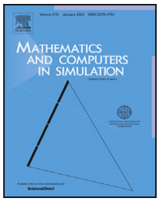
Mathematics and Computers in Simulation 240 (2026) 303–316

Contents lists available at ScienceDirect

Mathematics and Computers in Simulation

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



Original articles

# Comparing peak electricity load forecasting models for an industrial and a residential building

Michael Wood, Silvana Matrone *, Emanuele Ogliari, Sonia Leva

*Politecnico di Milano, Department of Energy, Via La Masa 24, Milano, 20156, Italy*

ARTICLE INFO

*Keywords:*
Peak load forecasting
Time series analysis
SARIMAX
LSTM

ABSTRACT

Electrical load forecasting models are becoming more and more essential for energy savings and effective energy management and the ability to forecast the load peak is a crucial feature for many applications. This paper presents a comparative analysis of various methods with low computational complexity, including Naïve Persistence, Statistical load forecasting, Seasonal Auto-Regressive Integrated Moving Average with eXogenous regressors, and Long Short-Term Memory enhanced by Empirical Mode Decomposition pre-processing. They are tested to forecast daily peak electricity load and peak hour in two distinct existing scenarios: residential and industrial.

The study investigates traditional statistical models and artificial intelligence techniques to determine the most effective method to obtain accurate load prediction. Through performance evaluation and data analysis, insights are provided into their applicability and effectiveness in peak load forecasting.

## 1. Introduction

As the world faces the growing challenges of climate change and its far-reaching impacts, the importance of load forecasting has become more pronounced [1,2]. The increasing frequency of extreme weather events has left nations in a vulnerable state, where the interplay between global warming and rising energy demands intensifies the situation. Variations in temperature lead to greater dependence on residential cooling and heating systems, while the expansion of global industries and population growth further drive up energy consumption [3].

Renewable energy's variability complicates load forecasting. Accurate peak demand prediction enables efficient resource allocation, minimizes outages, and optimizes energy distribution. Peak load forecasting is critical for ensuring the reliability, efficiency, and economic operation of power systems. Accurate predictions of when the peak will occur and how big the peak will be enable optimizing generation scheduling, reducing reserve margins, avoiding penalties under peak pricing schemes, and preventing grid overload. In particular, short-term peak load and peak hour forecasts are essential for demand response programs, real-time energy management, and the integration of distributed energy resources.

In scientific literature, Load Forecast (LF) has been tackled since years, and in the last decade the number of papers covering this issue is gradually increasing. Load forecasting methods are generally categorized based on their time horizon [4,5] and different acronyms could be employed for short [6], medium [7] and other time scales [8] as detailed in Table 1.

Relevant papers in the literature, where diverse methodologies and applications that address the peak LF are reported in Table 2.

---

[*] Corresponding author.
*E-mail address:* silvana.matrone@polimi.it (S. Matrone).

https://doi.org/10.1016/j.matcom.2025.06.029
Received 7 March 2025; Received in revised form 20 June 2025; Accepted 26 June 2025

**Table 1**
LF time horizon definition.

| Acronym | Load forecast | Time horizon description |
|---|---|---|
| VSTLF | Very short-Term | Periods shorter than a day |
| STLF | Short-Term | Spans from a day to a few weeks |
| MTLF | Medium-Term | Extends from a few weeks to months |
| LTLF | Long-Term | Forecasts a year or more |

**Table 2**
Representative scientific papers on peak load and electrical load forecasting.

| Ref. | Horizon | Load | Method | Topic |
|---|---|---|---|---|
| [9] | STLF | Utility | GRU with DTW pre-processing | Deep learning for daily peak load forecasting |
| [10] | LTLF | National | SARIMAX, ANN, SVR, ANN | Comparative study of traditional, machine learning, and hybrid models for peak load forecasting |
| [11] | STLF | Small industrial facilities | Ensemble models with feature selection and modified isolation forest | Peak-load forecasting under limited information environments |
| [33] | VSTLF | EV charging stations | ANN, RNN, GRU, SAE, Bi-ANN | Super-short-term EV charging load forecasting using deep learning |
| [34] | STLF | EV charging stations | GRU, ANN, RNN, ANN | Short-term EV charging station load forecasting with deep learning |
| [35] | STLF | Utility | Regression based with transformation function | Regression-based peak load forecasting with transformation |
| [12] | STLF | Residential | Deep recurrent neural network with ANN | Optimal load dispatch in microgrid using deep learning |
| [13] | STLF | Residential | Decision-support system | Intelligent home energy management for demand response |
| [14] | STLF | Industrial | Linear programming routine | Grid-connected PV-battery storage system optimization |
| [15] | STLF | Utility | ANN | Artificial neural network short-term load forecasting |
| [16] | LTLF | National | ANN | Peak load forecasting in Taiwan with ANN |

Here, [9] proposed a novel approach combining Dynamic Time Warping (DTW) with a Gated Recurrent Units (GRU) to improve daily peak load forecasting accuracy by capturing both trend and local variations in load curves. Meanwhile, [10] conducted a comprehensive comparative study between traditional, machine learning, and hybrid models for national-scale peak load forecasting in Korea, demonstrating that Artificial Neural Networks (ANN) models and hybrid ANN with Seasonal Autoregressive Integrated Moving Average with eXogenous regressors (SARIMAX) models outperform traditional time series models. The authors in [11] developed a Machine Learning (ML) based peak load forecasting method for small industrial facilities, and they addressed the challenges posed by highly irregular and nonstationary load data. Furthermore, a modified isolation forest approach was applied to compensate for prediction uncertainties, substantially improving forecasting accuracy even in the absence of rich feature data.

More recently, [12] explored DL techniques, using a deep recurrent neural network with ANN units for optimal load dispatch in a microgrid community.

Decision-support systems have also been employed, with [13] developing an intelligent home energy management system for Demand Response (DR). Optimization techniques, such as linear programming, were used by [14] to optimize grid-connected PV-battery storage systems. Further advancements in ANN were showcased by [15], who introduced an adaptive combination of forecasts. Finally, [16] applied ANN models for regional load forecasting in Taiwan, demonstrating the broader applicability of these methods. Electrical load forecasting uses traditional statistical models (ARIMA, regression, Kalman filtering) or Artificial Intelligence (AI) models (ANN, SVM, DL) [17]. Data-driven methods, particularly DL [18], are now dominant due to increased data availability [19,20]. Peak forecasting uses time-series (Holt-Winters, SARIMAX) for simplicity, or ML (ANN, NARX) for complex data [21–24], including hybrid ML [25]. ANN (FFNN, ESN) improves DR accuracy in smart grids, used in commercial buildings with HVAC and clustering [26]. DR programs optimize peak load, aiding grid and building operators [27]. Building type impacts load: residential (diurnal, weather-driven), commercial (weekday peaks), industrial (complex) [28–30]. ANN or hybrid models excel for national peak forecasts [25], but ANN needs extensive data, limiting building-scale use [31]. A support vector machine with parameters optimized by particle swarm optimization outperformed other methods in long-term national forecasting [32]. Generally there is a gap in the literature regarding peak load forecasting, especially short-term and at building scales.

In this work, we compare four forecasting models of increasing complexity. The aim is to check whether more complex methods improve peak load and hourly forecasting, and to try balancing simplicity, interpretability, and accuracy. The selected models are a Naïve Persistence (NP) model, a statistical model, SARIMAX, and a Long Short-Term Memory Network (LSTM) neural network enhanced by Empirical Mode Decomposition (EMD) for better feature extraction. They were chosen for their lightweight computational requirements, making them ideal for deployment on small industrial PCs and even edge devices like sensors [36,37]. These models balance efficiency and performance, ensuring real-time processing without the need for heavy computational resources.

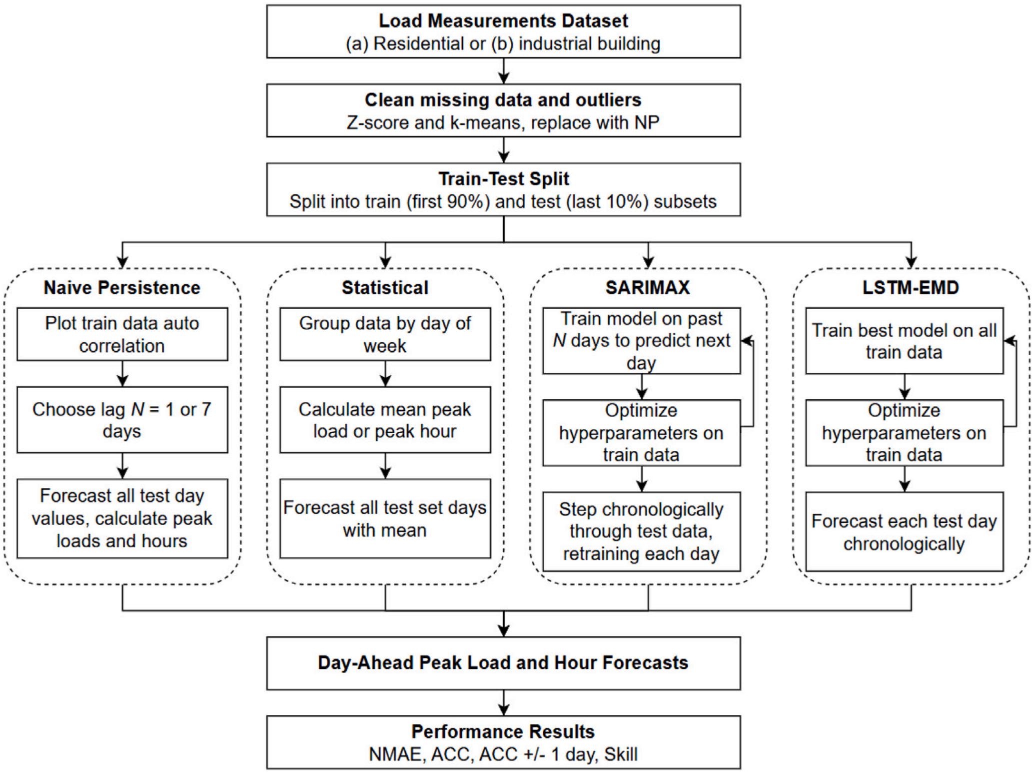


**Fig. 1.** Flow diagram of methodology used for peak load forecasting.

Their streamlined architecture allows for easy implementation in industrial environments, enabling seamless integration into existing systems with minimal power consumption and hardware constraints. This makes them a practical choice for applications where space, energy, and processing capabilities are limited. While LSTM-based models have been explored in the literature for load forecasting, their application to peak load forecasting is limited. For instance, in [38] the authors introduce a deep Bi-LSTM sequence-to-sequence model for day-ahead peak demand forecasting, focusing on residential settings and specific events such as holidays. However, their work did not include a comprehensive comparison with simpler or classical forecasting methods. Similarly, the work by Rubasinghe et al. [39] implemented a CNN-LSTM model for long-term monthly peak load forecasting, targeting multi-year horizons suitable for strategic grid planning, rather than short-term operational needs. Both studies highlight the potential of LSTM-based architectures but do not address short-horizon hourly or daily peak time predictions. Additionally, there are very few studies targeting peak hour forecasting [40], a task that is crucial for energy optimization and peak shaving. Our work provides a systematic comparison of forecasting models for both real-time peak load and peak hour prediction. This comparison provides guidance on selecting interpretable and computationally efficient models. Results are also analyzed using tools such as: statistical analysis, time series decomposition, and the Augmented Dicky Fuller (ADF) stationarity test. This is to assess whether advanced models significantly improve forecasting performance or if well-designed simple statistical methods, tuned on the basis of a robust data analysis, are still sufficient.

The paper is structured as follows: Section 2 describes the employed forecasting models. Section 3 investigates residential and industrial case study scenarios, and Sections 4 and 5 present the comparative analysis results and conclusions, respectively.

## 2. Peak load forecasting models

The flow diagram of Fig. 1 summarizes the methodology used to identify the models parameters and forecast peak load. To perform peak load and peak hour forecasting for both residential and industrial scenarios, we employ four models with an increasing complexity: Naïve Persistence (NP), Statistical Model, SARIMAX, and LSTM-EMD for day-ahead (that is in STLF) forecasting, comparing their performance.

The Naïve Persistence model serves as the simplest benchmark in forecasting. This method, while basic, provides a reference point against which more complex models can be compared. It helps determine whether more advanced models truly improve forecasting accuracy or if a simple assumption suffices.

We introduce a Statistical Model, which offers a more refined benchmark by incorporating seasonal averages of historical data. As a more traditional time series approach, we employ SARIMAX, a widely used regression forecasting model. SARIMAX

**Table 3**
Characteristics of the datasets after cleaning of bad data and outliers.

| Characteristic | Residential building | Industrial building |
|---|---|---|
| Units | kW (p.u.) | p.u. |
| Maximum | 4.45 (1.0) | 0.76 |
| Mean | 0.69 (0.15) | 0.55 |
| Standard Deviation | 0.94 (0.14) | 0.11 |
| Autocorrelation: 1 day/1 week 8 | 0.49/0.40 | 0.43/0.8 |
| ADF: Statistic/1% critical value 4 | −10.2/−3.4 | −19.5/−3. |
| Length | 4 years | 3 years |
| Interval | 1 h | 1 h |
| Location | USA | Italy |

builds upon ARIMA, a well-established method that combines autoregression, differencing, and moving averages to model complex dependencies in time series data. Although SARIMAX is more advanced than simple heuristic models, it does not incorporate ML techniques. Finally, we develop an LSTM, an ML-based method designed to model nonlinear and complex relationships within the data, with an additional pre-processing step using EMD, developed in [18]. Unlike SARIMAX, which relies on predefined statistical assumptions, LSTMs learn directly from the data through interconnected layers of artificial neurons in a recurrent structure. By fine-tuning hyperparameters such as the number of neurons, layers, and length of input sequence, the LSTM-EMD can provide superior forecasting accuracy, particularly in scenarios where relationships between variables are not strictly linear or stationary.

This research centers on two separate electric load datasets, each from a different scenario: a residential and an industrial building. The case study and the dataset are described in Section 3 The data, which are measurements of load real power averaged over 1-h intervals, underwent cleaning of missing indices and statistical outliers that could affect the accuracy of the forecast. This step ensured that the models were trained on relevant and high-quality data, minimizing the risk of skewed predictions due to anomalies or inconsistencies. Missing indices or data are replaced with the value according to NP (see Section 2.1) with the lag equal to one day. Statistical outliers are detected by a hybrid z-score and k-mean clustering method developed in [18]. Details about the datasets are shown in Table 3.

Next, each dataset was split into train (first 90%) and test (last 10%) sub-sets to ensure effective model training and generalization. The training set was used to develop the models, while the testing set assessed their predictive performance on unseen data. Finally, the performances of the different models were compared to determine their effectiveness in forecasting peak load and peak hour demand.

### 2.1. Naïve Persistence model

The NP model is the simplest forecasting model to implement and is typically used as a benchmark. It operates on the principle that the forecasted value for a given day ($\hat{y}_t$) is estimated by the value from some previous measured value, often that of $N$ days previous ($y_{t-N}$) as shown in Eq. (1).

$$\hat{y}_t = y_{t-N} \tag{1}$$

The only model parameter $N$ is chosen to be either 1 or 7 days, depending on which is associated with a higher autocorrelation value. This resulted in an $N$ of 1 for the residential scenario and 7 for the industrial scenario. The differing $N$-values is due to the different seasonalities present in each dataset. The industrial weekly seasonality is clearly evident in the autocorrelation plot in Fig. 2 which shows strong peaks at hourly lags corresponding to the same day of the week, whereas the residential load mostly exhibits a daily seasonality.

### 2.2. Statistical model

The statistical model used in this study operates by determining the most likely value for each forecast day through a simple average that incorporates daily and weekly seasonality, as shown in Eq. (2). While the natural logarithm transformation is commonly applied in statistical load forecasting to stabilize variance and address potential skewness, we opted to model the load in its original scale. This decision was made to preserve the original scale of the load, ensuring more interpretable results, which is important for practical deployment in operational settings.

This statistical model is more complex than the NP model and is based on the assumption that the peak load and hour of day tends to follow a random walk over months and years.

$$\begin{cases} \hat{y}_{L,d} = \dfrac{1}{N_w} \displaystyle\sum_{w=0}^{N_w-1} y_{L,d,w} \\ \hat{y}_{H,d} = \dfrac{1}{N_w} \displaystyle\sum_{w=0}^{N_w-1} y_{H,d,w} \end{cases} \tag{2}$$

where:

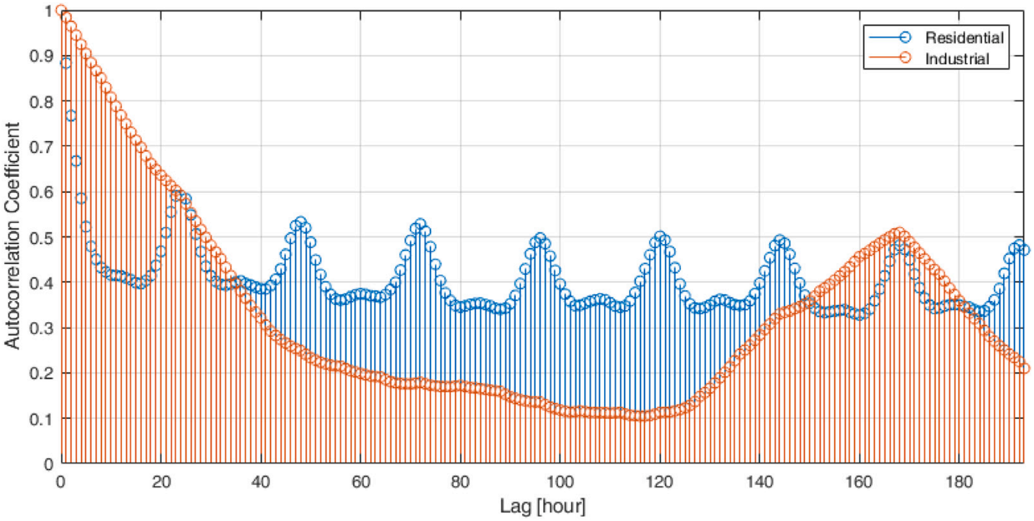


**Fig. 2.** Autocorrelation coefficient for the entire datasets of the load scenarios.

- $y_L, \hat{y}_L \in \mathbb{R}_{\geq 0}$ are the measured and estimated daily peak loads
- $y_H, \hat{y}_H \in \{0, 1, 2, \dots, 23\} \subset \mathbb{N}$ are the measured and estimated daily peak hours
- $d \in \{0, 1, 2, \dots, 6\} \subset \mathbb{N}$ is the day of the week beginning on Monday
- $w \in \{0, 1, 2, \dots, N\} \subset \mathbb{N}$ is the number of the week of the total data period

### 2.3. SARIMAX model

The first time series model examined in this paper is SARIMAX [41], a forecasting method closely related to ARIMA. ARIMA combines autoregression and moving averages with a differenced time series. Specifically, ARIMA is defined by three parameters:

- $p$ = autoregressive order, the number of lagged measurements of the differenced series;
- $d$ = differencing degree, the number of times the data is differenced to achieve stationarity;
- $q$ = moving average order, the number of lagged forecast errors.

SARIMAX builds upon the ARIMA model by adding a seasonal component to account for patterns that recur at regular time intervals. It introduces four additional parameters, known as the seasonal parameters $P$, $D$, and $Q$, which function similar to the $p$, $d$, and $q$ from ARIMA but apply only to the seasonal component of the data. Finally a lag value $s$ indicates the seasonal period of the data, which is 24 for the residential scenario 168 for the industrial one.

The complete set of parameters has been optimized separately for both peak load and peak hour in each case. The input is a sequence of length $L$ of past measured values (peak loads or hours) to predict the next day's peak value or hour. Training is performed using the modified Powell's method to minimize mean squared error. The SARIMAX forecaster is re-trained every day since training is computationally inexpensive and very effective in practice. Parameter tuning is performed by random search and only on the train data set so as to not overfit the test set. For testing on the test data set the parameters are fixed.

### 2.4. Long short-term memory with EMD model

The proposed novel method is based on the LSTM, which simulates the structure of human neurons and their many interconnections (synapses). These artificial neurons receive input data ($x$), process it by multiplying with a specific weight ($W$) and adding a bias ($b$), which are all unique to each neuron. The neuron's output is then passed through an activation function ($\sigma$), which introduces non-linearity to the neural response (as shown in Eq. (3)). The two primary activation functions used are the Rectified Linear Unit (ReLU) and the sigmoid function, which helps minimize backpropagation errors. The LSTM, specifically, is based on the recurrent neuron, which recursively processes each element of the input sequence in addition to the output of the same neuron. This provides a kind of memory effect that is useful in processing data time series. The LSTM algorithm adds additional so-called gates to selectively remember and forget data, as well as an additional output called the cell state. These functions reduce this short-term memory and increase trainability by reducing the exploding gradient problem. The outputs from the first layer of neurons can be passed to subsequent LSTM or normal Multi-Layers Perceptron layers as inputs, where this second layer of neurons will generate its own outputs. This process can be repeated multiple times depending on the complexity of the data and the level of precision required in the forecast, creating deep neural networks.

$$y = \sigma(W \cdot x + b) \tag{3}$$

The novel LSTM peak forecasting model proposed in this work includes multiple LSTM and ANN layers, with a pre-processing step of EMD for enhanced feature extraction. It takes an input sequence of past hourly measurements and outputs a single peak load or hour value for the next day. It is trained on the data set, minimizing mean squared error via the Adam optimizer and learning rate of 0.01, usually for several to dozens of epochs, or repetitions through the entire train data set. Batch size was limited to 32. In addition certain hyper-parameters of the LSTM model must be tuned: the length of each input sequence, the number, type, and dimensions of each layer, and any other inputs given to the model besides the historical measurements, which are called features. To determine the optimal set of hyper-parameters, a two-stage random search was conducted. In the first stage a very large search space is included but each model only trains for until the validation set, which is a random 20% of the train set, loss increases in any single epoch. In the second stage the search space is limited to the top 100 best performing models from the first stage, but they are trained until the validation stops decreasing for 25 consecutive epochs. The final optimized hyper-parameters are given in Table 6, any other hyper-parameters were fixed as described in this paragraph or were the default values in Keras version 2.10.0.

### 2.5. Evaluation framework

To assess which model better predicts peak load and peak hour, three metrics were considered:

- NMAE (Normalized Mean Absolute Error) is defined here as the Mean Absolute Error (MAE) divided by the maximum value of the dataset ($Y_{max}$), where MAE is the mean of the absolute value of the differences between the predicted values ($\hat{Y}_i$) and actual values ($Y_i$) and where $n$ represents the total number of values as shown in the following Eq. (4):

$$NMAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_{max}}\right| \tag{4}$$

- Accuracy (abbreviated sometimes as ACC) is the ratio of the total correct predictions to the total number of predictions, and can be expressed as a value from 0.0 to 1.0 or as a percentage as shown in the following Eq. (5):

$$\text{ACC} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{5}$$

- Skill (or often Forecast Skill) is a performance factor which typically ranges from 0 (the forecast is equal to the chosen benchmark, persistence in this work) to 1 (the forecast is equal to the perfect forecast, the true values), but can also go negative if the forecast is worse than the benchmark forecast, Eq. (6) express the Skill calculated based on the NMAE, while, in Eq. (7), the Skill calculated on the accuracy is expressed:

$$\text{Skill}_{NMAE} = \frac{NMAE_{forecast} - NMAE_{persistence}}{NMAE_{perfect} - NMAE_{persistence}} \times 100 \tag{6}$$

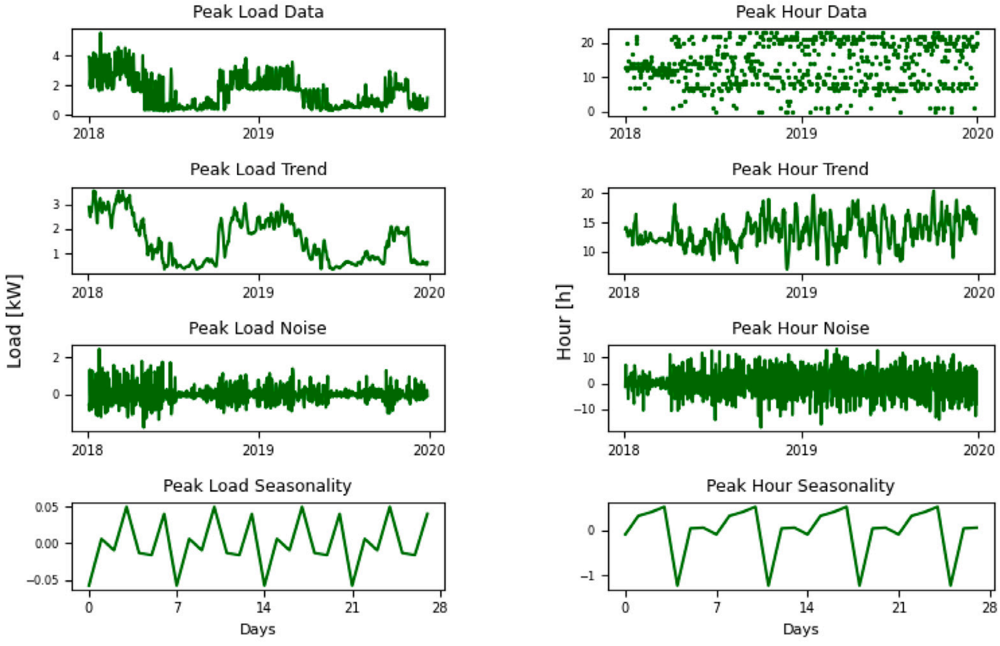
$$\text{Skill}_{ACC} = \frac{ACC_{forecast} - ACC_{persistence}}{ACC_{perfect} - ACC_{persistence}} \times 100 \tag{7}$$

The metric NMAE is used for peak load values, while accuracy was used for peak hour. Accuracy is also used as a core metric during the training phase, where the model is framed as a classification problem. Skill instead is a representation of the errors normalized to the benchmark persistence and perfect error values.

### 2.6. Data analysis

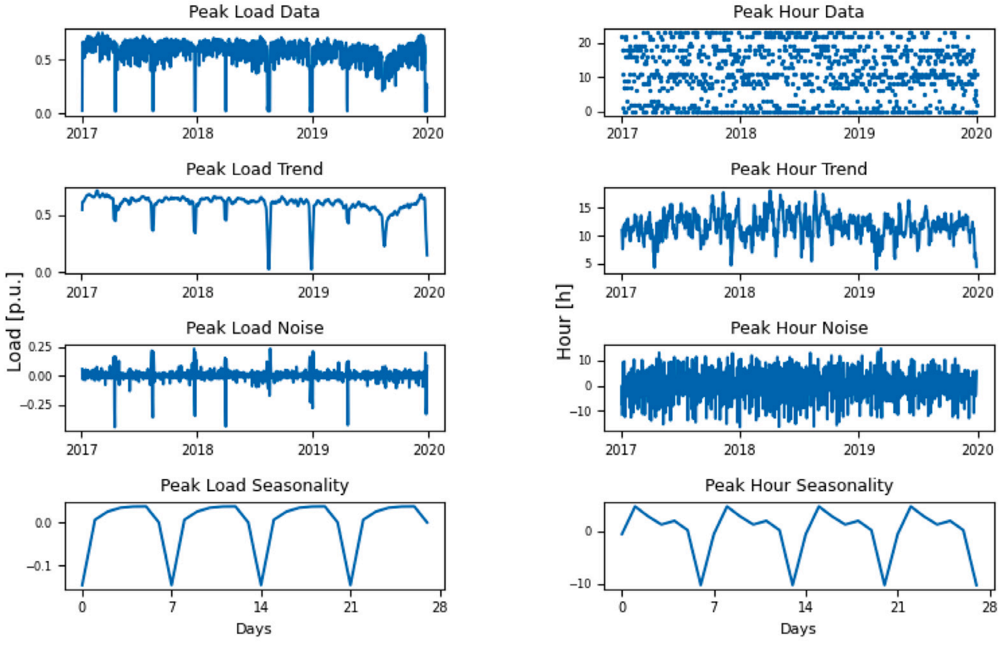
The power time series for both the residential and industrial datasets were decomposed using a convolutional filter to extract the trend and a moving average of the de-trended series to find the seasonal component. This is helpful to better understand why each forecast model succeeded or failed at the prediction task. The components in which the time series have been divided are:

- Trend: The trend component reflects the long-term direction or pattern in the data. Analyzing the trend helps uncover underlying patterns.
- Seasonality: The seasonality component represents recurring and predictable patterns that occur over shorter time periods, typically within a year. It reflects regular variations influenced by factors such as the calendar, weather, or other recurring events.
- Noise: The noise component represents random and unpredictable fluctuations in the time series that cannot be linked to the trend or seasonality. It encompasses factors like measurement errors, random events, or unexplained influences. The noise component is typically assumed to follow a stationary and uncorrelated distribution.



(a) (b)

**Fig. 3.** Decomposition of the peak load (left) and peak hour (right) time series for the residential scenario.



(a) (b)

**Fig. 4.** Decomposition of the peak load (left) and peak hour (right) time series for the industrial scenario.

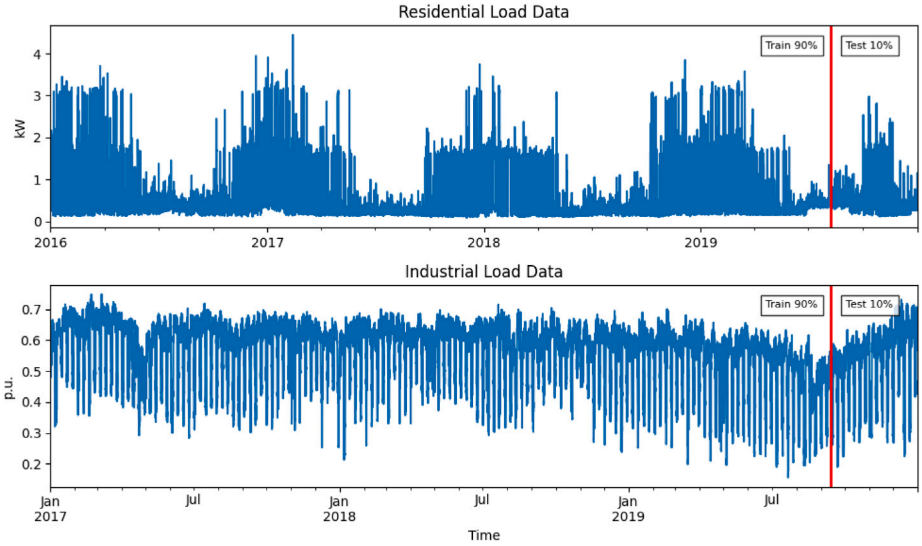


**Fig. 5.** Load scenarios analyzed: residential (above) and industrial (below).

## 3. Case study

Both residential and industrial datasets consist of hourly average power measurements data capturing the electric load in these buildings. To forecast the daily peak load 24 h in advance, we identified the maximum load values for each day and noted the specific hour at which they occurred. A detailed summary of the features of these datasets is provided in Table 3.

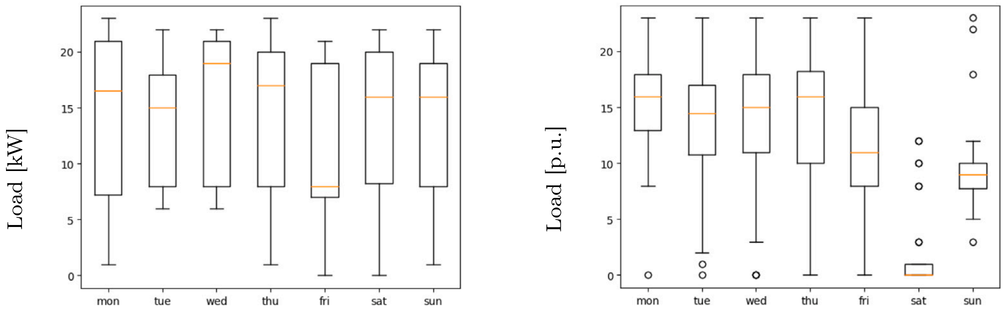
The difference in data volume between the residential and industrial cases is significant. The residential dataset used in this study was collected from a Habitat for Humanity Zero Energy Home located in Wheat Ridge, Colorado. It was in 2005 through a collaboration with NREL, and has been continuously monitored for 17 years, making it one of the longest dataset for buildings, particularly valuable for residential energy analysis and forecasting applications since it is publicly available [42]. Measurements include electric power consumption from 9 channels, plus other consumptions as natural gas usage and hot water flow. The data is available at both 60 min and 5 min resolutions. While this vast dataset offers valuable insights into temporal trend fluctuations, it also reveals several abrupt changes that could compromise model performance. To address this, our analysis focuses on the two most stable years, 2018 and 2019. Of this period, one and a half years serve as the training set, with the remainder designated for testing, as illustrated in Fig. 5. Data from earlier years exhibit distinct trend patterns, while later years are influenced by the unprecedented effects of the COVID-19 pandemic.

The industrial dataset was collected from a paper mill located in northern Italy. The dataset includes detailed measurements of electrical power consumption, recorded at a high temporal resolution. The dataset is confidential and cannot be publicly shared. It includes all three available years.

The time-series components outlined in Section 2 are presented in Fig. 3(a) and (b) for the residential case and Fig. 4(a) and (b) for the industrial case.

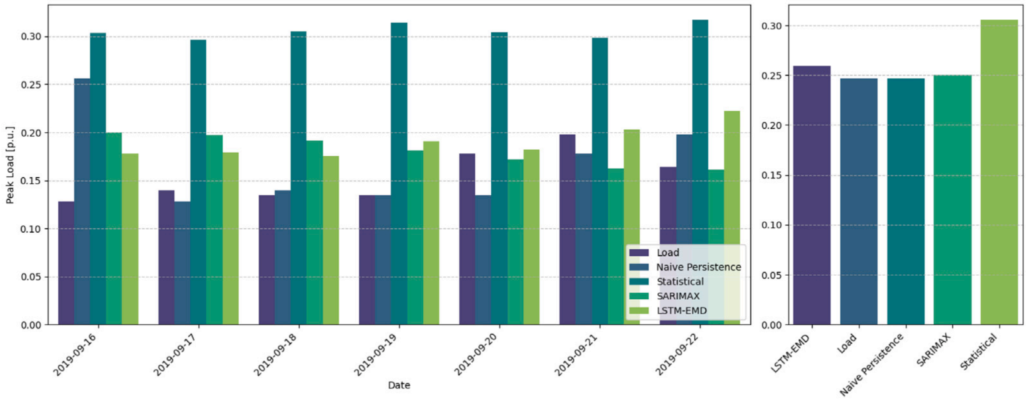
One notable observation is the presence of weekly seasonality (only the first four weeks of data are plotted, for clarity). The seasonal component is more pronounced in the industrial case. To evaluate the level of noise in the data, we applied a moving average smoothing technique with a window of 24 time steps, corresponding to the hourly resolution of the dataset. The noise component was computed as the difference between the original time series and the smoothed signal. For the industrial case, the resulting noise had a standard deviation of 0.04 and a variance of 0.0017, indicating a low level of variability. In contrast, the residential case exhibited a standard deviation of 0.82 and a variance of 0.68, reflecting a significantly higher degree of variability. Outlier detection was also performed based on the distribution of the estimated noise. Specifically, data points exceeding three standard deviations from the mean were classified as outliers. Using this approach, we identified 1.62% of the data points as outliers in the residential dataset and 0.28% in the industrial dataset.

Additionally, both datasets for the peak hour exhibit a higher noise component, which results in reduced forecast accuracy. For the residential case, the estimated noise has a standard deviation of 4.20 and a variance of 17.60, while the industrial dataset showed a higher standard deviation of 7.49 and a variance of 56.08. This is further evident in the two peak hour box plots (Figs. 6(a) and 6(b)), where the data distribution, broken down by the days of the week, shows wide quartile ranges. This indicates that the measurements can vary significantly each day, except on weekends for the industrial case. In the residential case, the seasonal component has a relatively low magnitude for both peak load and peak hour. Although the SARIMAX model cannot capture the noise, it performs well in predicting the trend. In contrast, the industrial case exhibits significant contributions from all three components for both peak load and peak hour, though the trend is only occasionally relevant for peak load. Accurate forecasting in this setting requires models capable of capturing both seasonality and stochastic noise, which SARIMAX is inherently unable to model effectively.

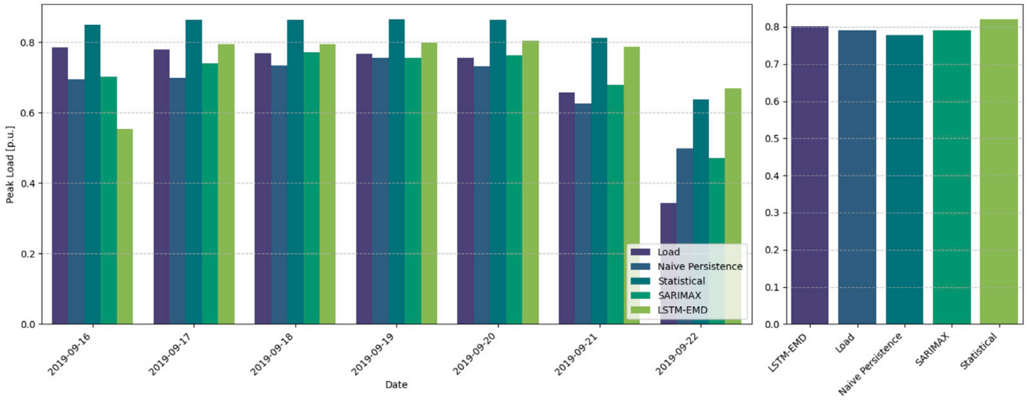


(a) Residential (b) Industrial

**Fig. 6.** Peak hour box plots for the two case studies.



(a) Residential



(b) Industrial

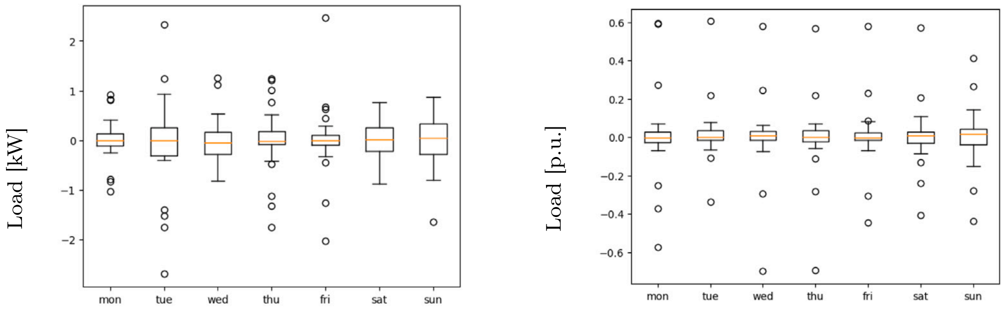
**Fig. 7.** Comparison peak load forecasted by the proposed models and true values for an example week (on the left) and averaged on the entire dataset (on the right).

## 4. Results

This section presents the results from each model along with a detailed analysis and discussion. Figs. 7(a) and 7(b) present the graph of a week and the average on the full dataset comparing the performances of all the models proposed in the study, illustrating the behavior of the forecasted peak load against actual values for both the industrial and residential cases.

**Table 4**
Comparison of model performance in terms of NMAE and its corresponding skill score for peak load forecasting, as well as accuracy (ACC) and skill score for peak hour prediction.

| Methods | Residential | | | | Industrial | | | |
|---|---|---|---|---|---|---|---|---|
| | Peak load | | Peak hour | | Peak load | | Peak hour | |
| | NMAE | Skill | ACC | Skill | NMAE | Skill | ACC | Skill |
| Persistence | 0.074 | – | 6.9% | – | 0.0478 | – | 15.7% | – |
| Statistical | 0.141 | −91.0% | 13.1% | 6.7% | 0.065 | −39.1% | 21.3% | 6.6% |
| SARIMAX | 0.067 | 10.1% | 9.0% | 2.2% | 0.043 | 7.1% | 18.7% | 3.5% |
| LSTM-EMD | 0.072 | 2.7% | 19.9% | 13.9% | 0.009 | 81.5% | 26.4% | 12.6% |



(a) Residential (b) Industrial

**Fig. 8.** Naïve residuals box plot for the two case studies.

### 4.1. Naïve Persistence model

Figs. 7(a) and 7(b) illustrate the forecasted peak load compared to actual values for both the industrial and residential cases over a typical week and the average on the full dataset. The NP model demonstrates a reasonably accurate prediction, especially considering its simplicity.

Meanwhile, the box plots (Figs. 8(a) and 8(b)) display the distributions of peak hours and forecast residuals across different days of the week. Notably, the industrial case exhibits lower uncertainty compared to the residential case. The dispersion increases for the predicted peak hour, and the residential dataset shows a significantly wider range between the first and third quartiles than the industrial dataset. In both cases, the residuals' box plots are reassuring, as they indicate a mean value close to zero and a narrow interquartile range.

As presented in Table 4, the NMAE are lower than expected for a model of this simplicity, although their values increase in the residential scenario due to the influence of weekly seasonality. Notably, the accuracy of peak hour predictions in the residential case is significantly lower, attributed to the more irregular data distribution and weaker seasonal patterns.

### 4.2. Statistical model

The Table 4 indicates that the statistical model provides a better estimation of the peak hour data than peak load data compared to the benchmark case. This finding is further supported by the plots in Figs. 9(a) and 9(b). Specifically, in the residential case, the peak load prediction exhibits quite a low skill.

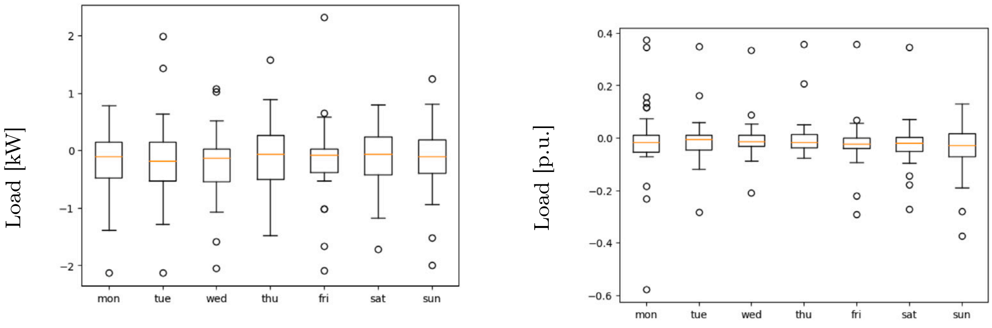
By examining the NMAE values in the error table, we observe an improvement in the industrial case compared to the benchmark. However, for the residential case, as previously discussed, the same conclusion cannot be made. On a positive note, the accuracy value has increased.

The slightly lower ADF statistic (Table 3) of the industrial scenario, which is to say a slightly less stationary time series, may explain the statistical model's advantage in the industrial scenario compared to residential. This is because the statistical model is more robust to changes in the data distribution over time, due to its simple nature of calculating the expected value of a given weekday's random variable of all previously measured values.

### 4.3. SARIMAX model

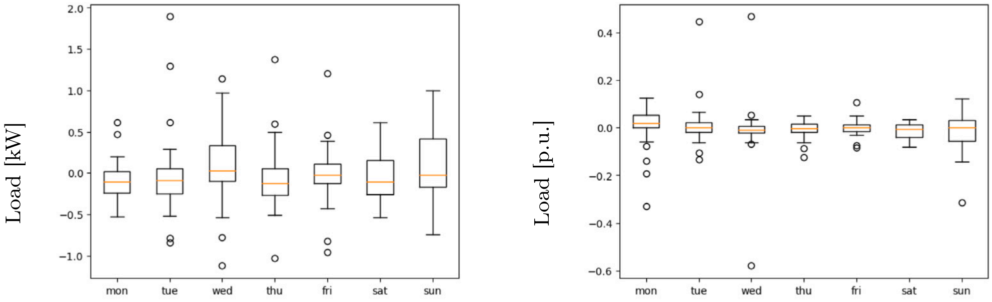
The SARIMAX parameters obtained from the random-search, minimizing train set MAE, are shown in Table 5.

A visual inspection of the graphs (7(a) and 7(b)) reveals that the SARIMAX model outperforms all other models for the residential but not industrial case. This conclusion is further supported by the box plots (Figs. 10(a) and 10(b)) and the error values in Table



(a) Residential (b) Industrial

**Fig. 9.** Statistical residuals box plot for the two case studies.



(a) Residential (b) Industrial

**Fig. 10.** SARIMAX model residuals box plot for the two case studies.

**Table 5**
SARIMAX parameters for different scenarios: (p,d,q)(P,D,Q)-s.

| Scenario | Time series | Parameters |
|---|---|---|
| Residential | Peak load | (3, 0, 2)(0, 0, 1)-24 |
| Residential | Peak hour | (7, 1, 4)(0, 1, 0)-20 |
| Industrial | Peak load | (1, 0, 7)(2, 0, 1)-30 |
| Industrial | Peak hour | (0, 0, 0)(1, 0, 1)-30 |

4, where NMAE is lower compared to the Naïve model, and accuracy is also improved. Since Table 3 confirms the stationary of the data in both scenarios, given that the ADF statistic is much lower than the 1% critical value, and therefor it is not implausible that SARIMAX could perform well. Additionally, SARIMAX can suffer when there are multiple strong seasonalities in the data, but the relative similarity between 1 day and 7 day autocorrelation factors suggests this is not the case.

### *4.4. LSTM-EMD model*

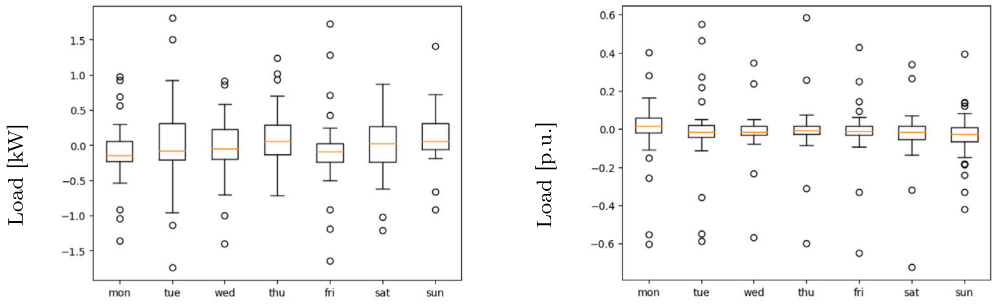
The novel LSTM-EMD model was trained to predict the peak load and the hour in which it would occur by giving as inputs, the month, the day of the month, the day of the week, the peak value of the previous $k$ days, and the correspondent peak load hours. Confidence in the results is high, given that several years of data were available for training and a distinct test data set was used to confirm the generalization results.

This method outperforms the NP and Statistical model benchmarks in all cases, and outperforms SARIMAX in three of the four cases. Only the residential peak load forecast skill is lower for LSTM than for SARIMAX, at 2.7% compared to 10.1% respectively. It may be at the very large LSTM dimension of 2096 contained enough parameters to overfit the training data despite using the early stopping mechanism to limit this effect. Oddly more dropout did seem to improve the test set loss, but this could be simply because the maximum dropout value during hyper-parameter optimization was 0.2, wheres 0.3 or greater may have been necessary to reduce overfitting.

The most encouraging result and strongest indication of the success of the novel LSTM-EMD structure is the 81.5% skill for the industrial peak load forecast, which is an order of magnitude of skill improvement over the next best model, SARIMAX. Also this

**Table 6**
Hyper-parameter values in this order: input number of days of historical peak value or load measurements, LSTM hidden dimension, ANN units, dropout.

| Scenario | Time series | Hyper parameters |
|---|---|---|
| Residential | Peak load | (5, 2096, 80, 0.0) |
| Residential | Peak hour | (1, 512, 16, 0.0) |
| Industrial | Peak load | (24, 96, 96, 0.1) |
| Industrial | Peak hour | (1, 256, 128, 0.0) |



(a) Residential

(b) Industrial

**Fig. 11.** LSTM residuals box plot for the two case studies.

is by far the largest skill of any model in the research, possibility indicating further improvements could have been made in the industrial peak hour forecast (12.6% skill) and both residential forecasts (see Fig. 11).

## 5. Conclusion

In conclusion, evaluating the feasibility of load forecasting for both scenarios is essential. The results indicate that the models achieve higher accuracy in predicting the industrial scenario. This is largely due to the more seasonal and pattern or event-driven energy demand of an industrial plant operating at full capacity during the workweek, in contrast to the more stochastic demand profile of a residential setting. Analyzing the box plots of actual samples from both scenarios reveals that peak hour predictions exhibit a high degree of randomness. These challenges highlight the difficulties the models face in accurately forecasting the peak hour. Additional observations can be made regarding the LSTM-EMD model: its relatively poor performance in forecasting the peak load for the residential case may be attributed to overfitting a large number of parameters, despite the substantial quantity of training data and efforts to limit overfitting. It is important to consider potential improvements for future work. The LSTM-EMD would likely require $k$-fold cross validation to increase generalization, by creating $k$ test data sets rather than just one. Furthermore increasing the data set size with transfer learning or the use of synthetic data would most likely be beneficial.

Given the poor results in hour accuracy for the all model in both the residential and industrial cases, noting that no model achieves more than 26.4% accuracy, more emphasis could be places on statistical methods such as Bayesian inference to increase accuracy. Additionally, it would be reasonable to explore the implementation of other ML models such as transformers and convolutional neural networks.

The best results overall come from the LSTM-EMD and SARIMAX models, which is likely due to the presence of all three time series components — trend, seasonality, and noise — but an otherwise well-behaved statistical variation in time, allowing these two models to learn patterns in the past and extrapolate them into the near future.

Finally, it is important to highlight that incorporating exogenous factors such as temperature, meteorological conditions, and others could enhance the accuracy of more complex models like LSTM-EMD and SARIMAX especially in the residential case, where certain load patterns could be temperature or sunlight dependent.

### CRediT authorship contribution statement

**Michael Wood:** Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Silvana Matrone:** Writing – original draft, Visualization, Software, Data curation. **Emanuele Ogliari:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Data curation. **Sonia Leva:** Writing – review & editing, Validation, Supervision, Formal analysis, Data curation.

### Generative AI

Generative AI was partially used in the creation of this text, though not at all in the study design, results, analysis, or findings.

## Acknowledgments

We would like to thank Fabio Casadidio, Giacomo Ramacciotti, and Marco Saldari for their contributions to this study as part of their coursework final project.

## References

[1] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy 55 (2013) 184–194.

[2] A.W. Yao, S. Chi, J. Chen, An improved grey-based approach for electricity demand forecasting, Electr. Power Syst. Res. 67 (3) (2003) 217–224.

[3] IEA, Electricity market report 2023, 2023.

[4] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, Int. J. Forecast. 32 (3) (2016) 914–938.

[5] Y. Kim, H. gu Son, S. Kim, Short term electricity load forecasting for institutional buildings, Energy Rep. (2019).

[6] K.G. Boroojeni, M.H. Amini, S. Bahrami, S. Iyengar, A.I. Sarwat, O. Karabasoglu, A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon, Electr. Power Syst. Res. 142 (2017) 58–73, http://dx.doi.org/10.1016/j.epsr.2016.08.031.

[7] e.a. Lindeberg, Long-term electricity load forecasting: Current and future trends, J. Up. (2019) http://dx.doi.org/10.1016/j.jup.2019.04.001.

[8] S. Matrone, E. Ogliari, A. Nespoli, S. Leva, Electric vehicle supply equipment day-ahead power forecast based on deep learning and the attention mechanism, IEEE Trans. Intell. Transp. Syst. 25 (8) (2024) 9563–9571, http://dx.doi.org/10.1109/TITS.2024.3391375.

[9] Z. Yu, Z. Niu, W. Tang, Q. Wu, Deep learning for daily peak load forecasting–A novel gated recurrent neural network combining dynamic time warping, IEEE Access 7 (2019) 17184–17194, http://dx.doi.org/10.1109/ACCESS.2019.2895604.

[10] J. Lee, Y. Cho, National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? Energy 239 (Part D) (2022) 122366, http://dx.doi.org/10.1016/j.energy.2021.122366.

[11] D.-H. Kim, E.-K. Lee, N.B.S. Qureshi, Peak-load forecasting for small industries: A machine learning approach, Sustain. 12 (16) (2020) 6539, http://dx.doi.org/10.3390/su12166539.

[12] L. Wen, K. Zhou, S. Yang, X. Lu, Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting, Energy 171 (2019) 1053–1065, http://dx.doi.org/10.1016/j.energy.2019.01.075.

[13] Y. Ozturk, D. Senthilkumar, S. Kumar, G. Lee, An intelligent home energy management system to improve demand response, IEEE Trans. Smart Grid 4 (2) (2013) 694–701, http://dx.doi.org/10.1109/TSG.2012.2235088.

[14] A. Nottrott, J. Kleissl, B. Washom, Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems, Renew. Energy 50 (2013) 1036–1043, http://dx.doi.org/10.1016/j.renene.2012.08.017.

[15] A. Khotanzad, R. Afkhami-Rohani, ANNSTLF - Artificial neural network short-term load forecaster - generation three, IEEE Trans. Power Syst. 13 (4) (1998) 1413–1422, http://dx.doi.org/10.1109/59.736275.

[16] C.-C. Hsu, C.-Y. Chen, Regional load forecasting in Taiwan - Applications of artificial neural networks, Energy Convers. Manage. 44 (12) (2003) 1941–1949, http://dx.doi.org/10.1016/S0196-8904(02)00225-X, Cited by: 210.

[17] J.W. Taylor, An evaluation of methods for very short-term load forecasting using minute-by-minute British data, Int. J. Forecast. 24 (4) (2008) 645–658, http://dx.doi.org/10.1016/j.ijforecast.2008.07.007.

[18] M. Wood, E. Ogliari, A. Nespoli, T. Simpkins, S. Leva, Day ahead electric load forecast: A comprehensive LSTM-EMD methodology and several diverse case studies, Forecast. 5 (2023) 297–314, http://dx.doi.org/10.3390/forecast5010016.

[19] J. Zhu, Z. Yang, Y. Chang, Y. Guo, K. Zhu, J. Zhang, A novel LSTM based deep learning approach for multi-time scale electric vehicles charging load prediction, in: 2019 IEEE Innovative Smart Grid Technologies - Asia, ISGT Asia, 2019, pp. 3531–3536, http://dx.doi.org/10.1109/ISGT-Asia.2019.8881655.

[20] Q. Gao, et al., Charging load forecasting of electric vehicle based on Monte Carlo and deep learning, in: 2019 IEEE Sustainable Power and Energy Conference, ISPEC, 2019, pp. 1309–1314, http://dx.doi.org/10.1109/iSPEC48194.2019.8975364.

[21] Y.H. Kareem, A.R. Majeed, Monthly Peak-load Demand Forecasting for Sulaimany Governorate Using SARIMA, IEEE, 2006.

[22] A.E. Clements, A.S. Hurn, Z. Li, Forecasting day-ahead electricity load using a multiple equation time series approach, European J. Oper. Res. (2015) http://dx.doi.org/10.1016/j.ejor.2015.12.030.

[23] B. Yildiz, J. Bilbao, A. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. (2019) http://dx.doi.org/10.1016/j.rser.2017.02.023.

[24] S.M. Awan, Z. Khan, M. Aslam, W. Mahmood, A. Ahsan, Application of NARX Based FFNN, SVR and ANN Fitting Models for Long Term Industrial Load Forecasting and Their Comparison, IEEE, 2012.

[25] J. Lee, Y. Cho, National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model, Energy (2021) http://dx.doi.org/10.1016/j.energy.2021.122366.

[26] M. Mansoor, F. Grimaccia, S. Leva, M. Mussetta, Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs, Math. Comput. Simulation 184 (2021) 282–293, http://dx.doi.org/10.1016/j.matcom.2020.07.011.

[27] D. Darwazeh, J. Duquette, B. Gunay, I. Wilton, S. Shillinglaw, Review of peak load management strategies in commercial buildings, Sustain. Cities Soc. 77 (2022) 103493, http://dx.doi.org/10.1016/j.scs.2021.103493.

[28] A.V. Kychkin, G.C. Chasparis, Feature and model selection for day-ahead electricity-load forecasting in residential buildings, Energy Build. (2021) http://dx.doi.org/10.1016/j.enbuild.2021.111200.

[29] Y. Kim, H.-g. Son, S. Kim, Short term electricity load forecasting for institutional buildings, Energy Rep. (2019) http://dx.doi.org/10.1016/j.egyr.2019.08.086.

[30] D.-H. Kim, E.-K. Lee, N.B.S. Qureshi, Peak-load forecasting for small industries: A machine learning approach, Sustain. 12 (16) (2020) 6539.

[31] I. Zapirain, et al., Short-term electricity consumption forecasting with NARX, LSTM, and SVR for a single building: small data set approach, Appl. Energy (2022) http://dx.doi.org/10.1080/15567036.2022.2104410.

[32] M.-R. Kazemzadeh, A. Amjadian, T. Amraee, A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting, Energy (2020) http://dx.doi.org/10.1016/j.energy.2020.117948.

[33] J. Zhu, Z. Yang, M. Mourshed, Y. Guo, Y. Zhou, Y. Chang, Y. Wei, S. Feng, Electric vehicle charging load forecasting: A comparative study of deep learning approaches, Energies 12 (14) (2019) 2692, http://dx.doi.org/10.3390/en12142692.

[34] J. Zhu, Z. Yang, Y. Guo, J. Zhang, H. Yang, Short-term load forecasting for electric vehicle charging stations based on deep learning approaches, Appl. Sci. 9 (9) (2019) 1723, http://dx.doi.org/10.3390/app9091723.

[35] T. Haida, S. Muto, Regression based peak load forecasting using a transformation technique, IEEE Trans. Power Syst. 9 (4) (1994) 1788–1794, http://dx.doi.org/10.1109/59.331433.

[36] S. Giazitzis, M. Sakwa, E. Ogliari, S. Badha, F. Rosetti, Tiny machine learning for li-ion battery state of health estimation, in: 2024 IEEE 22nd Mediterranean Electrotechnical Conference, MELECON, 2024, pp. 1019–1024, http://dx.doi.org/10.1109/MELECON56669.2024.10608784.

[37] D. Pau, E. Ogliari, M. Sakwa, Tiny machine learning for dynamic line rating of the overhead lines, in: 2024 IEEE 8th Energy Conference, ENERGYCON, 2024, pp. 1–6, http://dx.doi.org/10.1109/ENERGYCON58629.2024.10488819.

[38] N. Mughees, S.A. Mohsin, A. Mughees, A. Mughees, Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting, Expert Syst. Appl. 175 (2021) 114844, http://dx.doi.org/10.1016/j.eswa.2021.114844.

[39] O. Rubasinghe, X. Zhang, T.K. Chau, Y.H. Chow, T. Fernando, H.H.-C. Iu, A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting, IEEE Trans. Power Syst. 39 (1) (2024) http://dx.doi.org/10.1109/TPWRS.2023.3271325.

[40] J. Liu, L.E. Brown, Prediction of hour of coincident daily peak load, in: 2019 IEEE 2nd International Conference on Artificial Intelligence and Big Data, ICAIBD, IEEE, 2019, 114844, http://dx.doi.org/10.1109/ICAIBD.2019.8837012.

[41] J. Brownlee, A gentle introduction to SARIMA for time series forecasting in python, 2019, Online, (Accessed 03 March 2024).

[42] P. Norton, C. Christensen, E. Hancock, G. Barker, P. Reeves, NREL/Habitat for Humanity Zero Energy Home: A Cold-Climate Case Study for Affordable Zero Energy Homes, Technical Report NREL/TP-550-43188, National Renewable Energy Laboratory (NREL), Golden, CO, United States, 2008.