Solar Energy
And Sustainable Development
Refereed, biannual scientific journal issued by:
The Libyan Center for Solar Energy Research and Studies

Towards Efficient Electricity Management in Benghazi: Forecasting Demand and Load Shedding with ARIMA Models

Asma Agaal [1*], Hend M. Farkash[2], Mansour Essgaer[3], Abbas Ahessin[4].

[1,3]Department of Artificial Intelligence, Faculty of Technical Sciences, Sabha, Libya.
[2]The Faculty of Electrical & Electronics Technology, Benghazi, Libya.
[4]Department of Computer Science, Faculty of Technical Sciences, Sabha, Libya.

E-*mail:* [1] asma.agaal@sebhau.edu.ly , [2] Hend.mf_ceet@ceet.edu.ly , [3] man.essgaer@sebhau.edu.ly , [4] abbas5052@ctss.edu.ly .

SPECIAL ISSUE ON:

The 1st International Conference on Technical Sciences, 2024.
"Investing in Renewable Energies"
11 November 2024 SEBHA, LIBYA.

KEYWORDS

Autoregressive Integrated Moving Average, Benghazi Electrical Grid, General Electricity Company of Libya, Load Shedding, Time Series Analysis.

ABSTRACT

In Libya, the general electricity company is tasked with managing peak electricity demand, often resorting to load shedding. This practice, while necessary, results in power outages, particularly impacting areas like the Benghazi Electrical Grid. This study aims to bring predictability to these events by exploring time series forecasting models namely: Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Dynamic Regression ARIMA (DRARIMA). The models were trained using data from May 2020 and 2021, and subsequently tested on May 2022.

Performance was evaluated using metrics such as mean squared error, mean absolute error, mean absolute percentage error, and mean absolute percentage accuracy. The ARIMA model achieved the highest accuracy at 78.88% mean absolute percentage accuracy with a mean absolute error of 0.9. The SARIMA model, which considers seasonal patterns, achieved an accuracy of 73.86% and mean absolute error of 0.11, but its complexity may lead to overfitting. The DRARIMA, which incorporates exogenous variables, demonstrated an accuracy of 65.36% and mean absolute error of 0.15. Future projections for May 2024 and 2025 using ARIMA models indicate potential improvements in load shedding management and highlight the importance of model selection for accurate forecasting. By improving load forecasting accuracy, this research aims to enhance the effectiveness of load shedding management, thereby reducing power outages and their socio-economic impacts in regions like Benghazi. These findings are particularly valuable for energy planners and managers in similar contexts, providing practical insights and data-driven strategies.

*Corresponding author.

DOI: https://doi.org/10.51646/jsesd.v14iFICTS-2024.446

نحو إدارة فعَالة للكهرباء في بنغازي: التنبؤ بالطلب والأحمال باستخدام نماذج المتوسط المتحرك الانحدار الذاتي

أسماء أعجال، هند فركاش، منصور الصغير، عباس إحسين.

ملخص: في ليبيا، تتولى الشركة العامة للكهرباء إدارة ذروة الطلب على الكهرباء، وغالبًا ما تلجأ إلى خفض الأحمال. تؤدي هذه الممارسة، على الرغم من ضرورتها، إلى انقطاع التيار الكهربائي، وخاصة التأثير على مناطق كثيرة مثل شبكة كهرباء بنغازي. تهدف هذه الدراسة إلى تحقيق القدرة على التنبؤ بهذه الأحداث من خلال استكشاف نماذج التنبؤ بالسلاسل الزمنية وهي: المتوسط المتحرك المتكامل الانحدار الذاتي (ARIMA)، و ARIMA الموسمي (SARIMA)، و ARIMA الانحداري الديناميكي (DRARIMA). تم تدريب النماذج باستخدام بيانات من مايو 2020 و 2021، وتم اختبارها لاحقًا في مايو 2022. تم تقييم الأداء باستخدام مقاييس مثل متوسط الخطأ التربيعي، ومتوسط الخطأ المطلق، ومتوسط الخطأ المطلق النسبي، ومتوسط دقة النسبة المئوية المطلقة. حقق نموذج ARIMA أعلى دقة بنسبة % 78.88 من متوسط دقة النسبة المئوية المطلقة مع متوسط خطأ مطلق قدره 0.9. حقق نموذج SARIMA، الذي يأخذ في الاعتبار الأنماط الموسمية، دقة بنسبة 73.86٪ ومتوسط خطأ مطلق قدره 0.11، لكن تعقيده قد يؤدي إلى الإفراط في الملاءمة. وقد أظهر نموذج DRARIMA، الذي يشتمل على متغيرات خارجية، دقة بنسبة 65.36% ومتوسط خطأ مطلق بنسبة 0.15. وتشير التوقعات المستقبلية لشهري مايو/أيار 2024 و 2025 باستخدام نماذج ARIMA إلى تحسينات محتملة في إدارة انقطاع التيار الكهربائي وتسلط الضوء على أهمية اختيار النموذج للتنبؤ الدقيق. ومن خلال تحسين دقة التنبؤ بالأحمال، يهدف هذا البحث إلى تعزيز فعالية إدارة انقطاع التيار الكهربائي، وبالتالي تقليل انقطاع التيار الكهربائي وتأثيراته الاجتماعية والاقتصادية في مناطق مثل بنغازي. وهذه النتائج ذات قيمة خاصة لمخططي ومديري الطاقة في سياقات مماثلة، حيث توفر رؤى عملية واستراتيجيات تعتمد على البيانات.

الكلمات المفتاحية – المتوسط المتحرك الانحدار التلقائي، شبكة كهرباء بنغازي، الشركة العامة للكهرباء في ليبيا، انقطاع التيار الكهربائي، تحليل السلاسل الزمنية.

1. INTRODUCTION

The power sector faces diverse challenges, including humanitarian crises, as seen in Palestine [1], and political instability, as in Libya. Libya's energy sector is particularly strained, with an operational power capacity of only 4.8 GW—42% of its 11.5 GW installed capacity—resulting in a 3 GW deficit during peak summer demand. Persistent power shortages due to aging infrastructure, lack of maintenance, and fuel scarcity have impacted vital sectors like oil and gas [2]. Frequent blackouts since 2011, exacerbated by subsidized tariffs that inflate demand, have led to scheduled outages and widespread summer protests.

In Benghazi, these issues are especially acute, with the Benghazi Electrical Grid (SEGL) facing regular outages due to supply-demand mismatches. Infrastructure damage from the 2011 and ongoing development projects add further strain [3, 4]. While electricity generation grew significantly from 2000 to 2010, it has not kept pace with the projected doubling of demand by 2030 [5]. While electricity generation grew significantly from 2000 to 2010, it has not kept pace with the projected doubling of demand by 2030 [5].

This widening gap necessitates frequent load shedding [6], a reactive measure that disrupts daily life and hinders economic activity [7]. Current reliance on real-time monitoring for load shedding decisions proves insufficient in anticipating short-term demand fluctuations and accounting for external factors. Additionally, domestic energy consumption—36% of total energy use [6, 8], influenced by socio-economic and behavioral factors [9]—complicates accurate load forecasting. These challenges highlight the critical need for advanced forecasting methodologies.

To address this, our research leverages time series analysis to develop predictive models for the East City Benghazi station. Time series forecasting has gained traction for predicting trends based on historical data. At the heart of this approach is pattern recognition, uncovering regularities or anomalies within datasets [10]. Traditional methods like ARIMA and SARIMA have shown effectiveness in identifying patterns and producing accurate forecasts [11]. However, gaps remain in understanding the limitations of these models and the potential for improved accuracy through hybrid approaches, especially in Libya's grid context.

This study aims to: (1) Evaluate the performance of ARIMA [12], SARIMA [13], and Dynamic Regression ARIMA (DRARIMA) [14] models in predicting peak loads for the East City Benghazi station; (2) Identify the strengths and weaknesses of each model; and (3) Provide recommendations to enhance load forecasting accuracy. This comparative analysis informs model selection for similar energy challenges, offering a first-of-its-kind examination within the specific context of Benghazi. Furthermore, this study delves into model resilience, computational efficacy, and forecast accuracy across real-world datasets. By focusing on short-term load demand forecasting, this research not only fills a notable gap in the literature but also provides practical insights for energy planners in load-shedding-prone regions.

The rest of this paper is organized as follows: Section 2 offers a comprehensive review of current research on time series modeling, pattern extraction, and the application of ARIMA and SARIMA models. Section 3 outlines the research methods employed in the study. Section 4 presents the results and discusses the findings. Section 5 provides future predictions using ARIMA models. Section 6 concludes the paper, and Section 7 offers recommendations for future research.

2. PREVIOUS STUDIES

This section reviews existing research on electricity load forecasting, emphasizing time series models and their applications in various contexts. This review highlights contributions, limitations, and gaps in the literature, ultimately justifying the focus and methodology of the current study.

2.1. ARIMA and SARIMA Models for Load Forecasting

Numerous studies have employed ARIMA and SARIMA models, showcasing their versatility across different forecasting horizons and applications:

• Model Robustness and General Reviews: Chodakowska et al. [15] explored the resilience of the ARIMA model to noise within the Polish power system, using ARIMA to examine sensitivity thresholds. They found that ARIMA remained relatively stable up to a specific noise threshold, but they did not quantify this impact or compare ARIMA's performance with alternative models, limiting the scope of their findings. Similarly, Czapaj et al. [16] reviewed a variety of autoregressive methods, including ARIMA, and identified promising alternatives, such as Fuzzy Logic, Artificial Neural Networks (ANNs), and hybrid approaches. However, they did not conduct a direct comparison between these methods, leaving open the question of ARIMA's relative performance.

• Short-Term Load Forecasting: López et al.[17] employed multiplicative SARIMA for short-term load forecasting in distribution systems, comparing frequentist and Bayesian parameter estimation, Their findings emphasized the role of estimation technique in enhancing forecast accuracy but did not explore the model's effectiveness with non-linear data. Amerise and Tarsitano [18] used a two-stage linear regression-SARMA approach for hourly forecasts in Italy, addressing serial correlation but potentially missing non-linear relationships, While their model performed well with linear data, it potentially overlooked non-linear relationships that could affect forecasting accuracy in more complex settings. Crujido et al. [19] found that FFNN and LSTM outperformed SARIMA for day-ahead forecasting (MAPE: 1.80%, 1.75% vs. 4.48%), This finding highlighted SARIMA's limitations when dealing with non-linear data patterns, suggesting that neural network-based models may be more suitable in such scenarios. Dubey et al. [20] also showed LSTM's superiority over ARIMA and SARIMA (average MAE: 0.23), incorporating

weather feature correlations, This result underscored the importance of exogenous variables in enhancing the predictive power of LSTM models in load forecasting.
• Medium- and Long-Term Forecasting: Yin et al. [21] demonstrated SARIMA's suitability for medium- and long-term forecasting in China, focusing on parameter tuning. Durmus Senyapar and Aksoz [22] found SARIMA superior to exponential smoothing (MAPE: 2.21% training, 2.44% testing vs. SSE: 0.469) but didn't consider exogenous variables.

2.2. Hybrid Models

Hybrid models combine different forecasting approaches to leverage their respective strengths:
• ARIMA-based Hybrids: Velasco et al. [23] achieved a 4.09% MAPE with an ARIMA-ANN hybrid for next-day forecasting, which highlighted the effectiveness of combining ARIMA with neural networks. Kaytez [24] proposed an ARIMA-LSSVM hybrid for long-term forecasting in Turkey, finding that the hybrid model outperformed standalone forecasting methods, demonstrating the robustness of hybrid approaches for longer forecast horizons. Somu et al.[25] introduced ISCOA-LSTM, a hybrid LSTM model optimized with an improved sine cosine algorithm, for building energy consumption forecasting, demonstrating improved accuracy. These studies showcase the potential of hybrid models, but further comparisons with other hybrid approaches are needed.

2.3. Other Time Series and Machine Learning Models

Beyond ARIMA and hybrids, various other techniques have been applied:
• Deep Learning: Gasparin et al. [26] comprehensively reviewed and evaluated deep learning models (feedforward, recurrent, sequence-to-sequence, TCN) for short-term load forecasting. Lara-Benítez et al.[27] showed that TCN outperforms LSTM for energy demand forecasting in Spain.
• African Context: Farkash et al. [28] used NARX neural networks for medium-term forecasting in Benghazi, Libya, focusing on local forecasting needs. Ali et al. [29] focused on voltage control in Libya, incorporating load growth projections. Alarbi [30] investigated demand-side response in Libya. Hamouda et al. [31] analyzed residential loads in Sawknah, Libya. Chaaraoui et al.[32] evaluated various algorithms, including SARIMA and LSTM, in a Ghanaian health facility. Guefano et al. [33] forecasted residential consumption in Cameroon (MAPE: 1.628%). While these studies offer valuable insights into regional energy challenges and solutions, their limitations in scope or methodology suggest the need for broader comparative analyses to fully address the complexities of regional demand and system requirements.
• Broader Reviews: Nti et al.[34] systematically reviewed electricity load forecasting algorithms and influencing factors. Hoffmann et al. [35] reviewed time series aggregation methods in energy system models.

2.4. Research Gap and Justification

Despite extensive research, several gaps necessitate the current study:
1. Limited focus on the Benghazi context: Existing studies on Libya [28-31] use different methodologies or address related but distinct problems. Studies from other African contexts [5, 24] may not be directly transferable.
2. Inadequate exploration of exogenous variables with ARIMA-based models: Many studies using (S)ARIMA neglect external factors .[32, 33], limiting their predictive power. This study will specifically address this by employing DRARIMA.
3. Need for direct model comparison on the Benghazi dataset: While [16] identifies promising models, a direct comparison on a common dataset is lacking. This study will rigorously compare ARIMA, SARIMA, and DRARIMA on the same data, enabling informed model selection.

By addressing these gaps, this research will contribute valuable insights into effective time series forecasting for the unique challenges of the Benghazi electrical grid, ultimately supporting improved load shedding management and enhanced energy security.

3. METHODOLOGY

Figure 1 shows the proposed load shedding prediction framework using ARIMA models:

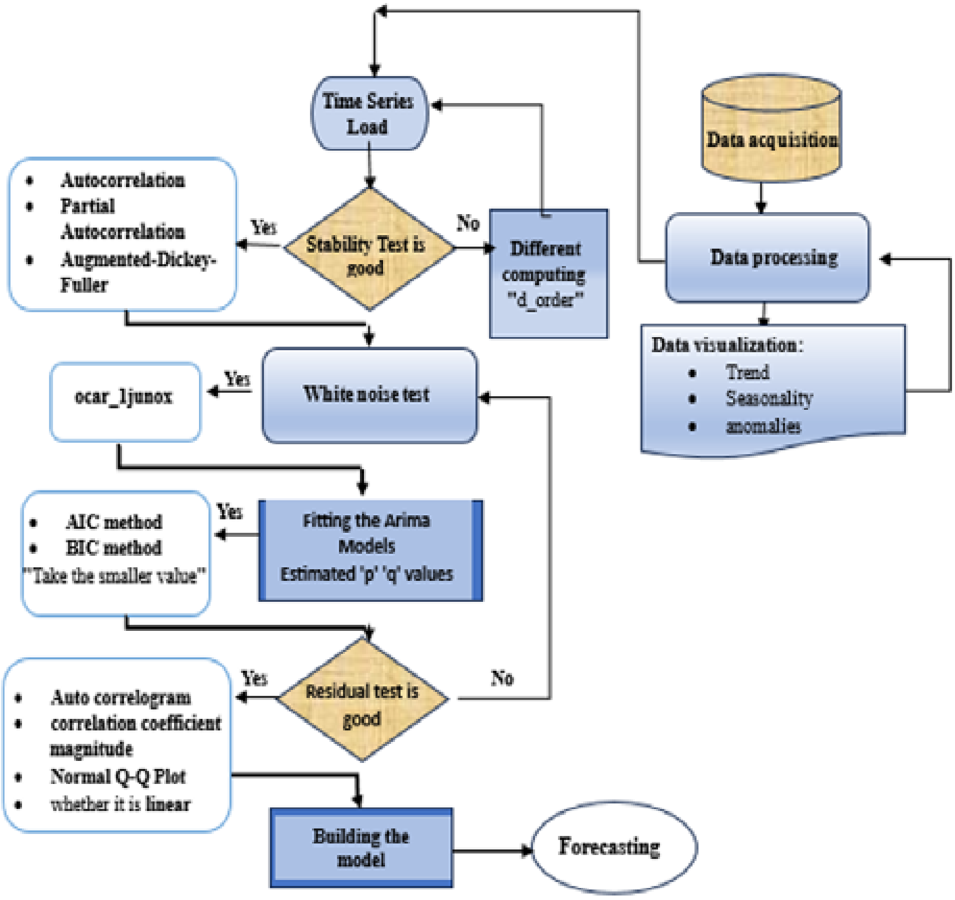

Figure 1.The proposed methodology.

Load Forecasting Methodology Pipeline, as shown below:

• Data Acquisition: Gather historical load/generation data (May 2020-2022) from the North Benghazi distribution station.

• Data Preprocessing: Clean data (handle missing values, remove outliers using Z-score), and normalize using Min-Max scaling.

• Time Series Analysis: Perform EDA (distributions, trends, seasonality), stationarity tests (ADF, ACF, PACF), and differencing if needed.

• Model Selection & Training: Choose ARIMA, SARIMA, DRARIMA models (using auto_arima), estimate parameters (MLE), and validate model assumptions (diagnostic plots).

• Forecasting & Evaluation: Generate forecasts (May 2022, 2024, 2025), and evaluate using MSE, MAE, RMSE, MAPE, MAPA.

3.1. Data Acquisition

3.1.1. Data collection

To forecast long-term peak loads for May, data was collected from the North Benghazi distribution station, located in the Eastern Region 220 control area. The dataset includes 2234 daily observations over three years, categorized into input and output data. Input data includes load, and generation for May 2020, May 2021, and May 2022.

The research team's affiliated institution submitted a formal request to the Electricity Company for data covering the month of May, due to significant load fluctuations observed during this period compared to other months. The Electricity Company approved the request and provided the data in electronic format, consisting of daily records for each day in May across the study years, as shown in Figure 2 of the report on daily electricity variables, dated May 1, 2020. Technical engineers from the company supervised the data collection and preparation process to ensure its accuracy and reliability, maintaining the highest standards and preventing any potential bias. The dataset is comprehensive, with detailed daily load records, making it ideal for thorough analysis and forecasting.

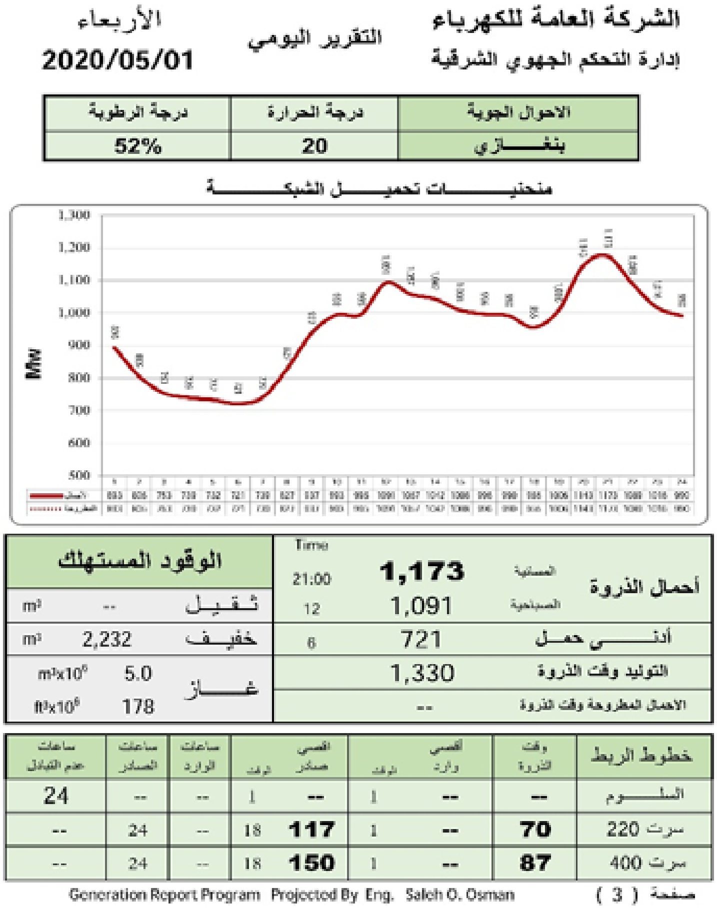

Figure 2. Daily report of electricity variables for the North Benghazi Distribution Station.

3.1.2. Characterization of the Data Set

Table 1 shows the features of the data set, which includes four attributes in addition to the target variable, representing the electrical load.

Table 1. Dataset characteristics.

Feature	Feature description
days	Number of days
hours	The number of hours
years	The value of years
generation	Value of electrical generation
Loads	Load value (Target)

3.2. Data Preprocessing

Upon the acquisition of raw data, the following processing steps were undertaken:

3.2.1. Data Cleaning and Outlier Removal

The raw data underwent a thorough cleaning process to remove any erroneous or redundant value. Outliers were identified and excluded using the Z-Score method [36]. In the case of continuous numerical features with missing values, the average value was used as a replacement.

3.2.2. Data Normalization

Post-cleaning, it was observed that different features exhibited different scales. To ensure equal weightage across all features, normalization of the dataset was deemed necessary. Data normalization refers to the process of adjusting all attribute values to a specified range [37]. Among the various data normalization techniques available, this study employed Min-Max Normalization, a widely accepted method. This technique transforms every attribute to a decimal value between 0 and 1, with the minimum value of the attribute mapped to 0, the maximum value to 1, and all other values proportionately scaled [38].

3.3. Autoregressive Integrated Moving Average

ARIMA is a popular time series forecasting method that combines three components [39]:

3.3.1. Auto Regressive (AR) Component

This section makes use of the dependence between several lagged observations (prior values) and an observation. The following equation (1) represents the AR model of order p:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t \tag{1}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model, c is a constant, and ε_t is white noise.

3.3.2. Integrated (I) Component

The time series is made stationary in this section by differencing the raw observations, which ensures that its mean and variance remain constant across time. The series is referred to as integrated of order d if it requires d difference steps to become stationary, as expressed by equation (2):

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_d \varepsilon_{t-d} \tag{2}$$

where $\theta_1, \theta_2, \ldots, \theta_d$ are the parameters of the model and ε_t is white noise.

3.3.3. Moving Average (MA) Component

This section makes use of the relationship between a lagged set of observations and a residual error from a moving average model. The following equation (3) represents the MA model of order q:

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_q \varepsilon_{t-q} \tag{3}$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model and ε_t is white noise.

The ARIMA (*p,d,q*) notation represents the general ARIMA model, where (*p*) represents the order of the AR component, (*d*) represents the number of differencing steps needed, and (*q*) represents the order of the MA part. To make sure time series are stable, Auto ARIMA verifies their stability. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are typically used as the criteria for selecting the optimal set of parameters [40].

3.4. Seasonal ARIMA

To handle seasonal data with a recurrent pattern, SARIMA extends ARIMA. It enhances the

ARIMA model by adding more seasonal elements [41]. The notation for the SARIMA model is ARIMA (p,d,q) (P,D,Q), where:

(p,d,q) are the non-seasonal parameters (as in ARIMA).

(P,D,Q) are the seasonal parameters for the seasonal AR, differencing, and MA parts, respectively.

(s) is the length of the seasonal cycle.

Although they are applied to lagged values at multiples of the seasonal period, the seasonal components function in a manner akin to that of the non-seasonal parts. The data's general trend and seasonal influences are both captured by the SARIMA model [42].

3.5. Dynamic Regression ARIMA

Regression in Dynamic Form is used with ARIMA, also known as RegARIMA, to manage the time series data and exogenous variables, or predictors. It entails modeling the residuals (errors) using an ARIMA model after regressing the time series on external variables. Equation (4) can be used to express the model [43]:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots\ldots + \beta_k X_{kt} + ARIMA_{(p,d,q)} \qquad (4)$$

where $X_{1t}, X_{2t}, \ldots, X_{kt}$ are the exogenous variables, $\beta_0, \beta_1, \beta_2 \ldots \beta_k$ are their coefficients, and $ARIMA_{(p,d,q)}$ models the residual errors from the regression.

When additional external factors impact the time series, this model is especially helpful since it incorporates such elements into the analysis, resulting in a more thorough knowledge and improved forecasts [44].

3.6. Model Evaluation

The assessment metrics are divided into two separate groups. While the second category is intended to calculate the accuracy and performance of the ARIMA models, the first category is devoted to evaluating the time series' stability.

3.6.1. Augmented-Dickey-Fuller (ADF) and p-values

Higher-order autoregressive processes are supported by the ADF test, which is a "augmented" variant of the Dickey-Fuller test. The test is predicated on the null hypothesis that a time series sample contains a unit root. A time series that has a unit root becomes non-stationary. We are unable to rule out the null hypothesis that there is a unit root if the p-value is greater than a certain size [45].

3.6.2. Autocorrelation Function (ACF)

It gauges a time series value's relationship to its historical values. For this reason, serial correlation is another name for it. Each bar in an ACF plot indicates the correlation's magnitude and direction. Bars that pass over into the red zone indicate statistical significance. Autocorrelations for random data should be close to zero for all lags. There is always at least one notable lag in non-random data [46].

3.6.3. Partial Autocorrelation Function (PACF)

After accounting for correlations at all shorter delays, it calculates the correlation between observations at various points in a time series. Stated differently, it provides the stationary time series' partial correlation with its own lagged values. When attempting to determine the order of an autoregressive model, the PACF is quite helpful. For example, the PACF of an AR(p) model ends at lag p. Accordingly, for lags up to p, the partial autocorrelation is considerable, whereas for lags longer than p, it is roughly nil [47].

3.6.4. Mean Squared Error (MSE)

The discrepancies between the expected and actual values are called errors, and it measures the average of the squares of the errors [48]. The calculation is expressed as equation (5):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i^{\sim})^2 \qquad (5)$$

where Y_i is the actual value, $Y_i^{\sim}$ is the predicted value, and n is the number of observations. Make use of Because the MSE squares the mistakes, greater errors add more to the MSE, making it susceptible to outliers. Due to its ability to offer a quadratic loss function, it is frequently utilized in regression analysis and model evaluation. A better match between the model and the data is shown by lower MSE values [49].

3.6.5. Mean Absolute Error (MAE)

It measures the average magnitude of the errors without considering their direction [50]. It is calculated as the equation (6):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i^{\sim})^2} \qquad (6)$$

Because MAE does not square the errors, it offers a simple way to evaluate average error magnitude that is less susceptible to outliers than MSE. A higher model fit is indicated by lower MAE values. Unlike MSE,MAE is easier to read because it is expressed in the same units as the original data.

3.6.6. Mean Absolute Percentage Error (MAPE)

It measures the average absolute percentage error between the predicted and actual values [51]. It is calculated as the equation (7):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^{\sim}}{Y_i}\right| \times 100 \qquad (7)$$

By expressing errors as a proportion of the actual values, MAPE allows for easy interpretation and cross-dataset comparison. A better match is indicated by lower MAPE values. When actual numbers are extremely close to zero, MAPE might cause very high or undefinable percentage errors, which can be troublesome.

3.6.7. Mean Absolute Percentage Accuracy (MAPA)

It's a metric for assessing a forecast's accuracy. It is based on the MAPE, which is a widely used indicator to assess how accurate forecasting models are. Although MAPE offers an error measure, MAPA converts it into an accuracy score that is simpler to understand as a percentage of correctness [41]. The following equation (8) is used to determine the MAPA:

$$MAPA = \left(1 - \frac{MAPE}{100}\right) \times 100 \qquad (8)$$

4. RESULTS AND DISCUSSION

This section shows the results derived from the implementation of the proposed framework. Data was collected from the North Benghazi electricity distribution station with the objective of forecasting load demands for May 2024 and 2025. This prediction was based on the recorded loads for the month of May in the years 2020, 2021, and 2022. The Python environment was chosen for conducting the experiments, owing to its comprehensive support for time series techniques. Subsequent to the pre-processing phase, the data was partitioned into two subsets: 85% of the data, corresponding to the years 2020 and 2021, was utilized for training the ARIMA models. The remaining 15% of the data, representing the year 2022, was allocated for testing and evaluation purposes.

The experiments conducted can be divided as follows:

- Exploratory Data Analysis was conducted to understand the characteristics of the data.
- Data visualization techniques were employed to identify trends, seasonality, and anomalies.
- Parameters of the ARIMA, SARIMA, and DRARIMA models were adjusted, and a predictive model was constructed for each.
- The performance of each model was evaluated using MSE, MAE, RMSE, MAPE, and MAPA.
- All ARIMA models were used to predict future values for May 2024 and 2025.

4.1. Exploration Analysis

The initial exploratory stage of our research encompassed an exhaustive analysis of the attributes of electrical load, and generation. This stage was instrumental in laying the foundation for the construction of ARIMA models designed to predict future values.

4.1.1. Descriptive Statistics and outliers' detection

Our analytical steps commenced with an in-depth exploration of the "loads" attribute, as illustrated in Figure 3. The density plot (on the left) unveiled an approximately normal distribution, with the mean load value hovering around 1325, as indicated by the red dashed line. This near-normality is beneficial, as many forecasting models assume a normal distribution, which could enhance model accuracy in predicting future loads. The 25th percentile (marked by the green dashed line) and the 75th percentile (denoted by the yellow dashed line) were approximately 1200 and 1420, respectively. These specific percentile values are critical, as they inform energy planners about the common range within which load values fluctuate, aiding in proactive load management. Moreover, the nearness of the mean and median hinted at a symmetrical distribution.

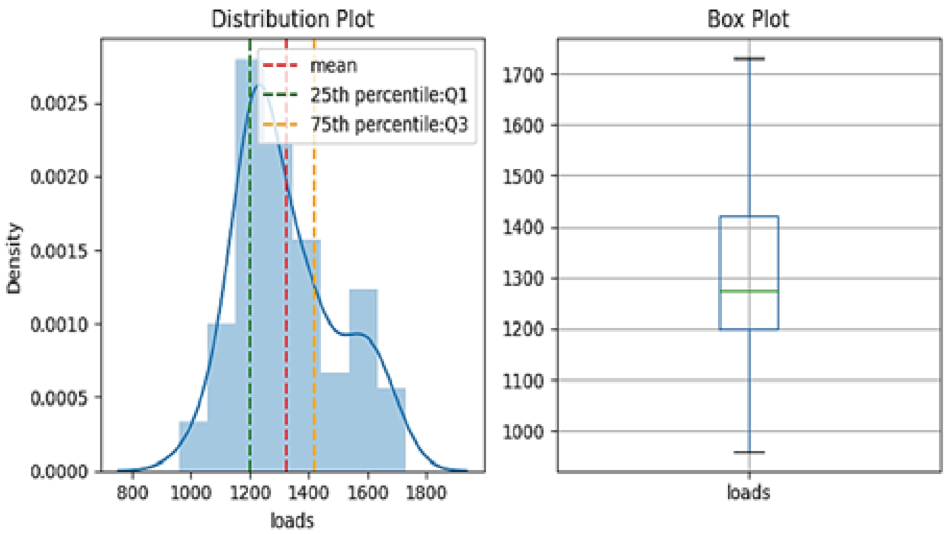

Figure 3. The density plot for loads attribute.

The box plot (right) summarizes the data distribution through its quartiles (Q1 to Q3) and further emphasizes that all points fall within 1.5 times the interquartile range (IQR). The absence of outliers suggests a stable and predictable load pattern, which may simplify the forecasting process by reducing the influence of extreme values. This moderate data dispersion, as indicated by the IQR and whiskers, reinforces the dataset's suitability for time series forecasting, potentially increasing the reliability of model predictions. Overall, these insights into load distribution are crucial for accurately projecting peak demand and effectively managing load shedding.

Moreover, the "generation" attribute, as depicted in Figure 4, also displayed a regular distribution with a symmetrical form. The proximity of the mean and median suggested a central concentration of data points for both attributes. suggesting a stable core of generation values, which is useful for forecasting reliability.

The average generation value of approximately 1257 closely aligns with the mean "loads" value of around 1275, indicating that typical generation levels are generally well-matched to typical load demands—a valuable insight for balancing supply and demand.

The minimum generation value of roughly 692 and maximum of approximately 1505 reveal a wider range than observed in "loads," reflecting the capacity fluctuations that may occur in response to varying demand levels or operational conditions. This broader range, paired with the presence of outliers (visible in the box plot), suggests higher variability in generation than in loads. This variability, as indicated by the outliers, points to occasional spikes or drops that might require more adaptive forecasting approaches to ensure system stability.

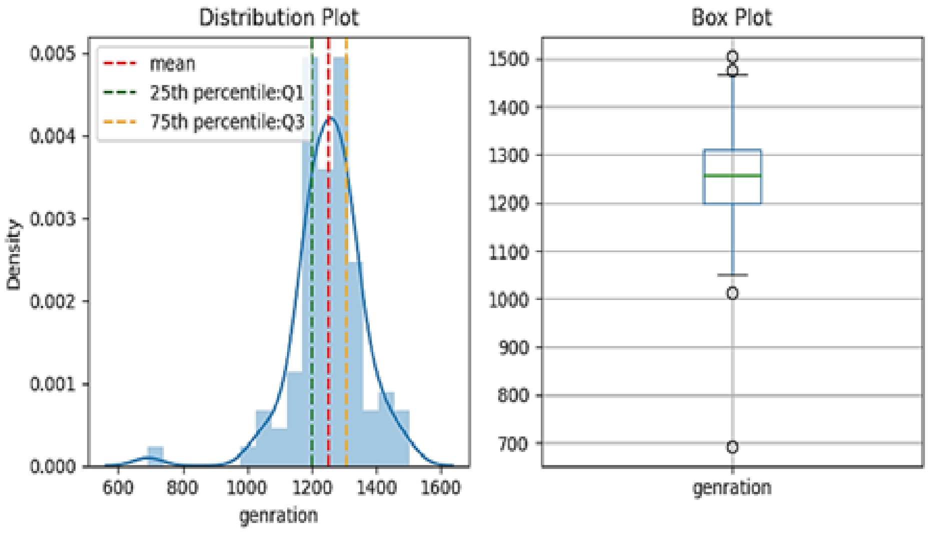

Figure 4. The density plot for generation attribute.

The moderate dispersion of the "generation" data, as indicated by the interquartile range (IQR), reinforces that while generation generally follows a predictable pattern, it has moments of deviation that could impact load management. Recognizing these outliers and variability is essential for planners, as it highlights the potential need for reserve capacity to cover occasional generation drops or surges.

4.1.2. Three Years Heatmap

The heatmap depicted in Figure 5 visually represents the load distribution over three consecutive years, focusing particularly on the month of May. In 2020, the load dynamics for May displayed concentrated activity on specific days, with a noticeable surge in loads towards late April and early May, reaching a peak just prior to mid-May, as indicated by a color intensity denoting a load magnitude of up to 17.5 units. Conversely, the load trend for May 2021 showed a more dispersed pattern, with an evident rise in load from mid-April and a peak around early May, albeit with a slightly lower maximum load value of approximately 15 units compared to the previous year. The distribution of loads across the initial half of May suggested a sustained, moderate load profile rather than sharp spikes. Moving to May 2022, the load distribution appeared more uniform throughout the days, with a gradual buildup leading into May and peak load values notably reduced compared to prior years, hovering just above 10 units. The May loads exhibited a consistent and moderate pattern, lacking the pronounced spikes observed in 2020 and 2021.

Indeed, the load feature in May over the years has shown a clear evolution. In 2020, the load was characterized by high intensity and sharp peaks. In 2021, while the load was slightly lower, it remained significant. By 2022, the load pattern had become more balanced and consistent. This progression could potentially indicate an adaptation in load management strategies or a shift in consumption patterns over the years.
This trend analysis provides valuable insights for future load forecasting and energy management planning.

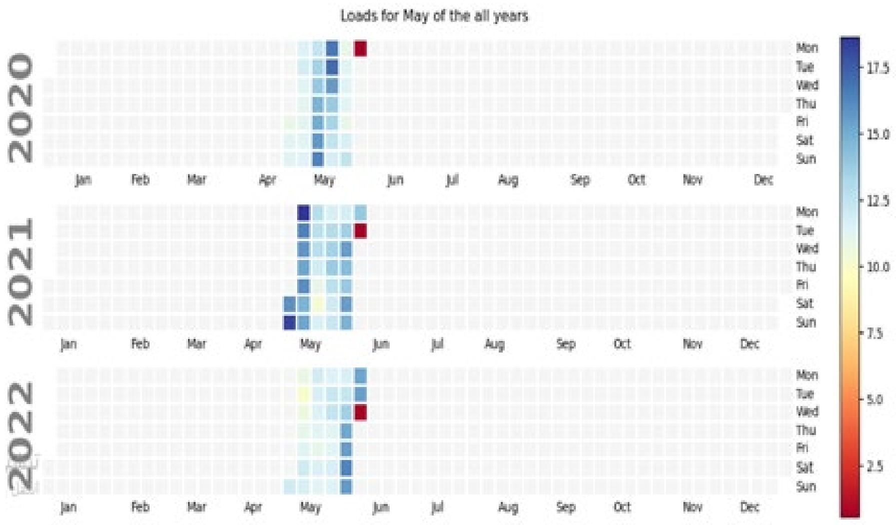

Figure 5. The load patterns for May over three years.

Figure 6 presents the generation patterns for May over three years, each exhibiting distinct trends. In 2020, generation activity mirrored the load pattern with high values concentrated on specific days, peaking at the end of April and early May with values reaching around 14 units. In contrast, 2021 showed a more distributed generation pattern, starting to increase in mid-April and peaking in early May with a maximum value of around 10 units, aligning with the lower peak loads of that year. The generation values were consistent and spread out, reflecting the distributed load pattern of 2021. By 2022, the generation activity was more evenly spread across the days with a moderate and consistent trend, peaking at 6 units.

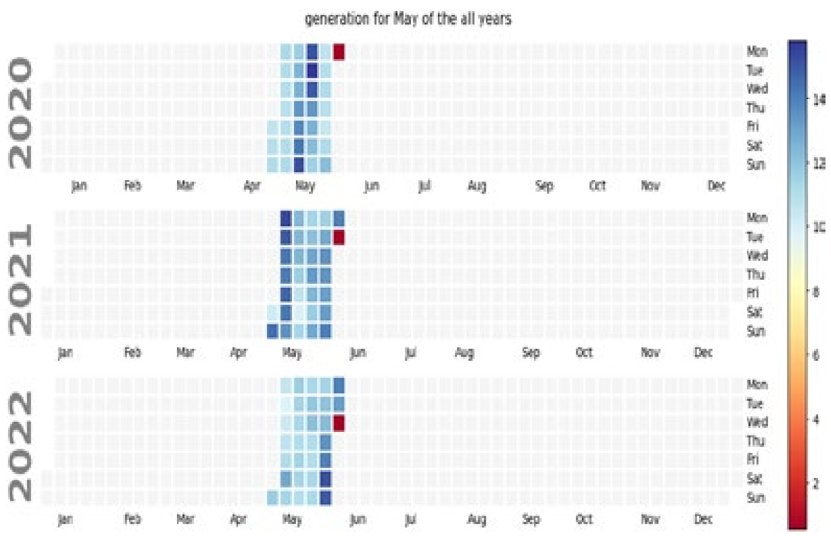

Figure 6. The generation patterns for May over three years.

This pattern aligned with the balanced load pattern of the same year, and the lower generation values reflected the lower peak loads observed. These patterns suggest a shift in generation strategies over the years, transitioning from high and concentrated peaks to a more balanced and consistent trend, indicative of advancements in load management and energy conservation practices.

The dynamic relationship between load and generation, as depicted in Figures 5 and 6, reveals a notable correlation and evolution over time. In 2020, both load and generation demonstrated high concentrations on specific days, reflecting a reactive approach to meet heightened load demands. However, by 2021, these patterns evolved towards greater distribution, indicating a transition towards a proactive and balanced strategy in load and generation management. This trend continued into 2022, where both load and generation patterns exhibited a more moderate and consistent trajectory, signaling an improved equilibrium between load demands and generation capabilities. The decrease in peak values for both load and generation across the years suggests a successful adaptation or transformation in load management strategies, moving away from reactive responses to a more balanced and sustained approach. This evolution indicates progress towards more efficient energy management, with generation becoming more predictable and closely aligned with load necessities.

4.1.3. Trend Analysis

Figure 7 illustrates the trend analysis of both load and generation from 2020 to 2022, revealing a distinct evolution in their relationship and management strategies. In 2020, both load and generation demonstrated high peaks and sudden increases, indicating a strong correlation between generation and load demand. However, the sharp load spikes were not always matched by corresponding increases in generation, which suggests a reactive and lagging response in the generation system to meet demand. This reactive approach highlights the challenges in synchronizing generation with fluctuating load needs.

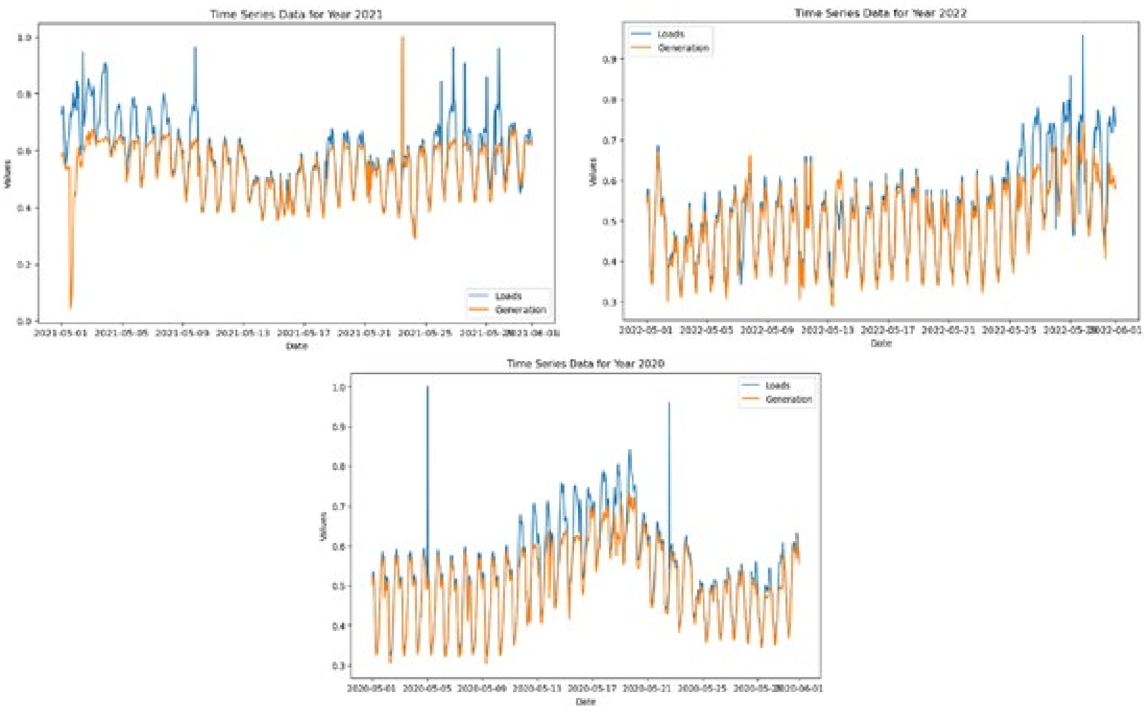

Figure 7. The trend analysis of both load and generation from 2020 to 2022.

In 2021, although the peaks in both load and generation were somewhat reduced, they remained significant, hinting at adaptations in response to the previous year's high demand. Daily cyclical patterns persisted, but more pronounced discrepancies emerged between load and generation, with anomalies appearing more frequently. These patterns suggest that, while there was an effort to balance load and generation, variability in daily load demands continued to pose challenges. By 2022, both load and generation exhibited more consistent and moderate patterns, signaling stabilization and improvements in the management of these parameters. The intensity of peak values for both load and generation was at its lowest across the three years, which could indicate enhanced management practices or a reduction in peak demand and generation requirements. The daily distribution became more balanced and dispersed, marking an improvement in the evenness of load and generation management. Additionally, the alignment and synchronization between load and generation improved significantly, with fewer anomalies and better handling of peak demands. This suggests a shift toward a more proactive and balanced management strategy, possibly due to refined load forecasting and generation planning techniques.

4.1.4. Seasonality analysis of the load

As illustrated in Figures 8, 9, and 10, distinct patterns emerge over the years 2020 to 2022. In 2020, energy consumption in May displayed a flat trend, while 2021 indicated a slight decrease, and 2022 exhibited a slight increase.

This dynamic trend in energy consumption suggests a temporal evolution over these years. All three years demonstrated non-stationarity, as indicated by the ADF and p-values of 0.65, 0.41, and 0.93 for 2020, 2021, and 2022 respectively. These values, exceeding the common significance level of 0.05, suggest the presence of underlying seasonal patterns or trends. Strong daily seasonality was observed across all three years, as evidenced by the sinusoidal patterns in the ACF plots and the significant PACF spikes at daily intervals. In terms of distribution, 2020 and 2021 had a more concentrated distribution around 0.5 to 0.6 kWh. However, 2022 showed a higher concentration of data points at slightly higher consumption levels, suggesting an overall increase in energy consumption in that year.

Lastly, each year exhibited notable spikes indicating occasional high-consumption days, which might correspond to specific events or anomalies. This underscores the importance of considering these outliers when forecasting future load demands.

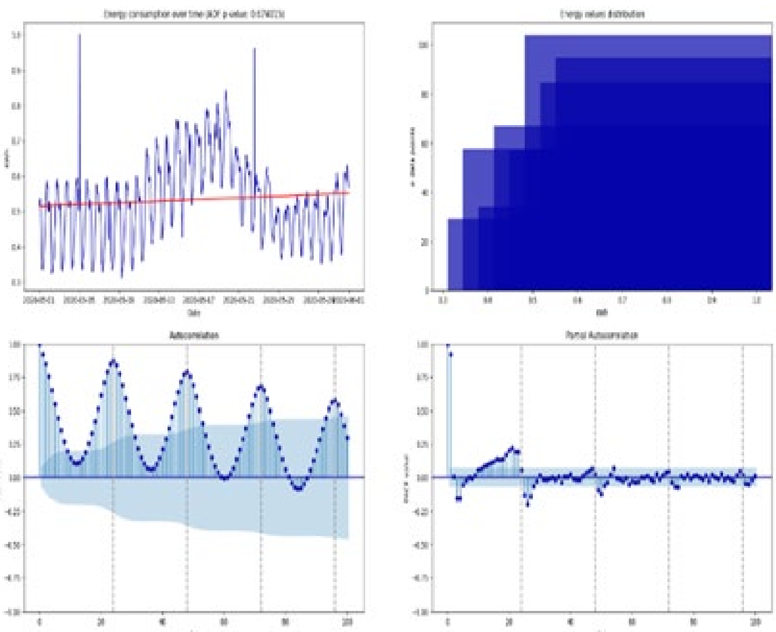

Figure 8. Load distribution, dynamic trend, and seasonal patterns in energy consumption during May 2020.

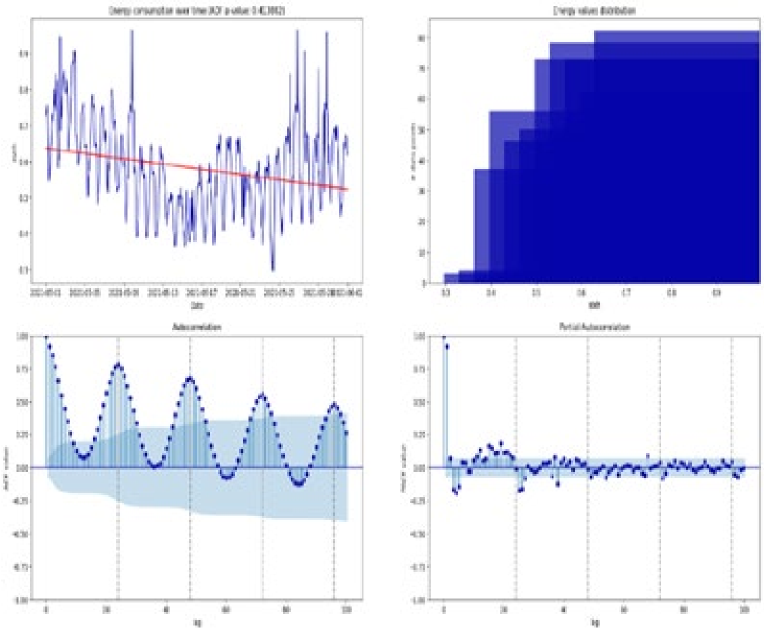

Figure 9. Load distribution, dynamic trend, and seasonal patterns in energy consumption during May 2021.

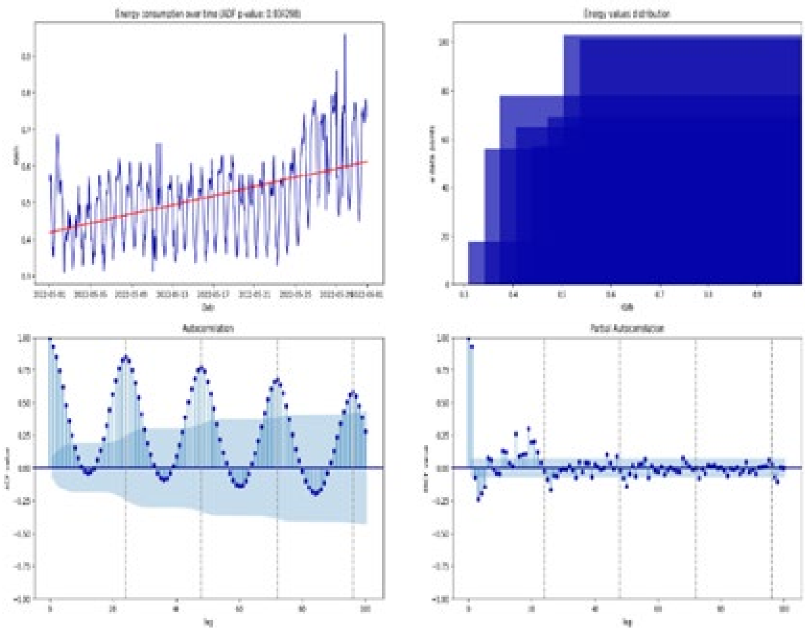

Figure 10. Load distribution, dynamic trend, and seasonal patterns in energy consumption during May 2022.

4.1.5. Seasonality Analysis for generation

As illustrated in Figures 11, 12, and 13, the seasonality analysis of the generation feature, as represented in Figures 6, 7, and 8, uncovers both similarities and disparities across the years 2020 to 2022. All three years exhibit clear seasonality with cyclical patterns discernible in the time series plots. This is further substantiated by the pronounced sinusoidal patterns in the ACF plots and the significant peaks at the first lag in the PACF plots, indicating a strong autocorrelation with the preceding value. However, when each year is examined individually, differences surface. In 2020, the generation data displays more variability and abrupt dips, suggesting potential outliers or anomalies within the data. Conversely, the seasonality pattern in 2021 is more regular and smoother. By 2022, an upward trend becomes evident in addition to the seasonal patterns, indicating an overall escalation in energy consumption over the period. In terms of trend analysis, 2020 and 2021 do not exhibit a strong trend component, with the variations primarily attributable to seasonal fluctuations. However, 2022 showcases a noticeable upward trend alongside the seasonal variations, suggesting an overall increase in energy consumption during this period.

Lastly, the stationarity analysis reveals that all three years have high ADF values, indicating non-stationarity. Specifically, the ADF values for 2020, 2021, and 2022 were 0.76, 0.16, and 0.65 respectively, all of which exceed the common significance level of 0.05. This non-stationarity could be attributed to the seasonal components, and addressing this will be crucial for accurate modeling.

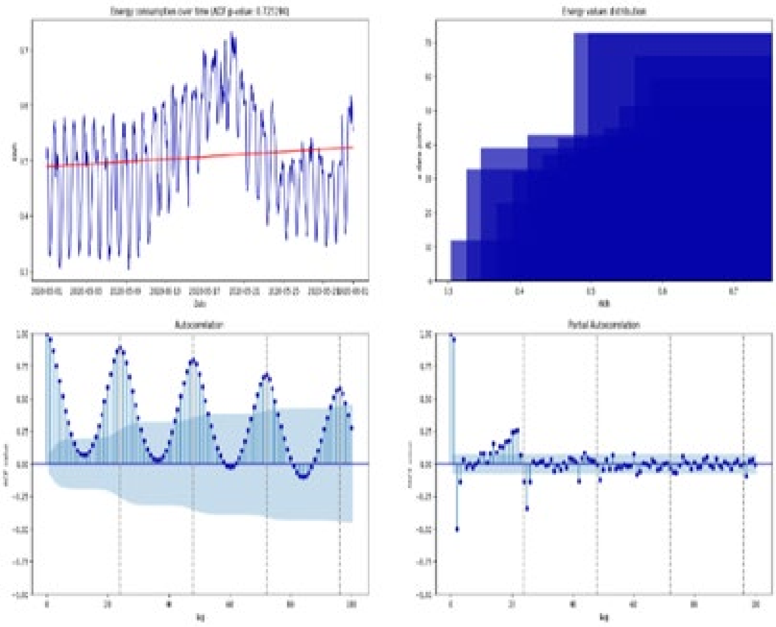

Figure 11. Electric generation characteristic distribution, dynamic trend, and seasonal patterns during May 2020.

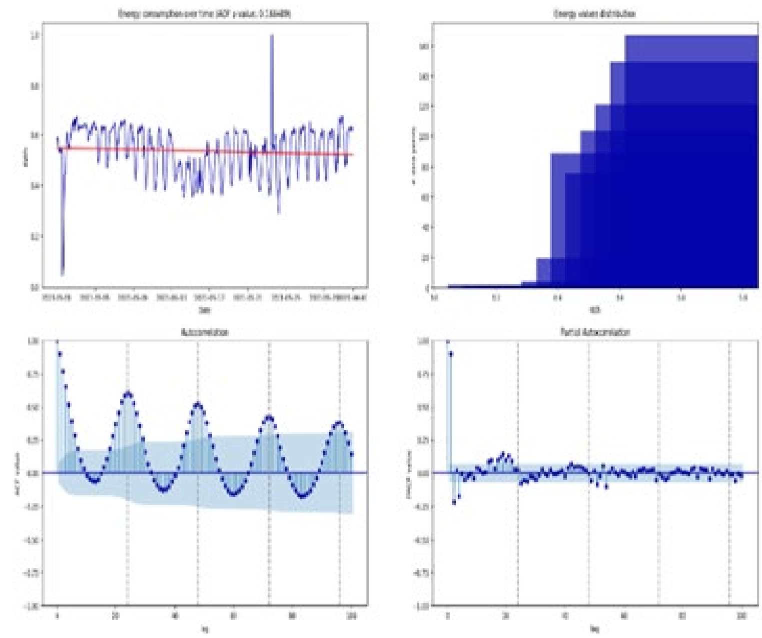

Figure 12. Electric generation characteristic distribution, dynamic trend, and seasonal patterns during May 2021.

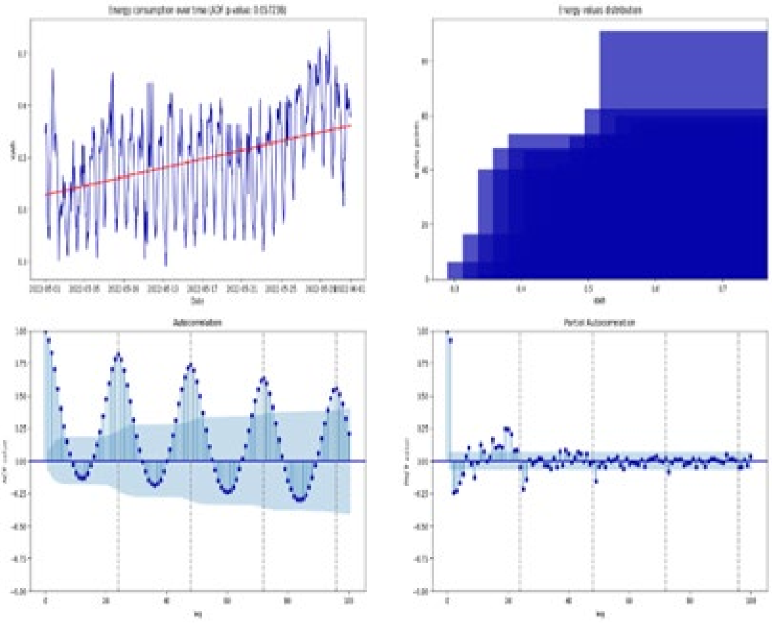

Figure 13. Electric generation characteristic distribution, dynamic trend, and seasonal patterns during May 2022.

Based on the previously discussed stability of the time series for both load and electrical generation, a differencing technique will be applied. The focus will primarily be on the time series related to the load. For this purpose, the Auto ARIMA model will be employed. This automatic model identifies the optimal order of differencing and the optimal orders for the Autoregressive (AR) and Moving Average (MA) components. The model aims to minimize a given information criterion to achieve optimal results. This approach allows for a more accurate and efficient analysis of the time series data, thereby enhancing the reliability of the forecasting process.

5. FITTING MODELS

To make the time series stationary and suitable for modeling, we will follow the approach described:

- Use auto_arima: Automatically determines the best ARIMA models parameters, including order of difference.
- Model Selection Criteria: Lower AIC and BIC values indicate a better model fit.
- Check the p-value to be less than 0.05.
- Time Series Diagnostics: plot_diagnostics is used to ensure that the time series is stationary and the residuals are white noise
- Autocorrelation analysis: Examination of plots of the autocorrelation function (ACF).

5.1. Fitting the ARIMA Model

The ARIMA model was set up with an order of (3,1,2). In this context, '3' denotes the autoregressive terms (p), '1' indicates the degree of differencing (d), and '2' represents the moving average terms (q). These parameters were selected based on a p-value of 0.000, an Akaike Information Criterion (AIC) value of -7493.368, and a Bayesian Information Criterion (BIC) value of -7459.104. Subsequently, an autoregressive (AR) model was fitted to the differenced training data using the method of maximum likelihood estimation. This approach allows for the efficient estimation of the model parameters, providing a robust framework for forecasting future values.
Figure 14 presents the diagnostic plots for the time series model generated by the ARIMA, which assist in evaluating the model's goodness of fit and the assumptions of the residuals:

1. Standardized Residuals (Top Left): The residuals from the model are standardized to have a

mean of zero and a standard deviation of one. They appear to be randomly scattered around zero, suggesting that the model has adequately captured the underlying structure of the data.

2. Histogram plus Estimated Density (Top Right): This plot shows a histogram of the residuals, the estimated density (KDE), and the normal distribution curve (N (0,1)). The residuals roughly follow the normal distribution, with some deviations in the tails indicating slight non-normality.
3. Normal Q-Q Plot (Bottom Left): This plot compares the quantiles of the standardized residuals to the theoretical quantiles of a normal distribution. The points lie approximately along the red line, indicating that the residuals follow a normal distribution, with slight deviations at the tails.
4. Correlogram (ACF of Residuals) (Bottom Right): This plot shows the autocorrelation function (ACF) of the residuals. Most of the autocorrelations fall within the 95% confidence interval, suggesting that the residuals are uncorrelated and resemble white noise, indicating that the model has adequately captured the time series dynamics.

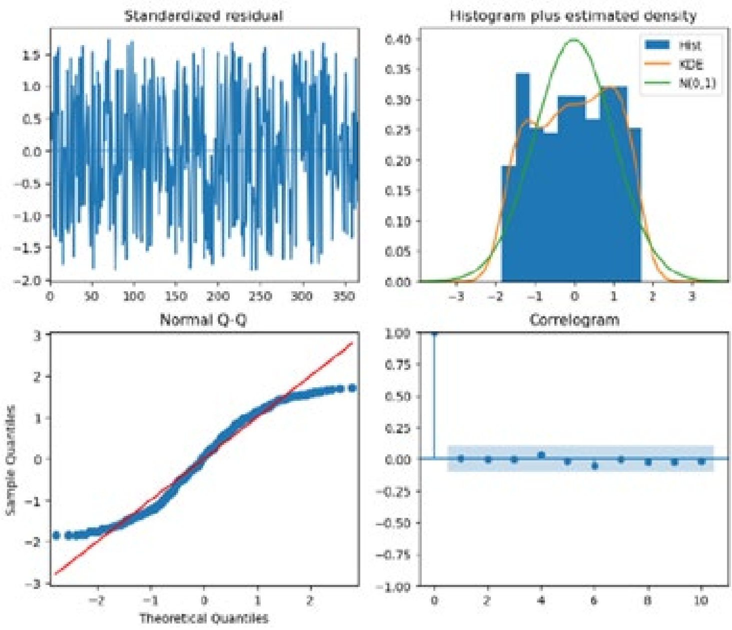

Figure 14. The diagnostic plots for the time series model generated by the ARIMA.

These diagnostics collectively suggest that the ARIMA model provides a good fit to the data. The absence of significant correlations among the residuals indicates that the model captures the underlying patterns in the data well and can be used for forecasting purposes.

5.2. Fitting the SARIMA Model

In this case, the model is configured with the order (2, 1, 0) x (1, 1, [1], 12). Here's what each parameter signifies:

- $p = 2$: This represents the number of autoregressive terms. Autoregressive terms are lags of the dependent variable, i.e., previous values of the time series.
- $d = 1$: This is the degree of differencing. Differencing is used to make the time series stationary.
- q = 0: This denotes the number of moving average terms. Moving averages use past errors to forecast future values.
- $P = 1$: This is the number of seasonal autoregressive terms.
- $D = 1$: This is the degree of seasonal differencing.
- $Q = 1$: This is the number of seasonal moving average terms.

- 12: This represents the length of the seasonal cycle, indicating monthly seasonality in this case.

The parameters were chosen based on a p-value of 0.000, AIC of -7611.031, and BIC of -7576.800. The model was fitted to the differenced data using maximum likelihood estimation, ensuring an accurate representation of the underlying patterns.

The diagnostic plots of the SARIMA model, as shown in Figure 15, are crucial for determining the validity of the model fit and assessing the quality of the forecast. Here's a breakdown of the **key aspects of these diagnostic plots:**

1. Standardized Residuals: The random pattern around zero suggests that the model is capturing much of the structure in the data. However, the presence of large spikes indicates that some patterns or irregularities may not be fully captured.
2. Histogram plus Estimated Density: A roughly normal distribution of residuals indicates that the model's predictions are generally unbiased and that it captures the underlying distribution of the data well. However, the deviations suggest that some aspects of the data distribution may be missed.
3. Normal Q-Q Plot: This plot indicates that while the model captures the central patterns in the data well, it struggles with the extremes (i.e., the tails). This could suggest the model is good for general patterns but might miss outliers or extreme values.
4. Correlogram: The significant autocorrelation at lag 1 suggests that the model does not fully capture short-term dependencies in the data. However, the lack of significant autocorrelation at higher lags indicates that longer-term patterns are well captured.

As a result, the SARIMA model appears to be reasonably good at capturing the main patterns in the data, although there are some areas where it could potentially be improved.

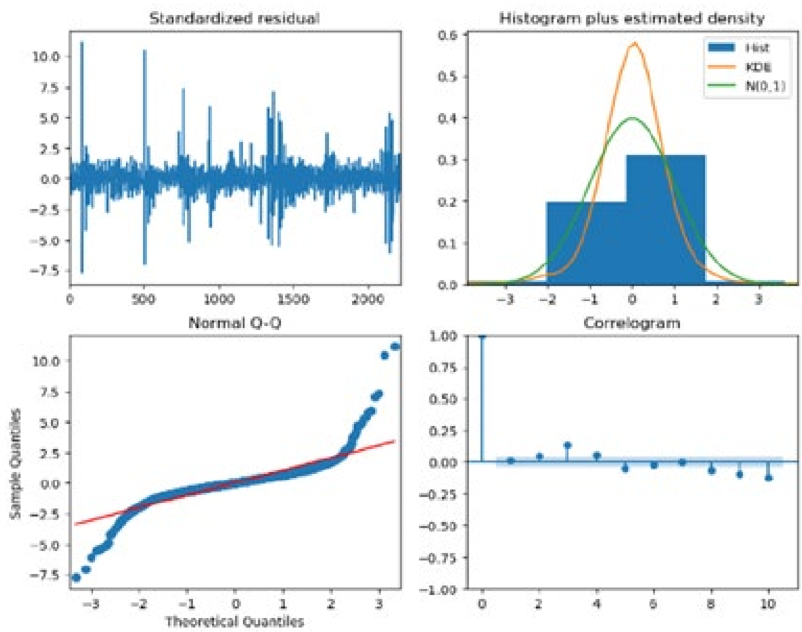

Figure 15. The diagnostic plots for the time series model generated by the SARIMA.

5.3. Fitting the DRARIMA Model

It is a type of time series model that expresses the dependent variable as a linear function of its previous values. In this case, the model is configured with an order of (2,1,0), which signifies that it is an autoregressive (AR) model. Here's what each parameter means:

- p = 2: This represents the number of autoregressive terms, which are lags of the dependent variable.
- d = 1: This is the degree of differencing, which is used to make the time series stationary.
- q = 0: This denotes the number of moving average terms, which are not present in this model as it is purely an AR model.

These parameters were chosen based on a p-value of 0.000, an AIC value of -5893.162, and a BIC value of -5111.100. The model was then fitted to the differenced training data using maximum likelihood estimation.

The diagnostic plots of the DRARIMA model, displayed in Figure 16, are essential for validating the model fit and evaluating the forecast quality. Here's a brief interpretation of these plots:

1. Standardized Residuals: The residuals fluctuate around zero with uniform variance, suggesting an unbiased forecast.
2. Histogram and Density Plot: The residuals appear to follow a Gaussian distribution with a mean of zero, indicating that the model's predictions are generally unbiased.
3. Normal Q-Q Plot: The residuals align closely with the red line, suggesting a proper and un skewed distribution.
4. Correlogram (ACF Plot): The absence of significant autocorrelation at lag 1 implies that the residuals are not auto-correlated, indicating that the model has captured the dependencies in the data.

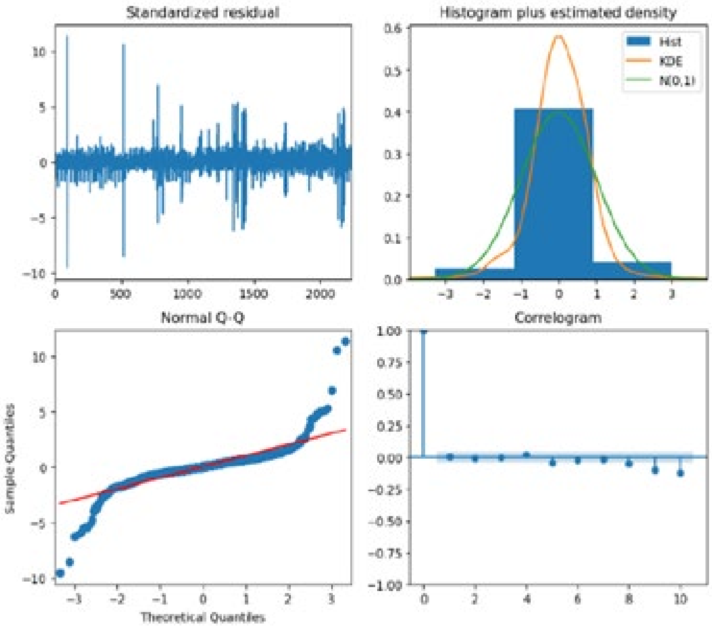

Figure 16. The diagnostic plots for the time series model generated by the RegARIMA.

The RegARIMA model is good but not perfect. It does a reasonable job of fitting the data, as evidenced by the random and uncorrelated residuals. However, the slight deviations from normality indicate that there might be room for improvement.

5.4. Model Training and Testing

This study aimed to develop and assess a range of ARIMA models for forecasting the electrical load time series in May 2022. The models were trained using electrical load data from May 2020 and May 2021, while the performance was evaluated on the data from May 2022. To account for external influences on the load, such as generation, exogenous variables were introduced into the DRARIMA model.

5.5. Performance Comparison of Models

The predictions from each model were compared with the actual load data for May 2022. As illustrated in Figure 17:

• Original Data: This is depicted by a blue line, which represents the actual electrical load.
• ARIMA Forecast: This is represented by a yellow dashed line. It captures the overall trend but tends to underestimate the peaks and valleys.
• SARIMA Forecast: This is shown as a green dashed line. It accounts for seasonal patterns but has higher errors compared to the ARIMA model.
• DRARIMA Forecast: This is displayed as a red dashed line. It incorporates exogenous variables but has the highest prediction error among all the models.

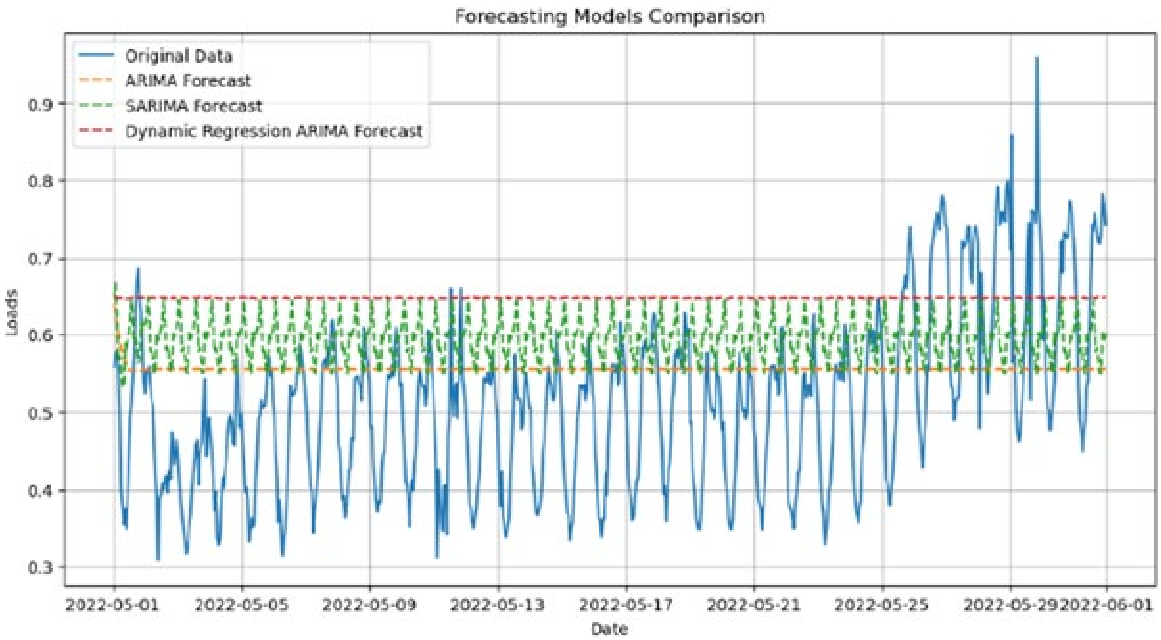

Figure 17. Forecasts from each model combined with actual load data for May 2022.

According to the evaluation metrics outlined in Table 2, the ARIMA model outperforms the other models, including SARIMA and DRARIMA, in terms of predictive performance. The ARIMA model, with the lowest values for MSE, MAE, RMSE, and MAPE, proves to be the most accurate in forecasting the electrical load for May 2022, achieving an estimated accuracy of 78.88%, as illustrated in Figure 17. In contrast, the SARIMA model, despite its incorporation of seasonality, did not perform as effectively as the ARIMA model, resulting in a prediction accuracy of 73.86%. Furthermore, the DRARIMA model, despite the inclusion of exogenous variables, demonstrated higher error rates and a diminished accuracy of 65.36%. This implies that the integration of the chosen exogenous variables did not significantly enhance the forecast accuracy.

Table 2. Evaluating the performance of arima models for electrical load demand for may 2022.

Models	MSE	MAE	RMSE	MAPE	Min Error	Max Error
ARIMA	0.014	0.09	0.12	21.12	-0.24	0.40
SARIMA	0.019	0.11	0.13	26.14	-0.30	0.40
DRARIMA	0.030	0.15	0.17	34.63	-0.33	0.30

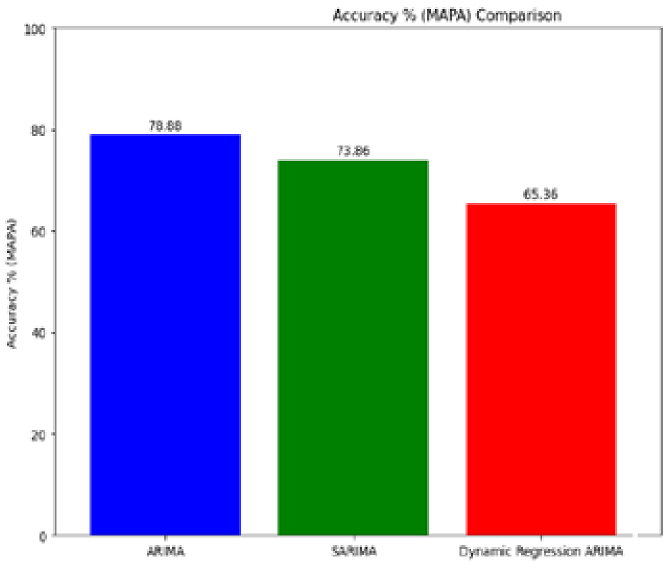

Figure 18. Estimating the performance of ARIMA models for forecasting electrical demand in terms of accuracy.

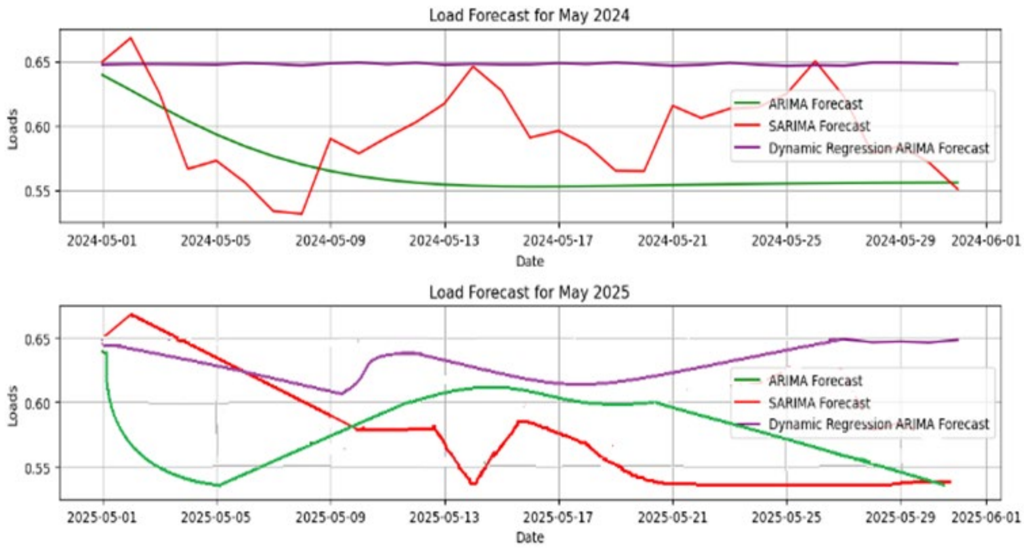

Figure 19. Future energy demand forecasting performance of three models for May 2024 and May 2025.

5.6. Future Forecasting

This section evaluates the forecasting performance of three proposed models for May 2024 and May 2025. The aim is to understand each model's predictive capabilities and assess their reliability and accuracy as show in Figure 18.

A. May 2024 Forecasts:

• ARIMA: Predicts a steady decline in load, showing a consistent downward trend.

• SARIMA: Captures more variability, reflecting short-term variations and seasonal patterns.

• DRARIMA: Forecasts a relatively stable load, suggesting that the inclusion of exogenous variables might stabilize predictions but possibly underestimate variability.

B. May 2025 Forecasts:

• ARIMA: Indicates a downward trend similar to the previous year, suggesting a continual reduction in load.

• SARIMA: Reflects a higher degree of variability, indicating its capacity to model seasonality and short-term dynamics effectively.

• DRARIMA: Predicts an overall upward trend, suggesting potential growth or stability in external factors impacting the load.

5.7. Strength and Limitation of all Models

The future forecasts highlight the different characteristics and strengths of each model: The ARIMA model offers a clear and straightforward forecast with a consistent trend, making it useful for understanding long-term tendencies. However, it falls short in capturing seasonal variations and short-term fluctuations, leading to potentially oversimplified predictions. On the other hand, the SARIMA model effectively addresses seasonality and short-term variability, enhancing its responsiveness to periodic changes in the load, although its increased variability might lead to overfitting and sensitivity to noise. The DRARIMA model incorporates exogenous variables, providing a nuanced perspective on future load predictions by stabilizing the forecast and accounting for external influences. Despite this, it may underestimate inherent fluctuations due to the specific exogenous variables used. Therefore, choosing the appropriate model depends on the specific requirements of the analysis: the ARIMA model is suited for long-term trend analysis, the SARIMA model for
detailed short-term predictions considering seasonality, and the DRARIMA model for scenarios where external factors significantly impact the load, with careful selection of exogenous variables essential for accurate forecasting.

6. THE EXPECTED DIRECTION OF THE DATA FROM THE PAPER IN LIGHT OF SIMILAR WORKS

The paper's focus on forecasting electricity load in Benghazi, Libya, using ARIMA models, allows us to anticipate certain data characteristics based on similar studies and the general principles of load forecasting:

• Seasonality: Electricity demand often exhibits strong seasonal patterns. We can expect higher loads during the summer months due to increased air conditioning use, a common finding in similar studies in hot climates. Daily and weekly seasonality are also anticipated, with daily peaks aligning with typical usage patterns and differences between weekday and weekend demand. The paper's mention of peak demand in summer reinforces this expectation.

• Trend: Given Libya's developing economy and growing population, an upward trend in electricity demand is likely. Similar studies in developing regions often observe increasing trends in energy consumption. However, the rate of this trend might not be uniform and could be influenced by economic fluctuations, energy efficiency improvements, or policy changes.

• Noise and Outliers: Real-world data is inherently noisy. Random fluctuations and outliers caused by unusual events (e.g., holidays, extreme weather) are expected. The paper's methodology includes outlier removal, which is crucial for accurate model training.

• Non-linearity: Electricity load data often exhibits non-linear behavior due to complex interactions between various factors. ARIMA models have limitations in capturing non-linearity. The paper should investigate the potential presence of non-linearity in the data and consider alternative models or hybrid approaches if necessary. The literature suggests that machine learning models or hybrid approaches might be better suited for capturing non-linear relationships.

In light of similar works, the paper's expected data should exhibit seasonal patterns, potentially an upward trend, inherent noise and outliers. Furthermore, the possibility of non-linear relationships should be investigated. These anticipated data characteristics inform the selection and evaluation of appropriate forecasting models, crucial for accurate load forecasting and efficient load shedding management in Benghazi.

7. CONCLUSION AND FUTURE WORK

In this study, we explored various time series forecasting models to predict the electrical load for May over several years. We employed ARIMA, SARIMA, and DRARIMA models, assessing their performance using historical data from May 2020 and 2021 to forecast loads for May 2022. Additionally, we extended the forecasting to predict loads for May 2024 and May 2025. The performance evaluation metrics, including MSE, MAE, RMSE, MAPE, minimum and maximum errors, and accuracy percentage, provided a quantitative basis for comparing these models. Among the models, ARIMA demonstrated a balance between simplicity and accuracy, while SARIMA showed strength in capturing short-term variability. The DRARIMA model highlighted the importance of considering external factors. Future research can build upon the findings of this study by addressing several key areas: Identify and integrate relevant exogenous variables (e.g., weather conditions, economic indicators) that can significantly impact electrical loads. This can enhance the predictive power of the DRARIMA model. Combine the strengths of multiple models (e.g., hybrid ARIMA-LSTM) to create more robust forecasting systems that leverage both statistical and machine learning techniques.

In light of our research findings and potential future work, we propose the following recommendations for those engaged in electrical load forecasting:

1) Data Quality and Preprocessing: Prioritize the collection of high-quality data and thorough preprocessing to reduce noise and improve model accuracy.

2) Model Selection: Choose models judiciously, considering the specific needs of the forecasting task. For example, opt for SARIMA when detailed short-term variability analysis is required, and consider DRARIMA when external factors play a significant role.

Author Contributions: All authors have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding: There is no funding for the article.

Data Availability: The data are available at request.

Conflicts of Interest: The authors declare that they have no conflict of interest.

REFERENCES

[1]. El-Khozondar, H.J., et al., A smart energy monitoring system using ESP32 microcontroller. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 2024. 9: p. 100666.

[2]. (EIA), I.-U.S.E.I.A. Country Analysis Executive Summary: Libya. 2024 9-5-2022 [cited 9-11-2024.

[3]. Almaktar, M., M.J.R. Shaaban, and S.E. Reviews, Prospects of renewable energy as a non-rivalry energy alternative in Libya. 2021. 143: p. 110852.

[4]. Yahya, W., et al., Future study of renewable energy in Libya. 2020. 7(10): p. 1-6.

[5]. Ayuk, N., Billions at Play: The Future of African Energy and Doing Deals. 2020: Made For Success Publishing.

[6]. Ihbal, A.J.J.o.P. and A. Sciences, Forecasting Short-Term Peak Load Demand in the Libyan Power Grid using Multiple Regression Model. 2023. 22(3): p. 124-128.

[7]. Alnaedh, A., A. Matoug, and S. Al-Behadili, Analysis of the Electrical Power Distribution System in Tarhuna City (Libya) to Avoid Future Loads Using NEPLAN Software. Wadi Alshatti University Journal of Pure and Applied Sciences, 2024. 2(1): p. 23-33.

[8]. El-Osta, W. and U.J.S.D.R. Elghawi, Assessment of Energy Intensity Indicators in Libya: Case Study. 2020. 2(1): p. p9-p9.

[9]. Li, L., et al., Impact of natural and social environmental factors on building energy consumption: Based on bibliometrics. 2021. 37: p. 102136.

[10]. Huang, Y., et al., Analyzing entropy features in time-series data for pattern recognition in neurological conditions. 2024: p. 102821.

[11]. Arumugam, V. and V.J.I. Natarajan, Mesures, Métrologies, Time Series Modeling and Forecasting Using Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average Models. 2023. 22(4).

[12]. Kontopoulou, V.I., et al., A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. 2023. 15(8): p. 255.

[13]. Kien, D.T., et al., Application of Sarima model in load forecasting in Hanoi city. 2023. 13(3): p. 164-170.

[14]. Shaqiri, F., R. Korn, and H.-P.J.E. Truong, Dynamic Regression Prediction Models for Customer Specific Electricity Consumption. 2023. 4(2): p. 185-215.

[15]. Chodakowska, E., J. Nazarko, and Ł.J.E. Nazarko, ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise. 2021.

[16]. Czapaj, R., J.W. Kaminski, and M.J.E. Sołtysik, A Review of Auto-Regressive Methods Applications to Short-Term Demand Forecasting in Power Systems. 2022.

[17]. López, J.C., M.J. Rider, and Q.J.I.T.o.P.S. Wu, Parsimonious Short-Term Load Forecasting for Optimal Operation Planning of Electrical Distribution Systems. 2019. 34: p. 1427-1437.

[18]. Amerise, I.L. and A.J.E.S. Tarsitano, Point and interval forecasts of electricity demand with Reg-SARMA models. 2021. 14: p. 89-109.

[19]. Crujido, L.J.C., et al., Day-Ahead Electricity Load Forecasting with Multivariate Time Series. 2023.

[20]. Dubey, A.K., et al., Study and analysis of SARIMA and LSTM in forecasting time series data. 2021. 47: p. 101474.

[21]. Yin, C., et al., SARIMA-Based Medium- and Long-Term Load Forecasting. 2023.

[22]. Durmus Senyapar, H.N. and A.J.S. Aksoz, Empowering Sustainability: A Consumer-Centric Analysis Based on Advanced Electricity Consumption Predictions. 2024. 16(7): p. 2958.

[23]. Velasco, L.C.P., et al., A hybrid model of autoregressive integrated moving average and artificial neural network for load forecasting. 2019. 10(11): p. 14-16.

[24]. Kaytez, F.J.E., A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. 2020. 197: p. 117200.

[25]. Somu, N., G.R. MR, and K.J.A.E. Ramamritham, A hybrid model for building energy consumption forecasting using long short term memory networks. 2020. 261: p. 114131.

[26]. Gasparin, A., S. Lukovic, and C.J.C.T.o.I.T. Alippi, Deep learning for time series forecasting: The electric load case. 2022. 7(1): p. 1-25.

[27]. Lara-Benítez, P., et al., Temporal convolutional networks applied to energy-related time series forecasting. 2020. 10(7): p. 2322.

[28]. Farkash, H.M., et al. Medium-term load forecasting for the city of benghazi using an artificial neural network based time series approach. in 2023 IEEE 3rd International Maghreb Meeting of the

Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA). 2023. IEEE.

[29]. Ali, S.B., A. Tahir, and B.T.M. Obadi. Application of Voltage Control Techniques in South Western Libyan Network to Enhance Voltage Stability with Future Load Forecasting. in 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA). 2022. IEEE.

[30]. Alarbi, A.A., An innovative residential demand side management method to the Libyan electricity network. 2022, Loughborough University.

[31]. Hamouda, A.A., et al., STUDY AND ASSESSMENT OF RESIDENTIAL ELECTRICAL LOAD: A CASE STUDY IN SAWKNAH (SUKNA) CITY-LIBYA. 2021. 5(2).

[32]. Chaaraoui, S., et al., Day-ahead electric load forecast for a ghanaian health facility using different algorithms. 2021. 14(2): p. 409.

[33]. Guefano, S., et al., Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models. 2021. 214: p. 118791.

[34]. Nti, I.K., et al., Electricity load forecasting: a systematic review. 2020. 7: p. 1-19.

[35]. Hoffmann, M., et al., A review on time series aggregation methods for energy system models. 2020. 13(3): p. 641.

[36]. Fan, C., et al., A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. 2021. 9: p. 652801.

[37]. Singh, D. and B.J.A.S.C. Singh, Investigating the impact of data normalization on classification performance. 2020. 97: p. 105524.

[38]. Herwanto, H.W., et al. Comparison of min-max, z-score and decimal scaling normalization for zoning feature extraction on Javanese character recognition. in 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE). 2021. IEEE.

[39]. Mulla, S., C.B. Pande, and S.K.J.W.R.M. Singh, Times Series Forecasting of Monthly Rainfall using Seasonal Auto Regressive Integrated Moving Average with EXogenous Variables (SARIMAX) Model. 2024. 38(6): p. 1825-1846.

[40]. Ramadan, A.I.H.A., et al., Energy communities from municipality level to province level: A comparison using autoregressive integrated moving average model. 2023. 17(8): p. 188-194.

[41]. Hossain, M.A., et al., Optimized forecasting model to improve the accuracy of very short-term wind power prediction. 2023. 19(10): p. 10145-10159.

[42]. Riaz, M., et al., Epidemiological Forecasting Models Using ARIMA, SARIMA, and Holt–Winter Multiplicative Approach for Pakistan. 2023. 2023.

[43]. Wang, L.J.O.J.o.S., Time Series Analysis and Prediction of COVID-19 Pandemic Using Dynamic Harmonic Regression Models. 2023. 13(2): p. 222-232.

[44]. Dey, B., et al., Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models. 2023. 9: p. 411-418.

[45]. Suleiman, M., et al., Modelling Nigeria Crude Oil Prices using ARIMA Time Series Models. 2023. 5(1).

[46]. Huang, X. and H.L.J.N.D. Shang, Nonlinear autocorrelation function of functional time series. 2023. 111(3): p. 2537-2554.

[47]. Weiß, C.H., et al., Partial autocorrelation diagnostics for count time series. 2023. 25(1): p. 105.

[48]. Katambire, V.N., et al., Forecasting the Traffic Flow by Using ARIMA and LSTM Models: Case of Muhima Junction. 2023. 5(4): p. 616-628.

[49]. Robeson, S.M. and C.J.J.P.o. Willmott, Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. 2023. 18(2): p. e0279774.

[50]. Xiao, R., et al., Predict stock prices with ARIMA and LSTM. 2022.

[51]. Ospina, R., et al., An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. 2023. 11(14): p. 3069.