Received: 7 August 2025 | Revised: 20 March 2026 | Accepted: 23 June 2026 | Published online: 14 July 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–16
DOI: 10.47852/bonviewAIA62027129

BON VIEW PUBLISHING

Optimizing Electricity Demand Forecasting Using ARIMA, SARIMA, and GRU with Weather and Calendar Variables

Emrah Aslan[1,2], Yıldırım Özüpak[3,*], Feyyaz Alpsalaz[4] and Hasan Uzel[4]

[1] *Department of Computer Engineering, Dicle University, Turkey*

[2] *Department of Computer Engineering, Mardin Artuklu University, Turkey*

[3] *Department of Electricity and Energy, Dicle University, Turkey*

[4] *Artificial Intelligence and Machine Learning Department, Amasya University, Turkey*

Abstract: Electricity demand forecasting is crucial for effective grid operation, planning, and decision-making. This study presents a comparison between classical time series models—AutoRegressive Integrated Moving Average (ARIMA) and Seasonal AutoRegressive Integrated Moving Average (SARIMA)—and a deep learning–based method, Gated Recurrent Units (GRU), for forecasting hourly electricity demand. The models are tested on a real-world dataset, enriched with weather and calendar variables to capture temporal and exogenous effects on electricity consumption. To ensure a fair and reproducible comparison, all models are trained and evaluated in a common experimental framework, including a well-defined chronological train-test split and rolling-origin (walk-forward) validation strategy. The forecasting performance is evaluated for short-term (24 h) and medium-term (168 h) horizons using standard error metrics, namely, root mean squared error, mean absolute error, and Mean Absolute Percentage Error (MAPE). The results of the experiment demonstrate that the GRU model performs better than ARIMA and SARIMA models especially for longer forecasting horizons due to its capability to learn nonlinear relationships and long-term temporal dependencies. The GRU approach gives better forecasting accuracy in the case of complex demand dynamics, but linear seasonal patterns can still be modeled by classical statistical models. The aim of this study is not to directly detect or predict system failures, nor does it depend on explicit fault or outage data. Its main contribution is instead in improving the accuracy of electricity demand forecasting, which can indirectly assist preventive grid operation and planning by reducing the uncertainty in expected load profiles.

Keywords: electricity demand, SARIMA, GRU, seasonal patterns

1. Introduction

Electricity demand forecasting is a crucial aspect of sustainable energy management, playing a significant role in the optimization of grid operations [1], the efficient use of resources, and the promotion of environmental sustainability [2]. The global increase in energy demand, together with the pressing necessity to decrease dependence on fossil fuels and incorporate renewable energy sources [3], has highlighted the importance of accurate and timely predictions to ensure grid stability and avoid fault-induced interruptions [4]. In this context, accurate forecasting models are increasingly important for balancing energy supply and demand, improving grid flexibility [5], supporting demand-side management strategies, and informing energy policy development [6]. Electricity demand has seasonal variations and nonlinear behaviors, which cannot be described by the traditional forecasting methods [7]. Recent advancements in statistical time series models (e.g., ARIMA, SARIMA) and deep learning methods (e.g., Gated Recurrent Unit [GRU]) have shown great potential to tackle these problems [8]. In particular, models that use high-frequency data [9], for example, hourly data [10], have the potential to capture complex patterns such as daily and weekly cycles and improve energy management and fault mitigation strategies [11].

The present research performs a thorough evaluation of time series forecasting methods for national electricity demand (nat_demand) using a dataset of 48,046 hourly entries of electricity demand, meteorological parameters (temperature, humidity, precipitation, and wind speed) for several locations (Tocumen, San Miguelito, and David), and calendar elements such as holidays and school periods. The aim of this work is to provide a comprehensive evaluation of time series forecasting methods for national electricity demand (nat_demand). The dataset used has 48,046 hourly records with electricity demand, weather (temperature,

***Corresponding author:** Yıldırım Özüpak, Department of Electricity and Energy, Dicle University, Turkey. Email: yildirim.ozupak@dicle.edu.tr

humidity, precipitation, and wind speed) from different locations (Tocumen, San Miguelito, and David), and calendar-related data (holidays and school periods). The research involves three different models, which are ARIMA, SARIMA, and GRU. These models are used to capture linear, seasonal, and nonlinear patterns, respectively. The models' performances are evaluated using a combination of statistical metrics (e.g., root mean squared error [RMSE], MAPE) and visual diagnostics (e.g., residual plots, autocorrelation function [ACF], and partial autocorrelation function [PACF]), supplemented by decomposition analyses to uncover underlying trends and seasonal components. A key finding is the better performance of the GRU model with an RMSE of 127, indicating its capability in modeling the nonlinear dynamics and abrupt changes typical of energy demand fluctuations and fault events. The study, however, recognizes practical constraints in real-world applications, including the difficulty of incorporating real-time fault data and the computational intensity of deep learning models such as GRU in resource-limited operational settings. This study contributes to the literature in many ways. It should be noted that the objective of the current study is not to detect, classify, or predict system failures directly. Instead, the contribution of the proposed framework is in enhancing the accuracy of short- and medium-term electricity demand forecasting. Accurate demand forecasts can indirectly assist preventive grid operation, maintenance scheduling, and decision-making processes by reducing uncertainty in expectations of loads.

1) Evaluation of Model Effectiveness: The study systematically assesses statistical and deep learning models (ARIMA, SARIMA, GRU) for capturing seasonal and nonlinear patterns in high-frequency electricity demand data.
2) Analysis of Model Applicability: It examines the strengths and limitations of ARIMA, SARIMA, and GRU, providing insights into their suitability for short- and long-term forecasting.
3) Impact of Exogenous Variables: The research investigates the influence of weather and calendar events on demand forecasting, offering a framework to optimize model performance amidst system faults.
4) Comparative Performance Insights: It delivers a comparative analysis of model performance using metrics, autocorrelation, decomposition, and residual diagnostics, yielding actionable recommendations for energy system operators.

The study contributes to the literature by providing a comprehensive analysis of statistical and deep learning models in a unified framework and by highlighting their practical implications for fault management and energy demand forecasting. It examines the effect of exogenous variables on the improvement of the accuracy of predictions, provides ideas for the application of the model for different forecasting horizons, and reinforces its conclusions with thorough diagnostics and comparative analyses. The outcomes of this research not only contribute to the theoretical and operational aspects of electricity demand forecasting but also improve the resilience of power grids and the effects of faults on them, opening the door to future hybrid modeling methods that can further improve sustainable energy management.

2. Literature Review

The optimization and sustainability of energy systems require electricity load forecasting as a basic requirement. ARIMA and SARIMA models have been widely used in the literature in time series analysis, and they are efficient in modelling the linear and seasonal patterns. On the other hand, deep learning methods such as GRU are very good at capturing complex nonlinear time dependencies [12]. The above models performed differently depending on the data structure and the application domain. In this context, our study aims to contribute to the literature by comparative analysis of SARIMA, ARIMA, and GRU.

Özen et al. [13] investigated deep learning methods for forecasting electricity consumption. They examined Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), ARIMA, and random forest on univariate data. For multivariate data, the hybrid models outperformed other models. The hybrid fusion model was better than ARIMA (0.2184) and CNN (0.1802) with an RMSE of 0.0871 in Chicago. Short-term load forecasting (STLF) is crucial for the supply–demand balance and power system stability. Anh et al. [14] devised techniques to process the data stream with real-time smart sensors. The proposed online SARIMA variant was used to forecast the Northern Vietnam hourly electricity load and obtained a MAPE of 4.57%. Amjadian et al. [15] discussed the importance of long-term electric load forecasting and the limitations of ARIMA in nonlinear trends. A new method was proposed using ARIMA, Support Vector Regression (SVR), and genetic algorithm. The prediction of TREC's annual peak load yielded reliable results with error criteria. Ullah et al. [16] evaluated time series, regression, neural networks, and hybrid models for the important role of STLF in reserve management in modern power grids. The proposed CNN-LSTM model resulted in better accuracy with 538.71 RMSE and 2.72 MAPE for one-step forecasting and 951.94 RMSE and 4.72 MAPE for 24-h forecasting on NTDC data. Ahamed et al. [17] worked on load forecasting techniques for the stability of the power grid and analyzed the influence of factors such as the generalization of algorithms and the complexity of models on efficiency. Discussion on the progress of STLF and the strengths, weaknesses, and future trends of STLF was performed.

Dorji et al. [18] explored the importance of electrical load forecasting for the management of power systems and the utilization of synthetic data to fill the missing data (MV). The synthetic data and hybrid method were compared with three datasets, generated with real data and linear interpolation. Fan et al. [19] pointed out the disadvantages of traditional models for complex fluctuations and emphasized the importance of accurate power load prediction for energy distribution and grid stability. A hybrid model of EEMD-SVR-WNN was proposed. The load series was decomposed by Ensemble Empirical Mode Decomposition (EEMD) and forecasted by SVR and WNN. High accuracies were achieved on New York data. El-Afifi et al. [20] studied the environmental advantages of electric vehicles (EVs) and the capacity challenges they pose to the grid. They proposed a deep learning–based two-stage optimization strategy for efficient placement of EV charging stations. Load forecasting was performed using an ARIMA-LSTM hybrid model. Losses were reduced by 33% using the Hybrid Archimedes-Genetic (HAG) algorithm. In their study on the impact of EV transition on grid management, Amara-Ouali et al. [21] developed probabilistic forecasting algorithms to predict charging behavior. Bottom-up and adaptive approaches were proposed directly. They were tested on the data of the Palo Alto dataset and were successful on the pinball loss and Ranked Probability Score (RPS) metrics. Kychkin and Chasparis [22] studied the complexity of the electricity load consumption and compared the short-term (daily/hourly) forecasting models. Autoregressive and deep learning models were presented. Persistence, Periodic AutoRegressive (PAR), and Seasonal Persistence Regression (SPR) models were proposed. Accuracy improvement in hourly forecasts was obtained up to 15–30%. Kyryk and Shatalov [23] compared SARIMA(X) and Decision Tree in their study of load forecasting models in the power grid. The models were trained on the 2021 load data of one substation in the Kyiv region. We tested the models on how they performed with forecasts for 01/01/2022.

Parkash et al. [24] have proposed an end-to-end (E2E) learning approach based on the top-down (TD) principle for STLF for hierarchical residential consumers. The model was tested with 9 years of data from 50 homes in the USA and outperformed SARIMA and SVR with a MAPE of 2.27%. Doddabasappa et al. [25] studied the effect of electricity price forecasting on Market Clearing Price (MCP) and the role of day-ahead forecasts in market and power plant planning. In terms of pen unit cost and test accuracy, benchmarking was used. MATLAB was used to forecast prices in the medium term. Bensalah and Hair [26] used the SARIMA model to study the power consumption forecasting of Tetouan city. The model captured seasonal patterns using 10-min data. SARIMA was efficient in smart grid management using training data, and its accuracy was calculated by mean absolute error (MAE), RMSE, MAPE, and R2. Kumar et al. [27] studied ARMA, ARIMA, and SARIMA models for long-term load forecasting with data from the mixed grid of NIT Patna. SARIMA was more accurate than ARMA and ARIMA due to considering seasonal changes. The efficiency of the model was illustrated by the achievement of 7% MAPE. Ningombam et al. [28] proposed a Box-Cox integrated sARIMA model for inertia estimation for low-inertia modern power systems from renewable energy. The inertial effect of wind generators and engine loads was considered. The model obtains good results with an annual prediction accuracy higher than 97.5%. Royal et al. [29] studied the importance of electricity demand forecasting for optimal system design. He proposed the Generalized Additive Model (GAM) and SARIMA hybrid model for long-term forecasting. The model was validated on 11 years of data and was able to capture seasonality and growth rates effectively with an Normalized Root Mean Squared Error (NRMSE) of 9.091%. This supports multi-year investments. Zhao et al. [30] proposed a multimodal fusion approach for renewable energy demand forecasting by fusing time series and text data using a hybrid model of CNN and Bi-GRU. The GRU, ARIMA, and EEMD-ARIMA models performed worse than the proposed model for short- and long-term forecasting. Heng [31] suggested the MTL-GAN model for multi-energy load forecasting (MELF) to account for energy demand diversity. GRU was employed to extract temporal dependencies, a shared layer for inter-load relations, and an attention module for spatial features. In ASU Tempe data, MAPE decreased by 43% and RMSE by 51%. Guo [32] studied the efficiency of power load forecasting and proposed the BO-GRU-LightGBM model to optimize the integration of renewable energy. The parameters of GRU and LightGBM were obtained using Bayesian optimization and were combined with Least Absolute Deviations (LAD). Simulation results in MATLAB were obtained with high accuracy, with an RMSE of 0.331. Chen et al. [33] considered the importance of MELF for integrated energy systems and proposed the DAM-MSHRN-GRU model. GRU performed well, as it captured time patterns.

3. Material and Method

In this study, we predict the national electricity demand (nat_demand) based on a dataset of 48,046 hourly records without missing values, as shown in Table 1. The dataset includes variables such as electricity demand, temperature, humidity, precipitation, and wind speed in different locations (Tocumen, San Miguelito, and David), and calendar-related features such as holidays and school periods. To capture the temporal dynamics and seasonal patterns embedded in the data, three different forecasting approaches were used: ARIMA, SARIMA, and GRU. Each method was systematically applied to capture linear, seasonal, and nonlinear dependencies, respectively, with performance evaluated using metrics such as RMSE and MAPE. The following

Table 1
Overview of dataset variables and temporal coverage

Variable	Description	Count	Mean	Std. dev.	Min	Max	Quantiles (25%–50%–75%)
DateTime	Timestamp of records	48,046	30-Sep-17	–	03-Jan-15	27-Jun-20	–
nat_demand	National electricity demand (MW)	48,046	1180	192	85.2	1750	1,020–1,170–1,330
T2M_toc	Temperature at Tocumen (°C)	48,046	27.4	1.68	23	35	26.2–27.1–28.6
QV2M_toc	Specific humidity at Tocumen (g/kg)	48,046	0.02	0	0.01	0.02	0.02–0.02–0.02
TQL_toc	Total precipitation at Tocumen (mm)	48,046	0.08	0.07	0	0.52	0.03–0.07–0.12
W2M_toc	Wind speed at Tocumen (m/s)	48,046	13.4	7.3	0.01	39.2	7.55–12.2–18.7
T2M_san	Temperature at San Miguelito (°C)	48,046	26.9	3.02	19.8	39.1	24.8–26.2–28.7
QV2M_san	Specific humidity at San Miguelito (g/kg)	48,046	0.02	0	0.01	0.02	0.02–0.02–0.02
TQL_san	Total precipitation at San Miguelito (mm)	48,046	0.11	0.09	0	0.48	0.04–0.09–0.16
W2M_san	Wind speed at San Miguelito (m/s)	48,046	7.05	4.1	0.06	24.5	3.96–5.99–9.41

subsections detail the data preprocessing steps, the model specifications, and the evaluation procedures.

3.1. Dataset

Accurate electric load forecasting plays an important role in efficient power system management and in balancing electricity supply and demand. The dataset used in this study provides a comprehensive basis for developing and testing forecasting algorithms because it includes historical electricity demand, official weekly pre-dispatch forecasts, weather variables, and calendar-related information. In the original dataset, electricity demand is recorded at hourly resolution, which allows detailed temporal analysis of short-term and medium-term demand variations. The dataset consists of 48,046 valid hourly records with no missing or mismatching values and covers the period from January 01, 2015, to June 27, 2020. It includes national electricity demand (nat_demand) and meteorological variables such as temperature, relative humidity, precipitation, and wind speed for Tocumen, San Miguelito, and David, as well as calendar variables including holidays and school periods. Historical load data were obtained from the grid operator CND through daily post-dispatch reports, while weekly forecast information was extracted from official pre-dispatch reports. School-period information was collected from the Ministry of Education in Panama, and holiday information was obtained from the "When on Earth?" database.

To preserve temporal consistency and avoid data leakage, the dataset was divided chronologically without shuffling. The training period covers January 01, 2015, to December 31, 2018, and the test period covers January 01, 2019, to June 27, 2020. The original dataset includes four files, namely, continuous dataset.csv, weekly pre-dispatch forecast.csv, train_dataframes.xlsx, and test_dataframes.xlsx, with a total of 355 columns. Since the dataset also contains weekly pre-dispatch forecasts, the weekly forecasting structure was considered in the statistical benchmark models. In this process, each forecast week begins on Saturday, and Friday is considered the last day of the previous forecast week. In addition, 72 h of unseen records, up to the last hour of Tuesday, are kept before the forecast period in accordance with the weekly pre-dispatch process.

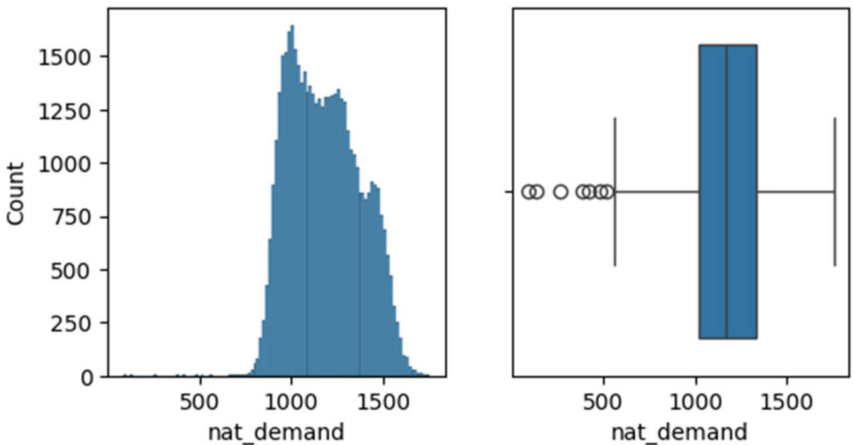
Although the complete dataset contains 48,046 hourly observations, the ARIMA and SARIMA models were fitted using a weekly aggregated benchmark series derived from the same hourly dataset. This aggregation resulted in 218 observations for the period from January 2015 to March 2019. This choice was made for two main reasons. First, it is consistent with the weekly pre-dispatch forecasting structure of the dataset. Second, direct estimation of ARIMA and SARIMA models on approximately 35,000 hourly training observations with long seasonal cycles would be computationally demanding and could lead to convergence instability. Therefore, ARIMA and SARIMA were used as classical statistical benchmark models based on the weekly aggregated representation of the same underlying demand series. In contrast, the GRU model was trained directly on the hourly-resolution data using a 168-h sliding-window structure in order to capture daily and weekly temporal dependencies. Thus, the difference in the number of observations does not indicate the use of different datasets; rather, it reflects model-specific preprocessing and input resolution required for computationally feasible and methodologically appropriate forecasting.

Figure 1 shows the distribution of demand for national electricity (nat_demand) in the form of a histogram and boxplot. The histogram shows a right-skewed distribution with a peak at about 1100–1200 MW and a range of 500–1750 MW, indicating that lower values of demand occur more often. This skewness is confirmed by the boxplot with a median near 1170 MW and an interquartile range (IQR) between 1020 MW and 1330 MW. Outliers below 500 MW suggest occasional low-demand events possibly caused by anomalies. This distribution enables the use of SARIMA to capture seasonal variations and GRU to model nonlinear patterns in demand data.

3.2. ARIMA model

ARIMA model is a statistical method widely used for predicting future values in a time series based on the model of dependencies between the consecutive observations [34]. It is especially effective for analyzing and forecasting time series data with trends and patterns, as long as the series can be made stationary. The ARIMA model combines three basic components, each of which deals with various aspects of time series behavior, and is represented as ARIMA (p, d, q), where p, d, and q are non-negative integers indicating the order of the model. The first component is AutoRegressive (AR), which predicts the current values based on past observations. The parameter p represents the number of lagged observations included in the model. For instance, an AR (1) model takes the immediately previous value into consideration to capture the AR relationship in the series.

Figure 1
Histogram and boxplot of national electricity demand distribution



The second part, Integrated (I), is responsible for stationarity, which is performed by differentiating the series, and d represents the number of times this process has been performed. Differencing is used to eliminate trends and to stabilize the mean of the series, which is a prerequisite for ARIMA modeling. The third component, the Moving Average (MA), includes past forecast errors to improve forecasts, and q is the number of lagged error terms. This takes account of the influence of past deviations of observed values from predicted values. In practice, ARIMA (p, d, q) combines these components to model both AR and MA dynamics and to account for non-stationarity through differencing. For example, an ARIMA (1, 1, 1) model includes one lag of the series, one step of differencing, and one lagged error term. Its strength is its flexibility to adapt to various time series models by adjusting p, d, and q, making it a fundamental tool for time series forecasting in areas such as energy load forecasting where capturing temporal dependencies is critical [35].

3.3. SARIMA model

The SARIMA model is an extension of ARIMA that can be used for time series data with seasonal patterns, so it can be used to forecast phenomena with recurring cycles, such as electric charge [36]. SARIMA is an extension of the ARIMA structure that adds seasonal components to capture both non-seasonal and seasonal dependencies and improve forecasting ability for datasets exhibiting periodicity. The model is written as SARIMA (p, d, q) (P, D, Q)m with lowercase parameters (p, d, q) denoting non-seasonal components, uppercase parameters (P, D, Q) denoting seasonal components, and m denoting the seasonal period such as 24 for daily cycles in hourly data or 12 for monthly seasonality. The non-seasonal components are similar to those of ARIMA: AR with parameter p, where the current values are predicted based on previous observations that indicate the short-term dependencies; I with parameter d, where the differencing is performed to obtain the stationarity by removing the trends; and MA with parameter q, where the previous forecast errors are used to correct the forecast errors due to previous mistakes. These ingredients are for the immediate (temporal) structure of the series. This framework is supported by seasonal elements: Seasonal AR (P) models the repeating patterns over periods by adding AR terms at seasonal lags (e.g., values from previous seasons); Seasonal I (D) uses seasonal level differencing to eliminate seasonal variations; and Seasonal MA (Q) uses lagged seasonal errors to improve forecasts by correcting seasonal biases [37]. The parameter m gives the length of the seasonal cycle, so that the model reproduces the periodic structure of the data.

3.4. GRU model

The GRU is a kind of recurrent neural network to model sequential data and capture long-term dependencies efficiently, which is a powerful tool for time series forecasting such as electrical charge forecasting [38]. GRU provides a less complex alternative than LSTM units, reducing the computational complexity without losing the ability to learn temporal patterns. The architecture uses two main gates—a reset gate and an update gate—that determine how the model handles and remembers information across time, reacting to both short-term and long-term patterns in sequential data. The reset gate determines how much the model should forget past information. This allows the model to ignore noise or irrelevant trends and focus on relevant historical data. It provides flexibility to capture changing patterns by adjusting the effect of the previous latent state. The update gate** determines the contribution of past and present information, deciding how much of the previous latent state to carry forward versus incorporating new input. This mechanism allows to save memory over long sequences for GRU without burdening computational demands [39].

3.4.1. GRU model configuration and training protocol

We constructed the GRU model in a supervised learning framework using a sliding-window approach. This subsection is explicitly added to guarantee a full reproducibility of the GRU-based experiments and to allow a fair comparison with classical statistical models. The input window was chosen to be of length 168 h to explicitly account for the daily (24 h) and weekly (168 h) seasonality in the data. All numerical input variables were normalized using Min–Max scaling, computed only on the training set to prevent data leakage.

The network architecture is composed of two stacked GRU layers with 64 and 32 hidden units, respectively. The output layer was a fully connected dense layer with linear activation. The optimizer used for optimization of the model was Adam with a learning rate of 0.001. The loss function was mean squared error (MSE).

Training was done with a batch size of 32 for a maximum of 100 epochs. To prevent overfitting, early stopping was employed with a patience of 10 epochs, monitored on a validation subset of 20% of the training data. The best GRU model was chosen as the one with the lowest validation loss. All hyperparameters reported above were fixed before evaluation and were not tuned on the test set.

3.5. Autocorrelation function

The plot of the ACF helps us to discover some key features of the time series we are analyzing. The peaks and troughs are cyclical, suggesting a strong seasonal pattern. This suggests a strong seasonal component. Such regular oscillations suggest a cyclic repeating pattern in the data, such as daily or monthly cycles. Furthermore, the decreasing trend in ACF values with increasing lags suggests that the series is not merely white noise but rather exhibits a persistent structure with interdependencies over several time steps. This decreasing trend suggests the existence of trends or AR patterns, requiring differencing or seasonal adjustments in models such as ARIMA or SARIMA to attain stationarity and appropriately capture the underlying dynamics [40].

3.6. Partial autocorrelation function

The PACF plot helps to understand the direct relationships in the time series. If there is a quick decay after some lags, the series is probably well described by an AR process where only a few past values directly influence the current observation. The first lag is significant, indicating that the series is strongly related to the immediate past, suggesting an AR(1) process, where the current value is strongly influenced by the previous one. Furthermore, seasonal lags with sporadic large spikes point toward a seasonal component, which implies that in addition to the immediate AR effect, there are recurrent patterns at specific intervals (e.g., every 12th lag for monthly data) that influence the series. This allows a SARIMA model to account for both short-term and seasonal AR effects [41].

3.7. Evaluation metrics

In order to ensure a comprehensive analysis of prediction accuracy, a number of statistical evaluation metrics are used to evaluate the performance of the SARIMA, ARIMA, and GRU models [42]. These metrics are the R^2 value that indicates the proportion of variance explained by the model, the MAE that gives the average size of the errors in the predictions, and the MSE that places more weight on larger errors through the squared differences between the true and predicted values. In addition, the RMSE is used to present an interpretable measure of prediction error in the same units as the original data. Together, these metrics offer a strong comparison of how well the models perform, showing their strengths and weaknesses in capturing seasonal and nonlinear patterns in energy consumption data [43].

$$MAE = \frac{\sum_{j=1}^{N} |y_i - x_i|}{n} \tag{1}$$

$$RMSE = \left[\sum_{j=1}^{N} \left(d_{fi} - d_d\right)^2 / N \right]^{\frac{1}{2}} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{3}$$

Forecasting performance was evaluated over two horizons: 24-h ahead and 168-h ahead forecasts, corresponding to one-day and one-week prediction intervals. RMSE, MAE, and MAPE values were computed over these horizons to ensure a consistent and fair comparison across models.

3.8. Forecasting horizon and evaluation protocol

Predictions were generated for two lead times: 24 h ahead and 168 h ahead, that is, one-day and one-week-ahead predictions at hourly resolution. To make the comparison between ARIMA, SARIMA, and GRU models equal, a rolling-origin (walk-forward) validation strategy was used. The models were trained at each step on all historical data available up to t and tested on the following forecasting horizon. This process avoids leakage of information and reflects real-world operational conditions of forecasting.

For fairness, the models were compared on aligned chronological evaluation periods using the same error metrics. However, the input resolution differed between model families. The GRU model used the full hourly-resolution training data with a supervised sliding-window representation, whereas ARIMA and SARIMA used the weekly aggregated representation of the same underlying electricity demand series. This design allowed the statistical models to remain computationally tractable while enabling the GRU model to exploit high-frequency hourly information. Accordingly, the comparison should be interpreted as a predictive performance comparison under model-appropriate preprocessing rather than as a direct comparison of identical input sample sizes.

4. Experimental Results

This section shows the results of using ARIMA, SARIMA, and GRU models to predict national electricity demand (nat_demand) from a dataset of 48,046 hourly records. The performance of each model was evaluated using a combination of visual inspections and quantitative metrics, such as RMSE and MAPE, to assess their ability to capture the temporal, seasonal, and nonlinear dynamics of the data. The results demonstrate the comparative effectiveness of each method and include detailed diagnostics of residuals, prediction accuracy, and model fit in the following subsections, which are accompanied by figures and tables to facilitate a complete evaluation.

4.1. Analysis of electricity demand dynamics

Figure 2 shows the national electricity demand (nat_demand) from January 2015 to June 2020. The demand fluctuates between 800 MW and 1600 MW, exhibiting periodic oscillations that suggest a strong seasonal component and sharp drops that may indicate anomalies like power outages. This supports the use of SARIMA for seasonal effects and GRU for nonlinear dependencies.

The time series of the nat_demand is decomposed into the components of trend, seasonal, and residuals, as shown in

Figure 2
Time series representation of national electricity demand

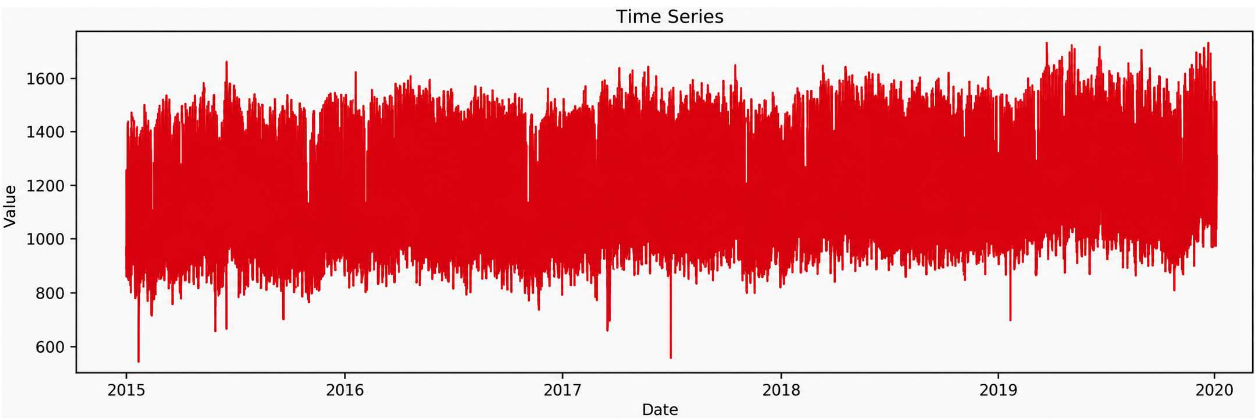


Figure 3. The trend increases from about 1100 MW to 1300 MW, the seasonal component oscillates between −250 MW and 250 MW, and the residuals contain outliers, which suggests that ARIMA is suitable for a linear trend while SARIMA is suitable for seasonality.

The residual is shown in Figure 4, varying from −500 MW to 500 MW. The distribution is largely random, and outliers are substantial. This demonstrates the capability of GRU to capture intricate nonlinear patterns that are not entirely addressed by ARIMA/SARIMA, leading to improved forecasting robustness.

Figure 5 shows the hourly national electricity demand (nat_demand) of 24 h. The seasonal effects are captured in the SARIMA model, the nonlinear dependencies are captured in the GRU model, and the linear effects are captured in the ARIMA model. The national electricity demand is 900 MW from 00:00 to 06:00, and the demand is at its maximum from 10:00 to 14:00,

Figure 3
Decomposition analysis of national electricity demand into trend, seasonal, and residual components

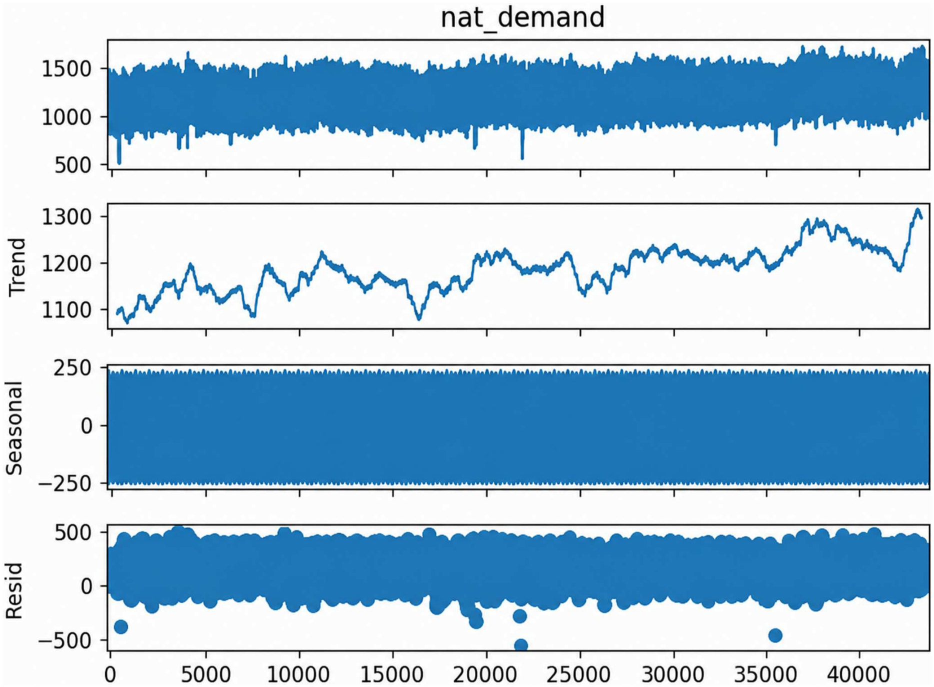


Figure 4
Residual component analysis of national electricity demand decomposition

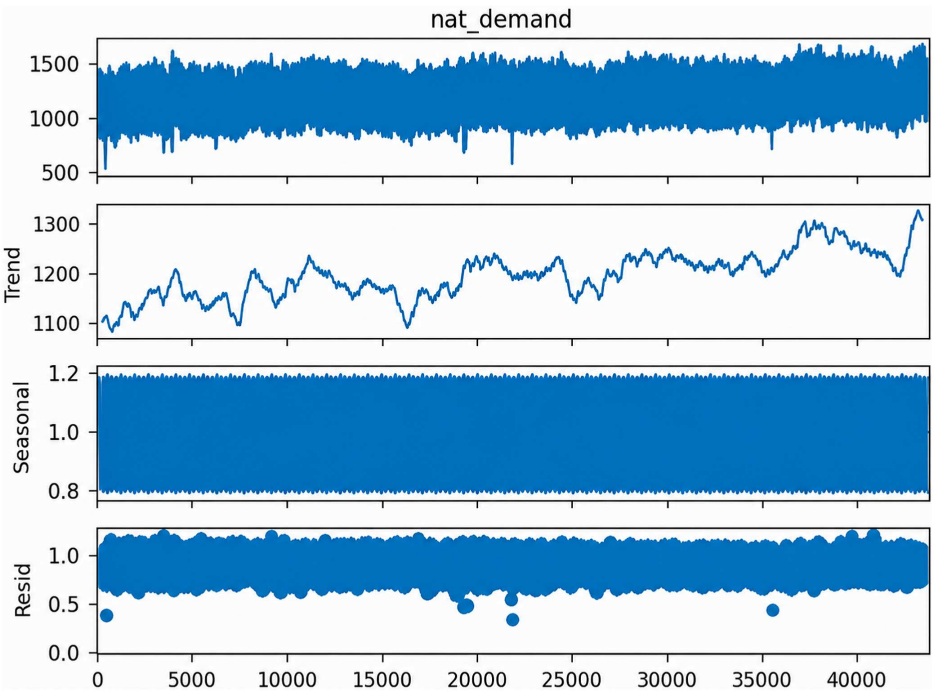
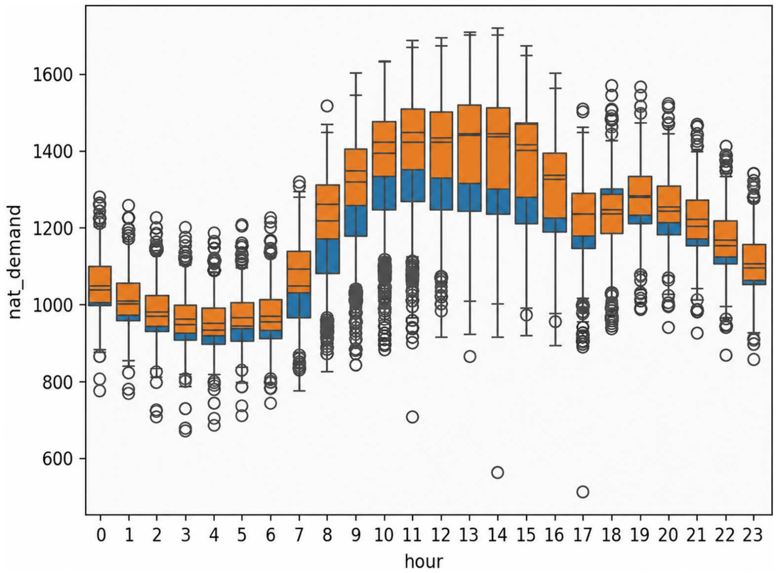
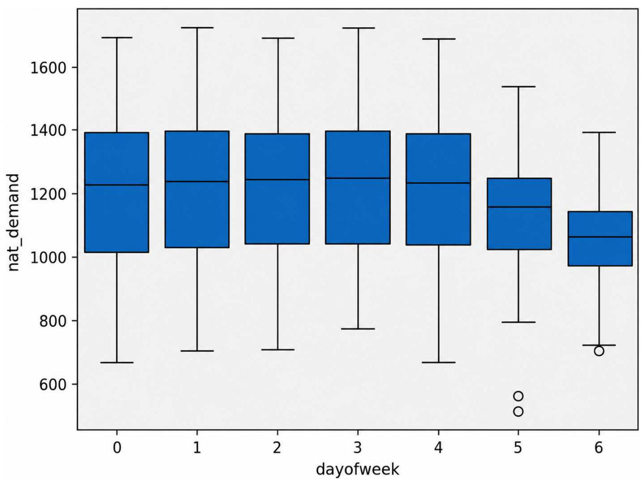


Figure 5
Analysis of hourly electricity demand patterns



ranging from 1400 to 1500 MW due to the industrial sector. By 16:00, the demand drops to 1200 MW, and from 18:00 to 20:00, the demand increases to 1300 MW due to the residential sector. From 21:00 to 23:00, the demand stabilizes at 1000 MW. The error bars show consistent variability. Figure 6 shows nat_demand over the week (Saturday–Friday). Median demand is consistent at 1200–1300 MW (Saturday–Thursday) with a small dip to 1100 MW on Friday. The IQRs and outliers indicate consistent variability. This supports ARIMA for linear trends and SARIMA for weekly seasonal effects.

Figure 6
Boxplot of national electricity demand across days of the week



4.2. Autocorrelation and partial autocorrelation analysis

Figure 7 shows ACF and PACF plots for the national electricity demand (nat_demand). The ACF plot exhibits a clear cyclical pattern with peaks approximately every 24 lags, suggesting a daily seasonal pattern in the hourly data. The slow decay in correlations points to persistent dependencies that signal the need to difference the series in ARIMA or SARIMA models to induce stationarity. On the other hand, the PACF plot shows a significant spike at lag 1, smaller spikes at seasonal lags (e.g., 24, 48), and almost zero correlation afterward. This suggests an AR(1) process with seasonal effects. This justifies the use of SARIMA to model short-term and seasonal AR components and GRU to capture any nonlinear dependencies.

Figure 8 shows the real (red) and predicted (green) electricity demand (nat_demand) for a short time period (35,000–36,000 time steps). The real demand varies between 800 MW and 1400 MW. The predicted demand follows the trend but is not able to capture some of the peaks (e.g., around 35400). This shows SARIMA and ARIMA model linear patterns efficiently, but GRU might be more suitable for such abrupt changes. Figure 9 shows the real (red) and predicted (green) electricity demand for a longer period (36,000–44,400 time steps). The real demand varies between 800 MW and 1600 MW, while the predicted demand remains constant at about 1200 MW, not accounting for peaks (e.g., around 42,000). This means SARIMA is good at capturing seasonal patterns, whereas GRU is better at capturing long-term complex dependencies.

The initial model performed poorly with an RMSE of 475 and a MAPE of 34%, suggesting large deviations from the true electricity demand values. The addition of trend and seasonality components helped to counter the constant decrease in predictions and thus produced a more stable forecast, but it still did not fit the true values very well. Later assessment produced a MAPE of 12% showing a 12% average difference from the true demand, indicating an increase in accuracy. The initial model was improved by adding a damped trend parameter and reducing the RMSE to 173. The best result was obtained by adding multiplicative trend and seasonality components over 3 days with an RMSE of 127. This is a significant improvement over the previous model (RMSE: 173) showing gradual improvement in the prediction accuracy.

Figure 7
ACF and PACF plots of national electricity demand

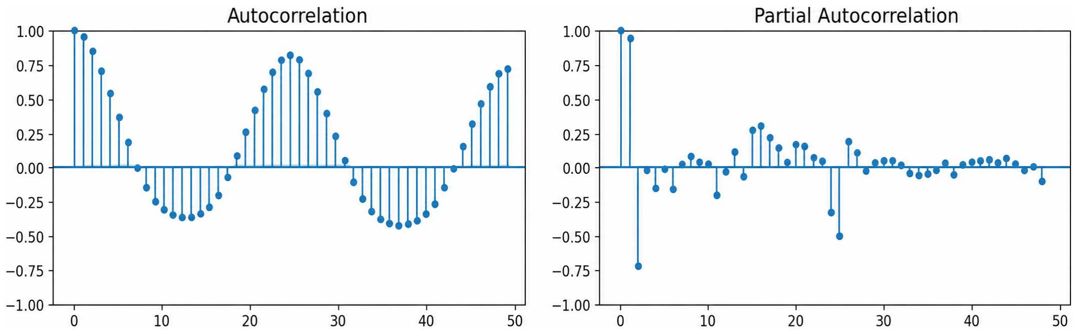


Figure 8
Comparison of real and predicted electricity demand (short-term)

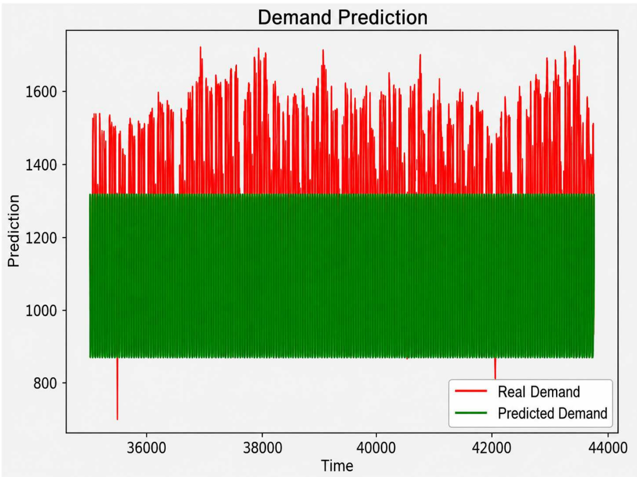
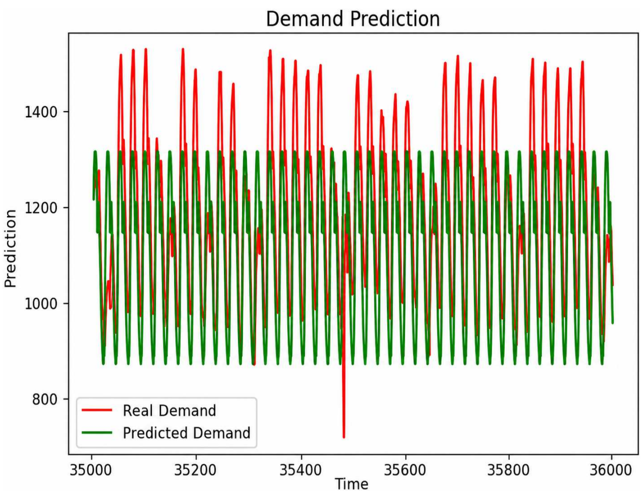


Figure 9
Comparison of real and predicted electricity demand (long-term)



4.3. ARIMA model forecasting performance analysis

The ARIMA model is a statistical approach to time series forecasting, modeling temporal dependencies in linear, stationary data with its components: AR (p), which uses p lags of the observation; I (d), which uses d differencing to make the series stationary; and MA (q), which uses q lags of forecast errors. ARIMA (p, d, q). Figure 10 shows diagnostic plots for the residuals of the ARIMAX model fitted to the national electricity demand (nat_demand) data for 2015–2019. The standardized residuals (top-left plot) seem to fluctuate around zero with some spikes (e.g., in 2017), and the histogram (top-right plot) is right-skewed, which suggests a departure from normality. The Normal Q–Q plot (bottom-left plot) further confirms the non-normality in the tails, and the correlogram (bottom-right plot) indicates no significant autocorrelation. Overall, the model appears to be successful in capturing the linear dependencies despite the non-normality and outliers. The ARIMAX results with ARIMA(1, 0, 2) are reported in Table 2. The number of observations is 218 (January 18, 2015–March 17, 2019). The AR(1) coefficient (0.9756) and the MA coefficients (−0.4420, −0.3202) are significant ($P>|z| < 0.05$). The Ljung-Box test (Prob(Q): 0.98) shows that there is no residual autocorrelation. The Jarque-Bera test (Prob(JB): 0.00) and the kurtosis (4.04) show non-normality. It implies that the model has good performance on linear patterns but needs some adjustments to address residual non-normality. The 218 observations reported in Table 2 correspond to the weekly aggregated series obtained from the original hourly dataset, not to the total number of available hourly records. This aggregation was necessary because direct ARIMA fitting on the full hourly training set would require modeling long seasonal cycles and a very large number of observations, which substantially increases computational burden and may lead to convergence instability. Therefore, the ARIMA model was used as a classical weekly benchmark model, while the GRU model was trained on the full hourly-resolution data.

4.4. SARIMA model forecasting performance analysis

The SARIMA model is an extension of the ARIMA model for time series data with seasonal patterns and can capture both short-term and seasonal dependencies in the data. It has the usual ARIMA components: AR part represented by p, which uses p lagged observations to predict current values; I part represented by d, which uses differencing d times to make the time series stationary; and MMA part represented by q, which uses q lagged forecast errors. Also, SARIMA has seasonal components: Seasonal AR (P) for seasonal autoregressive terms, Seasonal I (D) for seasonal differencing, Seasonal MA (Q) for seasonal error adjustments, and the Seasonal Period (m), which specifies the number of observations per seasonal cycle (e.g., 24 for daily seasonality in hourly data). SARIMA is written as ARIMA (p, d, q) (P, D, Q)m, where p, d, and q are the non-seasonal terms,

Figure 10
Diagnostic plots for ARIMA model residuals

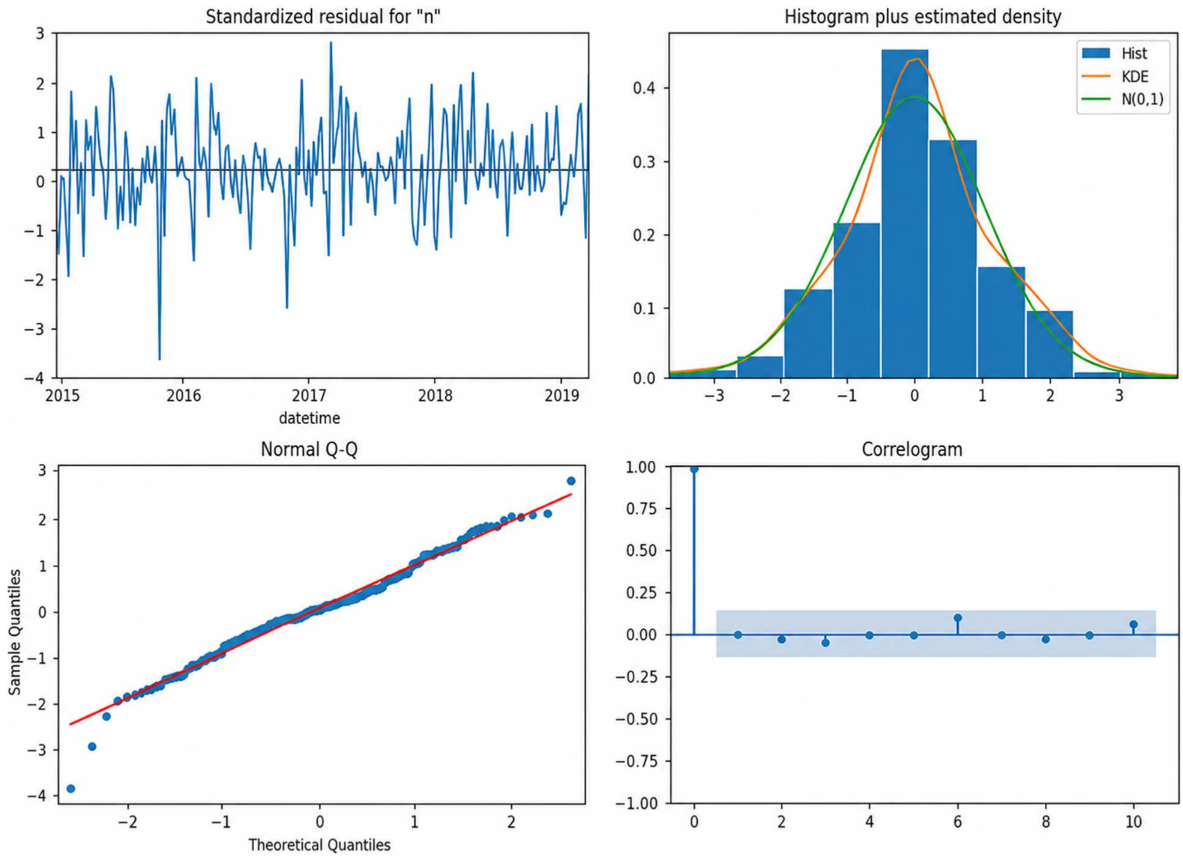


Table 2
Results of ARIMA model (ARIMA (1, 0, 2)) results for national electricity demand

Parameter	Value	Description
Dependent variable	nat_demand	National electricity demand
Model	ARIMA(1, 0, 2)	ARIMAX model with AR(1), I(0), and MA(2) components
No. observations	218	Total number of data points
Log likelihood	–2200.850	Log likelihood of the model
AIC	4411.700	Akaike Information Criterion
BIC	4428.622	Bayesian Information Criterion
HQIC	4418.535	Hannan-Quinn Information Criterion
Date	Thu, 20 Mar 2025	Date of model fitting
Time	14:53:44	Time of model fitting
Sample	01-18-2015 to 03-17-2019	Time range of the data
Covariance type	opg	Covariance type used for standard errors
Coefficients		
const	196,800	Constant term (std err: 3712.061, z: 53.026, P>
ar.L1	0.9756	AR(1) coefficient (std err: 0.022, z: 44.091, P>
ma.L1	–0.4420	MA(1) coefficient (std err: 0.067, z: -6.631, P>
ma.L2	–0.3202	MA(2) coefficient (std err: 0.069, z: -4.641, P>
sigma2	33,150,000	Variance of residuals (std err: 0.053, z: 6.29e + 08, P>

(*Continued*)

Table 2
(*Continued*)

Parameter	Value	Description
Diagnostics		
Ljung-Box (L1) (Q)	0.00	Ljung-Box test statistic for residual autocorrelation (Prob(Q): 0.98)
Jarque-Bera (JB)	11.37	Jarque-Bera test for normality (Prob(JB): 0.00)
Heteroskedasticity (H)	0.73	Test for heteroskedasticity (Prob(H): 0.18)
Skew	–0.21	Skewness of residuals
Kurtosis	4.04	Kurtosis of residuals

P, D, and Q are the seasonal terms, and m is the seasonal period. SARIMA is a good choice for datasets with recurring patterns like electricity demand. Figure 11 displays four diagnostic plots for the residuals of the SARIMAX model fitted to the national electricity demand (nat_demand). The top-left plot shows standardized residuals over time, hovering around zero with a notable spike around mid-2019, suggesting possible outliers. The top right histogram with KDE overlays (HSS, KDE, N(0,1)) is skewed to the right, not normal, as the residuals do not follow the standard normal curve closely. In the bottom-left Normal Q–Q plot, the points do not follow the diagonal line, especially at the tails, which confirms that the data are not normal. The bottom-right correlogram does not show any significant autocorrelation as the values are within the confidence bounds, which means that the model is able to capture linear and seasonal dependencies well. However, the non-normality and outliers suggest that the SARIMA model may need further refinement to accommodate these residual characteristics. Similar to the ARIMA implementation, the 218 observations used in the SARIMA model represent the weekly aggregated form of the original hourly electricity demand data. This modeling choice was made to capture the dominant medium-term seasonal structure while keeping the statistical model computationally feasible. Since SARIMA models with high-frequency hourly seasonality can become computationally intensive and difficult to estimate reliably, the weekly aggregated benchmark was preferred for the statistical comparison. The underlying data source remained the same; only the temporal resolution differed between SARIMA and GRU.

The results of the SARIMA (3, 0, 1) model are summarized in Table 3 for the national electricity demand (y) for 218 observations from January 18, 2015, to March 17, 2019. The AR coefficients (ar.L1: 1.3563, ar.L2: −0.5465, ar.L3: 0.1902) and MA coefficient (ma.L1: −0.8141) are statistically significant (P>|z|<0.05), indicating strong AR and MA effects. The Ljung-Box test (Prob(Q): 0.83) does not show the significant autocorrelation of residuals at lag 1, which shows that the linear dependencies are fitted well. The Jarque-Bera test (Prob(JB): 0.03) shows a slight non-normality as well as the kurtosis (3.84). The heteroskedasticity test (Prob(H): 0.14) does not show any significant variance instability. These results indicate that the SARIMAX model can capture linear patterns well enough, but that slight non-normality in residuals can be investigated.

Figure 11
Diagnostic plots for SARIMA model residuals

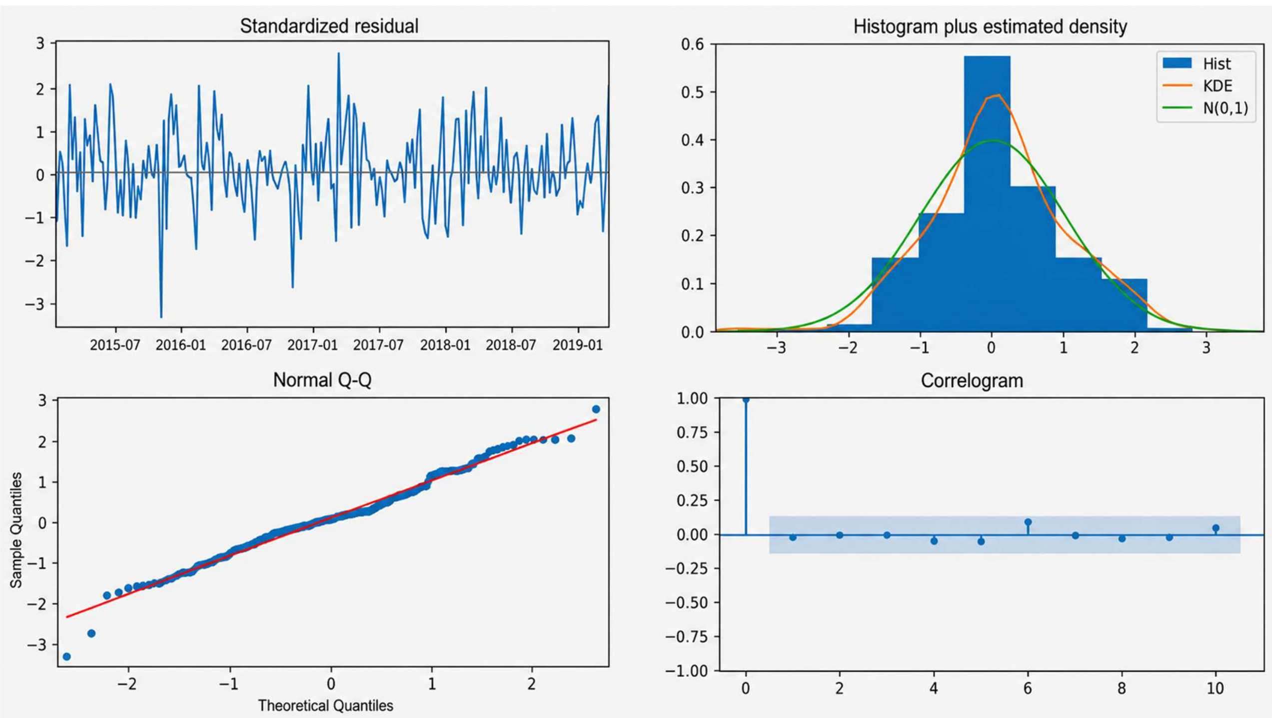


Table 3
SARIMA model results for national electricity demand

Parameter	Value	Description
Dependent variable	y	National electricity demand (denoted as *y*)
Model	SARIMAX(3, 0, 1)	SARIMAX model with AR(3), I(0), and MA(1) components
No. observations	218	Total number of data points
Log likelihood	–2205.857	Log likelihood of the model
AIC	4421.714	Akaike Information Criterion
BIC	4438.636	Bayesian Information Criterion
HQIC	4428.549	Hannan-Quinn Information Criterion
Date	Thu, 20 Mar 2025	Date of model fitting
Time	14:53:48	Time of model fitting
Sample	01-18-2015 to 03-17-2019	Time range of the data
Covariance type	opg	Covariance type used for standard errors
Coefficients		
ar.L1	1.3563	AR(1) coefficient (std err: 0.086, z: 15.827, P>
ar.L2	–0.5465	AR(2) coefficient (std err: 0.122, z: -4.484, P>
ar.L3	0.1902	AR(3) coefficient (std err: 0.089, z: 2.126, P>
ma.L1	–0.8141	MA(1) coefficient (std err: 0.068, z: -12.038, P>
sigma2	39,650,000	Variance of residuals (std err: 5.94e-10, z: 6.68e + 16, P>
Diagnostics		
Ljung-Box (L1) (Q)	0.05	Ljung-Box test statistic for residual autocorrelation (Prob(Q): 0.83)
Jarque-Bera (JB)	6.83	Jarque-Bera test for normality (Prob(JB): 0.03)
Heteroskedasticity (H)	0.71	Test for heteroskedasticity (Prob(H): 0.14)
Skew	–0.11	Skewness of residuals
Kurtosis	3.84	Kurtosis of residuals

4.5. GRU model forecasting performance analysis

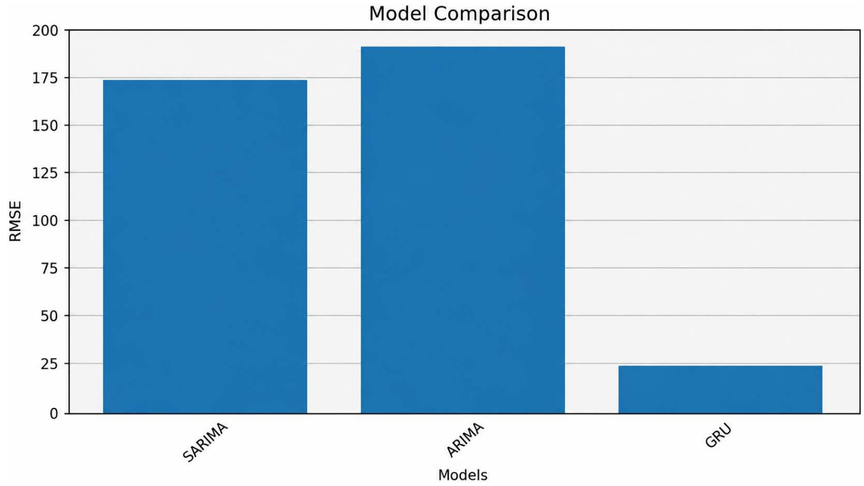
GRU is a recurrent neural network architecture designed specifically for sequential data and for effectively capturing long-term temporal dependencies, making it especially suitable for time series forecasting tasks. The GRU has two gates: a reset gate and an update gate. These gates control the flow of past information and control how much new information is added to the hidden state representation. GRU is a more simplistic architecture compared to LSTM networks, as it combines the hidden and cell states into a single representation and has fewer trainable parameters. The reduction in architectural complexity reduces computation cost and speeds up training, yet still provides strong modeling capability for nonlinear and temporally dependent patterns. These features have led to the widespread application of GRU models, for example, in electricity demand forecasting, speech recognition, and other sequential prediction problems. Unlike classical statistical models like ARIMA and SARIMA, GRU does not use explicit AR or MA structures; thus, it does not lead to interpretable model coefficients or parameter estimates. Therefore, the performance of the GRU model in this study is evaluated only by the error metrics of the forecast, such as RMSE, MAE, and MAPE, over the defined forecast horizons. Such an evaluation strategy guarantees a fair and methodologically consistent comparison with the classical approaches, based solely on the predictive accuracy and not on the interpretability of parameters.

The GRU model demonstrates better performance over traditional methods like ARIMA and SARIMA due to a few key advantages. First, GRU can capture complex and nonlinear patterns in the data efficiently, which is often overlooked by the linear models of ARIMA and SARIMA. Moreover, GRUs are very efficient in modeling long-term dependencies in sequential data, which gives a huge advantage, especially in cases with complex temporal dynamics. In addition, the gating mechanisms in GRUs allow them to learn adaptively, that is, adapt to different data patterns and provide more accurate predictions than the static structures of SARIMA. GRU is more efficient in terms of parameters and training time than the computationally intensive models like SARIMA, thereby making sure both performance and computational efficiency. Finally, GRUs are shown to be more robust to capture complex temporal dynamics, leading to better prediction accuracy compared to traditional statistical methods, making them a preferred choice for applications such as electricity demand forecasting. Figure 12 shows the comparison results.

5. Discussion

This paper investigates the effectiveness of ARIMA, SARIMA, and GRU models in forecasting national electricity demand (nat_demand) using a dataset of 48,046 hourly observations. Results highlight specific advantages and disadvantages of each model in modeling the temporal, seasonal, and nonlinear structure of electricity demand data. ARIMA was able to model linear trends but had trouble capturing seasonal and nonlinear patterns, as observed in the predictions around 20,000 MW that were almost constant, with an initial RMSE of 475 and an initial MAPE of 34%. This is consistent with the findings of Amjadian

Figure 12
Comparison results



et al. [15], who identified the limitations of ARIMA in capturing nonlinear trends and suggested hybrid approaches to improve accuracy. SARIMA addressed some of these constraints by incorporating seasonal terms and achieved an RMSE of 173 and a MAPE of 12% after accounting for additive trend and seasonality as presented in Section 4.2. Its predictions were better in capturing seasonal fluctuations but still did not reflect the peaks and troughs of the test data, consistent with the observation by Kumar et al. [27] that SARIMA does better than ARIMA in capturing seasonal variations, achieving a 7% MAPE in their study on NIT Patna's mixed grid. However, the best performance was observed with the GRU model, which closely followed the variability and trends of the test data, achieving an RMSE of 127 through multiplicative trend and seasonality components over a 3-day period. This superior performance discussed in Section 4.5 is because of GRU's capability to model nonlinear patterns and long-term dependencies, which is in line with the findings of Ullah et al. [16], where a hybrid model of CNN-LSTM obtained 538.71 RMSE and 2.72% MAPE for STLF on NTDC data. Diagnostic plots also suggested that ARIMA and SARIMA were not normally distributed, while GRU had stable residual behavior with no significant autocorrelation, but slight non-normality still exists, indicating that there is still room for improvement. Table 4 provides a comparative analysis of the models' performance, confirming the advantage of GRU in accuracy over ARIMA and SARIMA, in line with the literature's focus on the effectiveness of deep learning models in complex forecasting tasks.

Table 4 shows a comparison of the performance of the forecasting models in this study and in the literature in terms of MAPE, RMSE, and the variables used, together with the references of the respective studies. In this study, GRU attained the lowest RMSE (127), outperforming SARIMA (RMSE: 173, MAPE: 12%) and ARIMA (RMSE: 475, MAPE: 34%), which

Table 4
Comparative performance metrics of ARIMA, SARIMA, and GRU models

Study	Model	MAPE (%)	RMSE	Variables
This Study	ARIMA	34.0	475	nat_demand, weather, calendar
This Study	SARIMA	12.0	173	nat_demand, weather, calendar
This Study	GRU	8.7	127	nat_demand, weather, calendar
[13]	ARIMA	–	0.2184	Electricity consumption (univariate)
[14]	SARIMA	4.57	–	Hourly electricity load
[16]	CNN-LSTM	2.72	538.71	Electricity load (NTDC data)
[16]	CNN-LSTM	4.72	951.94	Electricity load (24-hour forecast)
[24]	TD-E2E	2.27	–	Residential load (50 homes, US)
[27]	SARIMA	7.0	–	Electricity load (NIT Patna grid)
[28]	sARIMA	2.5	–	Inertia estimation (renewable energy)
[29]	GAM-SARIMA	–	9.091 (NRMSE)	Electricity demand (11 years)
[31]	MTL-GAN	43% reduction	51% reduction	Multi-energy load (ASU Tempe)
[32]	BO-GRU-LightGBM	–	0.331	Renewable energy load

Note: The MAPE value for the GRU model was computed on the inverse-scaled test predictions using the same chronological evaluation period as the other models. The GRU model achieved RMSE = 127 and MAPE = 8.7%, confirming its superior forecasting accuracy compared with ARIMA and SARIMA in the present experimental setting.

demonstrates the effectiveness of GRU in capturing nonlinear dynamics in electricity demand forecasting with weather and calendar variables. In contrast, Ullah et al. [16] proposed a CNN-LSTM model, which achieved the RMSE and MAPE of 538.71 and 2.72% for short-term prediction, respectively, with higher errors than the GRU in this study, possibly due to the difference in data complexity. A low RMSE of 0.0871 was reported by Özen et al. [13] on multivariate data using a hybrid model that outperformed ARIMA (0.2184) and CNN (0.1802), suggesting that hybrid deep learning models are efficient in processing multiple types of input data. Anh et al. [14] reported a MAPE of 4.57% for SARIMA, which is better than the SARIMA of this study, likely due to the integration of real-time sensor data. Parkash et al. [24] achieved a MAPE of 2.27% with the TD-E2E model, demonstrating high accuracy in predicting residential load, and Kumar et al. [27] obtained a MAPE of 7% for SARIMA, reflecting its robustness in seasonal modeling. In summary, studies across a variety of domains have shown that GRU and hybrid deep learning models consistently outperform traditional methods like ARIMA and SARIMA, particularly when dealing with complex, multivariate data.

In response to the difference in sample size between the statistical and deep learning models, it should be emphasized that ARIMA and SARIMA were not trained on a different dataset. Instead, they were fitted on a weekly aggregated version of the same 48,046-record hourly dataset, resulting in 218 weekly observations. This preprocessing step was applied because direct fitting of ARIMA/SARIMA models to the full hourly training series with long seasonal cycles is computationally inefficient and can produce convergence-related instability. By contrast, the GRU architecture was specifically designed to process sequential high-frequency data and was therefore trained on the hourly data using a 168-hour sliding window. Thus, the reported comparison reflects the best practical implementation of each model family under a common chronological evaluation framework. The GRU model achieved RMSE = 127 and MAPE = 8.7%, outperforming SARIMA and ARIMA in both error magnitude and percentage error.

6. Conclusion

This study investigated the performance of classical time series models (ARIMA and SARIMA) and a deep learning–based approach (GRU) for hourly electricity demand forecasting using weather and calendar variables. Our study showed, by a systematic and fair comparison on a unified evaluation protocol, that data-driven deep learning models can successfully capture complex temporal dependencies existing in electricity demand data. The experimental result shows that the GRU model shows better forecasting accuracy than ARIMA and SARIMA models especially for longer forecasting horizons. This enhancement can be credited to the GRU's capability to capture nonlinear relationships and long-range temporal dependencies that are challenging to model using conventional linear statistical methods. This work does not aim at directly detecting, classifying, or predicting system failures, nor does it rely on explicit fault or outage data. The main contribution of this study is instead to improve the accuracy of short- and medium-term electricity demand forecasting. Reliable demand forecasts can indirectly support preventive grid operation, capacity planning, and data-driven decision-making by reducing uncertainty in expected load profiles. Despite these promising results, the present study has some limitations. The analysis is limited to one dataset and one geographical context. The GRU architecture was tested with a fixed set of hyperparameters. Future work will be directed toward the extension of the proposed framework to multiple regions, consideration of other deep learning architectures, and the integration of a probabilistic forecasting approach to better quantify uncertainty in electricity demand predictions.

The sample-size difference between ARIMA/SARIMA and GRU should also be interpreted as a limitation of the study. While ARIMA and SARIMA were implemented on a weekly aggregated benchmark series of 218 observations for computational feasibility, the GRU model used the full hourly-resolution data. Future studies should further investigate optimized statistical or hybrid approaches capable of directly modeling large-scale hourly data with complex seasonal structures.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting.

Author Contribution Statement

Emrah Aslan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Yıldırım Özüpak:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Feyyaz Alpsalaz:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Hasan Uzel:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision.

References

[1] de Klerk, J. (2025). Modelling time series structure, identifying outliers and forecasting ESKOM electricity production data using singular spectrum analysis. *Studies in Economics and Econometrics*, *49*(1), 34–52. https://doi.org/10.1080/03796205.2025.2458853

[2] Zeng, X., Ji, G., Zhou, Y., Li, H., & Wei, T. (2025). Multi-load forecasting for integrated energy systems based on GAT-MTL. *The Journal of Engineering*, *2025*(1), e70050. https://doi.org/10.1049/tje2.70050

[3] Yu, J., Hao, S., & Han, J. (2024). Optimizing virtual energy hub's for enhanced market participation and operational resilience. *Maintenance and Reliability*, *27*(1), 193171. https://doi.org/10.17531/ein/193171

[4] P, J. S., D, A. D., C, R. L., & R, N. N. (2025). Efficiency and reliability: Optimization of energy management in electric vehicles apply monarch butterfly algorithm and fuzzy logic control. *Maintenance and Reliability*, *27*(3), 200691. https://doi.org/10.17531/ein/200691

[5] Özüpak, Y., & Aslan, E. (2027). An enhanced interpretable ML approach for forecasting long-term CO2 and N2O emissions via optimized random forest modeling. *Fuel*, *427*, 139723. https://doi.org/10.1016/j.fuel.2026.139723

[6] Hao, X., Yin, H., Gao, J., & Lan, H. (2025). Ultra short-term forecasting for the propulsion energy consumption of all-electric ships based on TCFFA-GRU-parallel network. *Results in Engineering*, *25*, 103767. https://doi.org/10.1016/j.rineng.2024.103767

[7] Shivkumar, P., S, J. E., K, B., & T, M. M. (2025). Improving reliability in electric vehicle battery management systems through deep learning-based cell balancing mechanisms. *Maintenance and Reliability*, *27*(3), 200714. https://doi.org/10.17531/ein/200714

[8] Hasanat, S. M., Ullah, K., Yousaf, H., Munir, K., Abid, S., Bokhari, S. A. S., . . . , & Ullah, Z. (2024). Enhancing short-term load forecasting with a CNN-GRU hybrid model: A comparative analysis. *IEEE Access*, *12*, 184132–184141. https://doi.org/10.1109/ACCESS.2024.3511653

[9] Moussa, A., & Aoulmi, Z. (2025). Improving electric vehicle maintenance by advanced prediction of failure modes using machine learning classifications. *Maintenance and Reliability*, *27*(3), 201372. https://doi.org/10.17531/ein/201372

[10] Hua, Q., Fan, Z., Mu, W., Cui, J., Xing, R., Liu, H., & Gao, J. (2024). A short-term power load forecasting method using CNN-GRU with an attention mechanism. *Energies*, *18*(1), 106. https://doi.org/10.3390/en18010106

[11] Cui, J., Kuang, W., Geng, K., Bi, A., Bi, F., Zheng, X., & Lin, C. (2024). Advanced short-term load forecasting with XGBoost-RF feature selection and CNN-GRU. *Processes*, *12*(11), 2466. https://doi.org/10.3390/pr12112466

[12] Gong, R., Wei, Z., Qin, Y., Liu, T., & Xu, J. (2024). Short-term electrical load forecasting based on IDBO-PTCN-GRU model. *Energies*, *17*(18), 4667. https://doi.org/10.3390/en17184667

[13] Özen, S., Yazıcı, A., & Atalay, V. (2025). Hybrid deep learning models with data fusion approach for electricity load forecasting. *Expert Systems*, *42*(2), e13741. https://doi.org/10.1111/exsy.13741

[14] Anh, N. T. N., Anh, N. N., Thang, T. N., Solanki, V. K., Crespo, R. G., & Dat, N. Q. (2024). Online SARIMA applied for short-term electricity load forecasting. *Applied Intelligence*, *54*(1), 1003–1019. https://doi.org/10.1007/s10489-023-05230-y

[15] Amjadian, A., Kazemzadeh, M.-R., & Amraee, T. (2025). Long-term electric peak load forecasting of Tehran Regional Electric Company using a combinatorial artificial intelligence approach. *Scientia Iranica. Advance online publication*. https://doi.org/10.24200/sci.2025.63459.8408

[16] Ullah, K., Ahsan, M., Hasanat, S. M., Haris, M., Yousaf, H., Raza, S. F., . . . , & Ullah, Z. (2024). Short-term load forecasting: A comprehensive review and simulation study with CNN-LSTM hybrids approach. *IEEE Access*, *12*, 111858–111881. https://doi.org/10.1109/ACCESS.2024.3440631

[17] Ahamed, A., Tarafder, M. T. R., Rimon, S. M. T. H., Hasan, E., & Al Amin, M. (2024). Optimizing load forecasting in smart grids with AI-driven solutions. In *2024 IEEE International Conference on Data and Software Engineering*, 182–187. https://doi.org/10.1109/ICoDSE63307.2024.10829903

[18] Dorji, K., Jittanon, S., Thanarak, P., Mensin, P., & Termritthikun, C. (2024). Electricity load forecasting using hybrid datasets with linear interpolation and synthetic data. *Engineering, Technology & Applied Science Research*, *14*(6), 17931–17938. https://doi.org/10.48084/etasr.8577

[19] Fan, G.-F., Wei, H.-Z., Huang, H.-P., & Hong, W.-C. (2025). Application of ensemble empirical mode decomposition with support vector regression and wavelet neural network in electric load forecasting. *Energy Sources, Part B: Economics, Planning, and Policy*, *20*(1), 2468687. https://doi.org/10.1080/15567249.2025.2468687

[20] El-Afifi, M. I., Eladl, A. A., Sedhom, B. E., & Hassan, M. A. (2025). Enhancing EV charging station integration: A hybrid ARIMA-LSTM forecasting and optimization framework. *IEEE Transactions on Industry Applications*, *61*(3), 4924–4935. https://doi.org/10.1109/TIA.2025.3540788

[21] Amara-Ouali, Y., Hamrouche, B., Principato, G., & Goude, Y. (2025). Quantifying the uncertainty of electric vehicle charging with probabilistic load forecasting. *World Electric Vehicle Journal*, *16*(2), 88. https://doi.org/10.3390/wevj16020088

[22] Kychkin, A., & Chasparis, G. C. (2025). *Hourly short term load forecasting for residential buildings and energy communities*. arXiv. https://doi.org/10.48550/arXiv.2501.19234

[23] Kyryk, V., & Shatalov, Y. (2023). SARIMA(X) and decision tree models comparison for load forecasting in electrical grids. *Modern Engineering and Innovative Technologies*, (33–01), 12–17. https://doi.org/10.30890/2567-5273.2024-33-00-008

[24] Parkash, B., Lie, T. T., Li, W., & Tito, S. R. (2024). End-to-end top-down load forecasting model for residential consumers. *Energies*, *17*(11), 2550. https://doi.org/10.3390/en17112550

[25] Doddabasappa, N., T, R., & Dash, R. (2024). Mid-term demand forecasting using SARIMA model in distributed electricity market for MCP. In *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems*, 1–5. https://doi.org/10.1109/ISTEMS60181.2024.10560181

[26] Bensalah, M., & Hair, A. (2024). Empowering smart cities: SARIMA forecasting of power consumption in Tetouan's urban grid. In *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology*, 1–6. https://doi.org/10.1109/IRASET60544.2024.10548574

[27] Kumar, V., Mandal, R. K., & Rai, S. (2024). Smart metered distribution system load forecasting using statistical approaches. In *2024 4th International Conference on Emerging Frontiers in Electrical and Electronic Technologies*, 1–4. https://doi.org/10.1109/ICEFEET64463.2024.10866475

[28] Ningombam, R., Singh, C., Sreekumar, S., Bhakar, R., & Padmanaban, S. (2025). Box–Cox integrated sARIMA model for day-ahead inertia forecasting. *Electrical Engineering*, *107*(7), 9135–9153. https://doi.org/10.1007/s00202-025-02965-4

[29] Royal, E., Bandyopadhyay, S., Newman, A., Huang, Q., & Tabares-Velasco, P. C. (2025). A statistical framework for district energy long-term electric load forecasting. *Applied Energy*, *384*, 125445. https://doi.org/10.1016/j.apenergy.2025.125445

[30] Zhao, S., Xu, Z., Zhu, Z., Liang, X., Zhang, Z., & Jiang, R. (2025). Short and long-term renewable electricity demand forecasting based on CNN-Bi-GRU model. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, *2*(1), 1–15. https://doi.org/10.62762/TETAI.2024.532253

[31] Heng, S. (2025). Enhanced multi-energy load forecasting via multi-task learning and GRU-attention networks in integrated energy systems. *Electrical Engineering*, *107*(6), 7673–7683. https://doi.org/10.1007/s00202-024-02942-3

[32] Guo, Y. (2025). Research on electricity load forecasting method based on GRU-LightGBM combined model with Bayesian optimization. *Journal of Physics: Conference Series*, *2975*(1), 012019. https://doi.org/10.1088/1742-6596/2975/1/012019

[33] Chen, W., Rong, F., & Lin, C. (2025). A multi-energy loads forecasting model based on dual attention mechanism and multi-scale hierarchical residual network with gated recurrent unit. *Energy*, *320*, 134975. https://doi.org/10.1016/j.energy.2025.134975

[34] Wang, L., Chen, L., Jin, S., & Li, C. (2025). Forecasting the green behaviour level of Chinese enterprises: A conjoined application of the autoregressive integrated moving average (ARIMA) model and multi-scenario simulation. *Technology in Society*, *81*, 102825. https://doi.org/10.1016/j.techsoc.2025.102825

[35] Singh, S., Parmar, K. S., & Kumar, J. (2025). Development of multi-forecasting model using Monte Carlo simulation coupled with wavelet denoising-ARIMA model. *Mathematics and Computers in Simulation*, *230*, 517–540. https://doi.org/10.1016/j.matcom.2024.10.040

[36] Mondal, S. K., Ahmmad, M. S., Khan, S., Chowdhury, M. H., Paul, G. K., Binyamin, M., . . . , & Chakrabortty, R. (2026). Modeling and forecasting monthly humidity in South Asia: A SARIMA approach. *Earth Systems and Environment*, *10*(1), 493–509. https://doi.org/10.1007/s41748-025-00607-0

[37] Zuo, T., Tang, S., Zhang, L., Kang, H., Song, H., & Li, P. (2025). An enhanced TimesNet-SARIMA model for predicting outbound subway passenger flow with decomposition techniques. *Applied Sciences*, *15*(6), 2874. https://doi.org/10.3390/app15062874

[38] Kosarkar, U., & Sakarkar, G. (2025). Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis. *Multimedia Tools and Applications*, *84*(7), 3841–3857. https://doi.org/10.1007/s11042-024-19220-w

[39] Jiang, H., Li, W., Liu, Y., Liu, R., Yang, Y., Duan, C., & Huang, L. (2025). A hybrid EMD-GRU model for pressure prediction in air cyclone centrifugal classifiers. *Advanced Powder Technology*, *36*(1), 104743. https://doi.org/10.1016/j.apt.2024.104743

[40] Yang, M., Zhou, Q., Huang, H., Liu, J., Pan, H., Cheng, Y., . . . , & Hu, Y. (2025). A fault prediction method for CMOR bearings based on parameter-optimized variational mode decomposition and autocorrelation function. *Fusion Engineering and Design*, *212*, 114863. https://doi.org/10.1016/j.fusengdes.2025.114863

[41] Sun, Z., Zou, Z., Wang, J., Zhao, X., & Wang, F. (2025). Study on fluctuating wind characteristics and non-stationarity at U-shaped canyon bridge site. *Applied Sciences*, *15*(3), 1482. https://doi.org/10.3390/app15031482

[42] Manzoor, A., Atif Qureshi, M., Kidney, E., & Longo, L. (2024). A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE Access*, *12*, 70434–70463. https://doi.org/10.1109/ACCESS.2024.3402092

[43] Salmanpour, M. R., Alizadeh, M., Mousavi, G., Sadeghi, S., Amiri, S., Oveisi, M., . . . , & Hacihaliloglu, I. (2025). Machine learning evaluation metric discrepancies across programming languages and their components in medical imaging domains: Need for standardization. *IEEE Access*, *13*, 47217–47229. https://doi.org/10.1109/ACCESS.2025.3549702

How to Cite: Aslan, E., Özüpak, Y., Alpsalaz, F., & Uzel, H. (2026). Optimizing Electricity Demand Forecasting Using ARIMA, SARIMA, and GRU with Weather and Calendar Variables. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027129