Evaluating Power Usage Patterns: A Case Study on Time Series Modeling Forecasting in Erbil City 2015–2024

Azhin M. Khudhur[1], Dler H. Kadir[1], Sakar A. Jalal[1], Rebaz O. Yahya[2]

[1]Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Kurdistan Region - F.R. Iraq
[2]Department of Business Administration, Cihan University-Erbil, Kurdistan Region, Iraq

Abstract—Precise electricity consumption predictions are essential for efficient energy management, resource allocation, and power system stability, particularly in expanding urban areas like Erbil. Time series models are crucial tools for capturing trends, seasonal variations, and structural shifts in energy use patterns. This study aims to forecast monthly electricity consumption in Erbil for 2025 using the seasonal autoregressive integrated moving average (SARIMA) approach. Historical monthly electricity consumption data from 2015 to 2024 (120 observations) were analyzed. Preprocessing involved logarithmic transformation to stabilize variance and appropriate differencing to achieve stationarity. Model selection prioritized evaluation criteria such as the Akaike information criterion and mean squared error (MSE). The SARIMA(1,1,1)×(0,1,1)12 model yielded the lowest MSE (0.0487347) and was identified as the optimal model, with statistically significant parameters. The resulting forecasts for 2025 indicate notable seasonality, with predicted monthly averages ranging from 955 MW in January to 916 MW in December. This study provides a validated time series model tailored to Erbil's consumption dynamics, offering a robust foundation for improved energy demand forecasting and resource planning.

***Keywords*—Akaike information criterion, ARIMA model, Electricity consumption forecasting, Mean squared error, Seasonal patterns, Statistical modeling, Time series analysis.**

I. Introduction

Modern society and technological advancements lead to daily fluctuations in electrical demand. These varying patterns in industrial, residential, and commercial loads can cause electricity systems to become overloaded, a reality that also affects Erbil's electrical power grid. Accurately forecasting future electrical load needs in the governorate is essential for determining the necessary capacity for power generation to meet this demand.

Erbil's energy infrastructure has demonstrated both resilience and challenges (Massoudi and Ahmed, 2021), particularly following the 1991 Gulf War and subsequent internal conflicts that left critical northern governorates devastated and disconnected from the national grid in 1994. In the face of these challenges, hydroelectric facilities in Sulaymaniyah have served as vital sources of support, helping to alleviate electricity shortages in both Erbil and Sulaymaniyah. A study conducted on electricity usage in Erbil from 2015 to 2024 aimed to develop a forecasting model for future demand. Time series analysis methods were selected to better understand energy dynamics.

Time series analysis aims to achieve two primary goals. First, it attempts to comprehend or explain the stochastic dynamics that support observable series. Second, it seeks to forecast or anticipate future events within a series by using prior data and maybe including other relevant variables or series. Time series projection is an important part of this study because it uses models to estimate future events according to recognized trends in existing data, making data points easier to predict before they are gathered (Chatfield, 1984).

A key idea in time series analysis is stationarity, which represents the consistency of a series' statistical characteristics throughout the sequence. The mean, variance, and autocorrelation of a stationary time series are all constant, facilitating analysis and improving forecast accuracy. A crucial concept in time series analysis is stationarity, which denotes that the statistical characteristics of a series remain constant over time. A stationary time series makes analysis

Cihan University-Erbil Journal of Humanities and Social Sciences (CUEJHSS)
Volume IX No. 1 (2025) 9 pages
DOI: 10.24086/cuejhss.vol9n1y2025.152-160
Received: 15 February 2025; Accepted: 09 April 2025; Regular research paper: Published 1 May 2025
*Corresponding author's e-mail: azhin.khudhur@su.edu.krd

easier and improves forecast reliability because its mean, variance, and autocorrelation are all stable (Brockwell and Davis, 2002; Kadir, 2018).

Statistical forecasting, including various techniques such as regressing analysis, classical decomposition, Box and Jenkins, and smoothing methods, offers varying levels of precision. The lowest error determines the precision of a forecast. The proper forecasting technique is chosen based on factors such as forecasting interval, duration, time series characteristics, and size (Makridakis et al., 1998). This study aims to collect useful information about Erbil's power use. This will help us obtain insight into the area's electrical distribution and make recommendations for eco-friendly energy alternatives.

II. Literature Review

Accurate electricity consumption forecasting is crucial for grid management and planning, and numerous methodologies have been explored in the literature. Time series analysis, particularly the autoregressive integrated moving average (ARIMA) model and its seasonal variant (SARIMA), has emerged as a prominent approach due to its ability to capture temporal dependencies, trends, and seasonality inherent in energy usage data. While this study focuses on aggregate time series, it is worth noting that other methodological frameworks, such as panel data analysis, are also employed for building efficient models when data across multiple entities over time is available (Al-Zawbaee and Mahmood, 2023).

According to statistical criteria, Mahmood (2016) found that the Seasonal Autoregressive Integrated Moving Average (SARIMA) Model (1,1,0)×(0,1,1)12 performed better than the Multilayer feed-forward neural network model in predicting energy consumption in the Kurdistan Region for 12 months in 2016. In contrast to the ANN model, which had an RMSE of 362, this model has the lowest RMSE of 153.9.

Kadir (2020) discovered that the ARIMA(1, 1, 1)×(0, 1, 1)12 model with an appropriate minimum Root Mean Square (RMSE) which was equal to 22.7532 could forecast the distribution of power usage for homeowners, businesses, and factories in the Erbil region station from 2014 to 2018, using the Box-Jenkins procedure for establishing the model. This model forecasts monthly usage for the forthcoming 2019 year, assisting decision-makers in establishing priorities for power demand management.

Mahia et al. (2019) proposed that ARIMA models are frequently used to forecast several fields involving economics, stock markets, marketing, industrial output, and social dynamics. The statistical analysis methodology is suitable for short-term forecasting and requires at least 40 previous data points. It outperforms standard forecasting approaches in terms of efficiency, robustness, and accuracy.

Sim et al. (2019) predicted University Tun Hussein Onn Malaysia power usage for 2019; the SARIMA approach was used in SPSS software with the Box-Jenkins method and Expert Modeler. To create the models, 120 data were collected between January 2009 and December 2018. SARIMA(0, 1, 1)×(0, 1, 1)12 is the best model of the two approaches.

Ravaz and Marwan (2020) found that the SARIMA(0.1,1)×(1,1,2) 12 time series was the best expression as it had the lowest values for the following criteria and passed a fit test. The models were stationary, which means that the errors of the particular model were random, and after testing the stationary of the suggested model, they predicted the amount of demand for electrical energy for a period of 12 months (2019–2020) to use them in the future planning process.

Omer et al. (2021) used the discrepancy between the prediction findings obtained with the optimal parameters and the assaying data to determine the forecast's viability by performing normalcy and randomness tests. According to the parameterization results, the optimal MAPE in DES Brown is 9.23616%, and the optimal value of is 0.22. The ideal value in DES Holt is 0.95, the ideal is 0.05, and the ideal MAPE is 8.08586%. A DES Holt approach has a smaller MAPE than a DES Brown technique. Trials of feasibility have shown that both methods are highly predictive. Based on the evaluation process and the MAPE value, DES Holt's was determined to be the main prediction model.

Elsaraiti et al. (2021) presented ARIMA models with various parameter values for predicting power consumption. The three ARIMA models, which are excellent and resilient for developing a trustworthy model, are explored to anticipate power consumption to meet the desired level of performance. The results suggest that the ARIMA(1,1,1) model is accurate, consistent, and adequate for forecasting power consumption.

To determine the best model for predicting the demand for electricity load over the next 1, 2, and 3 months at a Malaysian university, Mansor et al. (2021) compared four error measures: Mean squared error (MSE), Root MSE, mean absolute percentage error, and geometric root MSE. To obtain the best results, they looked at two forecasting techniques. The SARIMA and Holt-winters exponential smoothing are the methods used. In terms of error measurement, the SARIMA(0,0,1)×(1,0,0) model performed better than the other 12 models.

Sosa et al. (2021) used the ARIMA approach and the Seasonal-ARIMA (SARIMA) pattern to predict long-term power usage using a database of Energy Sold (T1) in kW from 2008 to 2017. The ARIMA(1,1,0) (0,1,1) conjecture model was the best fit for training and testing data with 70:30 scenarios. The MAPE (%) error rate is quite low at 7,966; however, the R-Square value approaches −0.024 due to the absence of observational data.

Mohammed et al. (2022) demonstrated that the mean absolute error and the RMSE error should be as low as possible to select the optimal model. The US Dollar/IQ Dinar series was found to fit best with the ARIMA(2, 1, 0) model. This is the expected interpretation of this exchange rate data series' future, suggesting that it will continue to grow over the next 2 years.

Parreño (2022) applied ARIMA models to predict power usage in the Philippines. The dataset consists of 48 data points, 43 for model building and 5 for forecast assessment. ARIMA(0, 2, 1) is the most effective model for forecasting yearly electrical usage in the Philippines. The model has been validated multiple times to provide accurate projections.

The smallest values based on the RMSE criterion were 19.061, AIC = 630.708 for 72 units in each series in the monthly death and injury rates from auto accidents in Erbil, Iraq, from January 2015 to December 2020, according to Mahmood et al. (2024), who presented VARMA (1,0) as the most suitable model. The statistical model was considered suitable for predicting accidents that would result in fatalities and injuries in 2024.

Ajlouni (2024) utilized the SARIMA model to anticipate Jordan's daily peak electrical usage using hourly peak load data from January 1, 2010, to December 31, 2022. The Box-Jenkins approach yielded the SARIMA model: ARIMA(1,0,1)×(2,1,2)7, which was used to anticipate power load for 7 days.

From April 2013 to February 2023, Mahanta and Talukdar (2024) examined the seasonal impact on power consumption in Assam using SARIMA, one of the most widely used time series techniques. The results of the Canova Hansen test indicated that no seasonal differencing was required for the period in question, and they employed the Augmented Dickey-Fuller test to note that the data series stabilizes at the first-order difference. The results of the Akaike Information Criterion (AIC) were used to determine the prediction for SARIMA(1,1,1)×(1,0,1)12.

Despite these applications, a specific analysis leveraging the SARIMA framework on the most recent decade of available monthly electricity consumption data (2015–2024) for Erbil City, aiming to provide an updated forecast for 2025 while explicitly addressing the observed patterns in this specific timeframe, represents an area for contribution. This study seeks to address this by developing and validating an optimal SARIMA model based on this updated dataset. The established success of SARIMA models in similar contexts, as reviewed above, supports its selection as the primary methodology for achieving this study's forecasting objective.

III. Methodology

In this research, our main aim is to find an appropriate model for electricity in the city of Erbil and then take advantage of these models for prediction and control usage of electricity sources for the future. We're focused on time series analysis using the most prevalent approach, which is the Box and Jenkins method. This method's final model is more precise than the previous one and can be used for all forms of data transportation. The research's forecasting method used the ARIMA. We used these approaches to find patterns and trends in electricity usage in Erbil using real-time series data. For model construction, we utilized STATGRAPHIC version 19. The lowest AIC and RMSE values were used to select the best forecasting approach and timeframe.

A. The Dataset

The dataset used in this paper was retrieved from government electricity organizations for the historical daily electricity consumption in Erbil city for 120 months from 2015 to 2024 is recorded in Megawatts (MW) which is 3652 observations that have been used for forecasting the consumption of electricity for 2025.

Fig.1. shows a Time series plot chart of Erbil's electricity consumption using average data over 1 year from 2015 to 2024. It is evident that during the first 5 years of demand, electricity consumption has declined, but electricity demand has been steadily rising. The total amount of electricity consumed in 2015 was 328 MW, as you can see, but this amount increased significantly until 2019, when the average required amount of electricity consumption was 813 MW. In other words, some trends are very evident during the first 5 years of the time series for electricity consumers. However, the amount of electricity consumed will remain relatively constant at about 1,000 MW annually for the next 5 years, from 2019 to 2024.

The 2D-Area Plot in Fig. 2 to determine which months have the highest electricity consumption between seasons, a time series plot is created for the average months over 10 years. The total amount of electricity used was at its lowest in October at 676 MW and at its highest in December at 891 MW.

The total electricity consumption for the months and years 2015–2025 is shown in Table I. The total amount of electricity used during the first 5 years and the second 5 years is displayed in the table. The overall amount of electricity used during the second 5 years is double that of the first five. Approximately 500 MW of electricity will be consumed in

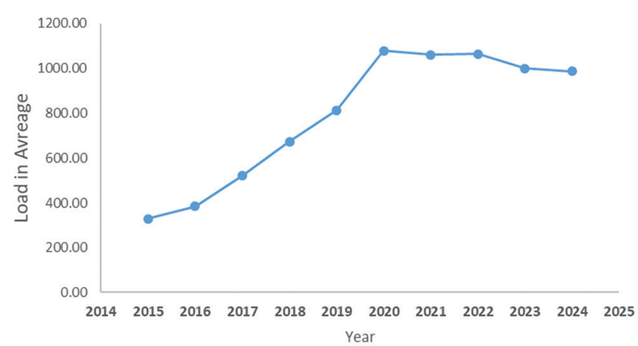

Fig. 1: Time series plot for electricity consumer in average per year data in Erbil from 2015–2024.

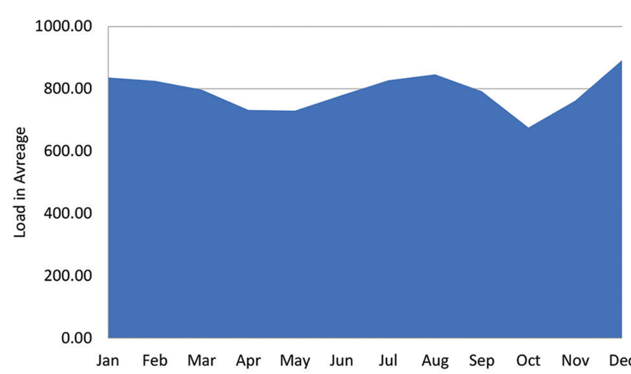

Fig. 2: Time series plot for electricity consumers in average per month data in Erbil from 2015–2024.

2015–2019, but approximately 1000 MW will be consumed between 2019 and 2024.

In Fig. 3, the Total electricity consumption (MW) for all years is demonstrated by a three-dimensional area chart. The chart explains that the total electricity consumption (MW) varies between months, but when taking the average, it seems that every month is about 800 MW of consumption. Through this chart, you can see that in the first 5 months. The amount of consumption the second 5 years. However, in terms of total annual consumption, the amount of consumption will be higher in 2020, with a total of 1079 MW.

IV. Results

A. ARIMA Model

ARIMA includes three variables (p, d, q). To correct a detectable trend in a time series, use differencing $\nabla^d X_t$. The purpose of employing differencing to reduce the trend is to transform the non-stationary time series into a stationary one. If the differenced model stabilizes, the model. The term "integrated" in ARIMA comes from the need to sum, or "integrate," the stationary model created from the differenced data. A single differencing operation (d = 1) is frequently sufficient to generate a stationary series. Three factors establish the nomenclature for an ARIMA process order of the autoregressive (AR) portion (p), the order of the moving average (MA) portion (q), and the number of differentiating procedures (d). As a consequence, an ARIMA(p,d,q) model includes these requirements and gives an intuitive summary of the model's structure and order when used with the time series in question (Brillinger, 2001).

$$\nabla^d X_t = (1-B)^d X_t \quad (1)$$

If $E(\nabla^d X_t) = \mu = 0$, then we can express the model as:

$$\emptyset(B)\,(1-B)^d X_t = \alpha + \theta(B) Z_t \quad (2)$$

Where $\propto = \mu(1-\emptyset_1-\ldots-\emptyset_p)$

$$\emptyset(B)=1-\emptyset_1 B-\emptyset_2 B_2-\ldots-\emptyset_p B^p \quad (3)$$

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q \quad (4)$$

TABLE I
Represented Average Load of Consumption Data (MW) per Year and Month

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2015	331	335	282	253	315	304	353	342	334	315	369	404	328
2016	416	405	400	355	345	395	403	413	351	335	382	423	385
2017	444	439	461	440	442	532	566	610	546	450	649	673	521
2018	683	675	677	542	581	681	742	766	700	572	661	798	673
2019	836	848	784	642	650	818	966	955	856	649	771	984	813
2020	1047	1044	1088	1107	1030	1139	1120	1146	1096	954	961	1216	1079
2021	1157	1111	1053	933	924	987	1141	1179	1061	920	1066	1184	1060
2022	1243	1230	1142	1075	1111	988	1009	1057	1024	852	959	1092	1064
2023	1097	1030	964	920	894	960	1015	1025	982	875	1010	1222	1000
2024	1107	1144	1124	1055	1006	987	958	966	971	833	791	917	987
Total	836	826	797	732	730	779	827	846	792	676	762	891	791

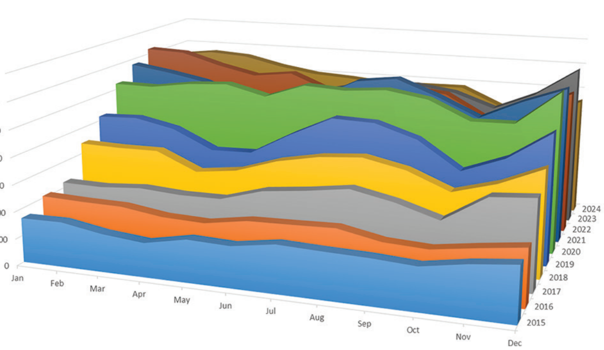

Fig. 3: 3D area chart between year and month for average consumption data load (MW).

B. Time Series Description

First, we need to plot the data and see its behavior. From the plot, it can be seen whether it has a trend or a seasonal shape. This is the most important step because all of the checking models are based on this step. If you go wrong, it will take a long time to gain the best model. Therefore, one has to be very careful at this stage. Since the data are daily.

C. Establishing ARIMA Models

The fundamental stages in fitting ARIMA models to time series data are:

Overview of data

The data are evaluated and analyzed to determine seasonal and overall patterns. The time series data were separated into trend, seasonality, and residual values. The trend represents the series' optional and frequently linear expanding or dropping activity over time, whereas seasonality represents its potential recurring trends or cycles of activity as time passes. The residual values successfully eliminate the data's trend and seasonality, resulting in time-independent values. We employed the seasonal deconstruct function in statistical models.

Stationarity test

Transforming Data If we need to alter the data in a time series plot, the analyst may observe that the disparity between lows and highs of seasonal peaks widens over time. This pattern implies that the variance does not follow a steady distribution. Before fitting the ARIMA model, we may apply the Box-Cox power transformation to ensure the variance remains stable.

Charting dynamic statistics

Make a graphic representation of the standard deviation and moving average and watch how it changes as time passes. A standard deviation and moving average is the average/standard deviation of the preceding year, or 12 months, at any given time (t).

Fig. 4 shows the time series plot of 3652 days of electricity consumption in Erbil from January 1, 2015 to December 31, 2024. Then, the appropriate model for the series is determined. However, due to the smell of outliers, these steps to determine the best model were unsuccessful, although differences were accepted. Therefore, to reduce outliers, we transformed the data by taking logs, as shown in Fig.5.

Table II displays the estimated Average Load Consumption (MW) autocorrelation at various delays. The delay k the autocorrelation coefficient assesses the connection between Average Load Consumption (MW) at t and time. It also has a probability limit of 95.0% about zero. If analysis does not restrict the possibilities to a set length of time. The computed coefficient shows a statistically significant link with the 95.0% delay (degree of confidence). In this situation, 24 of the 24 coefficients are statistically significant. A 95.0% confidence level indicates that the time series may not be fully random (white noise).

Fig. 6 shows that all autocorrelation coefficients are statistically significant at the 95.0% confidence level, indicating that the time series for the average Load Electricity user in Erbil by megawatt from 2015 to 2024 is not stationary.

Several partial autocorrelation coefficients in Fig. 7 are statistically significant at the 95.0% confidence level.

Dickey-Fuller test

Stationarity is determined by this statistical test. The time series is not stationary, according to the null hypothesis. Test statistics and important values for various confidence levels are included in the test results. The series is considered stationary, and the null hypothesis is rejected if the "Critical Value" is smaller. The alternative hypothesis is frequently stationarity or trend-stationarity, though it varies depending on the type of test used. In 1979, statisticians Wayne Fuller and David Dickey created the test, which bears their names. We estimate and eliminate seasonality and trends to stabilize the series. Logarithmic transformations and differential methods are used (Dickey and Fuller, 1979).

According to Table III, it can be stated that the data is not stationary and hence, as a first stage, the first difference was applied and the data became stationary as per the results provided in Table IV. Following that, we must examine the differenced series' ACF and PACF to ensure that it is stationary.

As a result of the above transformation, we now test our data to see the differences between those results and the test.

The ACF and PACF for the differenced log transform data are virtually steady, as shown in Figs 6 and 7, supporting the assumption that the series is stationary in both the mean and variance following a first-order difference. As a result, an ARIMA(p,1,q) model may be developed for the differenced electricity data. After identifying the ARIMA model, we must determine our model's p, q, P, and Q parameters.

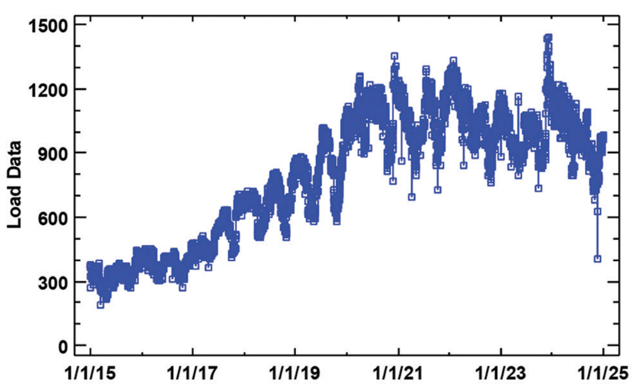

Fig. 4: Time series plot for electricity consumer data in Erbil.

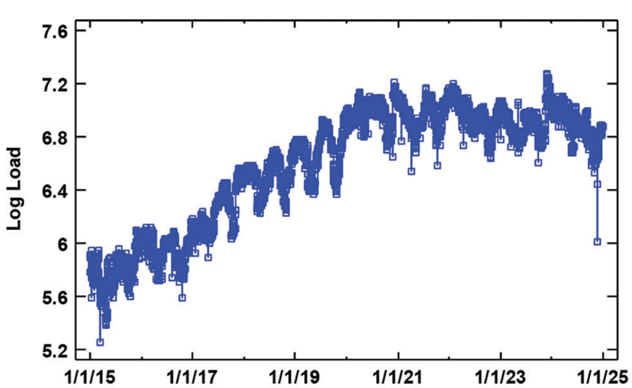

Fig. 5: Time series plot log transformation for electricity consumer data in Erbil.

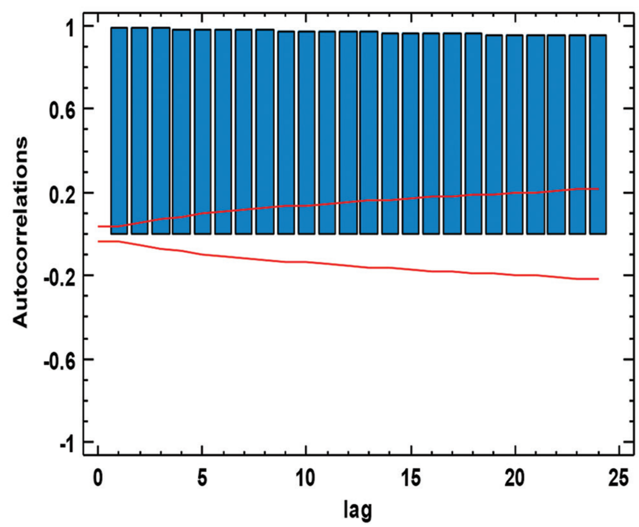

Fig. 6: ACF for the log transformation load of electricity consumer data.

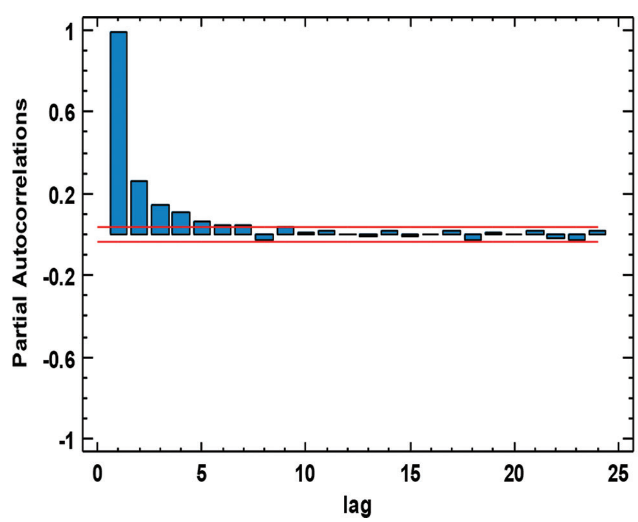

Fig. 7: PACF for the Log transformation load of electricity consumer data.

TABLE II
Estimated Autocorrelations for Average Log Transference Load of Electricity Consumer Data

Lag	ACF	SE	LPL 95.0%	UPL 95.0%
1.	0.992594	0.016548	−0.03243	0.032433
2.	0.989077	0.02852	−0.0559	0.055898
3.	0.986557	0.036731	−0.07199	0.071991
4.	0.984672	0.043384	−0.08503	0.085031
5.	0.982797	0.049124	−0.09628	0.096281
6.	0.981008	0.054241	−0.10631	0.106311
7.	0.979425	0.058899	−0.11544	0.115441
8.	0.977124	0.063202	−0.12387	0.123874
9.	0.975508	0.067211	−0.13173	0.131732
10.	0.973709	0.070983	−0.13912	0.139123
11.	0.972125	0.07455	−0.14612	0.146116
12.	0.970366	0.077944	−0.15277	0.152768
13.	0.968531	0.081185	−0.15912	0.159119
14.	0.967036	0.084289	−0.1652	0.165204
15.	0.965268	0.087274	−0.17106	0.171055
16.	0.963548	0.09015	−0.17669	0.176692
17.	0.962038	0.092927	−0.18214	0.182135
18.	0.960095	0.095616	−0.1874	0.187404
19.	0.958397	0.09822	−0.19251	0.192508
20.	0.956696	0.100748	−0.19746	0.197463
21.	0.955224	0.103206	−0.20228	0.20228
22.	0.953371	0.105599	−0.20697	0.206971
23.	0.951317	0.10793	−0.21154	0.21154
24.	0.949751	0.110202	−0.21599	0.215993

TABLE III
Randomness Test for the Transformed Data

Data	Stage	Hypotheses	Dicky-Fuller (P-Value)
Load	Before first difference	H_0: Not stationary H_1: Stationary	−3.2766 (0.0747)
	After first difference	H_0: Not stationary H_1: Stationary	−17.301 (0.0100)
Log Load	Before first difference	H_0: Not stationary H_1: Stationary	−2.814 (0.2337)
	After first difference	H_0: Not stationary H_1: Stationary	−17.177 (0.0100)

Choosing the appropriate model

After stabilizing the time series around the mean and variance, we need to define models by setting several values into the *p,q,P* and *Q*. The following are the MSE values for 19 different models demonstrated in Table V.

Several ARIMA models were tested as part of the time series data analysis process to determine the best model to capture the city's trends in electricity demand. MSE, parameter significance, and residual diagnostics using the Ljung-Box test were among the evaluation criteria. The seasonal model ARIMA(1,1,1)×(0,1,1)12 was found to be the best fit among the models that were examined. With the lowest MSE (0.0487347), this model demonstrated exceptional prediction accuracy. With a $P = 0.147027$, the residuals passed the Ljung-Box goodness-of-fit test, indicating that they were uncorrelated and that the model represented the underlying data structure well.

In addition, it was shown that the parameters AR(1), MA(1), and SMA(1) were relevant in predicting the behavior of the data since they were statistically significant ($P < 0.05$). This model's seasonal component (SMA) made it possible for it to identify the recurring patterns in the data successfully.

TABLE IV
Estimated Partial Autocorrelations for Log Transform Data Electricity Consumer

Lag	ACF	SE	LPL 95.0%	UPL 95.0%
1.	0.992594	0.016548	−0.03243	0.032433
2.	0.259807	0.016548	−0.03243	0.032433
3.	0.144486	0.016548	−0.03243	0.032433
4.	0.11007	0.016548	−0.03243	0.032433
5.	0.060565	0.016548	−0.03243	0.032433
6.	0.042772	0.016548	−0.03243	0.032433
7.	0.042024	0.016548	−0.03243	0.032433
8.	−0.02762	0.016548	−0.03243	0.032433
9.	0.034369	0.016548	−0.03243	0.032433
10.	0.00542	0.016548	−0.03243	0.032433
11.	0.019038	0.016548	−0.03243	0.032433
12.	0.001318	0.016548	−0.03243	0.032433
13.	−0.00509	0.016548	−0.03243	0.032433
14.	0.022224	0.016548	−0.03243	0.032433
15.	−0.00487	0.016548	−0.03243	0.032433
16.	−0.00219	0.016548	−0.03243	0.032433
17.	0.017521	0.016548	−0.03243	0.032433
18.	−0.02502	0.016548	−0.03243	0.032433
19.	0.005797	0.016548	−0.03243	0.032433
20.	0.001543	0.016548	−0.03243	0.032433
21.	0.014686	0.016548	−0.03243	0.032433
22.	−0.01496	0.016548	−0.03243	0.032433
23.	−0.02619	0.016548	−0.03243	0.032433
24.	0.018649	0.016548	−0.03243	0.032433

Several models, including ARIMA(1,0,0), ARIMA(2,0,0), and ARIMA(0,0,1), were fitted without accounting for any differences. The initial time series data, however, were found to be non-stationary based on the results. The increased performance and diagnostic outcomes of the differenced models, including ARIMA(1,1,1) and its seasonal versions, demonstrate that the data passed the stationarity test after applying differences. Despite being simpler, models without differencing were unable to explain the patterns, which led to worse performance indicators like residual autocorrelation and higher MSE values.

To sum up, the seasonal model ARIMA(1,1,1)×(0,1,1)12 was chosen because it performed better on all evaluation criteria, could deal with seasonality, and was reliable in capturing the stationary differenced data structure. This model is the best option for the specified time series analysis since it strikes a compromise between simplicity and accuracy.

Optimization Parameter Deduction

Yule-Walker equations, the Maximum Likelihood approach, and the Least Squares method can all be used to estimate both parameters ($\emptyset$ and θ). Typically, estimations are determined quantitatively. We may also construct confidence intervals for parameters.

Testing the fitted model

After identifying the model, it has to be tested to determine whether it is appropriate for the data or not. The most important part of doing this is to check the capability of the model and how far it goes with the original series.

TABLE V
Fitted Models and their MSE for Transform Log Data

Fitted models	MSE	Parameter	Estimate	Parameter test $P<0.05$	Goodness of fit test ljung-Box Test $P>0.05$
ARIMA (1,0,0)	0.0524447	AR (1)	0.992696	0	0
ARIMA (2,0,0)	0.0504603	AR (1)	0.734564	0	0
		AR (2)	0.260025	0	
ARIMA (0,0,1)	0.242344	MA (1)	−0.91387	0	0
ARIMA (1,1,0)	0.0505127	AR (1)	−0.27582	0	0
ARIMA (0,1,1)	0.0496007	MA (1)	0.39808	0	6.146E-11
ARIMA (0,0,2)	0.16215	MA (1)	−1.32296	0	0
		MA (2)	−0.78413	0	
ARIMA (1,1,1)	0.0490988	AR (1)	0.327295	0	0.197228
		MA (1)	0.68833	0	
ARIMA (1,1,2)	0.049218	AR (1)	−0.11654	0.377767	0.00352812
		MA (1)	0.251075	0.053702	
		MA (2)	0.174844	0.001729	
ARIMA (2,1,1)	0.0491039	AR (1)	0.339029	0	0.00273251
		AR (2)	0.010478	0.661772	
		MA (1)	0.701393	0	
ARIMA (2,1,2)	0.0491115	AR (1)	−0.09873	0.964282	0.1295
		AR (2)	0.145689	0.831998	
		MA (1)	0.263535	0.9048	
		MA (2)	0.295661	0.844524	
ARIMA (1,1,1)×(1,1,1) 12 without constant	0.0487549	AR (1)	0.334261	0	0.154904
		MA (1)	0.691729	0	
		SAR (1)	−0.01022	0.537754	
		SMA (1)	0.989732	0	
ARIMA (1,1,1)×(1,1,1) 12 with constant	0.0487606	AR (1)	0.33438	0	0.155422
		MA (1)	0.691869	0	
		SAR (1)	−0.01027	0.536054	
		SMA (1)	0.989719	0	
ARIMA (1,1,2)×(0,1,1) 12	0.0487481	AR (1)	0.338574	0.00136	0.111019
		MA (1)	0.697648	0	
		MA (2)	−0.00475	0.927089	
		SMA (1)	0.994298	0	
ARIMA (1,1,1)×(0,1,1) 12	0.0487347	AR (1)	0.331696	0	0.147027
		MA (1)	0.690244	0	
		SMA (1)	0.993945	0	
ARIMA (0,1,0)×(1,1,1) 12	0.0519537	SAR (1)	−0.00549	0.740796	0
		SMA (1)	0.992293	0	
ARIMA (1,1,0)×(1,1,0) 12	0.061058	AR (1)	−0.26225	0	0
		SAR (1)	−0.51692	0	
ARIMA (0,1,1)×(0,1,1) 12	0.0493088	MA (1)	0.405584	0	1.41E-13
		SMA (1)	0.994246	0	
ARIMA (2,1,2)×(1,1,1) 12	0.0487833	AR (1)	−0.08873	0.990683	0.092964
		AR (2)	0.149488	0.952113	
		MA (1)	0.269287	0.971739	
		MA (2)	0.296646	0.955192	
		SAR (1)	−0.01012	0.547302	
		SMA (1)	0.98895	0	
ARIMA (0,1,2)×(0,1,1) 12	0.0488167	MA (1)	0.365569	0	0.0129089
		MA (2)	0.135208	0	
		SMA (1)	0.99442	0	

ARIMA: Autoregressive integrated moving average

Thus, it can be reliable to predict future cases. Now, we test the residual autocorrelation. Since all of the values are inside the confidence interval, it means the variable of the residual autocorrelation function is randomness, and thus, the model is appropriate for the data.

$$-1.96\ S(r_\varepsilon) \leq \rho_\varepsilon \leq 1.96\ S(r_\varepsilon)$$

This process will forecast future megawatt values for the average electricity load user in Erbil at various time delays. Currently in use is the seasonal model ARIMA(1,1,1)×(0,1,1)12, which is demonstrated in Table VI. According to this model, the best forecast for future data is produced by a parametric model that compares the most recent

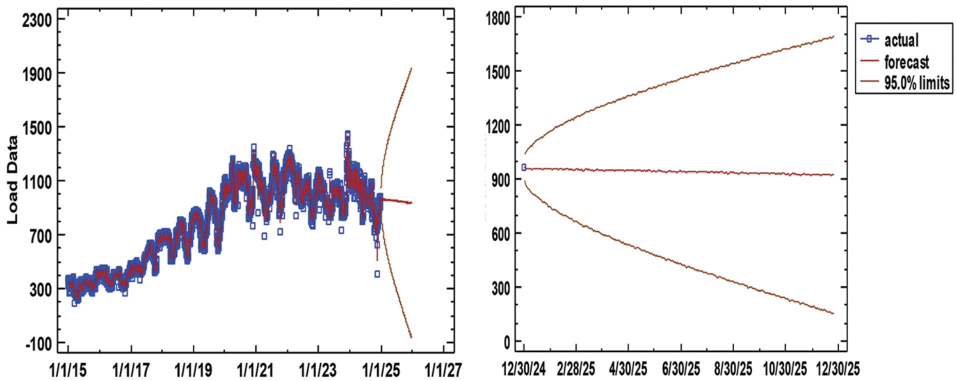

Fig. 8: Time sequence and forecast plot for electricity consumer data in Erbil [ARIMA(1,1,1)].

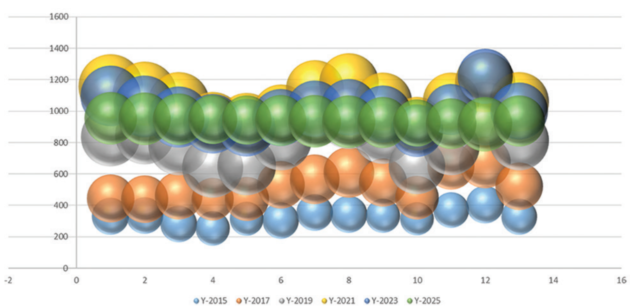

Fig. 9: Comparison of electricity consumption for the forecast year with previous years.

data value to previous data values and noise. The following form can be used to express the previously mentioned model:

$$(1-B^{12})Y_t = (1+\Theta_1 B_{12})\epsilon_i$$

Where B^{12} represents the backshift operator applied with a lag of 12 periods (seasonal period), Θ_1 is the seasonal moving average parameter. The $(1-B^{12})$ term is the seasonal differencing operator, which removes seasonal patterns from the data.

Combining both the non-seasonal and seasonal parts, the full model is:

$$(1-\phi_1 B)(1-B)(1-B^{12})Y_t = (1-\theta_1 B) = (1+\Theta_1 B^{12})\epsilon_i$$

The non-seasonal AR coefficient is denoted by ϕ_1, the non-seasonal moving average by θ_1, and the seasonal moving average by ϕ_1. Moving averages and seasonal differencing are reflected in the terms involving B^{12}.

Forecasting

Forward forecasting is the last step after the fitted model has been diagnosed and chosen as the best one. In time series analysis, this is the final step in the modeling process.

The predicted values for Erbil's electricity consumer data are displayed in Table VII, which shows the residuals (data forecast) and the expected values from the fitted model. Erbil City will require approximately 937 MW on average during the year 2025. Furthermore, Fig. 8 also displays this information. In Fig. 8, there is a comparison of electricity consumption for the forecast year with previous years. The figure shows that the forecast for electricity consumption in 2025 will be lower than in 2021–2023

TABLE VI
Result of ARIMA (1,1,1)×(0,1,1) 12

Parameter	Estimate	SE	T	*P*-value
AR (1)	0.331696	0.0335205	9.8953	0.000000
MA (1)	0.690244	0.0259273	26.6223	0.000000
SMA (1)	0.993945	0.0000496	20031.0	0.000000
Mean	−2.3E-06	0.0000050	−0.445931	0.655644
Constant	−1.5E-06			

ARIMA: Autoregressive integrated moving average

TABLE VII
Forecast Table for Electricity Consumer Data in Erbil

2025	Average of forecast	Average of upper 95% limit	Average of lower 95% limit
Jan	955.226	783.476	1167.277
Feb	952.058	690.902	1312.900
Mar	949.547	631.384	1428.827
Apr	946.213	583.794	1534.225
May	943.096	544.495	1634.056
Jun	939.707	510.508	1730.226
Jul	935.902	480.375	1823.869
Aug	932.659	453.266	1919.533
Sep	928.314	428.573	2011.194
Oct	924.866	406.400	2105.201
Nov	920.131	385.425	2197.033
Dec	916.451	366.434	2292.462
Total	936.969	521.656	1764.378

Fig. 8 illustrates the forecast for the electricity consumption load in Erbil for 2025, with a maximum average not exceeding 955 MW and a minimum average of 916 MW.

Fig. 9 provides a detailed visualization of the anticipated data consumption for each month in 2025, contrasted with the consumption levels from previous years. The graph highlights a significant downward trend, indicating that data consumption in 2025 is projected to be lower than in 2021–2023. This downward shift suggests a potential change in user behavior or improvements in data efficiency, as illustrated by the comparative analysis of the monthly figures.

V. Conclusion

An essential tool for forecasting and modeling is time series analysis. The information provided by the

ARIMA(1,1,1)×(0,1,1)12 model can assist decision-makers in Erbil in setting priorities, strategies, and the efficient use of electricity resources. This information is highly relevant and suitable for forecasting the precise monthly electricity data requirements. As a result, it is important to note that relying solely on our model for decision-making should not be done with individual data. We can test an intervention time series analysis to see if we can enhance our model's ability to predict the peak values of electricity data. In addition, Erbil's average total electricity consumption in 2025 will be around 937 MW. It has been demonstrated that the restriction policies, such as the lack of adequate electricity imposed by the KRG, cause people to reduce electricity consumption, which eventually leads to improvements in data efficiency.

References

Ajlouni, S.A. (2024). Using ARIMA model to forecast electricity load in Jordan. *Jordan Journal of Economic Sciences*, 11(2), 127-139.

Al-Zawbaee, O.M., & Mahmood, N.H. (2023). Building an efficient model by using panel data. *Cihan University-Erbil Journal of Humanities and Social Sciences*, 7(1), 82-87.

Brillinger, D.R. (2001). *Time Series: Data Analysis and Theory*. New Delhi: SIAM.

Brockwell, P.J., & Davis, R.A. (2002) *Introduction to Time Series and Forecasting*. 2nd ed. Berlin: Springer.

Chatfield, C. (1984). *The Analysis of Time Series*. London: Chapman and Hall.

Dickey, D.A., & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.

Elsaraiti, M., Ali, G., Musbah, H., Merabet, A., & Little, T. (2021). Time Series Analysis of Electricity Consumption Forecasting Using ARIMA Model. In: *2021 IEEE Green Technologies Conference* [GreenTech].

Kadir, D. (2018) *Bayesian Inference of Autoregressive Models*. England: University of Sheffield. White Rose eTheses Online.

Kadir, D. (2020). Time series modeling to forecast on consuming electricity: A case study Analysis of electrical consumption in Erbil City from 2014 to 2018. *Journal of Al-Rafidain University College For Sciences*, 1(46), 472-485.

Mahanta, N., & Talukdar, R. (2024). Forecasting of electricity consumption by seasonal autoregressive integrated moving average model in Assam, India. *International Journal of Energy Economics and Policy*, 14(5), 393-400.

Mahia, F., Dey, A.R., Masud, M.A., & Mahmud, M.S. (2019). Forecasting Electricity Consumption Using ARIMA Model. *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh.

Mahmood, N.H., Kadir, D.H., & Alzawbaee, O.M.M. (2024). Building a statistical model to forecast traffic accidents for death and injuries by using bivariate time series analysis. *Zanco Journal of Human Sciences*, 28(1), 278-289.

Mahmood, S.H. (2016). Using artificial neural network model with time series monthly average electricity demand in kurdistan region. *Journal of Humanity Sciences*, 20(6), 421-432.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998) *Forecasting: Methods and Applications*. United States: John Wiley & Sons.

Mansor, R., Zaini, B.J., & May, C. (2021). Medium-term electricity load demand forecasting using holt-winter exponential smoothing and SARIMA in university campus. *International Journal of Engineering Trends and Technology*, 69(3), 165-171.

Massoudi, A., & Ahmed, M. (2021). Assessing the sustainability of facility management in the food sector in Kurdistan Region of Iraq. *Cihan University-Erbil Journal of Humanities and Social Sciences*, 5(1), 106-110.

Mohammed, P.A., Obed, S.A., Ali, I.M., & Kadir, D.H. (2022). US Dollar/IQ Dinar currency exchange rates time series forecasting using ARIMA model. *Cihan University-Erbil Scientific Journal*, 6(1), 12-19.

Omer, A.W., Blbas, H.T., & Kadir, D.H. (2021). A comparison between brown's and holt's double exponential smoothing for forecasting applied generation electrical energies in Kurdistan region. *Cihan University-Erbil Scientific Journal*, 5(2), 56-63.

Parreño, S.J. (2022). Forecasting electricity consumption in the Philippines using ARIMA models. *International Journal of Machine Learning and Computing*, 12(6), 279-285.

Ravaz, M.S., & Marwan, T.H. (2020). Utilizing of seasonal time series forms in predicting the quantity of electric power demand in Erbil city between the year (2010-2019). *Zanco Journal of Human Sciences*, 24(5), 26-38.

Sim, S., Tay, K., Huong, A., & Tiong, W. (2019). Forecasting electricity consumption using SARIMA method in IBM SPSS software. *Universal Journal of Electrical and Electronic Engineering*, 6(5B), 103-114.

Sosa, G.R., Falah, M.Z., Dika Fikri, L., Wibawa, A.P., Handayani, A.N., & Hammad, J.A.H. (2021). Forecasting electrical power consumption using ARIMA method based on kWh of sold energy. *Science in Information Technology Letters*, 2(1), 7.