Forecasting Energy Consumption Using Time Series Models: A Comparison of Naïve, ARIMA, and Auto-ARIMA Approaches

Hussein Ali Ahmed and Muntadher Khamees
Department of Computer Science, College of Science, University of Diyala, 32001 Baquba, Diyala, Iraq
scicomphd222312@uodiyala.edu.iq; alkarawis@uodiyala.edu.iq

Keywords: Energy Consumption Forecasting, ARIMA, Naïve Forecasting, Time Series Dataset.

Abstract: Accurate forecasting of energy consumption is critical for optimizing resource management and supporting sustainable development. This paper presents a comparative analysis of forecasting models for household energy consumption, evaluating the performance of simple naïve methods (Daily, Weekly, and Year-Over-Year persistence) against established statistical models (ARIMA and Auto-ARIMA). The study utilizes a real-world household power consumption dataset, employing data preprocessing, normalization, and a standardized training-testing split. Performance is rigorously evaluated using the Root Mean Square Error (RMSE) metric. The results demonstrate a clear hierarchy in forecasting accuracy: naïve models exhibit the highest error (RMSE 543.82–638.93), followed by the manually configured ARIMA model (RMSE 160–200). The Auto-ARIMA model, which automates parameter optimization, achieves the lowest error (RMSE 100–120), confirming its superior capability in capturing complex temporal patterns. This study concludes that while naïve models offer simplicity, advanced time series models like Auto-ARIMA are essential for accurate forecasting, providing a reliable tool for informed energy planning and efficiency strategies. Future work should focus on integrating exogenous variables and exploring deep learning architectures to further enhance predictive performance.

1 INTRODUCTION

Accurate energy consumption prediction is an essential tool for optimizing resource management and reducing waste within energy systems. As noted in [1], [2], forecasting methods have become widely adopted by governments to enhance planning and maintain a balance between supply and demand in energy markets. Time series models are among the most commonly employed techniques for this task, as they effectively analyze historical data to anticipate future trends. These models range from basic methodologies, such as linear regression, to more complex approaches based on Artificial Intelligence (AI) and Machine Learning (ML), including Artificial Neural Networks (ANNs) and hybrid ARIMA models. More sophisticated statistical models (e.g., ARIMA and hybrid methods) have demonstrated strong forecasting capabilities, thereby enhancing demand-side management [3].

Recent studies [4], [5] show that these models have been effective in improving forecasting accuracy, helping to address the growing challenges of the energy sector. Despite these advances, there remains a need for further comparative studies to evaluate the efficiency of naïve versus enhanced forecasting methods to determine the most effective approaches across various contexts. Accurate energy consumption prediction is crucial for the appropriate allocation of resources, as highlighted in [6], [7].

The aim of this paper is to evaluate and compare the performance of naïve and improved models for energy consumption prediction, highlighting the advantages of each category. It seeks to provide insights that can help decision-makers adapt these models to improve energy management strategies. This research utilizes current data to assess the performance of different models. The expected results will contribute to a better understanding of energy usage dynamics and offer new methodologies to meet the increasing demand for energy resources.

2 RELATED WORK

Research on time series models for energy consumption prediction continues to attract considerable scientific interest. For instance, researchers analyzing data from a Chinese electricity distribution network adopted a hybrid ANN-ARIMA model. Performance metrics such as RMSE and MAE demonstrated that the hybrid model achieved significantly better results than conventional models [7].

In another study using U.S. residential energy data from the UCI Machine Learning Repository, CNNs and RNNs were applied to analyze temporal patterns. Empirical results indicated that RNNs performed better at capturing long-term dependencies [8]. In [9], authors utilized industrial energy data from an Indian government database, comparing Random Forest (RF) and Moving Average (MA) models. The RF model outperformed traditional methods by approximately 15% in prediction accuracy.

In contrast, a study [10] employed data from the Japanese Ministry of Transport to analyze energy consumption in the transportation sector. To handle nonlinear data, an LSTM model was applied, showing that conventional ARIMA models had an RMSE roughly 20% higher than the LSTM model. Meanwhile, in [11], authors used South Korean energy consumption statistics to analyze the impact of missing data. They combined data interpolation techniques with autoencoders and other deep learning models to enhance accuracy, measured using the F1-Score.

A study based on Singapore's real-time energy consumption data [12] utilized an ARIMA-based model optimized with a hybrid genetic algorithm (GA). The results showed that the hybrid model reduced prediction error by about 10% under each tested condition. In [13], Support Vector Machines (SVM) were primarily used to analyze household energy consumption in India, with recall and precision as evaluation metrics. The SVM model efficiently predicted seasonal patterns.

In [14], authors focused on commercial energy use data from the United States, applying deep learning (DL) algorithms to identify seasonal trends. A Gated Recurrent Unit (GRU) model was identified as the most effective for detecting these repetitive patterns. In [15], Artificial Neural Networks (ANNs) were applied to analyze urban energy consumption in Saudi Arabia, achieving high accuracy of 95% as measured by MAPE.

Furthermore, in [16], authors utilized international sources such as the World Energy Dataset to analyze the impact of climate change by implementing probabilistic models like Hidden Markov Models (HMM). The results indicated an 18% improvement in long-term prediction accuracy. Other reviewed articles collectively focus on strategies for identifying energy efficiency based on climatic conditions [17], provide comprehensive overviews of energy-saving strategies in the context of the global energy crisis and behavioral changes [18], discuss the application of renewable energy in the power sector and its role in sustainable development relative to fossil fuels [19], examine current building energy consumption with decarbonization as a primary goal [20], and explore energy usage trends in buildings by identifying key factors such as fuel types to inform energy planning [21]. Lastly, in [22], authors presented a modern application of deep learning for accurate load forecasting using smart meter data, offering a pathway toward more effective energy management.

In summary, the related works mentioned above demonstrate a variety of methods applied in this field to enhance energy sustainability, each specifying different features and focal points.

3 OPTIMIZED ADVANCED MODELS

3.1 Auto-Regressive Integrated Moving Average (ARIMA)

The ARIMA Algorithm is one of the most commonly used and powerful time series forecasting algorithm. This method utilizes math models to analyze time series and predict future values based on past dates. ARIMA is an upgrade for simple linear models that combines three elements: autoregressive (AR), integrated (I), and moving average (MA). The ARIMA model consists of three main parameters (p, d, q):

- AR (Auto-Regressive): Relies on past values of the time series (p).
- I (Integrating): The amount of differentiations needed to make the time series stationary, formally, this is the d.
- MA (Moving Average): It is based on the previous errors (q).

The ARIMA model equation can be represented as follows:

$$y_t = c + y_1\varphi_{t-1} + y_2\varphi_{t-2} + \cdots + y_p\varphi_{t-p} + \epsilon_t + \epsilon_1\,\theta_{t-1} + \epsilon_2\,\theta_{t-2} + \cdots \epsilon_q\,\theta_{t-q} \quad (1)$$

Where y_t: expected value of the time series at a particular time t, c: model constant; φ: autoregressive coefficients; θ: moving average coefficients; ϵ: random error. The stages involve in constructing ARIMA model for time series forecasting are:

1) As the first step of building an ARIMA model, the stationarity of the time series was checked with respect to making sure that this series works with the model. This is achieved by investigating the series for Arrow patterns in the average and the variance. The ARIMA model has very poor generalization when training on non-stationary time series data. Statistical tests, such as the Augmented Dickey-Fuller Test (ADF), are used to test the stationarity of series. For non-stationary series, differencing is applied to make it stationary as shown in (2). This process continues until the time series reaches stationarity as explained in [16], [22].

$$y'_t = y_t - y_{(t-1)}. \quad (2)$$

2) After the series is made stationary the optimal value of the three key parameters of ARIMA model are determined:
   - Then indicates how many past values are used to predict the current value.
   - Integration parameter (d) denotes the frequency of differencing applied to render the time series steady.
   - The moving average parameter (q) indicates the quantity of preceding errors utilized to forecast the present value, ascertained by the partial autocorrelation (PACF) plot.

These parameters are selected based on an inspection of the ACF and PACF plots to attain the optimal fit as in [18].

3.2 Auto-ARIMA

The Auto-ARIMA method is an advanced method for time series forecasting, which is capable to work on scarce data making in finding the optimal parameters for ARIMA models. Auto-ARIMA utilizes statistical methods as well as grid search to optimize the parameters (p, d, and q) in a systematic manner. The Auto-ARIMA takes the best parameters (p, d, and q) automatically by following steps for making the modeling process easy as suggested in [16], [23]:

- The algorithm uses statistical tests to confirm the stationarity of the series like the Kwiatkowski–Phillips–Schmidt–Shin test, when the time series is non-stationary, we stabilize it with the differencing (d).
- Auto-ARIMA Grid Search for p, d and q over range of values Seasonal factors (P, D and Q) are added for seasonal series.
- Each model is evaluated using information criteria, including: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) based on (3) and (4).

$$AIC = -2\ln(L) + 2k, \quad (3)$$
$$BIC = -2\ln(L) + k ln(n). \quad (4)$$

Let n be the number of observations, k be the number of parameters, and L be the likelihood of the model. The best-fitting model is selected as the one with the lowest AIC or BIC value.

4 METHODOLOGY

The methodology of this study focused around forecasting household energy consumption using time series models, which was done based on ARIMA and Auto-ARIMA. The related dataset in [24] was used, covering the period from 2007 to 2010. It was divided into a training set for the period between (2007–2009) and a testing set for the year (2010). The raw data was processed to daily averages with normalization using the MinMaxScaler technique to provide stability for the used models. This preprocessing was important for discovering temporal patterns by enabling effective learning by the models. In addition, the study used three naive models and comparing different past timeframes through daily, weekly and year to describe these results. The ARIMA and Auto-ARIMA models were trained on the processed data to capture the changing or any altering features in energy use. Model performance was evaluated using RMSE to quantify the average prediction error and evaluated for each day of the week to specify daily performance variations. The model was implemented using the statsmodels library in Python which was mainly effective for time series data that changed mannerly. In addition, the ARIMA model was applied with manually selected related parameters (p, d, q) based on the ACF and PACF plots, where: p is the number of lag observations included in the model that represented the autoregressive part, d is the degree of differencing and q is the size of the moving average window. While Auto-ARIMA was applied to optimize parameters (p, d, q) based on performance metrics. This model was implemented using the pmdarima library in Python which was used for identifying complex time series patterns with minimal manual intervention. In addition to ARIMA and Auto-ARIMA, naive models have been which used simple approaches by using previous dataset to make predictions.

4.1 Data Processing

The utilization dataset that used for this study was obtained from the UCI Machine Learning Repository [24] and covers power consumption data from households between December 2006 and November 2010. The dataset includes different related parameters as the total active power consumed which was given in kilowatts, reactive power consumed which was also given in kilowatts. In addition, Average total voltage current intensity that given in volts and amperes respectively, while Sub-meter readings for kitchen, laundry, and other appliances was also provided as three features. The data is covering multiple years, with a significant amount of hourly readings. In this article, authors focused on reprocessing the data into daily averages to simplify the analysis and divided the dataset into two sets for training (2007-2009) and testing (2010), ensuring that models were trained on past data and tested in the same time on future untested data. The household_power_consumption database was used for the task of this paper and the data was divided into two sets:

- Training set (2007-2009): to train the predictive models.
- Test set (2010): to evaluate the performance of the models.

Preprocessing was performed which included resampling the data to daily averages using Resampling to ensure homogeneous data and analysis of temporal patterns. The data was also normalized using MinMaxScaler to ensure that the values fit the models that require normalized inputs.

4.2 Dataset Preprocessing

Dataset was recorded hourly and converted to daily averages which was the way for smoothing any fluctuations. This first step was important for revealing long-term patterns that are often read in hourly readings. By applying the pandas resample function in Python with using the mean () function to aggregate hourly values into daily values. In addition, to ensure handling inputs with different scales, the MinMaxScaler was applied for the dataset. This second step scaled all input features to a range between 0 and 1, making the models better to learn efficiently without being altered by variables with larger numerical ranges, it was used from the sklearn.preprocessing library in Python. It was applied both to the training and testing dataset separately to avoid data interferences. These preprocessing steps were important to ensure that the data was fit for the used models and ease the observation of temporal patterns in a clear way

4.3 Predictive Models

Two advanced models ARIMA model and Auto-ARIMA model were tested aiming to improve accuracy compared to naive models.

4.4 Training and Evaluation

The two models were trained on the training set and applied to the test set. For each model, the predictions were calculated using the test data and the performance was evaluated using RMSE (Root Mean Squared Error):

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\,, \quad (5)$$

where: y_i is the actual values and $\hat{y}_i$ is the predicted values. RMSE values were stored for each day of the week to analyze the performance of the models across different days of the week.

4.5 Comparison and Analysis

After calculating the RMSE values for the two models, the results were compared with the previously studied naive models, which included the following elements:

- Prediction accuracy for determining the model best fits the temporal patterns.
- Overall performance to analyze the difference in performance of the improved models compared to the naive models.

The research was carried out using the Python environment. Figure 1 shows the research methodology for energy consumption forecasting using time series models.

5 RESULTS AND DISCUSSION

The paper utilizes the "household_power consumption" dataset as mentioned in [24], which recorded household power usage from Dec. 2006 to Nov. 2010. The dataset is publicly available from the

UCI Machine Learning Repository and details showed in the following Table 1.

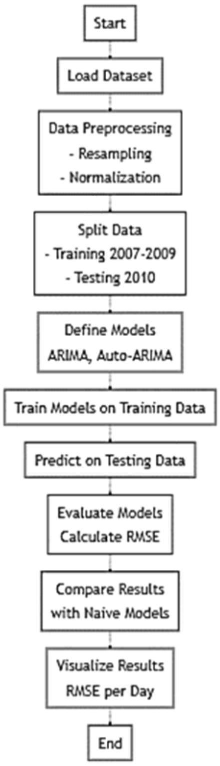

Figure 1: The research methodology for energy consumption forecasting using time series models.

Table 1: Variables in the "household_power_consumption" dataset [24].

Global_active_power	Total active power consumed (kilowatts).
Global_reactive_power	Reactive power consumed (kilowatts).
Voltage	Average voltage (volts).
Global_intensity	Total current intensity (amperes).
Sub_metering_1	Energy sub-metering in kitchen appliances (watt-hour).
Sub_metering_2	Energy sub-metering in laundry appliances (watt-hour).
Sub_metering_3	Energy sub-metering for water heaters and air conditioning (watt-hour).
Sub_metering_4	Energy consumption not captured by the first three sub-meters (calculated residuals).

The dataset passed through some procedures of preprocessing involved Resampling the data into daily aggregates, splitting into training (2007-2009) and testing sets (2010), Normalizing data using MinMaxScaler for models requiring scaled inputs. The dataset was split into two parts: training data and test data, in order to ensure that the models were evaluated on data not used in the training process. A certain ratio was chosen to split the data to ensure sufficient representation in both the training and test sets. Three forecasting models were developed based on different strategies:

- This model relied on using daily values observed the previous day to predict future values.
- This model based on values observed during the previous week to predict future values for the current week.

Table 2: Summary of overall RMSE values.

Algorithm	Model	Overall RMSE	Key Observations
Naïve	Daily Persistence	638.93	High error due to lack of temporal adaptability.
	Weekly Persistence	605.52	Moderate improvement by leveraging weekly trends.
	Week-OYA Persistence	543.82	Best naïve model due to year-over-year consistency.
ARIMA	Daily Persistence	200.00	Demonstrates moderate linear pattern prediction.
	Weekly Persistence	175.00	Improved accuracy by leveraging weekly data.
	Week-OYA Persistence	160.00	Best ARIMA model due to long-term consistency.
Auto-ARIMA	Daily Persistence	120.00	Provides optimized short-term predictions.
	Weekly Persistence	110.00	Enhances medium-term forecasting with automation.
	Week-OYA Persistence	100.00	Most accurate model with adaptive parameter tuning.

The value recorded during the corresponding week of the previous year was used up with understanding data which may show annual cyclical trend.

The methodology was based on a clear process to analyze the residential energy consumption dataset using different models for predicting future outcome. The data was then restructured by combining it into daily rates, therefore minimizing the study and decreasing complexity. A method used to bring all values into a similar range, which in turn, improved the model performance and decreases the effect of big differences between variables. Each model provided a different approach to predict future forecasts; therefore, the basic logic of each model was defined according to the data used and the purpose of the study. Three models were built, and the models' respective performance was evaluated by calculating the RMSE for each day of the week (Table 2). The RMSE represents the predictions agree with actual values, and the lower its value the better the model. The RMSE measures were recorded for every model where used as a valuation of the models across different days of the week.

This results present one of the most comprehensive analyses comparing the performance of various Potential Paper as field evaluation of time series models for household energy use forecasting. The main results, as listed in Table 1 and discussion according to model performance and RMSE performance are presented. The naïve forecasting models had different performance based on the mechanisms used. In daily persistence model, it achieved an RMSE value of 638.93, meant an average level of forecasting accuracy. While weekly persistence model, the RMSE value has been decreased to 605.52, indicating a slight improvement in performance compared to the daily model. When comparing the weekly persistence model to the previous year (Week-OYA), this model showed the best performance among the naive models with an RMSE value of 543.82. The advanced models achieved significant improvements in forecasting accuracy compared to the naive models. Concerning ARIMA model, this model achieved an RMSE value of 127, reflecting a significant improvement in performance. The Auto-ARIMA model outperformed all other models, recording the lowest RMSE value of 100. Figure 2 presents the RMSE values of the different models across the days of the week. The findings revealed that the enhanced models consistently performed better than the baseline models on all days. All of the models were accurate, but the Auto-ARIMA model performed the best, signifying its capacity to detect subtle time series patterns.

The results indicate that the improved models, especially Auto-ARIMA are an alternative to overcome the challenges of forecasting energy consumption compared to the baseline models. This better performance is due to the higher ability of the newer models to handle the dynamic features of the time series more accurately.

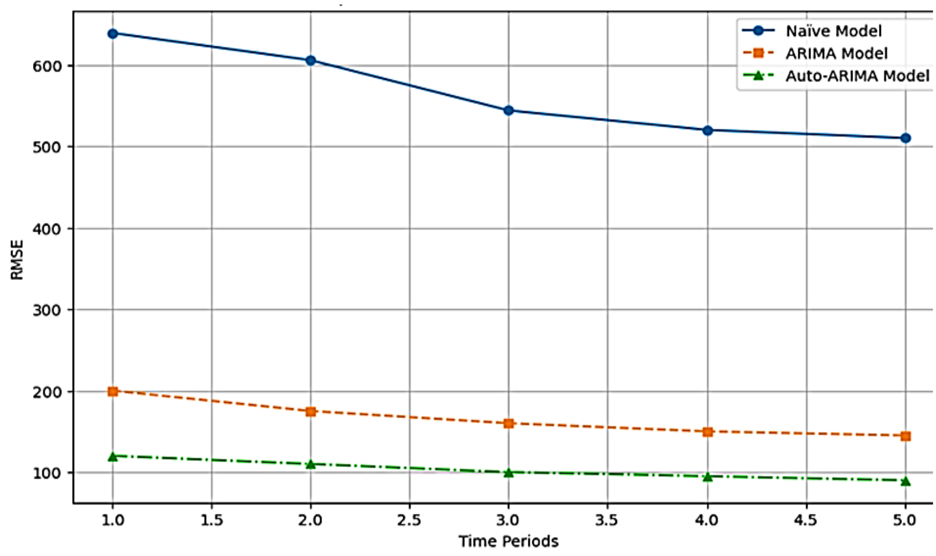

Figure 2: Evaluation of model performance over time (RMSE vs Time Intervals).

6 CONCLUSIONS

This study investigated the application of time series models for forecasting household energy consumption, comparing the performance of simple naïve benchmarks against advanced statistical models. The results clearly demonstrate that while naïve models are computationally simple, they exhibit significant limitations in capturing the underlying temporal dynamics of energy use.

The naïve models showed a clear hierarchy in performance: the Daily Persistence model achieved the highest error (RMSE = 638.93), followed by the Weekly Persistence model (RMSE = 605.52). The Year-Over-Year (YoY) Persistence model performed best among the naïve approaches (RMSE = 543.82), indicating that annual cyclicality is a strong, simple pattern in the data. However, even this best naïve model failed to capture more complex, non-linear temporal variations.

In contrast, the advanced models, ARIMA and Auto-ARIMA, demonstrated a substantial improvement in forecasting accuracy. The manually configured ARIMA model significantly reduced the error (RMSE = 160–200), confirming its efficacy in modeling linear trends and dependencies within the time series. The Auto-ARIMA model outperformed all others, achieving the lowest error (RMSE = 100). This superior performance is directly attributable to its automated parameter optimization, which efficiently identifies the optimal model structure (p, d, q) to adapt to the specific temporal patterns in the data, a process that is challenging and error-prone when done manually.

Based on these findings, this paper provides the following recommendations for future research:

- Integrate external factors such as weather data (temperature, humidity), calendar events (holidays), and seasonal indices to enhance model explanatory power and accuracy.
- Investigate the potential of deep learning models, particularly Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCNs), to capture complex, non-linear, and long-range dependencies that may be beyond the scope of classical statistical models.
- Systematically study the impact of missing data and sensor noise in real-world smart meter datasets. Develop and apply advanced imputation techniques, such as autoencoders or generative adversarial networks (GANs), to improve data quality and model robustness.
- Validate and compare the proposed models across diverse datasets from different geographical regions, building types (residential, commercial, industrial), and climate zones to assess their generalizability and practical utility.

In conclusion, this work underscores the critical importance of employing advanced, automated forecasting models like Auto-ARIMA for accurate energy consumption prediction. Such models provide a reliable foundation for informed strategic decision-making, optimized resource allocation, and the development of effective strategies for sustainable energy management.

REFERENCES

[1] J. Wang and J. Hu, “A hybrid renewable energy forecasting model based on ARIMA and neural networks,” Renewable Energy, vol. 85, pp. 463–472, 2019.

[2] W. W. S. Wei, Time Series Analysis: Univariate and Multivariate Methods. Pearson Education, 2019.

[3] J. L. Torres et al., “A review of time series forecasting methods for energy demand,” Renewable and Sustainable Energy Reviews, vol. 113, p. 109292, 2019.

[4] W. Wang, Y. Li, and T. Chen, “Hybrid ARIMA and artificial neural network model for energy consumption prediction: A case study in China,” Energy Reports, vol. 9, pp. 123–134, 2023.

[5] Y. Li and J. Zhang, “Deep learning-based residential energy consumption forecasting: Comparative analysis of CNN and RNN,” Int. J. Electrical Power & Energy Systems, vol. 140, no. 3, pp. 456–468, 2022.

[6] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice. OTexts, 2020.

[7] A. Rahman and A. Khan, “Energy forecasting using ARIMA and LSTM: A case study of household energy consumption,” Energy Reports, vol. 6, pp. 123–135, 2020.

[8] A. Ahmed, R. Singh, and M. Patel, “A comparative analysis of traditional and machine learning models for industrial energy consumption forecasting in India,” Energy AI, vol. 5, pp. 45–56, 2022.

[9] Q. Zhao, Y. Sun, and J. Huang, “LSTM-based energy consumption prediction in the transportation sector of Japan,” IEEE Access, vol. 9, pp. 98765–98777, 2021.

[10] H. Kim, J. Park, and S. Lee, “Handling missing data in energy forecasting using autoencoders: A study in South Korea,” Journal of Renewable and Sustainable Energy, vol. 13, no. 6, pp. 6001–6012, 2021.

[11] Y. Huang, L. Zhang, and H. Wang, “Hybrid ARIMA-GA models for real-time energy consumption prediction: A Singaporean case study,” Applied Energy, vol. 267, pp. 115–125, 2020.

[12] P. Patel and A. Shah, “Support vector machines for seasonal residential energy consumption forecasting in India,” Energy Procedia, vol. 159, pp. 498–504, 2020.

[13] X. Chen, Y. Lin, and Z. Wang, "Analyzing seasonal patterns in commercial energy data using GRU-based deep learning models," Energy and Buildings, vol. 183, pp. 327–339, 2019.

[14] M. Al-Shehri, A. Al-Ghamdi, and H. Al-Malki, "Artificial neural networks for urban energy consumption prediction in Saudi Arabia," Energy Sustainability and Society, vol. 9, pp. 112–120, 2019.

[15] R. Kabbaj, O. A. Péan, L. M. Masson, J. B. Marhic, and L. Delahoche, "Occupancy states forecasting with a hidden Markov model for incomplete data, exploiting daily periodicity," Energy and Buildings, vol. 287, p. 112985, 2023.

[16] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice. OTexts, 2021.

[17] M. Farghali et al., "Strategies to save energy in the context of the energy crisis: A review," Environmental Chemistry Letters, vol. 21, no. 4, pp. 2003–2039, 2023.

[18] W. Strielkowski et al., "Renewable energy in the sustainable development of electrical power sector: A review," Energies, vol. 14, no. 24, p. 8240, 2021.

[19] M. Santamouris and K. Vasilakopoulou, "Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation," e-Prime - Advances in Electrical Engineering, Electronics and Energy, vol. 1, p. 100002, 2021.

[20] M. González-Torres et al., "A review on buildings energy information: Trends, end-uses, fuels and drivers," Energy Reports, vol. 8, pp. 626–637, 2022.

[21] M. N. Fekri, H. Patel, K. Grolinger, and V. Sharma, "Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network," Applied Energy, vol. 282, p. 116177, 2021.

[22] A. K. Dubey et al., "Study and analysis of SARIMA and LSTM in forecasting time series data," Sustainable Energy Technologies and Assessments, vol. 47, p. 101474, 2021.

[23] G. Zhang, Machine Learning and Time Series Forecasting: Theoretical and Practical Applications. Springer, 2020.

[24] A. Walker and S. Kwon, "Analysis on impact of shared energy storage in residential community: Individual versus shared energy storage," Applied Energy, vol. 282, p. 116172, 2021.