



**Tikrit Journal of Administrative
and Economics Sciences**

مجلة تكريت للعلوم الإدارية والاقتصادية

EISSN: 3006-9149

PISSN: 1813-1719



**Nonlinear Autoregressive Neural Network and SARIMA Model for
Forecasting Peak Electricity Demand in the Kurdistan Region**

Huda Kardagh Yalda*, Saman Hussein Mahmood

College of Administration and Economics/Salahaddin University-Erbil

Keywords:

Time series, NAR neural network, SARIMA,
Forecasting, Peak Electricity Demand

ARTICLE INFO

Article history:

Received 29 Oct. 2024

Accepted 09 Dec. 2024

Available online 30 Jun. 2025

©2023 THIS IS AN OPEN ACCESS ARTICLE
UNDER THE CC BY LICENSE

<http://creativecommons.org/licenses/by/4.0/>



***Corresponding author:**

Huda Kardagh Yalda

College of Administration and
Economics/Salahaddin University-Erbil



Abstract: With the great economic and social development witnessed by the Kurdistan Region of Iraq, the demand for electrical energy has increased significantly, causing an imbalance as the current generation and distribution infrastructure struggles to keep pace. The researchers aim to study and address this issue by forecast peak demand to support future planning in the sector to ensure efficient electricity supply during peak times and avoid overloading the network, using artificial neural networks methods that are characterized by their ability to learn and adapt to complex data and traditional methods (Box-Jenkins method) known for their accuracy in analysis. We will analyze the time series data of monthly peak electricity demand for ten years in the Kurdistan Region (January 2014 to July 2024) with a total of 127 observations.

The results of the Box-Jenkins method identified the SARIMA (1,0,1)(0,1,1)₁₂ model as the most suitable for time series analysis, as it showed high predictive accuracy and outperformed other models in terms of RMSE, AIC and SBIC for forecasting electricity demand increase, While the results of the nonlinear autoregressive neural network (NARNN) model, which was structured by adjusting the hidden layer neurons and delay numbers through trial and error, showed that the optimal NARNN model (1:12,10) achieved the lowest RMSE, MAE, MAPE, and R values when compared to the other models tested.

When comparing the performance of the NAR neural network with the SARIMA method, the SARIMA (1,0,1)(0,1,1)₁₂ model showed superiority over the NAR neural network in terms of accuracy, making it the better choice for forecasting peak electricity consumption and working to reduce the gap between demand and production in the future.

الشبكة العصبية غير الخطية الانحدارية ونموذج SARIMA لتوقع الطلب على ذروة الكهرباء في إقليم كردستان

سامان حسين محمود

هدى قرداغ يلدا

كلية الإدارة والاقتصاد/جامعة صلاح الدين - أربيل

المستخلص

مع التطور الاقتصادي والاجتماعي الكبير الذي يشهده إقليم كردستان العراق، ازداد الطلب على الطاقة الكهربائية بشكل كبير، مما تسبب في اختلال التوازن، حيث البنية التحتية الحالية تكافح للتوليد والتوزيع لمواكبة الطلب. يهدف الباحثون إلى دراسة ومعالجة هذه القضية من خلال التنبؤ بذروة الطلب لدعم التخطيط المستقبلي في القطاع لضمان إمداد فعال بالكهرباء خلال أوقات الذروة وتجنب التحميل الزائد على الشبكة، باستخدام أساليب الشبكات العصبية الاصطناعية التي تتميز بقدرتها على التعلم والتكيف مع البيانات المعقدة والأساليب التقليدية (طريقة بوكس-جينكينز) المعروفة بدقتها في التحليل. سوف نقوم بتحليل بيانات السلسلة الزمنية للطلب الشهري على الكهرباء لمدة عشر سنوات في إقليم كردستان (يناير 2014 إلى يوليو 2024) بإجمالي 127 ملاحظة. أظهرت نتائج طريقة بوكس-جينكينز أن نموذج SARIMA (1,0,1) (0,1,1)₁₂ هو الأنسب لتحليل السلاسل الزمنية، حيث أظهرت دقة تنبؤية عالية وتفوقها على النماذج الأخرى من حيث (RMSE، AIC و SBIC) في التنبؤ بزيادة الطلب على الكهرباء. بينما أظهرت نتائج نموذج الشبكة العصبية الذاتية الانحدار غير الخطية (NAR)، الذي تم هيكلته عن طريق ضبط عدد العصبونات في الطبقة المخفية وأعداد التأخير من خلال التجربة والخطأ، أن النموذج الأمثل NAR (1:12, 10) حقق أدنى قيم لـ (RMSE، MAE، MAPE، R) مقارنة بالنماذج الأخرى التي تم اختبارها.

عند مقارنة أداء الشبكة العصبية NAR مع طريقة SARIMA، أظهر نموذج SARIMA (1,0,1) (0,1,1)₁₂ تفوقاً على الشبكة العصبية NAR من حيث الدقة، مما يجعلها الخيار الأفضل للتنبؤ باستهلاك الذروة للكهرباء والعمل على تقليص الفجوة بين الطلب والإنتاج في المستقبل.

الكلمات المفتاحية: السلاسل الزمنية، SARIMA، NAR NN، التنبؤ، ذروة الطلب على الكهرباء.

المقدمة

1. Introduction

The demand for electricity has grown significantly to meet the demands of the manufacturing, service and consumer sectors as a result of the remarkable economic and social development in the Kurdistan Region of Iraq. Since energy is a major factor in development, effective planning and management are essential. Accurate estimation of peak electricity consumption is an essential part of energy management and electrical infrastructure performance.

This research aims to study this problem and forecast peak demand by use SARIMA model and nonlinear neural regression networks (NARNN) in order to support future planning in the sector to ensure efficient electricity supply during peak times and avoid overloading the grid. This would help improve grid efficiency and reduce operating expenses.

Time series analysis is a valuable method for understanding and extracting insights from sequential data collected over time, across various fields such as finance, economics, signal processing, and environmental research. While multiple models exist for time series forecasting, spectral analysis and the autoregressive integrated moving average (ARIMA) model proposed by Box and Jenkins are the most widely used. ARIMA models offer flexibility in linear modeling, but their limitations can be addressed by artificial neural networks (ANNs). (Box, Jenkins, Reinsel, & Ljung, 2015) (Brockwell & Davis, 2002) (Kadir, 2020)

ANNs are powerful tools for time series forecasting due to their strength in handling nonlinear patterns. These networks are composed of interconnected nodes organized in layers, resembling artificial neurons. The structure includes input, hidden, and output layers. Connections between these neurons are assigned weights, and the network learns by adjusting these weights based on input data. This makes ANNs particularly effective for working with non-stationary time series data (Zhang & Qi, 2005) (Mahmood, 2016).

The results obtained from using SARIMA and NARNN models were very effective, with the Box-Jenkins time series analysis model outperforming the NARNN model and being used for electricity demand forecasting.

2. Material and Methodology: The ARIMA forecasting model setup involves three main steps: model identification, model parameter estimation, and diagnostic confirmation. The initial phase involves analyzing data to determine the appropriate class of ARIMA processes, determining the sequence of autoregressive and moving average polynomials and the order of consecutive and seasonal differencing.

2-1. Time series: A time series is a set of apparent values arranged over time, such as annual, quarterly, monthly, or daily. It can be analyzed to understand changes in the phenomenon's value over time. There are two types of changes: general changes over the long term and seasonal trends. Periodicity

is determined by developments, economic and political factors, and occasional changes, other random or systematic fluctuations (Allende, Moraga, & Salas, 2002).

Components of Time Series: Time series data can be decomposed into four main components, which are essential for understanding its underlying structure (Sarhan, 2018): Long-term movements in the mean are referred to as a trend (T_t). Seasonal effects (I_t), or calendar-related cyclical changes, Cycles (C_t): other cyclical fluctuations (such as business cycles), and the value of the irregular component (Residuals), other random or systematic fluctuations (Shumway & Stoffer, 2006).

$$Z_t = T_t + S_t + C_t + I_t \dots \dots (1)$$

2-1-1. Stationary Time Series: A stationary time series is a series of time data whose statistics between different time periods are stationary and unchanged with time, this means that the time-homogeneous series, (Koenecke, 2020), sometimes it is similar to plotting a series in period $\{t, t+h\}$ where we can say that the series is stationary and its properties do not change over time. In other words, a time series is "stationary" if the mean, variance, and autocovariance meet the following conditions: $\{s, s+h\}$ (Yaffee & Mcgee, 2000):

$\mu(t) = E(Z_t)$ $t = 1, 2, \dots, n$ $E(Z_t)$ is the expected value of the random variable X_t , and $\mu(t)$ is the expected value (or mean)

$Var(Z_t) = E(Z_t - \mu)^2$ The variance of the process Z_t remains constant

over time, indicating that the variability of the data points does not change. The covariance function measures the relationship between values at different points in time for a given time series, Where $t = 1, 2, \dots, n$ & $h = 1, 2, \dots, n - t$

$$\gamma(t, t+h) = Cov(Z_t + \mu_t) = E[Z_t - \mu_t * (Z_{t+h} - \mu_{t+h})]$$

The correlation function can detect non-stationarity in a time series when its coefficients remain high over different time periods, indicating that the values of the autocorrelation function are not yet zero (Zhang, Time series forecasting using a hybrid ARIMA and neural network model, 2003).

2-1-2. Types of Box-Jenkins Models:

1. Autoregressive of order (P) Model

Autoregressive models arise when the current value of the series is a function of its value in previous periods, in addition to some errors. If Z_t is the time series' current value and $(Z_{t-1}, Z_{t-2}, \dots, Z_{t-n})$ are its values from earlier periods, and we discover that Z_t depends on or is impacted by its earlier values, we may characterize this relationship using an autoregressive model of order p Written AR(P) if (Sarhan, 2018).

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + e_t \dots \dots (2)$$

Where: (e_t) Independent random variables follow a normal distribution with zero mean and variance σ^2 . $\{\varphi_1, \varphi_2, \dots, \varphi_p\}$ are unknown parameters.

2. Moving Average Model of order (q)

The current observation Z_t can be expressed as a linear function of the current random change at a_t and the previous random change at $(e_{t-1}, e_{t-2}, \dots)$ then the resulting process is called a moving average (MA) process of order q and it is formulated as follows (Tsay, 2005)

$$Z_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \dots \dots \dots (3)$$

Where: $\{\theta_1, \theta_2, \dots, \theta_q\}$ are unknown parameters. $\{e_t\}$ white noise purely random variable

3. Mixed Autoregressive Moving Average Models: The autoregression and moving average procedure can be described using the ARMA (p, q) model, where higher-order combinations are represented as ARMA. Here, p refers to the order of the autoregressive component, and q denotes the order of the moving average component. (Brockwell & Davis, 2002)

$$Z_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \dots \dots (4)$$

$$\varphi(B)Z_t = \vartheta(B)e_t \dots \dots (5)$$

Where: $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$, is (AR) polynomial

$\vartheta(B) = 1 + \vartheta_1 B + \vartheta_2 B^2 + \dots + \vartheta_q B^q$, is (MA) polynomial

The AR component uses past values of the series to predict future values, while the MA component uses past errors to predict future values. The ARIMA (p, d, q) model extends ARMA to handle non-stationary data by applying differencing (d) to the series before fitting the ARMA model.

$$\varphi(B)(1 - B)^d Z_t = \vartheta(B)e_t \dots \dots (6)$$

4. SARMA (Seasonal Autoregressive Moving Average) Model; A non-seasonal model is sometimes combined with seasonal models to create a suitable representation, referred to as ARMA (p, q)(P, Q).

$$\phi_P(B^S)\phi_p(B)Z_t = \theta_Q(B^S)\vartheta_q(B)e_t \quad \dots \dots (7)$$

The SARMA model is extended to handle non-stationary data by applying differential (d) to the previous series, thus combining the seasonal and non-seasonal components and adding differentials to both, which is represented by ARIMA (p, d, q) (P, D, Q). The operator for this model is created using the forward shift operator (Zhang, 2003) (Kadir, 2020).

$$\phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Z_t = \theta_Q(B^S)\vartheta_q(B)e_t \quad \dots \dots (8)$$

2-1-3. Stages of building the model: The Box-Jenkins method for building time series models involves four steps (Ahmed & Mahmood, 2023):

1. **Identification:** This involves using autocorrelation functions and partial autocorrelation functions to determine the appropriate ARMA model.
2. **Estimation:** Once the model is identified, its parameters are estimated using methods like maximum likelihood, least squares, or Yule-Walker.
3. **Diagnostic:** The model's validity is checked by analyzing residuals, adjusting parameters, and comparing criteria.
4. **Forecasting:** The model is used to forecast the time series' future values.

2-2. Artificial Neural Networks: Neural networks are a form of artificial intelligence designed to simulate the workings of the human brain, enabling tasks such as sorting, comparing, and predicting without relying on a specific data model. Researchers are exploring their flexibility and performance in comparison to statistical methods. Researchers compare these networks to statistical methods (Bishop, 1995) (Fausett, 1994).

The term "artificial neural networks" is derived from the idea of replicating the neural structure of the brain. Living organisms are often described as biological computers, electronic brains, analogous systems, or models of connectivity and parallelism. Technically, some researchers aim to replicate the cognitive abilities of the brain, using learning and trial-and-error techniques to achieve human-level problem solving. They use non-algorithmic methods to solve incomplete problems and work with noisy data (Fausett, 1994) (Bishop, 1995).

2-2-1. Artificial Neural Networks Architecture: Neural networks are made up of neurons, which are interconnected processing units. The structure

and function of these neurons is determined by their organization, or network architecture. Regardless of the input data, this architecture is fixed and computation-free (Fausett, 1994).

Three layers usually make up a neural network: the input layer receives data, the hidden layer processes it, and the output layer generates the results. Each layer contains nodes, which serve as connection points, and levels, which are groups of nodes that receive inputs and generate outputs. The strength of the connections between layers is represented by weights. (Fausett, 1994) (Mahmood & Ahmed, 2023)

A neuron consists of four components: weighting coefficients, a summation function, a transfer function, and an output function. Its operations include receiving input signals, modifying them, summing the inputs, using the activation function to generate the output, and passing the result either to other neurons or as the final output of the network.

The architecture of a neural network describes how its units and connections are structured. There are two primary types: feedforward neural networks (FFNN) and feedback neural networks. In FFNN data moves in one direction, while in feedback networks, information can flow in both directions; these basic architectures form the basis for more specialized architectures such as NAR neural network, convolutional neural networks (CNNs) and competitive generative neural networks (GANs). Neural networks are also categorized as either single-layer or multi-layer (Galushkin, 2007).

Neural networks can be single-layer or multilayer. Multilayer perceptrons (MLPs) were initially introduced to address complex classification challenges. However, due to their global approximation property, they quickly became widely used as nonlinear regression models and later for time series modeling and forecasting. Despite their versatility, estimating and specifying these models requires advanced techniques, and identifying the correct structure is not easy. By nature, these models are over-parameterized, and their minimization error functions often have many local minima, and their implementation can be very difficult (Fausett, 1994) (Galushkin, 2007).

2-2-2. Nonlinear autoregressive (NAR) neural network: The NAR neural network is a powerful feed-forward network model that excels at identifying patterns and nonlinear features in time series data. It works as a dynamic

network, where connections that span its layers, meaning the current output is influenced by previous output value $y(t-1)$, $y(t-2)$, ..., $y(t-d)$. The NAR network is capable of effectively forecasting time series and can be represented in the following manner (Benrhmach, Namir, Namir, & Bouyaghroumni, 2020) (Ali, Mahmood, & Wahdi, 2022):

$$y(t) = f(y(t-1) + y(t-2) + \dots + y(t-d)) + \varepsilon(t) \dots \dots (9)$$

Where, y represents the combined interference values from $N-1$ interferers. The function $f(\cdot)$ is a nonlinear function that can be approximated using a neural network. The variable d represents the time delay for the interference time series and ε_t represents the error between the predicted and actual values. This equation shows how the predicted value at time t depends on the past values of the time series (Taherdangkoo, Tatomir, Taherdangkoo, Qiu, & Sauter, 2020).

2-2-3. Network Architecture: The NAR network used is a feed-forward neural network consisting of three layers: the input layer, the hidden layer, and the output layer (Figure 1). The sigmoid function, a continuous nonlinear function, is the most commonly used activation function in neural network design with backpropagation training. The hidden layer uses a logistic sigmoid activation function, while the output layer uses a linear function. The optimal number of time delays and hidden nodes can often be determined by trial and error (Ratur & Sargsyan, 2018).

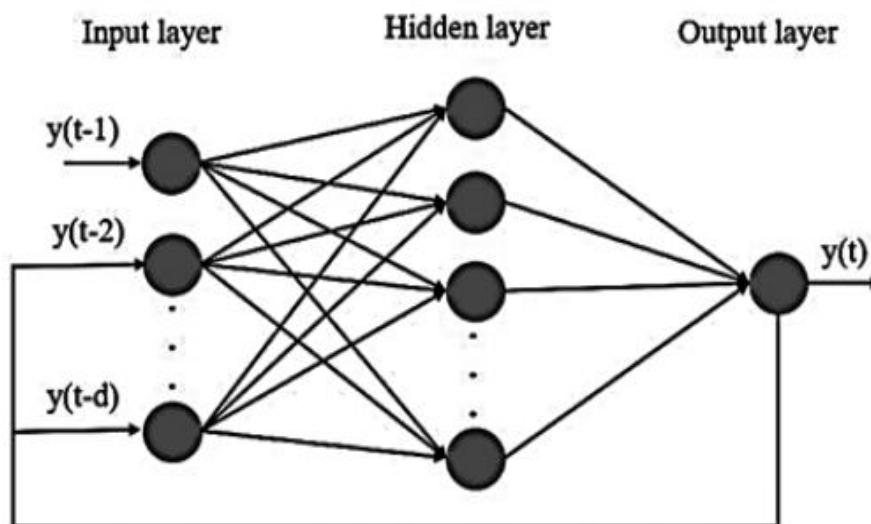


Figure (1): The architecture of the NAR model

Thus, the NAR neural network is specifically defined by an equation of the following form:

$$y(t) = \alpha_0 + \sum_{j=1}^k \alpha_j \Phi \left(\sum_{i=1}^{\alpha} \beta_{ij} y_{t-i} + \beta_{0j} \right) + \varepsilon_t \quad \dots \dots (10)$$

The NAR network's architecture equation (10) includes: **a**: The number of data points, **k**: The number of hidden layers, **Φ** : The activation function used in the hidden layers, **β_{ij}** : The weights connecting input units to hidden units. **α_j** : The weights connecting hidden units to the output unit, **β_{0j}** and **α_0** : Bias value of hidden units and output unit.

The optimization of the architecture focuses on minimizing the number of synapses (weights) and neurons to decrease the network's complexity, enhance computation times, and preserve generalization capabilities. Two primary approaches to optimizing network architecture have been proposed in the literature:

Selection Approach: This method begins with the construction of a complex network containing a large number of neurons. The goal is to reduce the number of unnecessary neurons and eliminate redundant connections either during or after the learning process.

Incremental Approach: This strategy starts with the simplest possible network and progressively adds neurons or layers until an optimal architecture is achieved.

An effective strategy for assessing prediction error involves using a dataset that was not employed in building the predictor, known as the test set. The dataset can be divided into three types of target time steps:

Training Set: These datasets are presented to the network during training, allowing adjustments based on the network's errors.

Validation Set: These datasets are used to evaluate the network's generalization capabilities and to stop training when further improvements in generalization cease.

Testing Set: These datasets do not influence the training process and provide an independent assessment of the network's performance both during and after training.

Evaluation Criteria: Using a set of standard statistical metrics, we assessed the predictive models' accuracy. These metrics include (Benrhmach, Namir, Namir, & Bouyaghroumni, 2020):

Mean Absolute Error (MAE): Expresses the average difference between actual and predicted values, and provides an indication of how close the predictions are to reality.

Root Mean Square Error (RMSE): Gives greater weight to large errors, and provides a more accurate estimate of overall model performance.

Correlation Coefficient (R): Measures the strength of the linear relationship between predicted and actual values, and indicates how similar the direction of changes in the data is.

Selecting the best model: We applied the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) to determine the best model among the options. (Allende, Moraga, & Salas, 2002)

3. Results In this research, we will use artificial neural networks to analyze the monthly time series data of electricity peak demand in the Kurdistan Region. We will compare the results of the neural networks with the traditional Box-Jenkins methodology.

The data analyzed in this study represents the average monthly peak electricity demand in megawatts (MW) for the Kurdistan Region. The data covers a period of five years, from January 1, 2014, to July 2024, totaling 127 observations. The data was obtained from the Kurdistan Region's Ministry of Electricity and analyzed using Statgraphics V.17 and Matlab(R2022b).

3-1. Analysis of the Time Series: We created a visual representation of the time series data prior to the analysis, the time series data shows the peak demand for electricity in the Kurdistan Region. The data exhibits a general trend and cyclical patterns, with spikes that occur regularly and increase over time. These patterns indicate a combination of a general trend and seasonality, as well as a degree of instability in the series.

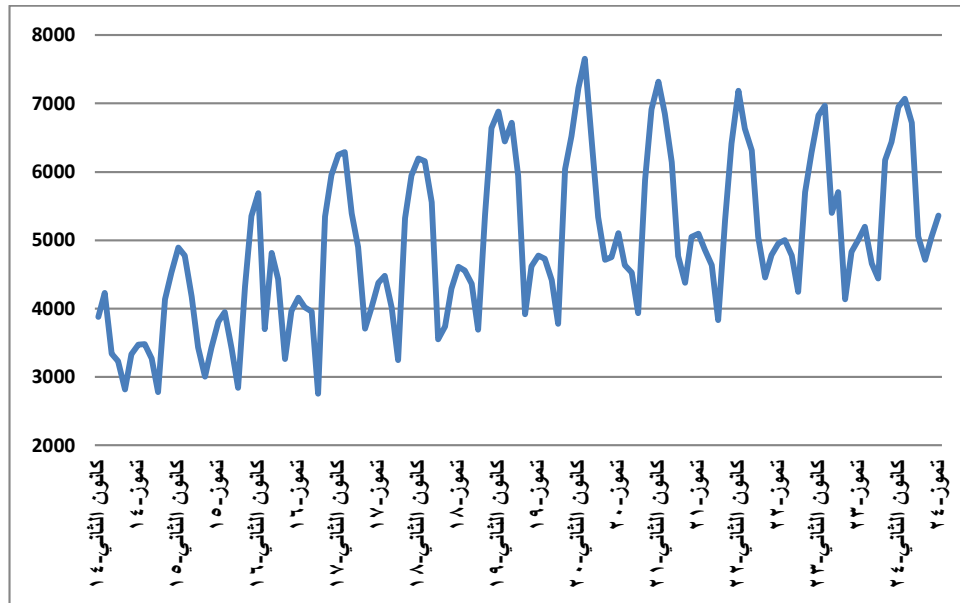


Figure (2): Original data plot peak electricity demand in Kurdistan Region
For greater accuracy, we draw both the autocorrelation function and the partial autocorrelation function, as in figure (3)

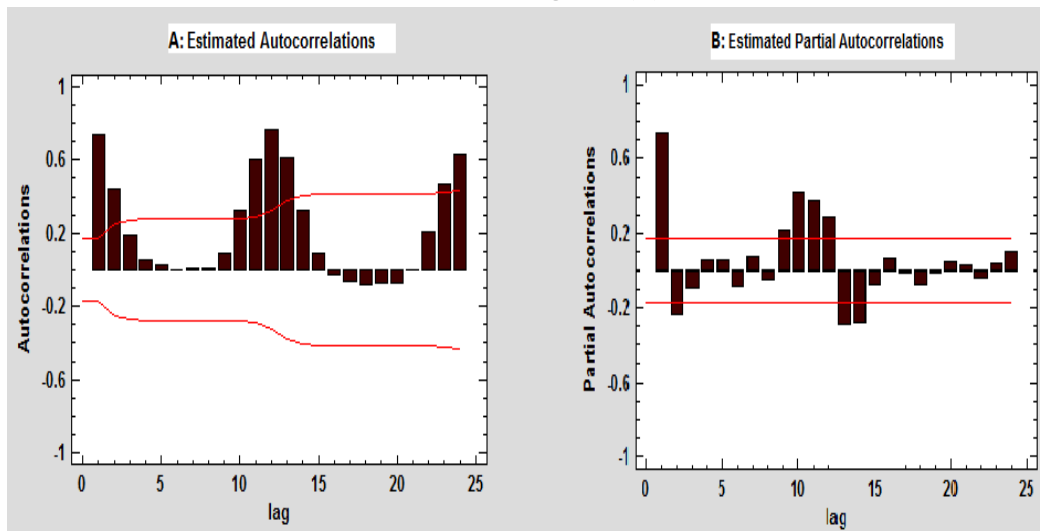


Figure (3): ACF & PACF for the original electricity data

Figure 3 shows that many of the autocorrelation coefficients (ACF) and the partial autocorrelation function (PACF) fall outside the confidence limits at the 95% level, which is an indication of the non-stationarity of the time series. This can be confirmed by the Augmented Dickey-Fuller (ADF) test, which is one of the most important unit root tests, to prove the stationarity of the series or not, as shown in the results of Table 1, as its value is equal to (-2.74) and value (0.0713), and therefore the null hypothesis cannot be rejected and indicates the non-stationarity of the time series.

Null hypothesis: There is a unit root in the data.

Alternative hypothesis: There is no unit root in the data.

Table (1): Results of Augmented Dickey-Fuller test for the original series

	Test Statistics	P-value	Critical Value	Null Rejected
Before Differencing	-2.74	0.0713	-2.580	False
After Differencing	-14.98	0.000	-2.587	True

To make the time series stationary, we applied several seasonal and non-seasonal differences, which is essential for preparing the data for modeling. We found that a first seasonal difference ($d = 1$) effectively stationary the series during peak electricity demand periods, making the data more suitable for the SARIMA model, where stationary is crucial. This was confirmed by the ADF test as shown in Table (1), where the absolute test statistic (14.98) and the p value (0.000) was less than 0.05. To further verify the stationarity, we will conduct ACF and PACF tests on different time series Figure3.

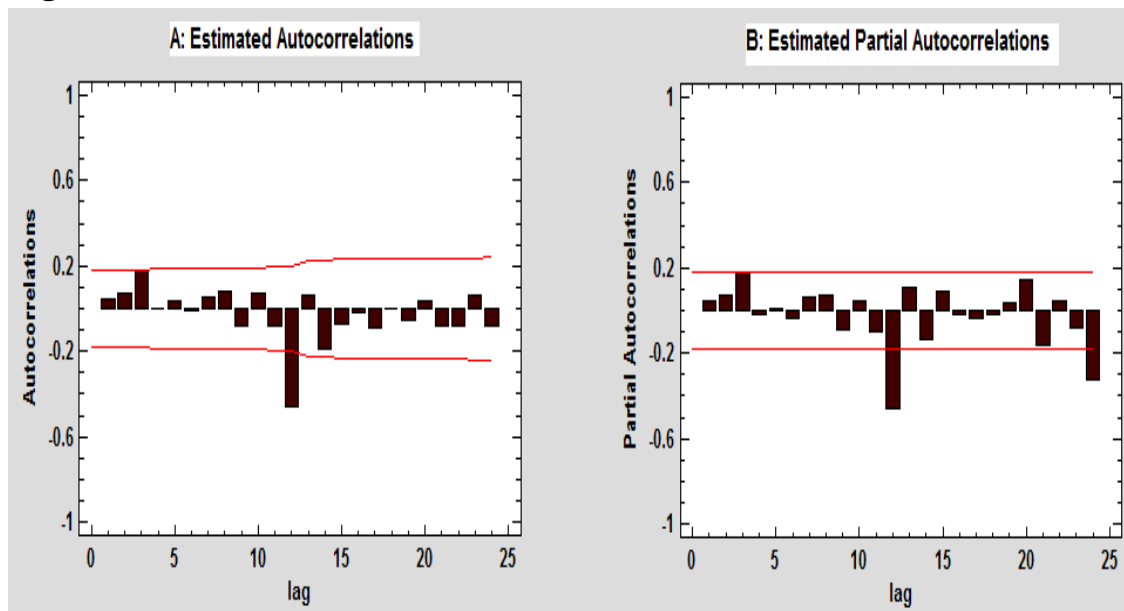


Figure (3): Adjusted ACF and PACF after first seasonal differences

To select the optimal SARIMA model, we analyzed several well-known models in Table 2. This requires experience in time series analysis. We evaluated the models based on their performance (RMSE, AIC, and SBIC) and selected the model that achieved the best overall results. It is SARIMA (1,0,1) (0,1,1)₁₂. This model has the lowest values for the comparison criteria compared to the other models.

Table (2): Proposed models with significant estimates

Model	RMSE	AIC	SBIC	Sig. Parameters
SARIMA (1,0,1) (0,1,1) ₁₂	393.303	11.996	12.064	Yes
SARIMA (1,0,1) (1,1,1) ₁₂	394.902	12.020	12.109	No
SARIMA (1,0,1) (1,0,1) ₁₂	401.583	12.054	12.143	No
SARIMA (1,0,0) (1,1,1) ₁₂	438.297	12.213	12.280	No
SARIMA (1,0,0) (0,1,1) ₁₂	441.973	12.214	12.259	No

We estimated the parameters of this model and selected the best estimation method in Table 3.

Table (3): Estimated Parameters Values of the SARIMA (1,0, 1) (0,1,1)₁₂ Model

Parameter	Estimate	Standard Error	t	P-value
AR(1)	0.99998	0.0040	252.075	0.00
MA(1)	0.8058	0.0546	14.7636	0.00
SMA(1)	0.756	0.0593	12.7579	0.00

Moreover, we can express the previous model's form as follows:

$$(1-0.99998B)Y_t=(1+0.8058B)(1-B^{12})(1-0.7556B^{12})\epsilon_t$$

Model Evaluation: Once we have identified and estimated the candidate SARIMA (1,0,1) (0,1,1)₁₂ models, we need to evaluate how well it fits the data. This involves checking both the model parameters and the residuals. To assess the model's suitability, we conducted a diagnostic check of the residuals for the SARIMA (1,0,1) (0,1,1)₁₂ model. We used ACF and PACF plots to analyze the residuals, as shown in Figure 6.

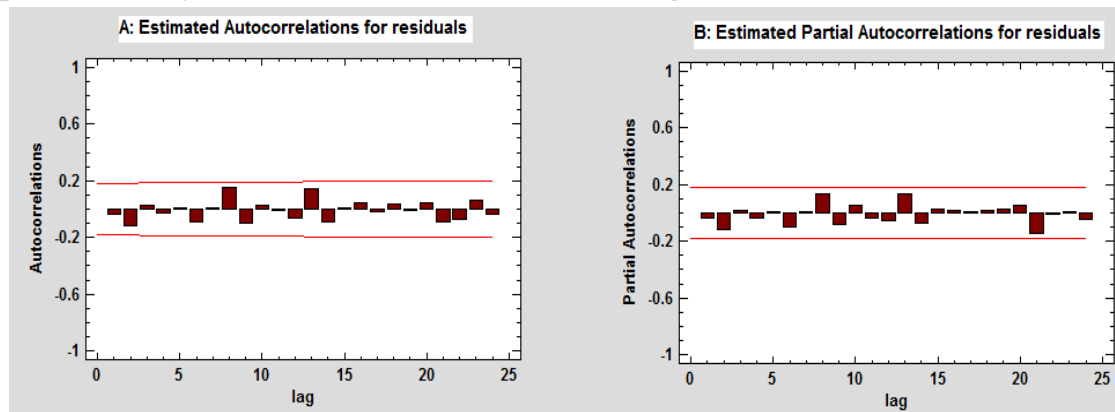


Figure (4): ACF and PACF of residual model

We found that all the ACF and PACF values of the residuals were within the confidence limits, further supporting the conclusion that the residuals are random and the model is appropriate.

We checked for autocorrelation in the residuals using the Box-Pierce test to make sure the model fit. The test yielded a p-value of 0.821767, which is greater than 0.05. This indicates that there is no significant autocorrelation in the residuals, confirming that the model is appropriate and effective.

Forecasting: Forecasting peak demand and managing seasonal loads are crucial for optimizing electrical grids, reducing costs, and promoting sustainable energy practices. By accurately forecasting peak demand, utilities can make informed decisions to optimize energy generation, distribution, and consumption.

The next step, after diagnosing and selecting the most suitable model, marks the conclusion of the time series analysis modeling.

Table (4): The forecasted monthly peak electricity demand

Period	Forecast	Limit Lower 95%	Limit Upper 95%
8/2024	5276.9	4389.36	6614.36
9/2024	4895.03	4109.76	6051.29
10/2024	4174.97	3580.89	5005.38
11/2024	6348.82	5052.1	8541.06
12/2024	7246.93	5583.58	10321.8
1/2025	7893.4	5936.02	11776.7
2/2025	7641.16	5770.26	11307.4
3/2025	6790.35	5253.55	9598.02
4/2025	5556.19	4470.34	7338.8
5/2025	4630.01	3841.31	5826.27
6/2025	5229.57	4233.96	6837.38
7/2025	5524.47	4413.27	7383.55

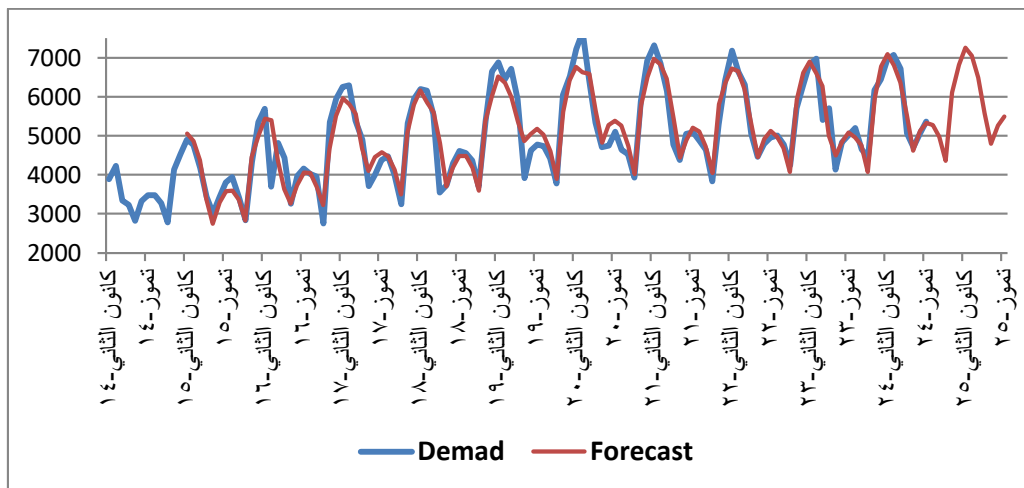


Figure (5): Forecasting using SARIMA (1,0,1) (0,1,1)₁₂ model from 8/2024 to 7/2025

3-2. Application of NAR Neural Network: Study and analysis of peak electricity demand using NAR NN model with the data collected for this study from 127 observations, divided as follows: 70% of the observations, which equals 89 observations, as the training set, and 15%, which equals 19 observations for both the validation set and the test set.

The prediction process starts by defining an input layer which is a single neuron representing the monthly time series of peak electricity demand. Determining the number of delays (lags) and hidden nodes, since the time series is seasonal and the PACF reveals large spikes in seasonal delays, this indicates the presence of seasonal dependencies at these points. As shown in Figure (3) and from the previous data analysis, both regular and seasonal delays will be taken into account when selecting the inputs for the neural network. Given the importance at delay (lags) 12, the preferred number of delays is 12. However, through the trial and error process, we confirmed this aspect. We conducted multiple experiments on peak power demand, varying the number of neurons in the hidden layer from 1 to 10 in each experiment, and varying the number of delays from 1 to 15. Finally, we determined the best NAR neural network model (1:12,10) which is a single input model with 12 delays and 10 hidden neurons, and some of the results are shown in Table 5. Forecasting peak demand and managing seasonal loads are crucial for optimizing electrical grids, reducing costs, and promoting sustainable energy practices. By accurately forecasting peak demand, utilities can make informed decisions to optimize energy generation, distribution, and consumption.

Table (5): Comparisons results for the NAR neural network

ID	Architecture	RMS	MAE	MAPE	R
1	(1:12,1)	7747.99	407.1 5	985.28	0.896
2	(1:12,2)	1712.9	374.43	911.08	0.884
3	(1:12,3)	1243.76	360.76	880.62	0.900
4	(1:12,4)	1351.01	386.34	940.98	0.880
5	(1:12,5)	1521.17	335.63	800.95	0.910
6	(1:12,6)	1360.53	337.36	805.78	0.908
7	(1:12,7)	1447.59	329.42	799.98	0.908
8	(1:12,8)	1232.21	371.64	906.64	0.891
9	(1:12,9)	1819.88	361.59	880.86	0.895
10	(1:12,10)	1356.72	315.07	764.79	0.918

From Table (5), we note that the best structure for the NAR neural network is (1:12,10) as can be seen in Figure 6, according to the criteria (RMS, MAE, MAPE, R).

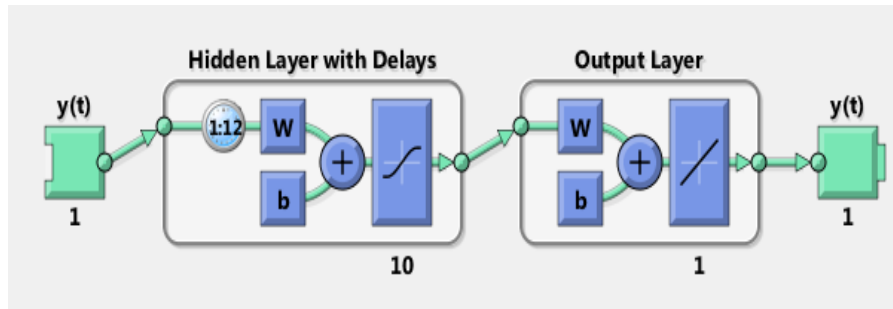


Figure (6): NAR Neural Network optimal architecture

Nonlinear autoregressive neural network training regression

Figure 7: shows the regression plot for the training, validation, and testing sets combined. The lines in each quadrant represent the model's predictions compared to the actual values. The closer the lines are to the diagonal line, the better the model's performance.

The correlation coefficient (R) measures the strength of the relationship between the predicted values and the actual values. A value closer to 1 indicates a stronger relationship.

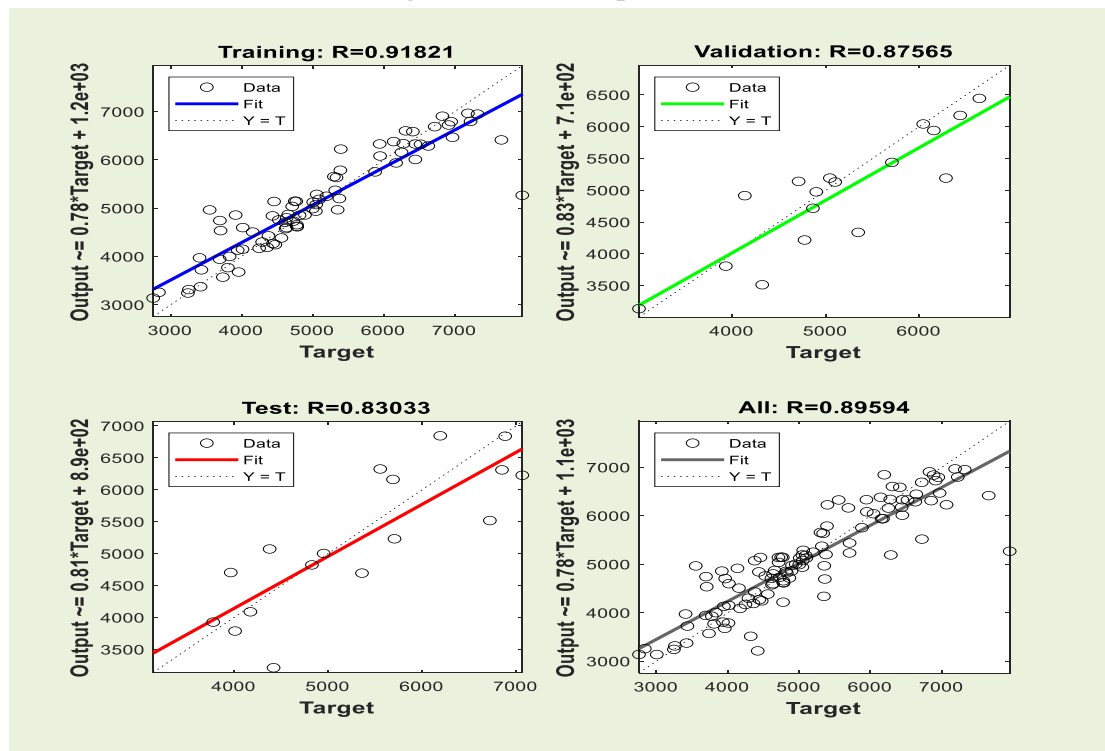


Figure (7): Regression plots showing NAR NN (1:12,10) for electricity demand

Error histogram and error autocorrelation function: Figure 13 shows the error histogram and error autocorrelation function to evaluate the suitability of the network to the training process for the four models.

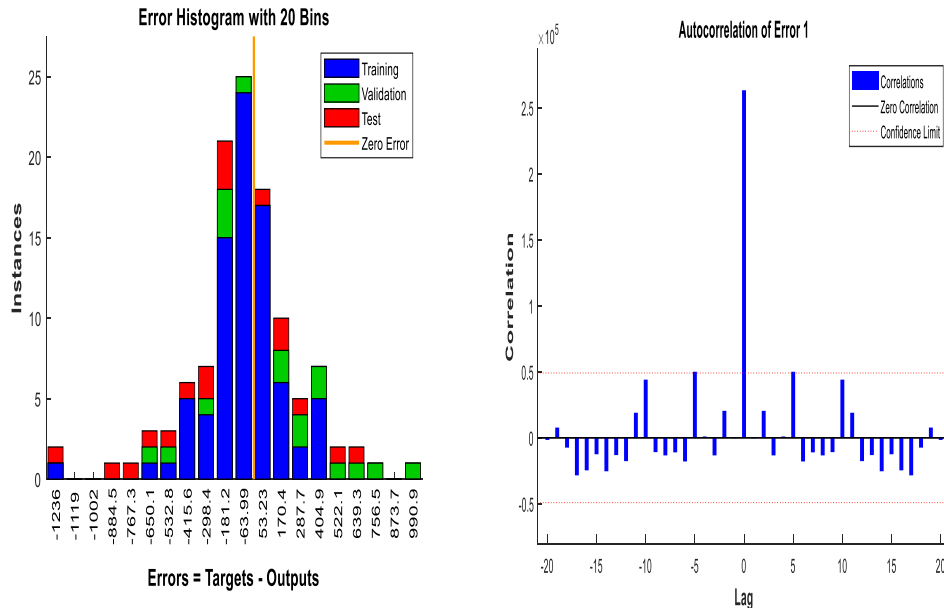


Figure (8): Shows the histogram of errors and the autocorrelation function of the error

Figure 8 reveals that the model's error histogram accurately represents the residuals. As the network trains, the errors between the target values and the expected values decrease and become symmetrical with respect to the zero axes, indicating a normal distribution and appropriate training of the network. Additionally, the error autocorrelation function verifies that all errors remain uncorrelated and fall within the 95% confidence interval around zero. This indicates the network's learning quality and its capacity to forecast the demand for electrical energy on a monthly basis using a daily time series, as evidenced by the correlation coefficient's value being close to one.

Nonlinear autoregressive Training Time-Series Response: Figure 4 presents the time-series response plot for the validated instances using the NAR neural network. It also highlights the time intervals allocated for the testing, validation, and training phases. Based on the test results, the NAR neural network model is identified as (1:12, 10). The plot demonstrates the consistent distribution of the response curve's outputs throughout the training, testing, and validation processes. Additionally, the accompanying

autocorrelation function plot supports this, as all its values fall within the acceptable range.

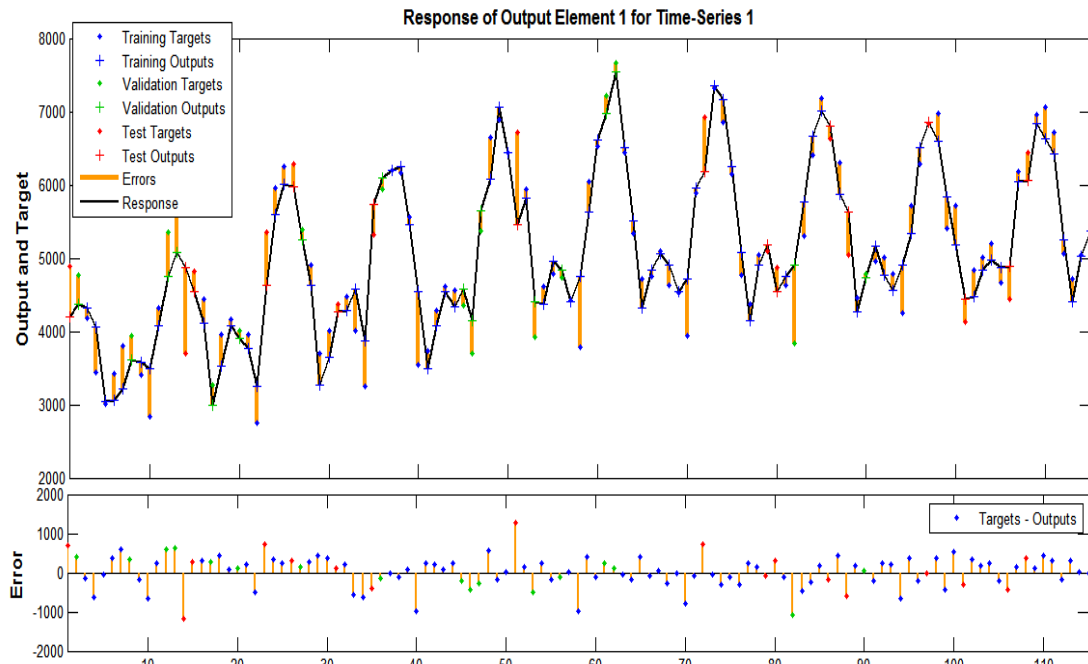


Figure (9): The Time-Series Response for the NARNN (1:12,10) Model

Comparison SARIMA model and NAR neural network: In order to choose the most appropriate model for predicting electrical energy demand, we compared the SARIMA model and NAR neural network based on statistical criteria (MAE, RMSE, and MAPE), and Table 6 shows the comparison results.

Table (6): Comparison between (SARIMA and NARNN) Models for prediction Peak Demand Electricity

Models	RMSE	MAE	MAPE
SARIMA (0,1,1) (0,1,1)12	393.30	279.99	5.72
(NAR NN)(1:12,10)	1356.72	315.07	764.79

Based on the results, the SARIMA (1,0,1) (0,1,1) model outperforms the NAR (1:12,10) model in terms of the three statistical criteria. This indicates that the SARIMA model provides more accurate predictions.

4. Conclusion: Based on the results, the following is a summary of the most important conclusions obtained from this study:

1. Using the Box-Jenkins method for time series analysis, the SARIMA (1,0,1) (0,1,1)12 model was identified as the most suitable model. This model demonstrated strong predictive power after comparing it to other models and evaluating its performance using criteria like RMSE, AIC, and SBIC.

2. Nonlinear autoregressive neural networks were used to predict the increase in electricity demand. By experimenting with different numbers of hidden nodes and time delays, the NAR NN (1:12,10) model was found to be the best fit for the data. This model achieved the lowest values for statistical criteria (RMSE, MAE, and MAPE).
3. We evaluated the performance of the NAR neural network and the traditional Box-Jenkins (SARIMA) method by comparing their results using statistical criteria (RMSE, MAE, MAPE). The SARIMA (1,0,1) model showed the lowest values for these metrics, making it the most accurate model for forecasting peak electricity consumption and highly effective for forecasting.

Reference:

1. Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. UK: Clarendon Press.
2. Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. New York: NY: Springer New York.
3. Ahmed, M. M., & Mahmood, S. H. (2023). Building SARIMA Models to Analysis and Forecast Time Series Data of Road Traffic Accidents in the Kurdistan Region of Iraq. *Zanco Journal of Human Sciences*, 27(1), 382-393.
4. Ali, T. H., Mahmood, S. H., & Wahdi, A. S. (2022). Using Proposed Hybrid method for neural networks and wavelet to estimate time series model. *Tikrit Journal of Administrative and Economic Sciences*, 18(57), 432- 448.
5. Allende, H., Moraga, C., & Salas, R. (2002). Artificial neural networks in time series forecasting: A comparative analysis. *Kybernetika*, 38(6), 685-707.
6. Benrhmach, G., Namir, K., Namir, A., & Bouyaghroumni, J. (2020). Nonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time Series. *Journal of Applied Mathematics*, 1-6.
7. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. New York: John Wiley & Sons.
8. Fausett, L. V. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. New York: Prentice-Hall.
9. Galushkin, A. I. (2007). *Neural Networks Theory*. New York: Springer.
10. 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 (Print ISSN: 1681-6870)*, 1, pp 472-485.
11. Koenecke, A. (2020). *Applying deep neural networks to financial time series forecasting*. Stanford, California, USA: Institute for Computational & Mathematical Engineering.
12. Mahmood, S. H. (2016). Using Artificial Neural Network Model with Time Series Monthly Average Electricity Demand in Kurdistan Region. *Using Artificial Neural Network Model with Time Series Monthly Average Electricity Demand in Kurdistan Region*, 20(6), 421-432.

13. Mahmood, S. H., & Ahmed, M. M. (2023). Application of Elman Neural Network and SARIMA Model to Modeling Road Traffic Accident in the Kurdistan Region of Iraq. *Iraqi Journal of Statistical Sciences*, 20(2), 43-56.
14. Ratur, R., & Sargsyan, H. (2018). A Nonlinear Autoregressive Scheme for Time Series Prediction via Artificial Neural Networks. *Journal of Computer and Communications*, 6(9), 14-23.
15. Sarhan, H. A. (2018). Using some statistical methods to predict the lost electrical energy (applied Study). karbala: University of KaralaIra.
16. Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications*. New York : Springer.
17. Taherdangkoo, R., Tatomir, A., Taherdangkoo, M., Qiu, P., & Sauter, M. (2020). Nonlinear Autoregressive Neural Networks to Predict Hydraulic Fracturing Fluid Leakage into Shallow Groundwater. *water*, 12(3), 1-14.
18. Tsay, R. S. (2005). *Analysis of financial time series*. Hoboken, New Jersey: John Wiley & sons.
19. Yaffee, R. A., & McGee, M. (2000). *An introduction to time series analysis and forecasting: with applications of SAS and SPSS*. USA: Academic Press.
20. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, pp 159-175.
21. Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160(2), 501-514.