be primarily due to loads from a refrigerator and standby-mode electronics that are present even when a resident is not in the apartment for up to several days. In such cases, predictions are difficult as the load pattern changes between a state where residents are generally at home and one where they are away for several days. This can lead to overfitting if this problem is addressed by using multiple previous time-step load values (as done by the aforementioned studies) and by relying only on the feature-selection process of the encoder-decoder layers of Seq2Seq and TCNN or the sum of weighted states of AM-based LSTM structure. This is because the inference by hidden states of the historical load of individual apartments sometimes cannot be used to build a correct or strong connection to the forecasting time step. Aiming to address these issues in the present study, we extend a previously introduced Convolutional LSTM framework (ConvLSTM) [38], whose built-in kernels allow the extraction of key information, by adding a dynamic feature-selection algorithm and a

model-simplification approach, which enables timely reactions to the rapidly changing states of various load profiles in case of overfitting. The resulting ConvLSTM-based neural network with selected autoregressive features (henceforth CLSAF model) is tested as a short-term load prediction in a multifamily residential setting over three different season types (winter, summer, and the shoulder seasons of spring or fall) and across three spatial granularities (apartment, floor, and building level).

To test the feasibility and forecasting performance of our approach, we use a residential apartment building in New York City, NY, USA as a case study. We use the actual, hourly apartment-level electricity load of 59 individual apartments across 11 floors and from three different seasons (2019 data) to train the forecasting models and evaluate their accuracy. This data-rich case study allows us to systematically evaluate the effects of season, spatial granularity, and model choice on the forecasting accuracy. Finally, we determine two key characteristics of the residential load data and how these affect the forecasting accuracy for different apartments or floors. Based on this analysis, we discuss basic elements of a possible data screening technique, which could aid in providing confidence levels of load predictions to facilitate more complex transaction schemes within TENS.

2. Data and methods

2.1. Overview of electricity data

Our electricity use dataset is of a building in Manhattan, New York, NY (IECC climate zone 4A [39]), a multi-story residential building built prior to 1940. The building is a pre-war construction with a steam-based, central heating system and electric window air conditioners for cooling. Therefore, air conditioning loads are reflected in the apartments' electricity use, whereas heating loads are not (except for the occasional supplementary heating via personal electric space heaters or heating blankets, for example). Electricity use for every apartment was separately metered by a Siemens® SEM3 micro-meter system with 50-amp split core current transformers and $\pm 1\%$ accuracy [40]. As the model training data at apartment-level, we used the incremental electricity consumption (kWh) from one hour to the next. For the floor and building level, we first aggregated the observed electricity load of the respective apartments at either the floor or building level, and then used the aggregated data as training data to forecast the aggregate level (as opposed to first predicting the apartment-level data and then aggregating the forecasted apartment-level data, as done by, e.g., Jain et al. [20]).

The dataset contains 59 individual apartments, eleven floors, and one building, for three time periods in 2019, to reflect various weather conditions during the year: a period in winter when the use of indoor lights and possible auxiliary use of electric space heaters is highest (Jan. 7th to Feb. 3rd); a period during a shoulder season when little or no auxiliary heating but also little or no air conditioning will be used (Apr. 1st to Apr. 28th); and a period in the summer when the use of air conditioning is high (July 15th to Aug. 11th). In order to ensure comparability of the 3 different time periods, each period was chosen to start on a Monday and to last exactly 28 days, such that the different periods would each comprise of the same number of weekdays and weekend days. For convenience, we henceforth refer to these three periods simply as 'January', 'April' and 'July', respectively.

Fig. 3(a) displays the diurnal load profiles averaged over all 59 apartments. Fig. 3(b) shows three examples of the hourly consumption of individual apartments during a one-day period. Data averaged over all apartments show systematic load patterns (e.g., high in the evening hours, low during the night), whereas some individual apartments do not, with volatile loads, partially caused by residents leaving the apartment for several days at a time. As reviewed in the Introduction, such idiosyncratic patterns render load forecasting more challenging. In response to such challenges, the model approach developed here aims at extracting the most correlated information from daily load profiles as

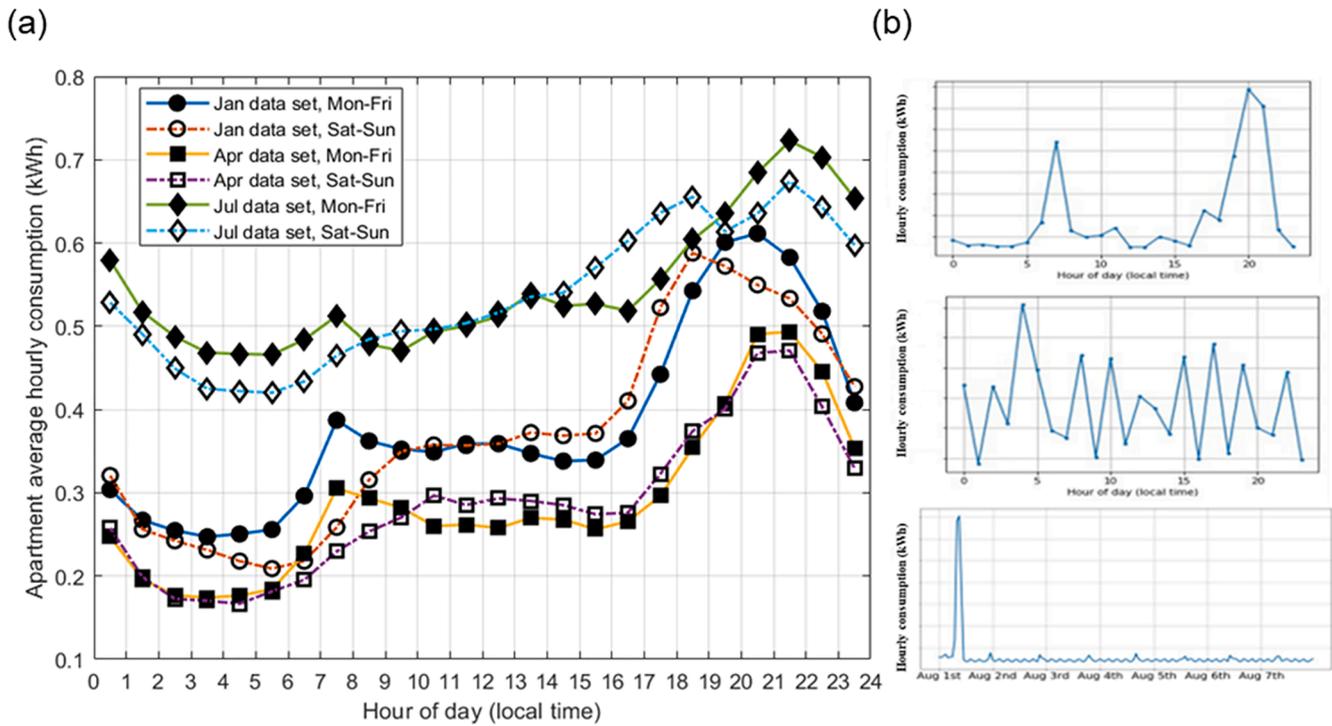


Fig. 3. (a) Diurnal patterns of average hourly electricity consumption in 59 apartments. The patterns are shown separately for weekdays and weekends, across the three time periods (Jan. 7th to Feb. 3rd; Apr. 1st to Apr. 28th; and July 15th to Aug. 11th). (b) Example daily electricity-load profiles of three sample apartments. The first two are daily profiles of 24 h in April and January, respectively, and the third is a daily profile of one week in August. For the sample apartments, actual hourly consumption data and the exact time period are not shown, due to privacy considerations. All data is from 2019.

prediction features, in order to mitigate the interference of idiosyncratic human behavior with prediction accuracy.

2.2. Metric to evaluate forecasting accuracy

In past studies, four types of metrics have been used to assess forecasting accuracy [41]: (i) scale-dependent measures, e.g., Root Mean Square Error or Mean Absolute Error; (ii) normalized metrics, e.g., Mean Absolute Percent Error (MAPE) or Coefficient of Variation (CV); (iii) relative metrics such as Mean Relative Error; and (iv) scale-free metrics such as Mean Absolute Scaled Error [21]. Since the scale-dependent measures cannot be used for comparing the accuracy of forecasting at different magnitudes and MAPE is not applicable when handling the case of zero load values, in the present study we use CV, as applied, e.g., by Jain et al. [20]. We refer to CV as CV-residual, in order to distinguish it from another, similarly defined metric in the following sections. CV-residual is defined as follows:

$$CV_{residual} = \frac{\sqrt{\frac{1}{N-1} \sum_{t=1}^N (y_t - \hat{y}_t)^2}}{\bar{y}} \quad (1)$$

where N is the number of individual hourly load observations for which the load is forecasted. In our study, N is equal to 504 ($24 \times 7 \times 3$), representing the hourly load over the last three weeks of each 4-week time period (the first week is used for training and the last three weeks are used for accuracy evaluation, as shown in Table 1). y_t and \hat{y}_t are the observed and predicted hourly load at time step t , respectively. \bar{y} is the mean value of the N observations of the hourly electricity load.

2.3. Forecasting models and features used in this study

In the present study, we firstly tried 4 benchmark models (introduced in Section 2.3.1) and a ConvLSTM model (introduced in Section 2.3.2) to

complete the forecasting task. Then, by overcoming some disadvantages of the ConvLSTM model, and combining it with some advantages of one benchmark model, a more accurate and robust CLSAF model (introduced in Section 2.3.3) was developed, which can carry out short-term load forecasts for all scenarios (i.e., for the three spatial granularities and three seasons). Table 1 provides an overview of the 6 models and the corresponding feature types used in the present study.

After an overview of the input data, the subsequent sections are organized as follows: Section 2.3.1 illustrates 4 benchmark models to establish some baselines for achievable accuracies. Section 2.3.2 presents the mechanism and implementation steps of a previously used ConvLSTM model. Section 2.3.3 presents the mechanism and implementation steps of the novel CLSAF model and demonstrates the advantages of the CLSAF model over both the ConvLSTM model and the Persistence model.

Input data, source, and accuracy

For the source and accuracy of the electricity data, please refer to Section 2.1. All weather data (i.e., both for training and for testing) was historical 2019 data as shown in Table 1, which was obtained from the National Oceanic and Atmospheric Association (NOAA) [42], NY Central Park Station. The temperature data is accurate to ± 0.3 °C, and the humidity and wind speed data to around $\pm 1\%$ [42]. The building is about 1 mile from the weather station. In practical applications of the forecasting, the weather conditions used as the exogenous features for the forecasting of each time-step will be the one-hour weather forecasts (however, for the initial training (“warm-up”) period, the models would still use actual, observed weather conditions, as shown in Table 1). As can be seen from the typically achievable load forecasting accuracies (Fig. 2), these relatively minor uncertainties in the input data do not materially contribute to the uncertainty of load forecasts.

Regarding the use of exogenous features shown in Table 1, dry-bulb temperature (henceforth “temperature”), absolute humidity (henceforth “humidity”), wind speed, binary weekday/weekend, and the sinusoid of local time were chosen as our predictors. Wind speed rather than wind

Table 1

Overview of the inputs, outputs, and initial training (warm-up) periods for the employed 6 models (4 benchmark models and 2 newly employed models). The definition of initial training (warm-up) period will be introduced in Section 2.3.2. Prediction horizon is single step forward. Weather conditions were obtained from the National Oceanic and Atmospheric Association (NOAA) [43], NY Central Park Station.

Model name	Autoregressive features (historical hourly electricity load), $y[t]$ denotes current time-step load)	Exogenous features (hourly granularity)	Initial training (warm-up) period	Output (one-step ahead hourly electricity load)
Persistence SW-ARIMA	$y[t-1]$ Selected by the default setting of “forecast” package in R	None None	None First 7 days of each 28-day period	$y[t]$ $y[t]$
SW-ETS	Selected by the default setting of “forecast” package in R	None	First 7 days of each 28-day period	$y[t]$
SW-SVR	$y[t-1]$	Temperature[t], absolute humidity[t], wind speed[t], binary weekday/weekend[t] and sin(local time [t])	First 7 days of each 28-day period	$y[t]$
ConvLSTM	$y[t-1]$	Temperature[t], absolute humidity[t], wind speed[t], binary weekday/weekend[t] and sin(local time [t])	First 7 days of each 28-day period	$y[t]$
CLSAF	Selected $y[t-p_i]$ or $y[t-1]$ (p_i denotes a selected lag from time index)	Temperature[t], absolute humidity[t], wind speed[t], binary weekday/weekend[t] and sin(local time [t])	First 7 days of each 28-day period	$y[t]$

direction was chosen as one of the features as the prevailing wind direction at a weather station is not indicative of the actual wind direction at a specific apartment in a dense urban setting. On the other hand, wind speed and solar radiation are more likely to be closely associated with cooling and lighting needs at the apartment in question and have been used in our analyses. However, when solar radiation was added as an additional exogenous feature, it was detrimental to accuracy; this may be because in multi-family high rise buildings, only the predominantly south facing apartments or the apartments on the higher floors are subjected to direct solar radiation, even if the sun shines. Therefore, we decided to remove irradiance from the list of exogenous features.

2.3.1. Benchmark models

Persistence models can be applied as a benchmark for time-series prediction applications [44]. In this study, as we aim at conducting a single-step hourly forecast, the persistence model we adopt uses the hourly load observed during the most recent time step:

$$\text{Singlestepforecasting} : \hat{y}_t = y_{t-1} \quad (2)$$

where \hat{y}_t is the predicted hourly load at time step t and y_{t-1} is the

observed hourly load at time step $t-1$.

Auto Regressive Integrated Moving Average (ARIMA) models and Exponential Smoothing (ETS) models are two strong, and well-established model types for time-series forecasting [45]. Therefore, they were selected among our series of benchmark models and used on all the aforementioned datasets, with default parameters automatically selected by using the “forecast” package in R [46]. In addition, as discussed in *Introduction*, the Support Vector Regression (SVR) has proven to be a well-performing model in residential load forecasting [20], so it was also selected as a benchmark model, using the same features as the ones for the employed ConvLSTM model to set up a SVR, as shown in Table 1.

Regarding the training process for the ARIMA, ETS, and SVR models in the study, an efficient sliding-window (SW) approach was used as in Khan et al. [47]: After the initial training via the first 7-day load data of each 28-day period (introduced in Section 2.3.2), a sliding window was employed for continuous training to update model parameters for each forecasted time-step. The window size was determined by exploiting spectral components of the historical load data using the Lomb–Scargle method [48]. The reason for using the sliding window is that the load patterns of the individual apartments change a lot due to the seasonal changes and idiosyncratic resident behavior, as discussed in Section 2.1. For convenience, we henceforth refer to these three benchmark models simply as SW-ARIMA, SW-ETS and SW-SVR, respectively, as shown in Table 1.

2.3.2. Convolutional long short-term memory neural network (ConvLSTM) model

2.3.2.1. Motivation for adopting ConvLSTM model

Long Short-Term Memory (LSTM) Neural Networks have been proven to be an efficient and powerful approach to short-term residential load forecasting tasks across multiple spatial granularities, as shown, e.g., by Kong et al. [21]. However, an LSTM model might not completely meet the requirements of our dataset, for two reasons: First, as shown in Section 2.1, electricity loads in some apartments are volatile without clear diurnal patterns. Second, the primary factors driving electricity load may vary between seasons. In particular, the ambient temperature will likely affect the electricity consumption of air conditioners during the summer time but will be less relevant in wintertime when a building is centrally heated. Consequently, it might be best to only use the exogenous features most correlated with electricity load. In this study, we tried a ConvLSTM layer to capture core information of the exogenous features that are highly correlated with electricity load, by taking advantage of the built-in kernels. Also, such a ConvLSTM layer can be easily combined with our developed dynamic feature-selection algorithm and “default” state, which will be discussed in Section 2.3.3.

2.3.2.2. Mechanism and implementation of ConvLSTM model

A ConvLSTM neural network, introduced by Xingjian et al. [38], is a variant of the LSTM neural network, which integrates a convolution operation into the LSTM cell. The convolution operation takes the place of a matrix multiplication at each of the LSTM cell’s gate, and thereby captures inherent spatial features by several convolution operators in multi-dimensional data. Xingjian et al. [38] applied their proposed ConvLSTM network to better capture the spatiotemporal correlations of their spatial data. They concluded that the ConvLSTM network outperforms an LSTM with fully connected layers for precipitation nowcasting.

A ConvLSTM cell consists of a series of operations that can store temporal information with a selection process by the built-in kernels, and timely erases the cell’s memory, like an LSTM cell, to prevent gradient vanishing [49]. Fig. 4 displays the basic mechanism of a ConvLSTM cell whose operations can be formulated as six core equations:

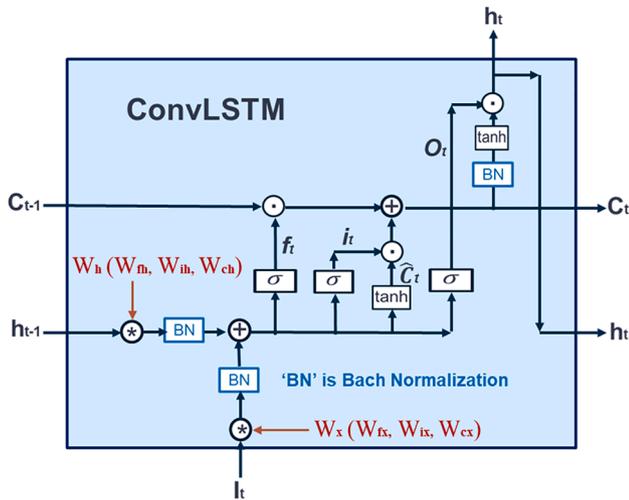


Fig. 4. Mechanism of a ConvLSTM cell. ‘ \odot ’ denotes the Hadamard product and ‘ $*$ ’ denotes the convolutional operator. Adapted from Xingjian et al. [38] and Marino et al. [17].

$$f_t = \sigma(W_{fx} * I_t + W_{fh} * h_{t-1} + b_f)$$

$$i_t = \sigma(W_{ix} * I_t + W_{ih} * h_{t-1} + b_i)$$

$$\hat{C}_t = \tanh(W_{cx} * I_t + W_{ch} * h_{t-1} + b_c)$$

$$o_t = \sigma(W_{ox} * I_t + W_{oh} * h_{t-1} + b_o)$$

$$C_t = \hat{C}_t \odot i_t + C_{t-1} \odot f_t$$

$$h_t = \tanh(C_t) \odot o_t \quad (3)$$

where ‘ \odot ’ denotes the Hadamard product and ‘ $*$ ’ the convolution operation. I_t is the input of the ConvLSTM cell at time step t . h_{t-1} and C_{t-1} are the output and state of the ConvLSTM cell at time step $t-1$, respectively. Similarly, h_t and C_t are the output and state of the cell at time step t . They are generated by several joint computations based on four intermediate vectors: f_t , i_t , \hat{C}_t , and O_t at time step t . W_{fx} , W_{fh} , W_{ix} , W_{ih} , W_{cx} , W_{ch} , W_{ox} , and W_{oh} are trainable weights that appear in pairs for each intermediate vector. b_f , b_i , b_c , and b_o are corresponding trainable biases.

As shown in Table 1, we employed the electricity load of the most recent time-step as the only autoregressive feature, and temperature, humidity, local time, wind speed, and a binary variable (1 for Mon-Fri and 0 for Sat-Sun) as the exogenous features for the prediction. In this case, the exogenous features vector $E_{[t]}$, the input vector $I_{[t]}$, and the predicted hourly load $\hat{y}_{[t]}$ for time step t , are defined as:

$$E_{[t]} = [\text{temperature}_{[t]} \text{humidity}_{[t]} \text{time}_{[t]} \text{windspeed}_{[t]} \text{weekday\&weekend}_{[t]}]$$

$$I_{[t]} = [y_{[t-1]} E_{[t]}]$$

$$\hat{y}_{[t]} = \text{ConvLSTM}\{I_{[t]}\} \quad (4)$$

where $y_{[t-1]}$ denotes the observed hourly load at time step $t-1$, $I_{[t]}$ is composed of $y_{[t-1]}$ and exogenous vector $E_{[t]}$, as the input vector at time step t , and $\hat{y}_{[t]}$ denotes the corresponding output (forecasted hourly load). As discussed by Xingjian et al. [38], the ConvLSTM cell expects the feature dimension of an individual input to be a two-dimensional array. In our case however, we treat $I_{[t]}$ as a single feature with a dimension of one by six.

As the model is employed to conduct single-step load forecasting, there is no need to separate the data into training and testing data.

However, the model needs a warm-up period for initial training to adjust itself to the best state (the warm-up period is necessary for the benchmark models as well, for the same reason). Therefore, for each of the three 28-day periods, the load data of the first-week (first 7 days) was used as the warm-up period for initial training, and after that the model was formally employed to make the forecast. After forecasting each time-step, the observed hourly load at the last predicted time step was used for parameter updating. The reason for choosing the first-week load data of each 28-day dataset as the initial training (warm-up) period was twofold: First, through multiple experiments, we found that if the warm-up period exceeds 2 days, the accuracy of the subsequent forecast will converge, meaning that the forecasting accuracy would not rise if a longer warm-up period were implemented (due to the varying load patterns as discussed in Sections 2.1 and 2.3.1). Second, choosing the previous one week as the warm-up period (instead of a longer period) can make our developed dynamic feature-selection algorithm (as introduced in Section 2.3.3) characterize the historical electricity diurnals of the targeted apartment quickly without high computational resources. Table 2 shows the hyper parameters of the ConvLSTM model, and Fig. 5 shows its un-rolled sequential architecture, which can be used for both the training (by back-propagation through time, BPTT) and for the one-step-ahead load forecasting.

2.3.3. ConvLSTM neural network with selected auto-regressive feature (CLSAF model)

2.3.3.1. Motivations for developing CLSAF model.

As shown in Fig. 3(b), the load profile in some apartments is characterized by idiosyncratic human behavior (an apartment’s temporary vacancy, for example), which could prompt overfitting in the traditional ConvLSTM model (see example shown in Fig. 10). In order to preempt such overfitting, a better-suited feature representation was devised which dynamically selects the most correlated lagged hourly load (rather than inflexibly using the previous one or several load values irrespective of their correlation with the next load value). We thus extended the ConvLSTM model by two additional strategies: An autocorrelation-function (ACF)-based algorithm to select the most correlated lagged load as the autoregressive feature, and a “default” state in which the Persistence Model was employed whenever the algorithm failed to obtain a lagged load with sufficient correlation to serve as an autoregressive feature. The “default” state, conceived as a model-simplification, is aimed at handling overfitting issues, which were mostly caused by an apartment’s load profile changing between periods of occupancy vs. vacancy, as shown in Section 3.2.2. The CLSAF model was set up by a combination of the above methods, which are described one by one below.

Table 2

Structure and hyper parameters of the ConvLSTM model. All hyper parameters were obtained by tuning manually, based on a large number of experiments, via a process similar to that described in, e.g., Xingjian et al. [38]. The hyper parameters shown here are specific to the present study and were chosen to balance forecasting accuracy and robustness across the three spatial granularities and seasons.

Property	Value
Structure	One ConvLSTM2D layer and two dense layers
Number of filters	36
Kernel size	1×2
Activation function	Relu
Nodes number of first dense layer	4
Nodes number of second dense layer	1
Activation function of dense layers	Relu
Epoch	20
Size of batch	1
Loss function	Mean Squared Error (MSE)
Optimizer	Adam
Training (warm-up) period	First week of each 28-day period
Training time (over CPU)	20 s (see Section 2.5)

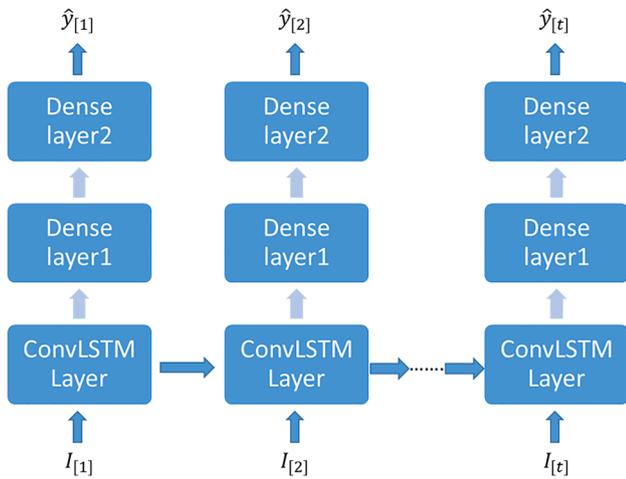


Fig. 5. Un-rolled sequential architecture of the ConvLSTM model. This architecture was designed specifically for the present study. It can be used both for the model training by backpropagation through time (BPTT) (“warm-up”) and for the one-step load forecasting. $I_{[t]}$ and $\hat{y}_{[t]}$ denote the input and output of the ConvLSTM model at time step t , respectively.

2.3.3.2. Autocorrelation-based algorithm. The proposed algorithm, based on an autocorrelation function (ACF), is aimed at selecting the lagged hourly load most correlated with the one-step-ahead load as the autoregressive feature for prediction. The ACF computes the correlation of the time-series lagged values with themselves, thus investigating the periodical nature of a time-series dataset. It is formulated as follows:

$$r_k = \frac{\text{Cov}(y_t, y_{t+k})}{\sqrt{\text{Var}(y_t) \cdot \text{Var}(y_{t+k})}} \quad (5)$$

where Cov and Var denote covariance and variance, respectively, and y is the observed hourly load at the given time step t or $t + k$. $\text{Var}(y_t)$ and $\text{Var}(y_{t+k})$ are two variances of the hourly loads with a separation by k time steps. r_k denotes the correlation of the hourly load values with a lag of k hours apart. The autocorrelations at these different lags together form the autocorrelation function [50]. In the present study, we set the range of lags returned by the ACF to be 24 ($k = 0, 1, 2, \dots, 24$), to capture diurnal patterns, with r_k measured over the previous 7 days’ hourly load data (starting with the one-week warm-up period, i.e., the longest period available in the dataset prior to the first model employment time step). Fig. 6 displays the ACF results of two example apartments.

The ACF is used in the implementation of the model as follows. First,

the ACF is employed to compute the autocorrelation in the previous week’s hourly load data to obtain the most correlated lag. Then, the corresponding lagged hourly load is selected as the autoregressive feature, but only for the next one-step-ahead hourly forecast. After moving to the next forecasting time step, the algorithm updates the previous one-week data by adding the latest hourly observation. It then repeats the autocorrelation computation and selection process to update the most correlated lagged load for the next-step prediction.

For example, it can be seen in Fig. 6(a) that the first example apartment exhibits a regular pattern with approximately 24 h periodicity. The highest auto-correlation is at the smallest lag considered, i.e., 1 h. (We ignore the correlation at lag 0 because it is a self-correlated result.) Therefore, in this case, the algorithm would select the load of the previous hour as the autoregressive feature for the next forecast (indicated as step “Select optimal lagged load ...” in the flowchart in Fig. 7). By comparison, Fig. 6(b) shows that the 2nd example apartment exhibits a much weaker diurnal electricity load pattern. In this example, the highest auto-correlation is at a lag of 22 h, but the correlation is small, suggesting that a 22-hour lag feature might lead to overfitting. This motivated us to improve the CLSAF model further by developing a “default” state, which we explain in the next section.

2.3.3.3. Default state and implementation in CLSAF model. Fig. 7 shows the role of the default state – which acts as a more robust option as a fallback for the load forecasting – and its dynamic implementation in the CLSAF model. We defined a new variable θ , referred to as the autocorrelation threshold, which determines at what time steps the prediction model switches back and forth between the neural network-based forecasting and the forecasting based on the default state. The optimal value of θ was determined by a calibration procedure based on experiments (see next section). As shown in Fig. 7, at every time step, the autocorrelation-based algorithm is employed to select the most correlated lag, as detailed above. Then the neural network state of the CLSAF model (left path in Fig. 7) is employed for the one-step ahead hourly forecasting by the selected lagged hourly load and the exogenous features stated earlier. However, the resulting model output is used as the CLSAF model’s forecast only if the correlation of the selected lag was larger than the threshold θ . Otherwise, the CLSAF model’s “default” state (right path in Fig. 7) is activated by using the Persistence Model to obtain the forecast for the next time-step. This procedure is repeated at every time step. It is important to note that even after the initial “warm-up” training over the first-week, the parameters related to the CLSAF’s neural network and the most correlated lag are still updated for each time-step during the forecasting, regardless of whether the actual forecast is taken from the CLSAF’s neural network or from its default state.

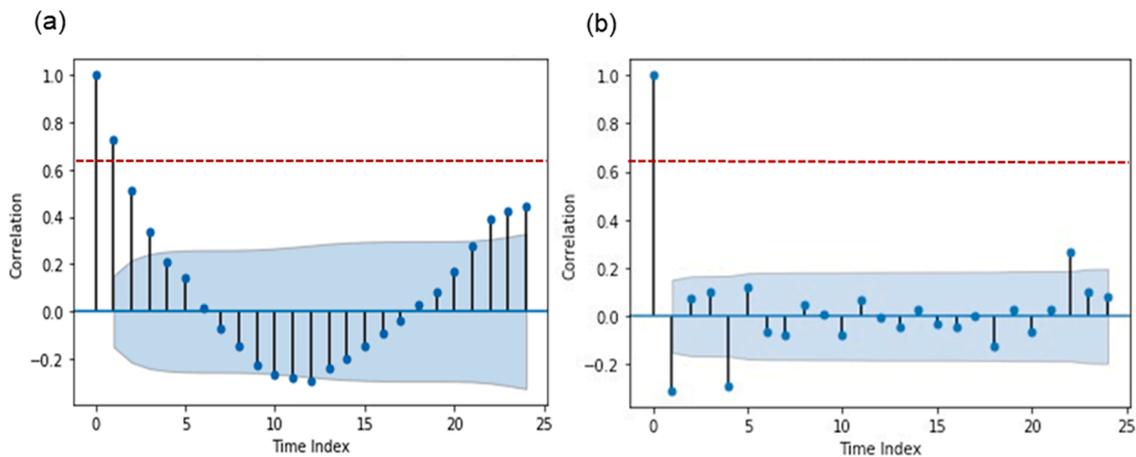


Fig. 6. Autocorrelation r_k of two example apartments with up to 24 lags (July data). Blue shading shows a 95% confidence level to indicate whether the correlation at each lag is significantly different from zero (using the `Plot_acf` function from the `Statsmodels` module [51] in Python 3). Red dashed line shows the autocorrelation threshold θ as defined in Fig. 7. “Time index” (k) is in hours.

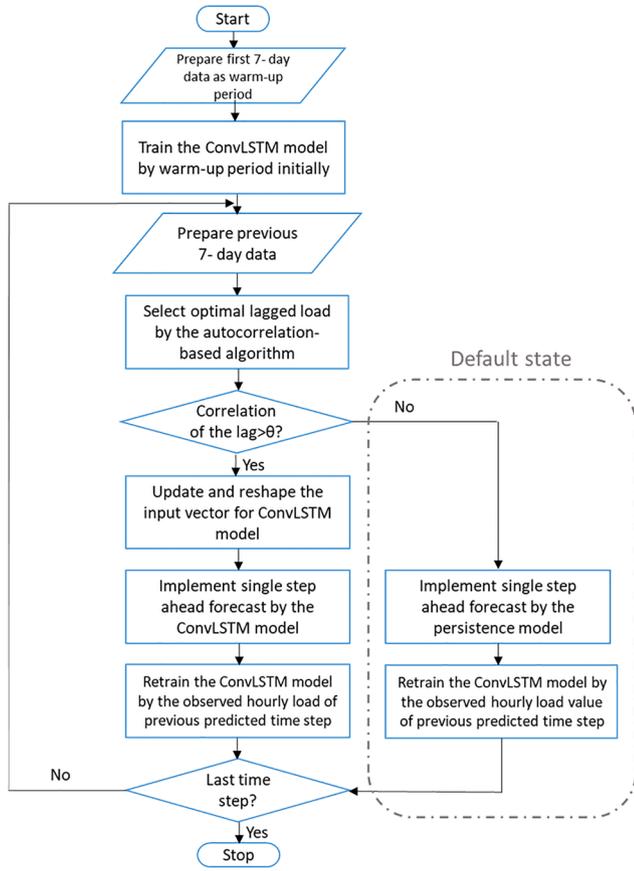


Fig. 7. Flowchart of CLSAF model. The CLSAF model consists of a standard ConvLSTM model with two additions: the autocorrelation-based algorithm, which dynamically selects the most-suited prior load, and the “default” state which ensures robust forecasts during periods of possible overfitting.

This ensures that the model can switch back seamlessly to the neural network-based forecast whenever the correlation for the most correlated lag is above θ . The exogenous-feature vector $E_{[t]}$, the input vector $I_{[t]}$, and the predicted hourly load $\hat{y}_{[t]}$ of the CLSAF model for time step t are defined as follows:

$$E_{[t]} = [\text{temperature}_{[t]} \text{humidity}_{[t]} \text{time}_{[t]} \text{windspeed}_{[t]} \text{weekday\&weekend}_{[t]}]$$

$$I_{[t]} = \begin{cases} [y_{[t-p_t]} E_{[t]}] & \text{Case one} \\ y_{[t-1]} & \text{Case two} \end{cases}$$

$$\hat{y}_{[t]} = \begin{cases} \text{ConvLSTM}(I_{[t]}) & \text{Case one} \\ y_{[t-1]} & \text{Case two} \end{cases} \quad (6)$$

where the input vector $I_{[t]}$ and the predicted hourly load $\hat{y}_{[t]}$ have two cases. The first case means the load is forecasted by the neural network of the CLSAF model. In case one, $y_{[t-p_t]}$ is the most correlated hourly load selected by the algorithm as the autoregressive feature at time step t . $I_{[t]}$ is the input vector that consists of the selected lagged load $y_{[t-p_t]}$ and the exogenous-feature vector $E_{[t]}$ at time step t . Similarly, $I_{[t]}$ needs to be reformatted into the required dimension of the ConvLSTM framework, as described in Section 2.3.2. The predicted hourly load $\hat{y}_{[t]}$ denotes the load forecasted for time step t . The second case means that the load is forecasted by the CLSAF’s default state (i.e., Persistence Model). In this case, the predicted hourly load $\hat{y}_{[t]}$ is equal to $y_{[t-1]}$ at time step t .

2.3.4. Calibration of optimum theta

As shown in Fig. 7, the threshold θ determines which state of the

forecasting model is used at which time steps and, therefore, impacts forecasting accuracy. To determine the optimal value of θ , experiments were carried out to quantify the average achieved CV-residual at apartment level (sample of randomly selected 20 of the 59 apartments, for 3 seasons). As shown in Fig. 8, the best average forecasting accuracy (i.e., lowest average CV-residual) of the CLSAF model is achieved with $\theta = 0.64$. As θ increases from 0.64 to 0.9, it is increasingly unlikely that the correlation of the selected lag is greater than θ , thus resulting in the default state of the CLSAF being employed more frequently. Similarly, when θ is decreasing from 0.64 to 0.3, it is increasingly likely that the correlation of the selected lag is greater than θ , thus favoring the neural network to produce the load forecast. $\theta = 0.64$ was used in all subsequent analyses.

2.4. Fast Fourier Transform (FFT) to assess strength of diurnal pattern

We used frequency spectrum analysis to characterize the daily electricity load profiles, using a Fast Fourier Transform (FFT) algorithm which uses periodicity and symmetry to significantly reduce the computation time [52]. For a sequence of electricity loads y_n at N time steps, the discrete Fourier transform (DFT) is formulated as follows:

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-\frac{2\pi k n i}{N}} \quad (7)$$

where N denotes the sequence length. Previous studies [38] have shown that the computation of the DFT can be separated into two sections, odd and even, reducing the computational complexity and thus allowing for a more favorable analysis of the spectrum [53]. Therefore, Eq. (7) can be written as:

$$Y_k = \sum_{m=0}^{\frac{N}{2}-1} y_{2m} e^{-\frac{2\pi k m i}{N/2}} + e^{-\frac{2\pi k i}{N}} \sum_{m=0}^{\frac{N}{2}-1} y_{2m+1} e^{-\frac{2\pi k m i}{N/2}} \quad (8)$$

where $k = 0, 1, 2, \dots, N-1$, and $0 \leq n < M \equiv N/2$. Y_k is the original amplitude by transformation, in terms of the frequency k .

Once the original amplitudes Y_k for all frequencies ($k = 0, 1, 2, \dots, N-1$) were determined, we used a scaling approach by standardizing the original amplitudes to generate comparable amplitudes of specific frequencies across the 3 spatial granularities (apartment, floor, and building levels). The standardization was formulated as follows:

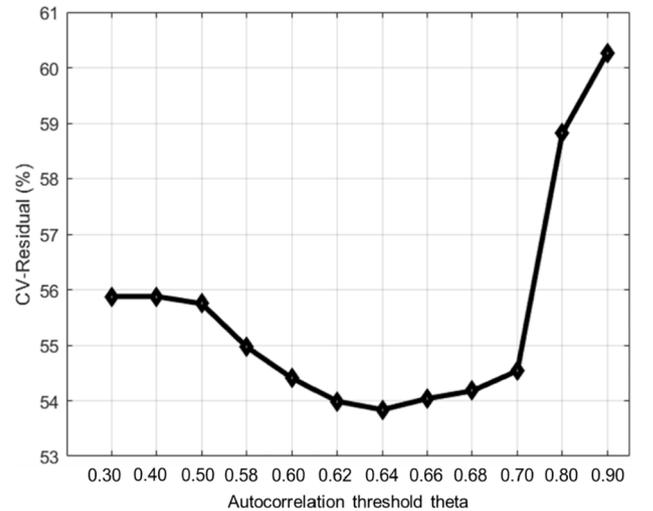


Fig. 8. Threshold theta vs. average CV-residual of 20 sampled apartments. The theta resulting in the lowest CV-residual across all apartments and 3 seasons, theta = 0.64, was used for all analysis shown in Results. (x-axis not drawn to scale in order to emphasize the behavior around optimum theta.)

$$Y_k^s = \frac{Y_k - \mu_Y}{\sigma_Y} + C \quad (9)$$

where Y_k^s denotes the standardized amplitude of frequency k . μ_Y and σ_Y are the mean and the standard deviation of the original amplitudes. The constant C prevents negative amplitude values and was set to 0.5.

In order to quantify the strength of diurnal patterns of the load profiles, we defined a new variable S , as shown in Eq. (10), which is the mean value of the standardized amplitudes (Eqs. (8) and (9)) at two specific frequencies, namely 1 cycle per day and 2 cycles per day:

$$S = \frac{Y_{k_1}^s + Y_{k_2}^s}{2} \quad (10)$$

where $Y_{k_1}^s$ and $Y_{k_2}^s$ denote the standardized amplitudes as per Eq. (9), and k_1 and k_2 represent the specific frequencies 1 cycle per day and 2 cycles per day, respectively.

2.5. Computational resource requirements

The ConvLSTM and CLSAF models evaluated in this study were run on a standard computer with Intel (R) core (TM) 1.99 GHz CPU and 16 Gb of memory. The code was written in Python. No significant computational resource or code was needed for the Persistence Model as the load forecast is simply executed by applying the previous observed hourly load. The ConvLSTM and CLSAF models require approximately the same computational resources because the CLSAF model is a combination of the ConvLSTM and Persistence models. The CPU time for the warm-up period for each model was about 20 s. Only 0.2 s were required for each subsequent time-step for parameter updating and prediction. Therefore, a standard machine with one CPU could easily provide the required computational power for a real-life application of the CLSAF model in a TEN, meaning that each next-hour-load could be forecasted near instantaneously as soon as the previous time step's load has been measured and exogenous variables have been collected.

3. Results

3.1. Best performing models

As discussed above, the principal challenge of load forecasting with respect to our dataset is the large volatility of loads in individual apartments. Thus, in selecting the best performing models, our priority was focused on the performance of all models in forecasting apartment-level load data. An overall summary of apartment-level forecasting accuracies by the 6 employed models (4 benchmark models and two new models) is provided in Table 3.

As shown in Table 3, the SW-ARIMA, SW-ETS, and SW-SVR models have worse accuracy (higher CV-residual) than the ConvLSTM and the CLSAF models when forecasting the load of individual apartments. Notably, their accuracies are even lower than the accuracy of the Persistence Model. Therefore, in the following, we only pay attention to the forecasting results of the Persistence Model, the ConvLSTM model, and the CLSAF model.

Table 3

Overall average CV-residuals (Eq. (1)), with corresponding minimum and maximum, of apartment-level load forecasting for all three datasets (January, April, and July) by the 4 benchmark models and the 2 newly employed models.

Model name	Mean value	Minimum	Maximum
Persistence	61.2	6.3	141.4
SW-ARIMA	64.1	6.4	201.5
SW-ETS	63.7	6.4	188.4
SW-SVR	62.0	6.3	162.3
ConvLSTM	57.9	6.2	131.1
CLSAF	53.3	5.9	115.8

3.2. CV-residuals by spatial granularity, models, and seasons

Fig. 9 shows the forecasting accuracies of the three models (Persistence, ConvLSTM, and CLSAF), evaluated by CV-residual (Eq. (1)), for all nine scenarios (three spatial granularities and three seasons). The forecasting accuracy varies as a function of spatial granularity, model type, and season, as analyzed in more detail in the following sections.

3.2.1. Effect of spatial granularity on forecasting accuracy

As shown in Fig. 9, the highest average accuracy is achieved at the building level (lowest CV-residual), followed by floor level and then apartment level (highest CV-residual). To test this result for statistical significance, we carried out t-tests (two-tailed, unequal variances).

For the floor vs. apartment level, nine such tests (Floor-Jan & Apt-Jan, Floor-April & Apt-April and Floor-Jul & Apt-Jul, across three models) were carried out. These showed that the average accuracies of all 9 combinations are significantly different ($p < 0.05$), confirming that the floor level forecasting outperforms that at apartment level.

For the building level, no further statistical tests were carried out because our dataset only contained one building. However, as seen in Fig. 9, the CV-residual at the building is smaller than even the minimum CV-residual of any of the floors. Consequently, the building level produces the highest forecasting accuracy for our dataset.

3.2.2. Effect of model type on forecasting accuracy

As shown in Fig. 9, the CLSAF model yields the highest average accuracy, followed by the ConvLSTM model, and then the Persistence model. To verify the statistical significance of this finding for floor and apartment level, nine paired t-tests (two-tailed, unequal variances) were carried out (Persistence & ConvLSTM, Persistence & CLSAF and ConvLSTM & CLSAF, across three seasons). The results show that the averages of all 9 combinations are significantly different ($p < 0.05$).

One typical example that illustrates an advantage of the CLSAF model compared to the ConvLSTM model is shown in Fig. 10: When either model is confronted with a period of vacancy in an apartment, the CLSAF model reacts to the change faster, regardless of whether the apartment changes from occupied to vacant (around August 2nd in Fig. 10) or vice versa (after August 10th). This is because the CLSAF model can switch its state back and forth between the neural network and the Persistence Model (see Fig. 7), thus mitigating overfitting due to volatile load data, as stated earlier. In contrast, the load forecasted by the ConvLSTM model shows a continuing diurnal variation for the full period of the vacancy, because it overfits to the pre-vacancy period, thus leading to smaller forecasting accuracy.

Furthermore, as can be seen in Fig. 11(a), the pattern of average forecasting accuracy shown in Fig. 9 does not hold for all apartments individually. While the average CV-residual of the ConvLSTM model for all three seasons are lower than the ones of the Persistence Model, the situation is reversed for some apartments (e.g., red circle in Fig. 11(a)). The reason is that the ConvLSTM model sometimes loses robustness leading to overfitting, as was illustrated for one example apartment in Fig. 10 (which shows the observed and forecasted load profiles of the same apartment as the one highlighted by the red circle in Fig. 11(a)). Overall, such possible overfitting is avoided by the CLSAF model which outperforms both the Persistence Model and the ConvLSTM model, not only on average, but for every apartment, floor, building, and season, individually.

3.2.3. Effect of season on forecasting accuracy

In addition to the above-mentioned effects, it can be noticed that, for any particular spatial granularity and model type, forecasting accuracies are considerably affected by the season, with July consistently exhibiting the lowest (i.e., best) average CV-residual, followed by January, and then April.

We again used t-tests (two tailed, unequal variances) to determine whether the observed differences in forecasting accuracy caused by

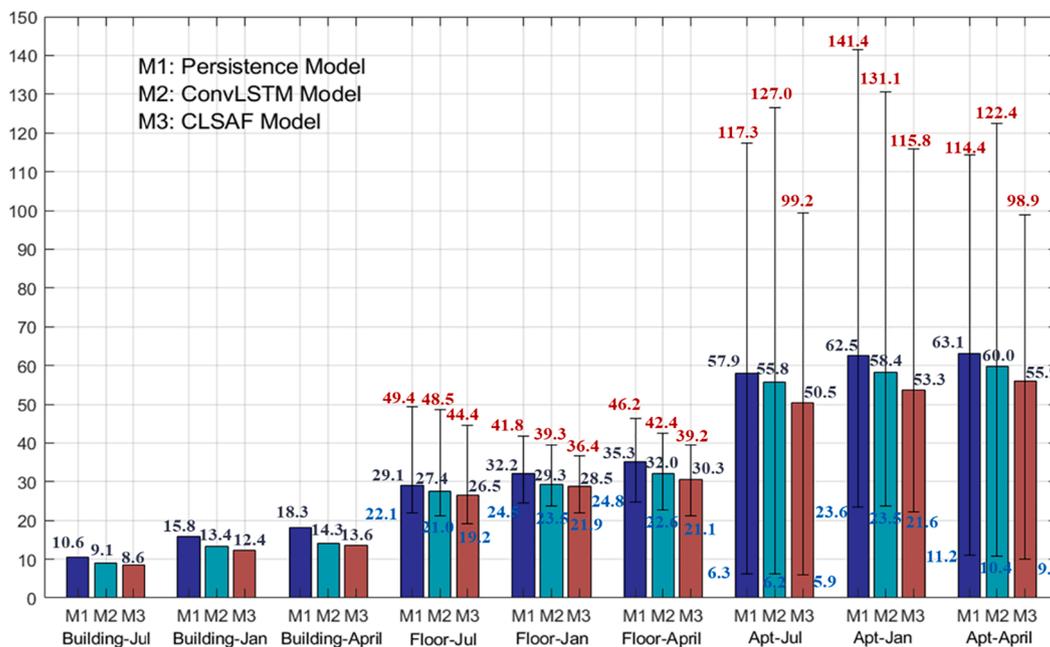


Fig. 9. Forecasting accuracy (CV-residual, in %) of building, floor, and apartment level over the 3 seasons. Error bars indicate the maximum and minimum CV-residuals of each group (red and blue numbers, respectively); black numbers give the averages. The building level has only one forecast accuracy for each model and season.

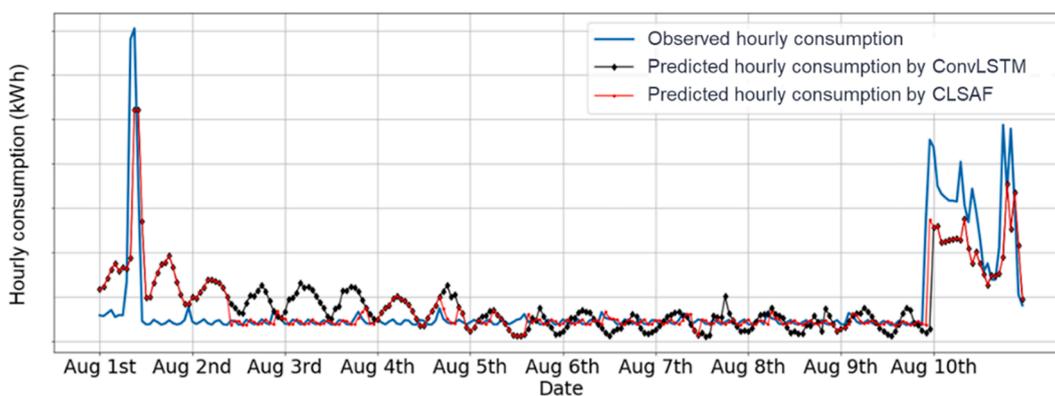


Fig. 10. Hourly forecasting results of the ConvLSTM and the CLSAF models for one example apartment from Aug 1st to Aug 10th, 2019. Exact hourly load values are not shown for privacy considerations.

seasonal changes are statistically significant. Nine pairs were set up for the floor and apartment level (January & April, January & July, and April & July, across three models). The results show that the differences in the average CV-residual of apartments [of floors] between the different seasons were not statistically significant ($p > 0.05$), owing to the large variation in each sample of 59 apartments [11 floors] and the limited sample sizes. Consistent with that, intra-group variance of CV-residual, determined via ANOVA, is substantially larger than inter-group variance, as evidenced by $(1 - \eta^2) = 0.98$ for apartments [0.88 for floors].

Such high level of intragroup variance in CV-residual – which is not explained by the spatial granularity or model type – points to the possible existence of other not yet identified characteristics in each observed load profile. This will be explored in the next sections.

3.3. Volatility of electricity consumption vs. forecasting accuracy

The large variations in CV-residual in Figs. 9 and 11(a) indicate different levels of forecasting- “difficulty” for different apartments and/

or floors. Therefore, we searched for underlying characteristic of the electric load profiles that impacted forecasting accuracy. One such characteristic was found to be the volatility of the load data, henceforth *CV-observation*. The definition of *CV-observation* is as follows:

$$CV_{observation} = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}}{\bar{y}} \tag{11}$$

where y_i denotes the observed hourly load at the i th time step. \bar{y} denotes the mean value of the observed loads, and N is the number of the observations (here 672, for hourly data over 28 days). CV-observation can be understood as a type of normalized standard deviation of electricity load over time. Therefore, a load profile with a larger mean value but similar absolute standard deviation has a smaller CV-observation. Fig. 11 shows the relationship between CV-observation and CV-residual (i.e., forecasting accuracy), along with the respective linear correlations and p-values. July was randomly chosen as the example to visualize the relationship. For January and April, results are similar to

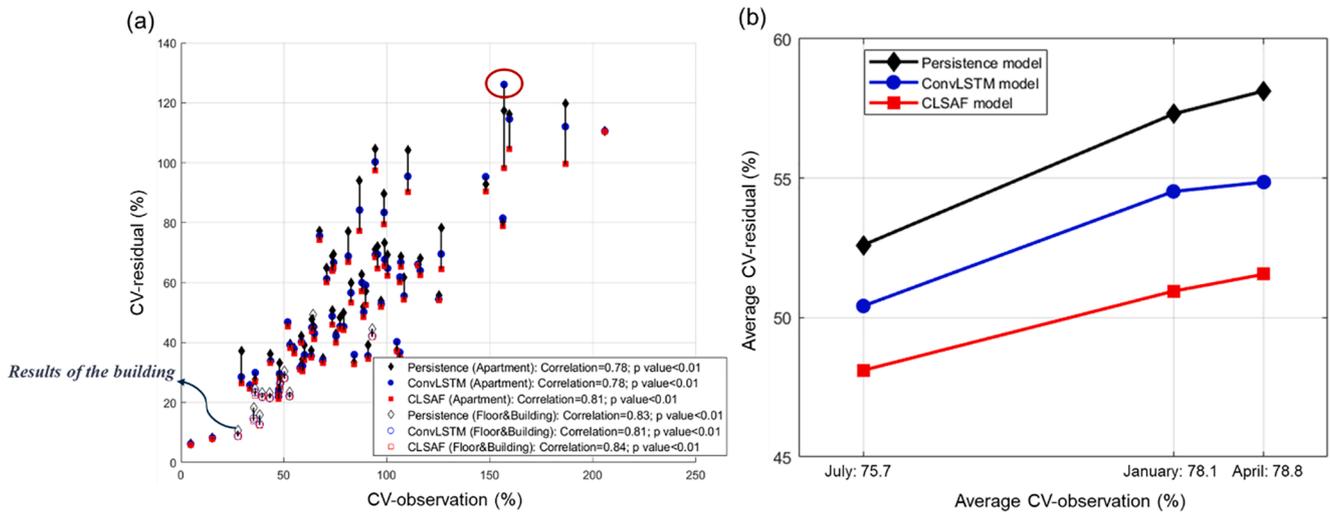


Fig. 11. (a) CV-observation vs. CV-residual, for three models and three spatial granularities (July data). Solid markers show results of the apartments, and open markers those of the floors and building. The result marked by a red circle highlights the example apartment also shown in Fig. 9, for which the ConvLSTM model performs worse than the Persistence Model. (b) Average CV-observation by three seasons vs. respective average CV-residual. Results of each dataset include all three spatial granularities (i.e., combined average of the CV-residuals of 59 apartments, 11 floors, and 1 building).

those in July, namely all correlations were between 0.62 and 0.69, and all p-values were smaller than $5e-6$. This shows that, regardless of model type, season, or spatial granularity, the achieved forecasting accuracy is driven to a considerable extent by CV-observation of the load profile, with forecasting accuracy the higher, the lower CV-observation.

The relationship between average CV-observation and average CV-residual across the three seasons is shown in Fig. 11(b). Observe that July data yields the lowest averages of the two metrics, and April data the highest. Since this is consistent with the pattern in Fig. 9, this provides a likely explanation for why the three seasons exhibit different average CV-residuals: As the average CV-observation increases from July to January and April, the average achievable forecasting accuracy decreases accordingly. In other words, the seasonal effect on forecasting accuracy seen in Fig. 9 is at least partially explained by a concurrent difference in CV-observation between the seasons.

3.4. Strength of diurnal patterns vs. forecast improvement

Examining Fig. 11(a) shows another effect which does not seem to be easily explained by CV-observation: The improvement in forecasting accuracy from the Persistence Model (benchmark) to the CLSAF model varies between apartments (as well as between floors), even if they have (nearly) the same CV-observation. This led us to search for an additional characteristic of the load profiles that affected this accuracy improvement. As illustrated in Section 2.3, the key difference between the Persistence Model and the CLSAF model is that the latter employs a feature-selection technique that can extract the core information of daily load profiles. Therefore, the difference in accuracy between these two models is likely mainly due to how much of such daily-profile information is present in a particular profile.

In order to test this hypothesis, we defined a new variable S to quantify the strength of diurnal patterns of the load profile, as defined in Methods. To illustrate graphically which load characteristic S is sensitive to, Fig. 12 shows the spectral analysis (see Section 2.4) of two load

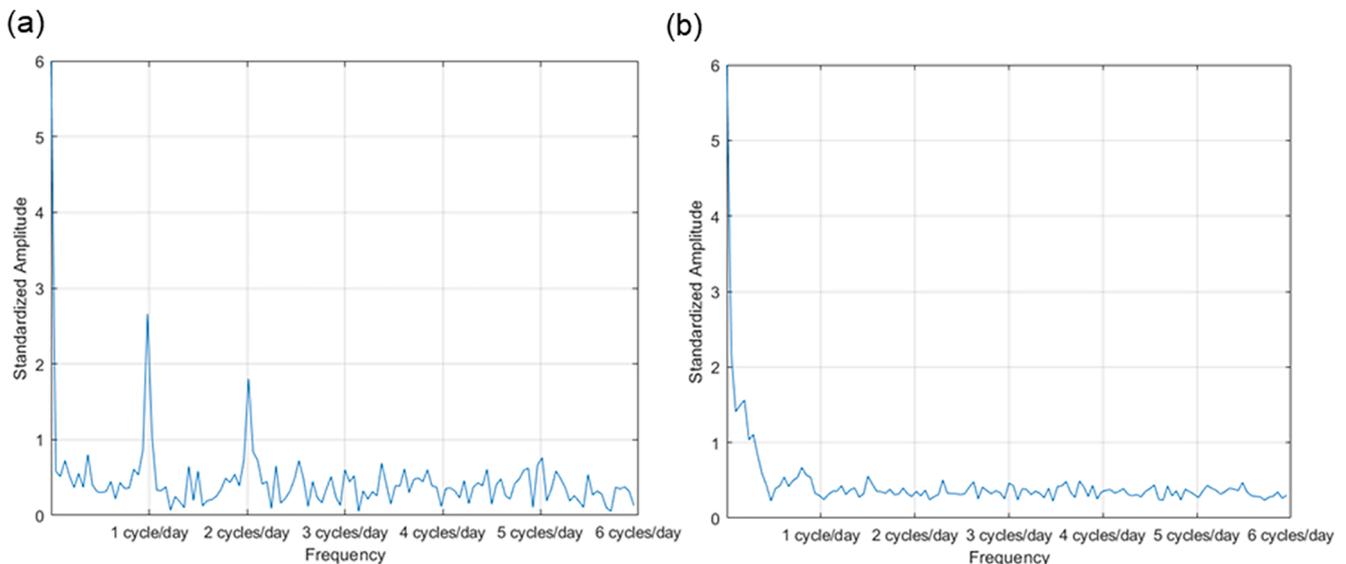


Fig. 12. Spectral analysis of two sample apartments by FFT with the standardized amplitude as defined in Eq. (9). The spectrum in (a) reflects a strong diurnal pattern, evidenced in the spikes at 1 and 2 cycles per day, respectively. The spectrum in (b) reflects few to none diurnal electricity patterns.

profiles, one with strong diurnal periodicity at the 12 h and 24 h mark, and one without.

The improvement in forecasting accuracy, i.e., reduction in CV-residual, R was defined as follows:

$$R = \frac{CV_{residual(persistence)} - CV_{residual(CLSAF)}}{CV_{residual(persistence)}} \quad (12)$$

where CV-residual is as defined in Eq. (1). Fig. 13 shows the relationship between the strength of the diurnal pattern (S) and the forecasting accuracy improvement (R) for all spatial granularities and seasons.

The results demonstrate that the strength of the diurnal pattern has a statistically significant impact on the forecasting accuracy improvement, with an improvement of up to 25% in some cases. The apartment-level has the smallest average improvement ($R = 11\%$), followed by floors ($R = 14\%$), and building ($R = 23\%$). The result underlines that, as outlined in Section 3.2, it is inherently more difficult to predict electricity load profiles whose diurnal profiles are either not present at all or masked by high volatility. In contrast, a stronger diurnal pattern, which tends to be more pronounced in the aggregated loads of an entire floor or building, facilitates the information extraction and learning process executed by more complex models such as the CLSAF model, thus resulting in larger forecasting accuracy improvement for such models vs. the benchmark Persistence Model.

3.5. Combination of CV-observation & diurnal pattern strength vs. CV-residual

Next, we sought to understand to what extent the above two underlying characteristics (CV-observation and strength of diurnal patterns) in combination can explain the achieved forecasting accuracy. This is shown in Fig. 14, which divides the parameter space of CV-observation and S into four areas representing four load profile categories, using the averages of CV-observation and S as the area separation points. We classified all 213 CV-residual obtained by the CLSAF model (59 apartments, 11 floors, and 1 building; each for 3 seasons) into the four categories according to their corresponding CV-observation and S .

We found that electricity load profiles with high S and low CV-observation yield the highest average forecasting accuracy (i.e., lowest CV-residual). The opposite is true for load profiles with low S and high

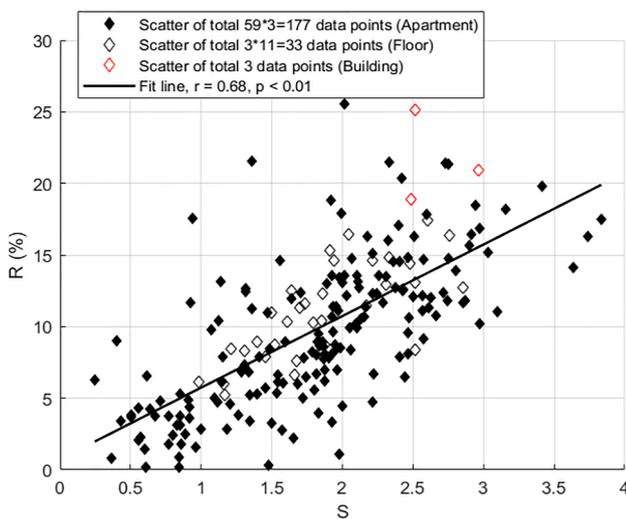


Fig. 13. Strength of diurnal pattern (S) vs. relative reduction of CV-residual (R) from Persistence Model to CLSAF model, covering the results of 3 spatial granularities and all 3 seasons. The substantial correlation of 0.68 ($p < 0.01$) shows that there is a statistically significant effect between the forecasting accuracy improvement (that the CLSAF model can achieve over the Persistence Model) and the diurnal pattern strength of the observed load profile.

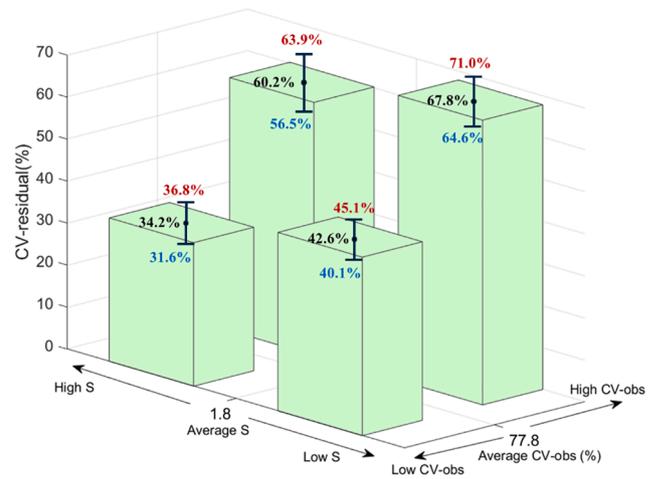


Fig. 14. CV-observation vs. strength of diurnal pattern (S) and average CV-residual for 4 load profile categories. Black numbers of each bar denote average CV-residual of corresponding category. Red numbers are equal to average plus corresponding Standard Error of the Mean (SEM) and blue ones are equal to the average minus corresponding SEM.

CV-observation. Furthermore, the effect of CV-observation on forecasting accuracy is stronger than that of S , as seen by changes of 26 percentage points in CV-residual along the CV-observation dimension but only 8 percentage points in the S dimension. To test for statistical significance of these effects, we carried out t-tests (two-tailed, unequal variances). We found statistical significance ($p < 0.01$) for 5 of the 6 pair-wise differences, and moderate statistical significance ($p = 0.06$) for one pair-wise difference (Table 4). A possible real-life application of these findings is explained in Discussion.

4. Discussion

Our results could serve as a starting point to set up a possible data screening method for time-series electricity-load datasets. The method would allow users of load forecasting models to make a preliminary assessment of the nature of a load profile dataset, providing two benefits: (i) reducing the modeling complexity for some apartments; and (ii) providing confidence levels for the predicted electricity use.

With regards to the first benefit, possible implementation steps would be as follows: First, one could use the previous 28-day electricity data of an apartment intended for forecasting to compute CV-observation. Using the relationship illustrated in Fig. 11, this would provide an approximation for the forecasting accuracy likely achievable by even a simple Persistence Model. Second, one could use the spectral analysis, described in Section 2.4, to determine the load profile's

Table 4
Results of t-tests (two-tailed, unequal variance) to determine statistical significance of the differences in average CV-residual per load profile category.

p values for corresponding t-test	High CV-observation & high S	High CV-observation & low S	Low CV-observation & high S	Low CV-observation & low S
High CV-observation & high S	N/A	$p = 0.06$	$p = 9e-12$	$p = 3e-21$
High CV-observation & low S		N/A	$p = 1e-5$	$p = 5e-10$
Low CV-observation & high S			N/A	$p = 2e-3$
Low CV-observation & low S				N/A

strength of diurnal pattern S , again using the previous 28-day data. Using the relationship illustrated in Fig. 14, CV-observation and S together would then provide an estimate of the forecasting accuracy of the CLSAF model. This information could be used to decide whether the accuracy likely available from a simple Persistence Model is sufficient or the higher computational cost of the CLSAF model warranted. Alternatively, one may find that the load profile of the particular apartment is such that none of the three models would likely provide sufficient accuracy for the specific application in question.

As for the second benefit, knowing not only the forecasted electricity use, for example for the next hour, but also the confidence levels of the prediction (inferred from CV-residual) would allow more sophisticated transaction schemes within the examples of TEN applications outlined in *Introduction*, as follows: Any such trading of electricity with others would carry risks – namely the risk of either not having enough electricity for one's own use or, alternatively, not being able to honor the transaction agreed to with another user. However, the ability to evaluate how accurate the forecast will likely be, makes these risks more manageable. For example, user A may be able to determine that despite having committed to selling a certain number of kWh from their own storage to user B, user A can still be 90% confident to have enough electricity for themselves. Alternatively, the transaction could be priced such that user B knows that there is a 10% risk that user A will not be able to provide the full amount of electricity that was agreed on.

5. Conclusions

A Convolutional Long Short-Term Memory neural network model with selected autoregressive features (CLSAF model) is found to improve single-step-ahead electricity load forecasting. This improvement is observed at all three spatial granularities: apartment, floor, and building level. The CLSAF model achieves higher forecasting accuracy (up to 25% improvement vs. the Persistence Model). The CLSAF model enables durable robustness by leveraging the advantages of its autocorrelation-based feature-selection algorithm and a model-simplification method to prevent overfitting when confronted with volatile load data caused by changes in resident behavior and/or temporary absences.

Based on the predictions across the three seasons, we identified two load profile characteristics that are statistically significantly correlated with forecasting accuracy, namely CV-observation and the strength of the diurnal pattern S . These characteristics capture the load profile volatility and the degree of learnable daily-profile information, respectively. The smaller the CV-observation and stronger the diurnal pattern, the higher is the forecasting accuracy that the CLSAF model can achieve. In real life applications, these two characteristics would allow a preliminary assessment of the nature of a load profile dataset and estimate, for every load profile individually, the expected forecasting accuracy by model type. Such a data screening technique is suitable for integration into a transactive energy network ecosystem and allows one to (i) reduce computational complexity by choosing the appropriate model and (ii) estimate confidence levels of the load forecasts.

6. Current limitations and possible improvements

In this study, the temporal granularity and forecasting horizon of the proposed CLSAF model were limited to a single time step (1 h resolution). This was primarily due to the volatile nature of the apartment level data which makes forecasting of multiple steps forward more difficult. Therefore, as a possible improvement, firstly one could try to apply random-walk information (such as white noise) to simulate random human behavior. Secondly, building on the forecasting advantages of the CLSAF model, one could expand the TEN application, e.g., by integrating electricity storage and associated dispatch algorithms, resulting in lower electricity trading risks and further environmental and commercial values of these applications.

CRedit authorship contribution statement

Lechen Li: Methodology, Visualization. **Christoph J. Meinrenken:** Supervision, Data curation, Investigation. **Vijay Modi:** Supervision, Formal analysis, Methodology. **Patricia J. Culligan:** Funding acquisition, Project administration.

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.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), under the Building Technologies Office's BENEFIT program, Award Number DE-EE-0007864. We thank Rongye Shi for critically reading earlier versions of the manuscript.

References

- [1] Wang N. Transactive control for connected homes and neighborhoods. *Nat Energy* 2018. <https://doi.org/10.1038/s41560-018-0257-2>.
- [2] Zheng M, Meinrenken CJ, Lackner KS. Smart households: Dispatch strategies and economic analysis of distributed energy storage for residential peak shaving. *Appl Energy* 2015 Jun;1(147):246–57. <https://doi.org/10.1016/j.apenergy.2015.02.039>.
- [3] Song Y, Ding Y, Siano P, Meinrenken C, Zheng M, Strbac G. Optimization methods and advanced applications for smart energy systems considering grid-interactive demand response. <https://doi.org/10.1016/j.apenergy.2019.113994>.
- [4] Gross G, Galiana FD. Short-term load forecasting. *Proc IEEE* 1987;75(12):1558–73. <https://doi.org/10.1109/PROC.1987.13927>.
- [5] Hernandez L, Baladron C, Aguiar JM, Carro B, Sanchez-Esguevillas AJ, Lloret J, et al. A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Commun Surv Tutor* 2014;16(3): 1460–95. <https://doi.org/10.1109/SURV.2014.032014.00094>.
- [6] Javed F, Arshad N, Wallin F, Vassileva I, Dahlquist E. Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short-term multiple loads forecasting. *Appl Energy* 2012;96:150–60. <https://doi.org/10.1016/j.apenergy.2012.02.027>.
- [7] Meinrenken CJ, Mehmani A. Concurrent optimization of thermal and electric storage in commercial buildings to reduce operating cost and demand peaks under time-of-use tariffs. *Appl Energy* 2019;15(254):113630. <https://doi.org/10.1016/j.apenergy.2019.113630>.
- [8] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. *Renew Sustain Energy Rev* 2018;81:1192–205. <https://doi.org/10.1016/j.rser.2017.04.095>.
- [9] Ghofrani M, Hassanzadeh M, Etezadi-Amoli M, Fadali MS. Smart meter based short-term load forecasting for residential customers. In 2011 North American Power Symposium 2011 Aug 4. IEEE, p. 1–5. <https://doi.org/10.1109/NAPS.2011.6025124>.
- [10] Munkhammar J, van der Meer D, Widén J. Very short term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model. *Appl Energy* 2021;282:116180. <https://doi.org/10.1016/j.apenergy.2020.116180>.
- [11] Iwafune Y, Yagita Y, Ikegami T, Ogimoto K. Short-term forecasting of residential building load for distributed energy management. In 2014 IEEE International Energy Conference (ENERGYCON) 2014 May 13. IEEE, p. 1197–1204. <https://doi.org/10.1109/ENERGYCON.2014.6850575>.
- [12] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: A machine learning case study. *Energy Build* 2012;1(49):591–603. <https://doi.org/10.1016/j.enbuild.2012.03.010>.
- [13] Hendron R, Engebrecht C. Building America house simulation protocols (No. DOE/GO-102010-3141). Washington, DC (United States): Office of Energy Efficiency and Renewable Energy (EERE); 2010. <https://pdfs.semanticscholar.org/5a1a/4e1f491ea597063451ca6f3ad7286aaf08b.pdf>.
- [14] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl Energy* 2014;15(127):1. <https://doi.org/10.1016/j.apenergy.2014.04.016>.
- [15] Zhang J, Wei YM, Li D, Tan Z, Zhou J. Short term electricity load forecasting using a hybrid model. *Energy*. 2018;1(158):774–81. <https://doi.org/10.1016/j.energy.2018.06.012>.
- [16] Zheng J, Xu C, Zhang Z, Li X. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In 2017 51st Annual Conference on Information Sciences and Systems (CISS) 2017 Mar 22. IEEE, pp. 1–6. <https://doi.org/10.1109/CISS.2017.7926112>.

- [17] Marino, D. L., Amarasinghe, K., & Manic, M. (2016, October). Building energy load forecasting using deep neural networks. In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. IEEE, p. 7046–51. <https://doi.org/10.1109/IECON.2016.7793413>.
- [18] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. *Sustain Energy Grids Netw* 2016;1(6):91–9. <https://doi.org/10.1016/j.segan.2016.02.005>.
- [19] Andriopoulos N, Magklaras A, Birbas A, Papalexopoulos A, Valouxis C, Daskalaki S, et al. Short Term Electric Load Forecasting Based on Data Transformation and Statistical Machine Learning. *Appl Sci* 2021;11(1):158. <https://doi.org/10.3390/app11010158>.
- [20] Lusi P, Khalilpour KR, Andrew L, Liebman A. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Appl Energy* 2017; 1(205):654–69. <https://doi.org/10.1016/j.apenergy.2017.07.114>.
- [21] 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. <https://doi.org/10.1109/TSG.2017.2753802>.
- [22] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl Energy* 2014;15(123):168–78. <https://doi.org/10.1016/j.apenergy.2014.02.057>.
- [23] Zheng Z, Chen H, Luo X. A Kalman filter-based bottom-up approach for household short-term load forecast. *Appl Energy* 2019;15(250):882–94. <https://doi.org/10.1016/j.apenergy.2019.05.102>.
- [24] Xu L, Wang S, Tang R. Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Appl Energy* 2019;1(237):180–95. <https://doi.org/10.1016/j.apenergy.2019.01.022>.
- [25] Chen H, Wang S, Wang S, Li Y. Day-ahead aggregated load forecasting based on two-terminal sparse coding and deep neural network fusion. *Electr Power Syst Res* 2019;1(177):105987. <https://doi.org/10.1016/j.epr.2019.105987>.
- [26] Jain RK, Damoulas T, Kontokosta CE. Towards data-driven energy consumption forecasting of multi-family residential buildings: feature selection via the lasso. In *Computing in Civil and Building Engineering* (2014); 2014, p. 1675–82. <https://ascelibrary.org/doi/abs/10.1061/9780784413616.208>.
- [27] Yu CN, Mirowski P, Ho TK. A sparse coding approach to household electricity demand forecasting in smart grids. *IEEE Trans Smart Grid* 2016;8(2):738–48. <https://doi.org/10.1109/TSG.2015.2513900>.
- [28] Kaneko N, Fujimoto Y, Kabe S, Hayashida M, Hayashi Y. Sparse modeling approach for identifying the dominant factors affecting situation-dependent hourly electricity demand. *Appl Energy* 2020;1(265):114752. <https://doi.org/10.1016/j.apenergy.2020.114752>.
- [29] Candanedo LM, Feldheim V, Deramaix D. Data driven prediction models of energy use of appliances in a low-energy house. *Energy Build* 2017;1(140):81–97. <https://doi.org/10.1016/j.enbuild.2017.01.083>.
- [30] Kwac J, Flora J, Rajagopal R. Household energy consumption segmentation using hourly data. *IEEE Trans Smart Grid* 2014;5(1):420–30. <https://doi.org/10.1109/TSG.2013.2278477>.
- [31] Benítez I, Quijano A, Díez JL, Delgado I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int J Electr Power Energy Syst* 2014;1(55):437–48. <https://doi.org/10.1016/j.ijepes.2013.09.022>.
- [32] Chaouch M. Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *IEEE Trans Smart Grid* 2013;5(1):411–9. <https://doi.org/10.1109/TSG.2013.2277171>.
- [33] Hernández L, Baladrón C, Aguiar JM, Calavia L, Carro B, Sánchez-Esguevillas A, et al. Experimental analysis of the input variables' relevance to forecast next day's aggregated electric demand using neural networks. *Energies*. 2013;6(6):2927–48. <https://doi.org/10.3390/en6062927>.
- [34] Amarasinghe K, Marino DL, Manic M. Deep neural networks for energy load forecasting. In 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE); 2017 Jun 19. IEEE, pp. 1483–8. <https://doi.org/10.1109/ISIE.2017.8001465>.
- [35] Wang S, Wang X, Wang S, Wang D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *Int J Electr Power Energy Syst* 2019;1(109):470–9. <https://doi.org/10.1016/j.ijepes.2019.02.022>.
- [36] Wilms H, Cupelli M, Monti A. Combining auto-regression with exogenous variables in sequence-to-sequence recurrent neural networks for short-term load forecasting. In 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); 2018 Jul 18. IEEE, p. 673–9. <https://ieeexplore.ieee.org/document/8471953>.
- [37] Wan R, Mei S, Wang J, Liu M, Yang F. Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics*. 2019;8(8):876. <https://doi.org/10.3390/electronics8080876>.
- [38] Xingjian SH, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*; 2015, p. 802–10.
- [39] International Code Council, Building Officials, Code Administrators International, International Conference of Building Officials, & Southern Building Code Congress International. *International energy conservation code*. International Code Council; 2000. <https://www.iccsafe.org/>.
- [40] Siemens. SEM3™ - Embedded Micro Metering Module - Selection & Application Guide. Siemens Industry, Inc; 2019. <https://usa.siemens.com/sem3/> (accessed on 25 May 2020).
- [41] Hyndman RJ. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: Int J Appl Forecast* 2006;4(4):43–6.
- [42] National Oceanic and Atmospheric Association. Details on Measured Parameters. <https://www.ncdc.noaa.gov/cm/measurements.html/> (accessed on 10 January 2021).
- [43] National Oceanic and Atmospheric Administration; 2019. <https://www.ncdc.noaa.gov/cdo-web/> (accessed on 15 August 2019).
- [44] Lü X, Lu T, Kibert CJ, Viljanen M. A novel dynamic modeling approach for predicting building energy performance. *Appl Energy* 2014;1(114):91–103. <https://doi.org/10.1016/j.apenergy.2013.08.093>.
- [45] Hewamalage H, Bergmeir C, Bandara K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int J Forecast* 2020. <https://doi.org/10.1016/j.ijforecast.2020.06.008>.
- [46] Khandakar Y, Hyndman RJ. Automatic time series forecasting: the forecast package for rj stat. Soft.
- [47] Khan IA, Akber A, Xu Y. Sliding window regression based short-term load forecasting of a multi-area power system. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE); 2019 May 5. IEEE, p. 1–5. <https://doi.org/10.1109/CCECE.2019.8861915>.
- [48] VanderPlas JT. Understanding the lomb–scargle periodogram. *Astrophys J Supp Ser* 2018 May 11;236(1):16. <https://iopscience.iop.org/article/10.3847/1538-4365/aab766/pdf>.
- [49] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8): 1735–80. <https://doi.org/10.1162/neco.1997.9.8.173>.
- [50] Sood R, Koprinska I, Agelidis VG. Electricity load forecasting based on autocorrelation analysis. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* 2010 Jul 18. IEEE, pp. 1–8. <https://doi.org/10.1109/IJCNN.2010.5596877>.
- [51] Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference* 2010 Jun 28 (Vol. 57, p. 61). Scipy.
- [52] Spyers-Ashby JM, Bain PG, Roberts SJ. A comparison of fast Fourier transform (FFT) and autoregressive (AR) spectral estimation techniques for the analysis of tremor data. *J Neurosci Meth* 1998;83(1):35–43. [https://doi.org/10.1016/S0165-0270\(98\)00064-8](https://doi.org/10.1016/S0165-0270(98)00064-8).
- [53] Li C, Tao Y, Ao W, Yang S, Bai Y. Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. *Energy* 2018;15(165):1220–7. <https://doi.org/10.1016/j.energy.2018.10.113>.