Jordan Journal of Economic Sciences, Volume **11**, No.**2, 2024**

Using ARIMA Model to Forecast Electricity Load in Jordan

Sameh A. Ajlouni [1]

[1] Associate professor of economics, Department of Economics, Yarmouk University Faculty of Business, Yarmouk University, Jordan. ajlouni.sameh@yu.edu.jo

Received: 9/9/2023
Revised: 2/11/2023
Accepted: 17/12/2023
Published: 1/7/2024

Citation: Ajlouni, S. A. . (2024). Using ARIMA Model to Forecast Electricity Load in Jordan. *Jordan Journal of Economic Sciences*, *11*(2), 127–139.
https://doi.org/10.35516/jjes.v11i2.1733

Abstract

Objectives: This study aims to forecast the daily peak electricity load in Jordan using a dataset of hourly peak load data for the period from January 1, 2010, to December 31, 2022, compiled by the National Electric Power Company (NEPCO).
Methods: This study employs the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to make forecasts. The data exhibits an upward trend, seasonality, and non-constant variance. To address these features, the SARIMA model is used to account for the trend and seasonality, while a Box-Cox transformation is applied to manage the non-constant variance.
Results: Following the standard Box-Jenkins methodology (identification, estimation, diagnostic checking, and forecasting) and utilizing the “auto.arima()” function in the RStudio software package, the resulting SARIMA model is ARIMA(1,0,1)(2,1,2)[7]. This model is used to forecast 7 future values of the electricity load. The Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE) values, as measures of forecast accuracy, support the precision of our forecasts.
Conclusion: Based on empirical results, electricity companies in Jordan are encouraged to use time series models for forecasting electricity loads instead of relying on simple spreadsheet models.

Keywords: ARIMA model, SARIMA model, Box-Jenkins method, electricity load forecasting, Jordan, auto.arima(), RStudio software package.

التنبؤ بالأحمال الكهربائية في الأردن باستخدام نموذج الانحدار الذاتي والمتوسطات المتحركة المتكاملة ARIMA

سامح عاصم العجلوني*1

[1] قسم الاقتصاد، كلية الأعمال، جامعة اليرموك

ملخّص
الأهداف: تهدف هذه الدراسة إلى التنبؤ بذروة حمل الكهرباء اليومي في الأردن باستخدام مجموعة بيانات أحمال الذروة اليومية للأربع وعشرين ساعة للفترة من 1 كانون الثاني 2010 إلى 31 كانون أول 2022 والمتوفرة لدى شركة الكهرباء الوطنية (NEPCO).
المنهجية: تستخدم هذه الدراسة نموذج الانحدار الذاتي والمتوسطات المتحركة المتكاملة ARIMA الموسمي أو ما يسمى نموذج SARIMA للتنبؤ. تظهر البيانات اتجاهًا تصاعديًا وموسمية وتباينًا غير ثابت. وللتعامل مع هذه الخصائص، تم استخدام نموذج SARIMA لمعالجة وجود الاتجاه والموسمية، في حين تم استخدام تحويل Box-Cox للتغلب على خاصية التباين غير الثابت.
النتائج: استناداً إلى نتائج التحليل القياسي تم تحديد نموذج SARIMA التالي: ARIMA(1,0,1)(2,1,2)[7] والذي تم استخدامه للتنبؤ بسبعِ قِيم مستقبلية للحمل الكهربائي، حيث تراوحت متوسطات هذه القيم ما بين 2530 ميجاوات كحد أدنى و 2938 ميجاوات كحد أعلى. وللتأكد من دقة التنبؤات تم استخدام معياري متوسط النسبة المئوية للخطأ المطلق (MAPE)، وقيمة جذر متوسط مربع الخطأ (RMSE) واللّذَيْن أكّدا دقة التنبؤات وبالتالي ملاءمة استخدام هذا النموذج للتنبؤ بقيم الأحمال الكهربائية المستقبلية في الأردن.
الخلاصة: استنادا إلى النتائج التجريبية، فإنّ شركات الكهرباء في الأردن مدعوة لتفعيل استخدام نماذج السلاسل الزمنية للتنبؤ بأحمال الكهرباء بدلاً من الاعتماد على استخدام النماذج البسيطة مثل دالة التنبؤ (forecast function) المتاحة في برنامج MS-Excel على سبيل المثال.

الكلمات الدالة: نموذج ARIMA، نموذج SARIMA، طريقة بوكس-جينكينز، التنبؤ بالأحمال الكهربائية، الأردن، ()auto.arima، حزمة برامج .RStudio.

1. INTRODUCTION

The recent Russian-Ukrainian war is seriously threatening global energy security, and Jordan is no exception. Given that Jordan imports more than 90% of its energy needs, the situation is even more precarious, particularly concerning electricity generation. Jordan relies heavily on natural gas and fuel oil to generate electricity, so hikes in the prices of these commodities and any supply disruptions would have serious negative impacts on electricity generation and the prices paid by end-users. Additionally, unexpected shifts in electricity demand due to recent global and local circumstances—such as climate change, global warming, the COVID-19 pandemic, and the influx of Syrian refugees starting in 2011—have disrupted electricity generation and supply plans, increased electricity load, and added extra pressure on the electricity grid.

Accordingly, reliable forecasts of future electricity loads are crucial to enable decision-makers to make informed plans and mitigate, as much as possible, the negative impacts of risks associated with the provision of natural gas and crude oil or potential price hikes. Therefore, the purpose of this study is to use the ARIMA model to forecast daily electricity load in Jordan, using hourly peak load data from January 1, 2010, to December 31, 2022, compiled by the National Electric Power Company (NEPCO). The importance of the study stems from the significance of energy security in general and electricity security in particular, given the increased reliance on electricity in all aspects of life.

This research is expected to provide reliable estimates of future electricity load in Jordan. This is deemed extremely important given that Jordanian electric utilities do not currently use state-of-the-art techniques to forecast electricity load, relying instead on spreadsheets (Alhmoud & Nawafleh, 2021). These estimates are of utmost importance for planning processes and for the economic regulation of the sector.

The rest of the paper is divided into sections. Section 2 contains the background, Section 3 introduces the model, Section 4 describes the data, Section 5 presents the empirical results, and Section 6 concludes the paper.

2. BACKGROUND

Reliable forecasts of electricity load are extremely important for both utilities and policymakers to design reliable energy infrastructure and avoid inefficiencies. Inefficiencies can take various forms, including but not limited to, building redundant generating units, incurring high operational costs, and unnecessarily high consumption of fuel (Bashir et al., 2022). Bashir et al. (2022) classify load forecasting into three categories: short-term, mid-term, and long-term, while Goswami and Kandali (2020) add “very short-term” as a fourth category. Generally speaking, electricity load forecasting methods are divided into three classes: (i) classical or traditional statistical techniques, (ii) machine learning or artificial intelligence techniques, and (iii) hybrid models that use two or more methods from one or both classes.

Statistical methods use time series techniques such as ARMA, ARIMA, SARIMA, ARIMAX, and the Kalman filter to forecast electricity load, while machine learning techniques include fuzzy logic, Artificial Neural Networks (ANN), and Support Vector Regression (SVR) (Goswami and Kandali, 2020; Chodakowska et al., 2021). Hong and Fan (2016), Hammad et al. (2020), and Hahn et al. (2009) thoroughly discuss various techniques used in load forecasting. The literature on forecasting electricity load using different approaches is truly extensive, and attempting to cite all relevant literature would be futile. Therefore, this study will review only a few select studies. Interested readers can consult Nti et al. (2020) and Kuster et al. (2017) for an extensive review of relevant literature.

Among the studies that forecast electricity load in Jordan is Alhmoud and Nawafleh (2021), which uses hourly load data for 2018 to forecast the hourly load in Jordan for one week ahead using three techniques: the nonlinear autoregressive exogenous model (NARX) recurrent neural network, the Elman neural network, and the autoregressive moving average (ARMA). The main conclusion of the study is that the Elman method is the most effective, and NEPCO is recommended to use it. Alasali et al. (2021) use ANN, ARIMAX, and a rolling stochastic ARIMAX forecast model to forecast electricity load in Jordan and capture the impact of the COVID-19 pandemic on electricity demand and consumption behavior. The rolling stochastic ARIMAX produces more reliable forecasts compared to the benchmark ARIMAX and the ANN model. Tawalbeh et al. (2021) use ARMA and ARIMA to forecast electricity load in Jordan. Although the authors do not clearly state the sample size, they mention that they used the study results to analyze the impact of the COVID-19 pandemic on electricity demand.

3. MODEL

3.1. ARIMA model

To capture the seasonality in the data, this study uses an extension of the ARIMA model called the Seasonal ARIMA, or SARIMA, model to forecast electricity load. The ARIMA model, commonly known as the Box-Jenkins (B-J) method, developed by Box and Jenkins (1976), is a simple yet powerful and popular statistical model that has been extensively used to forecast electricity load in different countries. The general notation of the ARIMA model is written as ARIMA (*p*,*d*,*q*), where *p* represents the number of the autoregressive (AR) orders, (*d*) reflects the degree of differencing after which the time series becomes stationary if it is originally non-stationary, and *q* denotes the order of moving average (MA) terms. An ARIMA(*p*,*d*,*q*) model is a generalization of the Autoregressive Moving Average or ARMA(*p*,*q)* model.

The general form of an ARMA(*p*,*q*) model can be written as (Asteriou & Hall, 2016):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

this can be written using summations as:

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + u_t + \sum_{j=1}^{q} \theta_j u_{t-j}$$

or using the lag operator as:

$$Y_t(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^P) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) u_t$$
$$\Phi(L) Y_t = \Theta(L) u_t$$

Note that an ARMA(*p*,*q*) can be represented as an ARIMA(*p*,*0*,*q*) model, so ARMA models can be used only if Y_t is stationary; otherwise ARIMA model should be used.

3.2. SARIMA model

Before proceeding further with explaining how the B-J methodology is empirically applied, we will introduce the SARIMA model. SARIMA is modelled by adding new seasonal terms into the ARIMA model as follows: ARIMA(*p*,*d*,*q*)×(*P*,*D*,*Q*)$_s$, where (*p*) denotes the order of autoregression, while (*P*) represents the order of seasonal autoregression. The order of integration is (*d*) and the seasonal integration is given by (*D*). The order of moving average is given by (*q*) and the order of seasonal moving average is given by (*Q*). Finally, the length of seasonal period is given by (s).

Traditionally, if the time series exhibits seasonality, seasonal differencing or Box-Cox transformation, among other methods, can be used to remove seasonality.

3.3. Application of B-J methodology

There are four steps for B-J methodology: identification, estimation, diagnostic checking and forecasting (Asteriou & Hall, 2016; Gujarati & Porter 2009).

3.3.1. Identification:

Identification means whether we will use ARMA(*p*,*q*) or ARIMA(*p*,*d*,*q*) and what the values of *p*,*d* and *q* are. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) can be used to determine *q* and *p* respectively. While *d* is determined based on the number of differences taken to make the series stationary. Therefore, testing for the stationarity of the series is a perquisite for the identification step.[1]

3.3.2. Estimation:

In the estimation stage, various models with different specifications (different values of *p, d, q*) are estimated, i.e., the parameters of the autoregressive and moving average components are estimated. This is usually carried out using the Ordinary Least Squares (OLS) method. Next, the various estimated models are compared to each other using the Akaike

[1] More precisely, the Box-Jenkins model requires the series to be stationary and the model to be invertible (Asteriou & Hall, 2016).

Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC).

3.3.3. Diagnostic Checking:

In this stage, the goodness of fit of the model is examined, i.e., assessing if the chosen model fits the data reasonably well. This is usually undertaken using the Ljung–Box Q-statistic to test for autocorrelations of the residuals. Alternatively, it can be checked whether the residuals are white noise.

3.3.4. Forecasting:

In the final stage, the estimated ARIMA model is used to forecast future values of the variable of interest. The accuracy of the predictions from the ARIMA model will be evaluated using the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE).

4. DATA

4.1. Description of data

This paper constructs a time series of maximum daily electricity load using hourly peak load data over the period from January 1, 2010, to December 31, 2022, provided by NEPCO. On any given day, peak electricity demand is defined as the maximum of 48 half-hourly demands during that day (As'ad, 2012).

Figure 1 plots the maximum daily peak over time. An initial look at the figure indicates that the time series has an intercept and is upward trending, suggesting it is likely nonstationary, and it seems that the variance is not constant. Moreover, a seasonality pattern can be readily observed. The boxplot reveals the presence of outliers and shows that the median is 2,500 megawatts (MW) (see Figure 2). Figure 3 shows the histogram and the descriptive statistics of the time series. Based on the Jarque-Bera test, the data is not normally distributed.

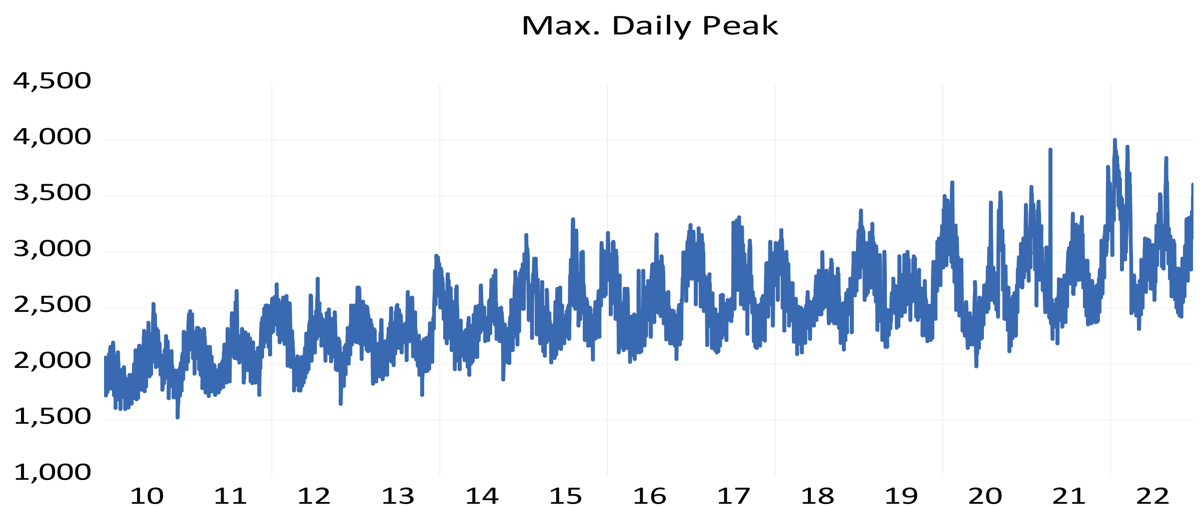


Fig. 1. Plot of the "Max. Daily Peak" series

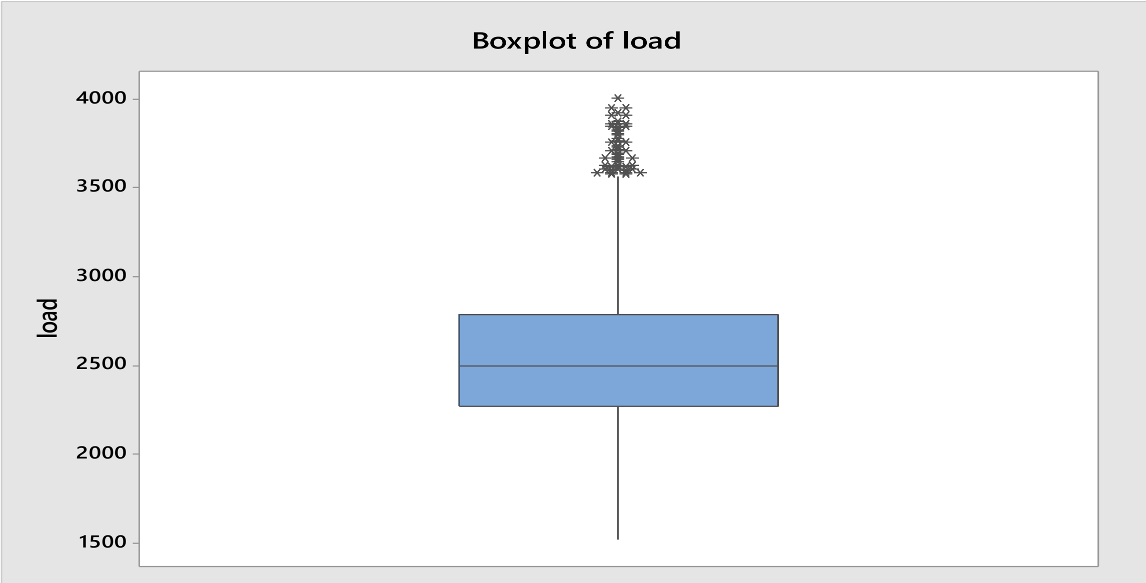


Fig. 2. Boxplot of the "Max. Daily Peak" series

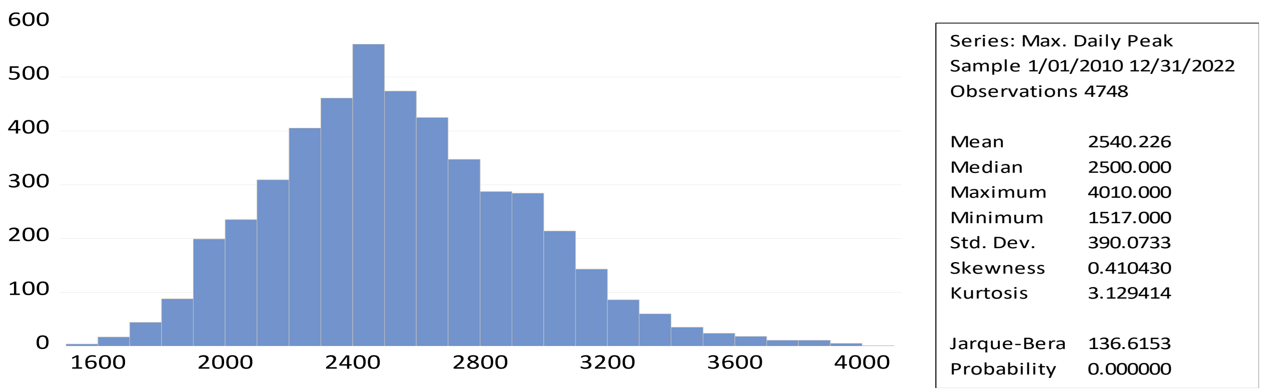


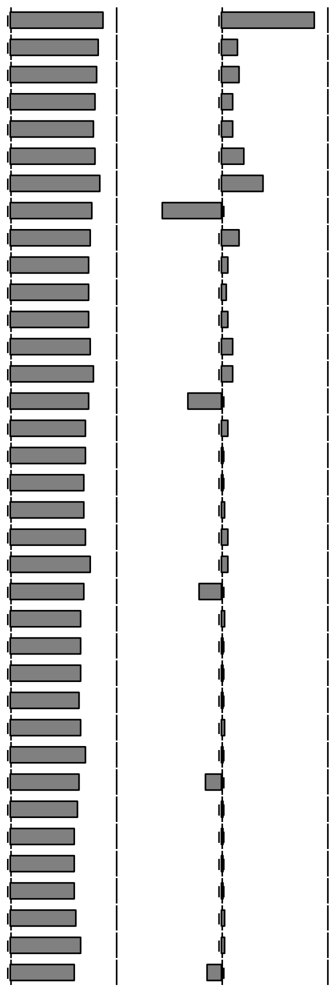
Fig. 3. Histogram and the descriptive statistics of the "Max. Daily Peak" series

4.2. Addressing time-series properties of the data

4.2.1. Stationarity

Figure 1 shows that the series is upward trending and contains a cyclical component, which might indicate nonstationary. However, this conclusion needs formal verification. Therefore, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) were plotted against lags, producing what is called a Correlogram, as shown in Figure 4. The shapes of the ACF and PACF do not allow for a decisive decision regarding the stationarity of the timeseries.

Sample: 1/01/2010 12/31/2022
Included observations: 4748



Autocorrelation / Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.923	0.923	4044.5	0.000
2	0.876	0.168	7693.6	0.000
3	0.853	0.174	11151.	0.000
4	0.837	0.114	14478.	0.000
5	0.827	0.112	17726.	0.000
6	0.838	0.226	21065.	0.000
7	0.883	0.419	24771.	0.000
8	0.819	-0.576	27959.	0.000
9	0.789	0.173	30923.	0.000
10	0.780	0.075	33819.	0.000
11	0.774	0.064	36673.	0.000
12	0.772	0.077	39510.	0.000
13	0.788	0.116	42467.	0.000
14	0.835	0.114	45788.	0.000
15	0.774	-0.313	48644.	0.000
16	0.746	0.067	51300.	0.000
17	0.738	0.026	53895.	0.000
18	0.732	0.027	56448.	0.000
19	0.728	0.043	58978.	0.000
20	0.744	0.074	61621.	0.000
21	0.791	0.067	64608.	0.000
22	0.731	-0.216	67161.	0.000
23	0.705	0.037	69531.	0.000
24	0.696	0.012	71847.	0.000
25	0.691	0.021	74125.	0.000
26	0.686	0.007	76374.	0.000
27	0.700	0.039	78718.	0.000
28	0.744	0.013	81363.	0.000
29	0.684	-0.139	83597.	0.000
30	0.656	0.001	85653.	0.000
31	0.647	0.002	87652.	0.000
32	0.640	0.011	89609.	0.000
33	0.635	0.013	91538.	0.000
34	0.649	0.035	93554.	0.000
35	0.694	0.043	95860.	0.000
36	0.635	-0.128	97793.	0.000

Fig. 4. ACF and PACF of the ““Max. Daily Peak” series

Yet, stationarity can be rigorously tested using standard stationarity tests such as the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests. As shown in Table 1, the series is stationary at the level in two cases (constant and constant with trend) and non-stationary in the third case (neither constant nor trend).

Table 1. Augmented Dickey Fuller (ADF) and Phillips–Perron (PP) unit root tests for the “load” series

Level	Test critical values			ADF test statistic	Decision	Test critical values			PP test statistic	Decision
	1% level	5% level	10% level			1% level	5% level	10% level		
Constant	-3.43	-2.86	-2.57	-4.29	stationary	-3.43	-2.86	-2.57	-14.11	stationary
Constant and trend	-3.96	-3.41	-3.13	-7.49	stationary	-3.96	-3.41	-3.13	-28.32	stationary
None	-2.57	-1.94	-1.62	-0.13	Non-stationary	-2.57	-1.94	-1.62	-0.43	Non-stationary

4.2.2. Trending and Seasonality

As is well known, the demand for electricity is highly volatile and subject to seasonal effects across various time spans (hours, days, weeks, months, etc.). To test for seasonality, we use a seasonal dummies model. The results in Table 2 show that both the constant and the trend are significant and different from zero, which aligns with the pattern shown in Figure 1.

On the other hand, the day effects of Thursdays and Fridays are significant. Upon closer examination of the load data for the entire sample, it is observed that, with few exceptions, the load often reaches its minimum on Fridays. This could be attributed to Friday being an official holiday in Jordan.[2]

Additionally, the data indicates that the next lowest load levels typically occur on Thursdays, possibly due to many firms in the private sector, particularly those operating on Saturdays, reducing their operations on Thursdays (Almuhtady et al., 2019).

The boxplots shown in Figure 5 confirm this observation, where the median load is lowest on Fridays followed by the next lowest median on Thursdays.

Table 2. Testing the seasonality of the “load” series using dummy variables.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2101.488	11.60434	181.0951	0.0000
T	0.205909	0.002678	76.89554	0.0000
D1	2.544228	13.73564	0.185228	0.8531
D2	3.683356	13.73564	0.268160	0.7886
D3	-0.013465	13.73564	-0.000980	0.9992
D4	-66.10443	13.73564	-4.812620	0.0000
D5	-270.2129	13.73058	-19.67964	0.0000
D6	-19.52295	13.73058	-1.421859	0.1551
R-squared	0.580276	Mean dependent var		2540.226
Adjusted R-squared	0.579656	S.D. dependent var		390.0733
S.E. of regression	252.8999	Akaike info criterion		13.90555
Sum squared resid	3.03E+08	Schwarz criterion		13.91644
Log likelihood	-33003.77	Hannan-Quinn criter.		13.90938
F-statistic	936.1612	Durbin-Watson stat		0.118806
Prob(F-statistic)	0.000000			

Note: “C” stands for constant, Di, i=1,…,6 are the weekdays dummy variables and “T” denotes trend

[2] Although Saturday is a weekly holiday too, but many private businesses work on Saturdays.

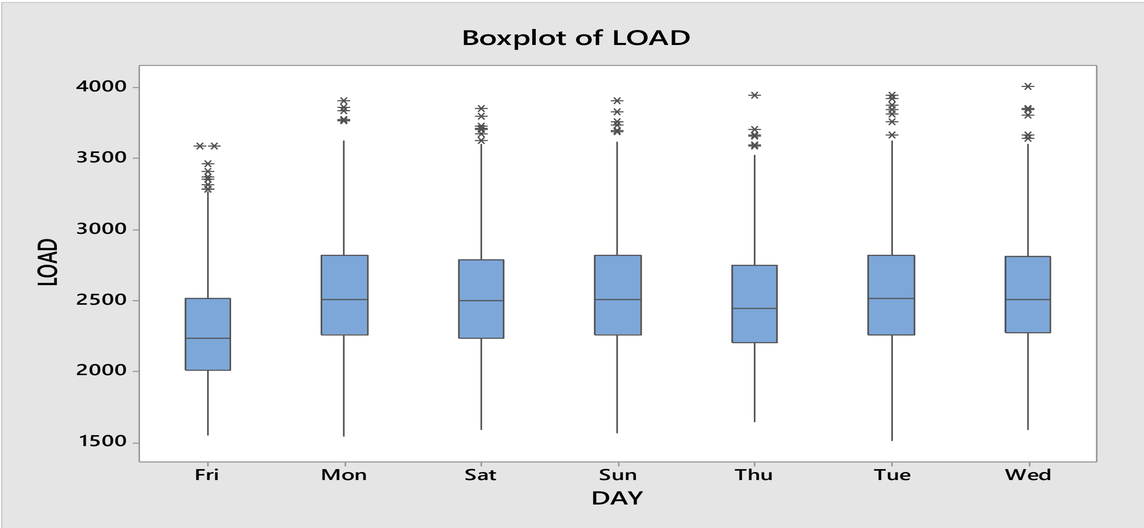


Fig. 5. Boxplot of the "load" series by day of the week.

To detrend and deseasonalize the series, we apply seasonal differencing to the data, giving rise to a new series called "sload," as shown in Figure 6.

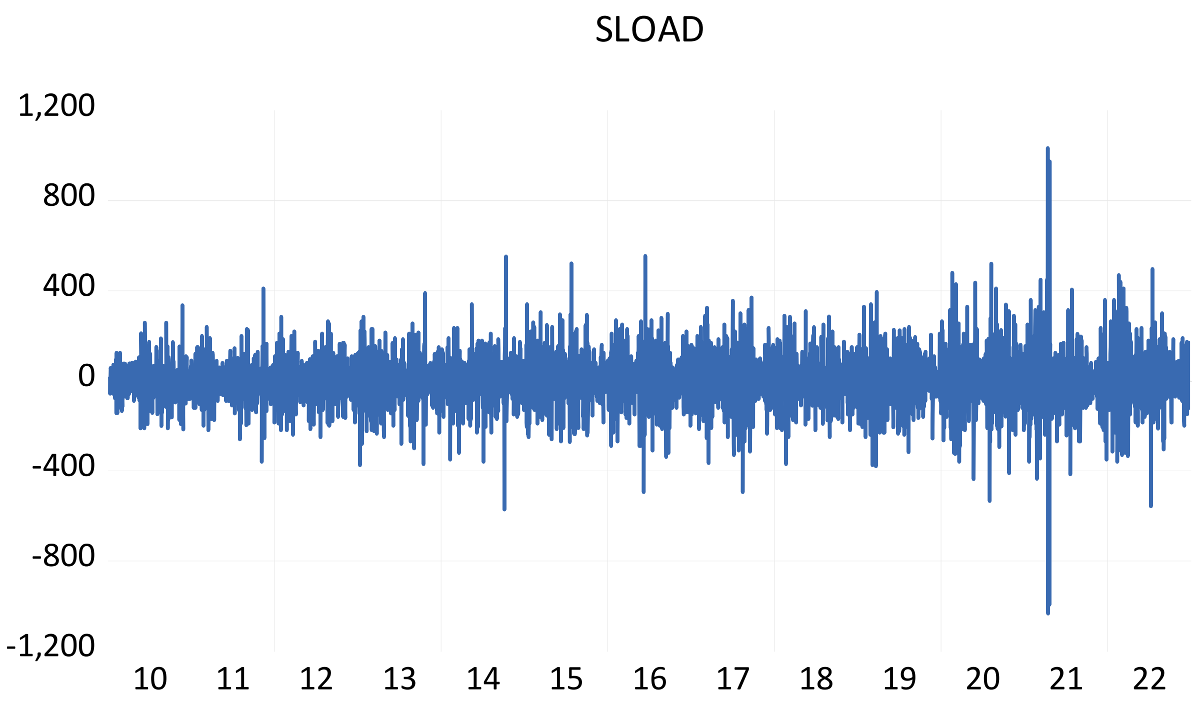


Fig. 6. Plot of the seasonally differenced "load" series

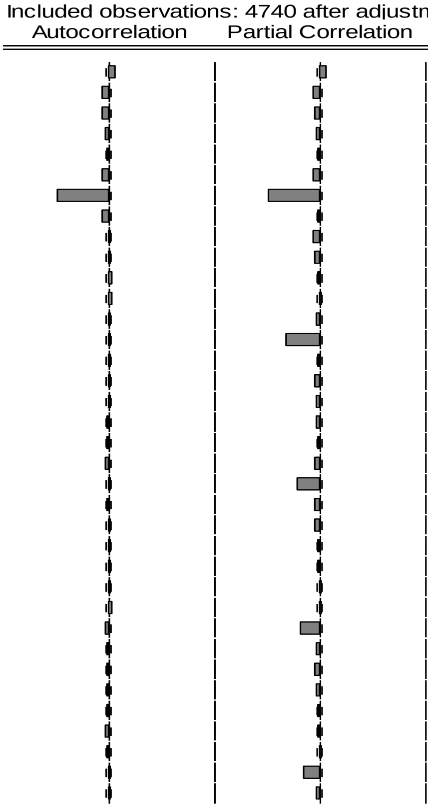
The stationarity of the new series was tested using ADF and PP tests and the results confirm that the series is stationary (see Table 3).

Table 3. Augmented Dickey Fuller (ADF) and Phillips–Perron (PP) unit root tests for the "sload" series

Level	Test critical values			ADF test statistic	Decision	Test critical values			PP test statistic	Decision
	1% level	5% level	10% level			1% level	5% level	10% level		
Constant	-3.43	-2.86	-2.57	-23.98	stationary	-3.43	-2.86	-2.57	-196.81	stationary
Constant and trend	-3.96	-3.41	-3.13	-23.97	stationary	-3.96	-3.41	-3.13	-196.76	Stationary
None	-2.57	-1.94	-1.62	-23.98	stationary	-2.57	-1.94	-1.62	-196.84	stationary

The ACF and PACF functions (Figure 7) of the newly detrended and de-seasonalized series clearly indicates to the presence of seasonality, thus there is a need to employ the SARIMA model rather than the ARIMA model to capture and model seasonality properly.

Included observations: 4740 after adjustments



	AC	PAC	Q-Stat	Prob
1	0.071	0.071	24.083	0.000
2	-0.052	-0.057	36.782	0.000
3	-0.050	-0.043	48.747	0.000
4	-0.030	-0.027	53.145	0.000
5	-0.008	-0.009	53.471	0.000
6	-0.059	-0.064	69.990	0.000
7	-0.492	-0.495	1221.2	0.000
8	-0.054	-0.011	1235.0	0.000
9	0.019	-0.053	1236.7	0.000
10	0.031	-0.040	1241.3	0.000
11	0.038	-0.015	1248.0	0.000
12	0.036	0.009	1254.2	0.000
13	0.028	-0.027	1257.8	0.000
14	0.004	-0.322	1257.9	0.000
15	0.015	-0.014	1258.9	0.000
16	0.005	-0.044	1259.0	0.000
17	0.006	-0.029	1259.2	0.000
18	-0.011	-0.028	1259.8	0.000
19	-0.013	-0.002	1260.6	0.000
20	-0.019	-0.043	1262.3	0.000
21	0.018	-0.207	1263.8	0.000
22	-0.007	-0.037	1264.0	0.000
23	0.005	-0.035	1264.1	0.000
24	0.010	-0.016	1264.6	0.000
25	0.017	-0.010	1266.0	0.000
26	0.014	0.011	1267.0	0.000
27	0.033	0.008	1272.2	0.000
28	-0.032	-0.182	1277.0	0.000
29	-0.001	-0.027	1277.0	0.000
30	-0.015	-0.043	1278.0	0.000
31	-0.013	-0.021	1278.8	0.000
32	-0.004	-0.005	1278.9	0.000
33	-0.017	-0.011	1280.3	0.000
34	-0.016	0.001	1281.4	0.000
35	0.008	-0.153	1281.8	0.000
36	0.012	-0.020	1282.4	0.000

Fig. 7. ACF and PACF of the "sload" series

5. EMPIRICAL RESULTS AND DISCUSSION

After describing the data and investigating its stationarity, seasonality, and trending properties, we next move to the empirical investigation where we apply the four-stage Box-Jenkins (B-J) methodology to the data.

5.1. Identification stage

Traditionally the correlogram (Figure 7) is used to specify or identify the model, i.e., determining the values of the parameters p, q, d, P, Q and D of the ARIMA(p,d,q)×(P,D,Q)$_s$ model or alternatively the SARIMA models. However, it is worth mentioning that determining the order of the SARIMA model using the correlogram is not a straightforward exercise and is largely subject to value judgments. Alternatively, Hyndman and Khandakar (2008) have proposed automatic forecasting algorithms that have been implemented using the forecast package in RStudio. These algorithms can accurately determine the order of the SARIMA model.

5.2. Estimation stage

Using the R command (auto.arima), various specifications of the SARIMA model were obtained. However, based on the Ljung-Box test, the residuals did not exhibit white noise properties. To address this issue and considering the non-constant variance of the series (see Figure 1), the data was transformed using the Box-Cox method. Subsequently, the model was re-estimated, resulting in the following SARIMA model: ARIMA(1,0,1)(2,1,2)[7]. This SARIMA model indicates one autoregressive term, one moving average term, two seasonal autoregressive terms, and two seasonal moving average terms, with a seasonal period length of 7 days. The estimation results shown in Table 4 will be used for forecasting purposes.

Table 4. Estimation results of the ARIMA(1,0,1)(2,1,2)$_{[7]}$ model

```
ARIMA(1,0,1)(2,1,2)[7]

Coefficients:
        ar1     ma1     sar1     sar2    sma1     sma2
     0.8852  0.2128  -0.9244  -0.3509  0.0552  -0.4915
s.e. 0.0567  0.1235   0.1870   0.1208  0.1991   0.1839

sigma^2 = 0.008297:  log likelihood = 85.2
AIC=-156.4   AICc=-155.02   BIC=-138.98
```

5.3. Diagnostic Checking

Diagnostic checking implies testing the residuals. As shown in Figure 8, the residuals are approximately white noise. This fact is confirmed using Ljung-Box Q statistic as shown in Table 5.[3]

Table 5. Ljung-Box test

```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,1)(2,1,2)[7]
Q* = 13.057, df = 8, p-value = 0.1099

Model df: 6.   Total lags used: 14
```

In addition to examining whether the residuals are white noise, diagnostic checking involves testing whether the estimated ARMA process is (covariance) stationary—meaning the inverse AR roots should lie inside the unit circle—and whether the ARMA process is invertible—meaning all inverse MA roots should lie inside the unit circle.[4] Figure 9 confirms that these two conditions are fulfilled, indicating that the model is valid for forecasting purposes.

[3] The null hypothesis: residuals are white noise.

[4] See (Hyndman & Athanasopoulos, 2018) for elaboration on these conditions and how they differ if we use the complex roots instead of inverse roots.

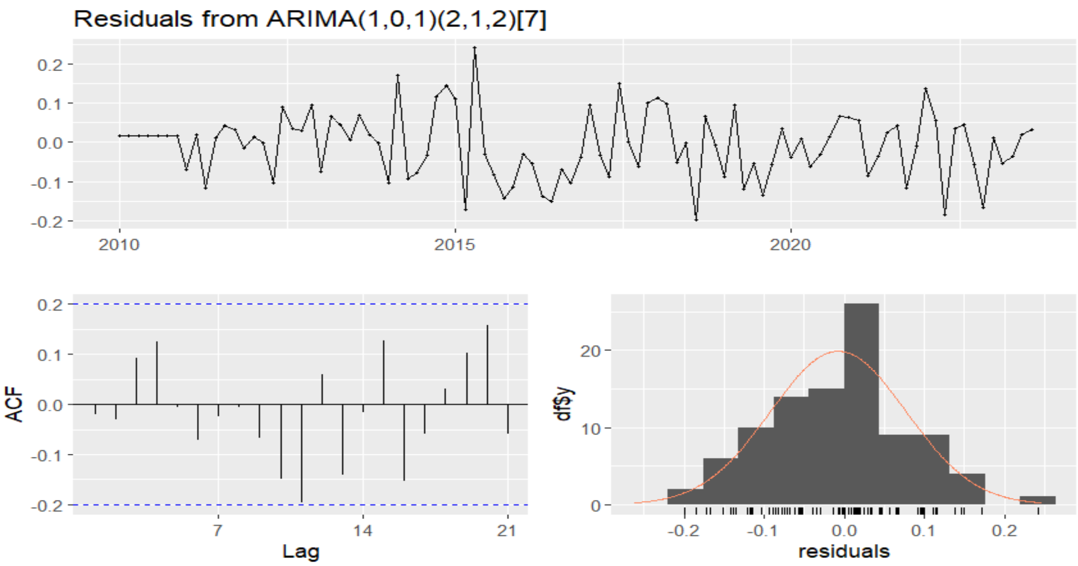


Fig. 8. Residuals from ARIMA(1,0,1)(2,1,2)[7]

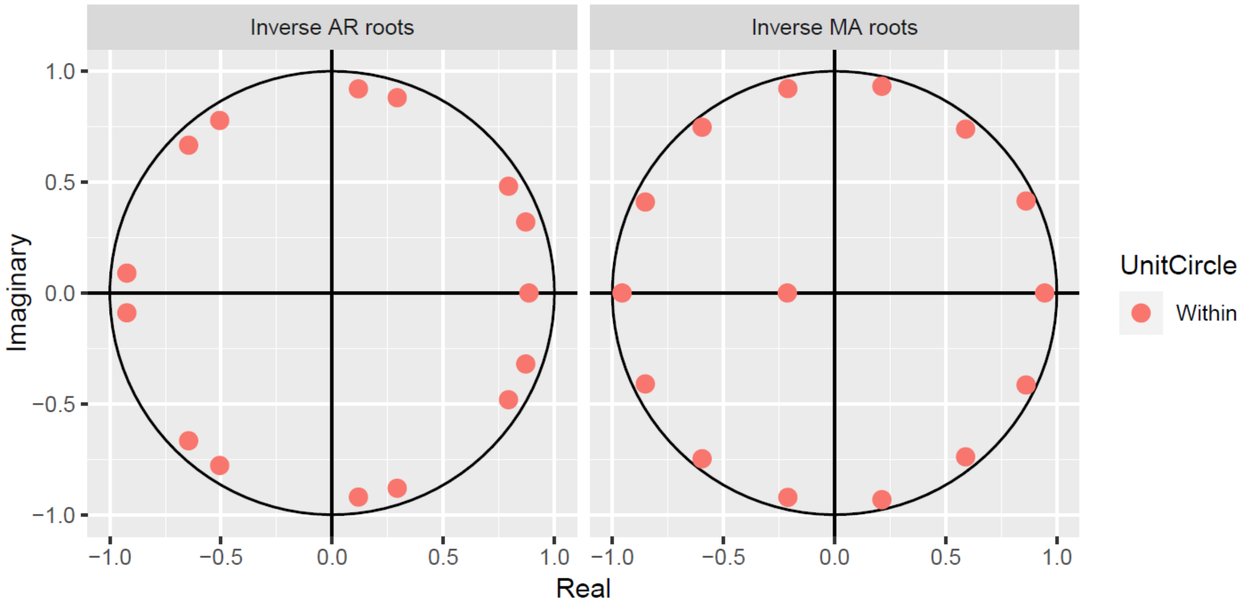


Fig.9. Inverse AR and MA roots

5.4. Forecasting stage

Now, after concluding the diagnostic checking phase, the model is deemed valid for forecasting, marking the final phase in the Box-Jenkins methodology. The shaded areas in dark blue and light blue around the forecasts in Figure 10 represent the 85% and 95% confidence intervals, respectively.

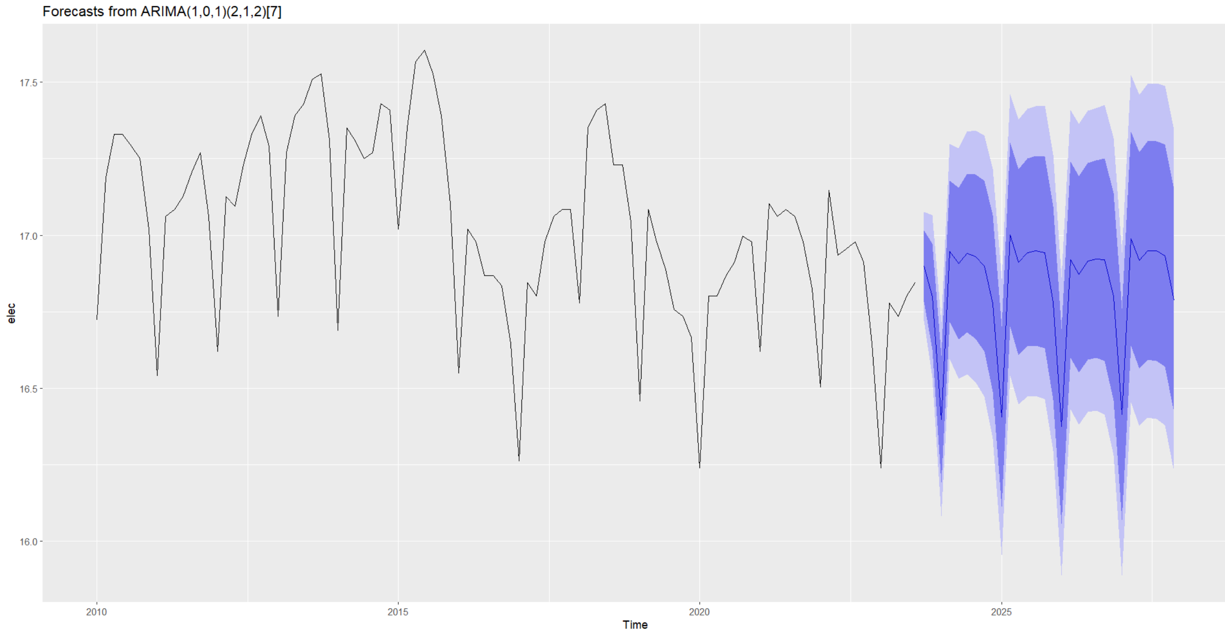


Fig.10. Forecasts from ARIMA(1,0,1)(2,1,2)[7] model

The SARIMA model was used to forecast electricity load for seven days ahead. After applying the inverse of the Box-Cox transformation to the data, the forecasted values are presented in Table 6. For comparison purposes, the table includes the actual minimum, maximum, and average values for the first 7 days of the year 2023 based on 24-hour peak load values.

Table 6. Forecasted value of load for 7 days using the ARIMA(1,0,1)(2,1,2)$_{[7]}$ model

Day	Forecasted value	Actual values 24-hour daily load		
		Minimum	Maximum	Average
1	1945.8	1790	3620	2695
2	1940.2	1780	3890	2915
3	1786.9	1880	3850	2938
4	2048.9	1860	3810	2857
5	2042.3	1890	3600	2689
6	2087.4	1810	3290	2530
7	2069.2	1750	3540	2615

To assess the accuracy of the forecasts, several measures are used, including the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE), among others. The small values of RMSE (0.08469665) and MAPE (0.2902158) confirm that the SARIMA model is a good fit for forecasting electricity load in Jordan.

6. SUMMARY AND CONCLUSION

This paper employs the Box-Jenkins methodology to forecast daily electricity load in Jordan based on 24-hour daily peak load data. The data exhibits several noteworthy features, including upward trending, seasonality, and non-constant variance. Therefore, the SARIMA model was utilized to address trending and seasonality, while the Box-Cox transformation was employed to mitigate the issue of non-constant variance. Following the formal steps of the Box-Jenkins method (identification, estimation, diagnostic checking, and forecasting) resulted in the following SARIMA model: ARIMA(1,0,1)(2,1,2)$_{[7]}$, which was used to forecast future values of electricity load for seven days. The small values of RMSE and MAPE confirm the precision of the forecasts. Hence, electricity companies are strongly encouraged to consider utilizing time series models for forecasting future loads instead of relying on simplistic spreadsheet models.

REFERENCES

Ajlouni, S. (2016). Price and income elasticities of residential demand for electricity in Jordan: An ARDL bounds testing approach to cointegration. ***Dirasat: Administrative Sciences***, 43(1), 335-349.

Alasali, F., Nusair, K., Alhmoud, L., & Zarour, E. (2021). Impact of the COVID-19 pandemic on electricity demand and load forecasting. ***Sustainability***, 13(3), 1435.

Alhmoud, L., & Nawafleh, Q. (2021). Short-term load forecasting for Jordan power system based on narx-elman neural network and ARMA model. ***IEEE Canadian Journal of Electrical and Computer Engineering***, 44(3), 356-363.

Almuhtady, A., Alshwawra, A., Alfaouri, M., Al-Kouz, W., & Al-Hinti, I. (2019). Investigation of the trends of electricity demands in Jordan and its susceptibility to the ambient air temperature towards sustainable electricity generation. ***Energy, Sustainability and Society***, 9(1), 1-18.

As' ad, M. (2012). Finding the best ARIMA model to forecast daily peak electricity demand. Applied Statistics Education and Research Collaboration (ASEARC) - Conference Papers. 12.

Asteriou, D., & Hall, S. G. (2021). ***Applied Econometrics***. Bloomsbury Publishing.

Bashir, T., Haoyong, C., Tahir, M. F., & Liqiang, Z. (2022). Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. ***Energy Reports***, 8, 1678-1686.

Box, G., & Jenkins, G. (1976). ***Time Series Analysis: Forecasting and Control***. San Francisco Holden-Day.

Chodakowska, E., Nazarko, J., & Nazarko, Ł. (2021). Arima models in electrical load forecasting and their robustness to noise. ***Energies***, 14(23), 7952.

Goswami, K., & Kandali, A. B. (2020, July). Electricity demand prediction using data driven forecasting scheme: ARIMA and SARIMA for real-time load data of Assam. In 2020 International Conference on Computational Performance Evaluation (ComPE) (pp. 570-574). IEEE.

Gujarati, D. N., & Porter, D. C. (2009). ***Basic Eeconometrics***. McGraw-Hill Education.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. ***European Journal of Operational Research***, 199(3), 902-907.

Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and models for electric load forecasting: a comprehensive review. ***Logistics, Supply Chain, Sustainability and Global Challenges***, 11(1), 51-76.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. ***International Journal of Forecasting***, 32(3), 914-938.

Hyndman, R. J., & Athanasopoulos, G. (2018). ***Forecasting: Principles and Practice***. OTexts.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. ***Journal of statistical software***, 27, 1-22.

Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. ***Sustainable Cities and Society***, 35, 257-270.

Nti, I. K., Teimeh, M., Nyarko-Boateng, O., & Adekoya, A. F. (2020). Electricity load forecasting: A systematic review. ***Journal of Electrical Systems and Information Technology***, 7(1), 1-19.

Tawalbeh, N. A., Al Mattar, S. S., Elhaija, W. S. A., & Khasawneh, M. A. (2021, October). Impact of COVID-19 on Electric Energy Consumption. In *2021 12th International Renewable Energy Congress (IREC)* (pp. 1-6). IEEE.