# A comparative study to obtain an adequate model in prediction of electricity requirements for a given future period

<sup>1</sup>Maysaa Abd Ulkareem ,<sup>2</sup> Wid Akeel Awadh , <sup>3</sup>Ali Salah Alasady

<sup>1,2,3</sup> Dep. of Computer Information System, College of Computer Science and Information Technology University of Basrah - Iraq

Email : 1 maysaakarem13@gmail.com, 2 umzainali@gmail.com, 3 Alishashim2009@gmail.com

Abstract- Recent years have seen an increasing interest in providing accurate prediction models for electrical energy consumption. In Smart Grids, energy consumption optimization is critical to enhance power grid reliability, and avoid supply-demand mismatches. Additional economic and environmental benefits can be obtained if a proper and reliable description and forecast of energy absorption is available. This research presents fits for neural network model, and comparative resulting forecasts with those obtained from Box- Jenkins Method. We use time series data of monthly electrical energy consumption, into Basra city (Iraq) from 2007-2016, to perform a comparative. The result of the data analysis show that the proper and efficient model for representing time series data is the multiplicative seasonal model of order: SARIMA (3, 1, 3) × (0, 1, 1). According to this model the Research forecast the monthly electricity consumption for January 2017 to December 2018. As for application, Box-Jenkins Method has given more appropriate forecasts than those given by Back propagation artificial neural network. We used Minitab program in the statistical aspect and Alyuda program in the neural network aspect.

#### Keywords— Electricity consumption, Prediction, Time Series Analysis, Error analysis, Artificial Neural Networks.

#### I. INTRODUCTION

In large cities, the extremely important thing is the efficient energy management, which has many benefits that can be present in the forecasting of electricity consumption and an accurate description. Firstly, which relates to the management of the energy system of the city. The energy provider needs each large consumer prior to declare needed energy for a given period of time. Secondly, that is related to the economic aspect, which the cost of energy consumed above a stated amount is much higher than the primary energy. Incorrect queries lead to economic losses for the transport company. The third is the ability to use the model for the development of various business strategies, such as to provide energy consumption when changing routes and number of vehicles [1] [2]. To assist these jobs, Information Technology (IT) providing units for adequate storage to store an up to date information and used this information to get the highest benefits from various methods. Away, to apply statistical knowledge to analyze data in the past which are related to the current event, is Planning by forecasting trends in the future [3]. Time series Analysis (TSA) is "the order of historical data, which resembles the group or observation of the data that have been collected over time according to the continuous period of time".

Data collected may be already on a daily, weekly, monthly, quarterly, or yearly format, depending on which one is appropriate to use. Time series data consists of four components: trend, seasonal effect, cyclical, and irregular effect [4].

The aim of this study is to obtain an adequate model describing the process to be used to predict the electricity required for a given future period. In this study, time series analysis (TSA) will be used as a practical method, and we will focus on the Box and Jenkins method [5] as the most popular method in TSA. The result, the model, of this study is quite accurate compared to other methods and can be applied to all types of data movement.

One of the problems that is causing problems for the development and use of the Box-Jenkins models is the requirement to model the model in terms of the probability distribution of the data as the method of the Box-Genghis is sometimes not suitable for specific cases. It fails to describe dynamically changing behavior with time [6].

Other nonlinear methods, such as Neural Networks, Fuzzy Logic, Distributed Logic, and Evolutionary Algorithm, have provided solutions for such situations [7]. Artificial Neural Networks (ANNs) have provided ARIMA with some more probabilistic features than time series models in dealing with Non Normal Non Linear and Nonlinear Data Problems. The relationship between neural network models and Box Jenkins models can be a non-competitive integration relationship to wellmatched models.

#### II. BASIC CONCEPTS

#### A. Time Series Analysis

The time series is a set of observations arranged according to their occurrence in time: The values of a specific phenomenon are given and these observations are written as follows: y1, y2, yn where y1 means the value of the observation that occurred in time t1 and the value of y2 the observation that in time t2.  $y = a t_2 + b t + c$ . When: a, b, c value constant.

According to the time series, the data set that is classified as a classification of the times of occurrence of the day, the week, the month, the year, or the period of time, which takes views. Examples of time series are: Deaf taken in the gulf produced annually by Sudan. Sales of a store for ten years in a row,

<sup>978-1-5386-7858-9/18/\$31.00 ©2018</sup> IEEE

successive years The time series are used to predict where prediction is defined as the process of predicting what will happen in the future and depending on whether positive or negative [8].

#### B. Artificial Neural Networks

One of the artificial intelligence topics and methods are Artificial Neural Networks (ANNs), has been used for prediction recently. Artificial Neural Networks have multiple variance and non-parametric structure that is lead to preferred in predisposition for application in non-linear time series. Especially, ANNs are suitable for solving the issues which cannot easily be in mathematically modelled, also they have the ability to learn new conditions by changing their weights and they can learn with different learning algorithms, in the learning aspect, Artificial Neural Networks used in solving similar problems as a result of such learning. A trained ANN can act as a specialist, analyses the data that is given and make projections [9].

ANN neurones are disributed in order layers, where the first is called the input layer and the last is called the output layer.the layer between these two layers are known as hidden layers. Each neurone in any particular layer is connected with all neurones in its immediate next layer. the connection between ith and jth neurone is characterized by the weight coefficient  $\omega_{ij}$  and the threshold coefficient  $\vartheta_i$ . The output of the ith neurone  $x_i$  is calculated by equation (1),(2) [10].

$$x_i = f\left(\xi_i\right),\tag{1}$$

$$\xi_i = \vartheta_i + \sum_{j \in \Gamma_i^{-1}} \omega_{ij} \, x_j \,, \qquad (2)$$

Where  $\Gamma$  is a mapping function which assigns each neurone *i* a subset  $\Gamma(i) \equiv V$  consisting of all ancestors of the given neurone. All preddecessors of the given neurone *i* is represented by the subset  $\Gamma_i^{-1} \equiv V$ .  $\xi_i$  is the potential of the *i*th neurone and  $f(\xi_i)$  is the transfer function. For the transfer function *is* calculated by equation (3) [10] :

$$f(\xi_i) = \frac{1}{1 + \exp(-\xi_i)} \tag{3}$$

ANN consists of the interrelation of a series of neural cells in back-propagation and forward-driven connection patterns. Now, numbers of Artificial Neural Networks have grown for several purposes and for usage in different fields such as MLP, LVQ, Perceptron, SOM, ART, Hopfield, and Recurrent. Among these, a Back Propagation Algorithm is an education algorithm which is most widely used for Multiple Layer Perceptron networks [11]. Based on these, a Back Propagation Algorithm will deployed in this study.



Fig. 1. Multiple Layer Back Propagation Artificial Neural Networks Algorithm

## III. STEPS OF ANALYSIS

### A. Data Collection

Data collected in this research represent monthly time series data (120 scenes) of electric energy consumption of Basra City, estimated in megawatts/hour for all types of consumption (domestic, commercial, governmental, street lighting and so on). Table (I) shows the consumption data from January 2007 to December 2016. Table (II) describes the data.

TABLE I . THE MONTHLY CONSUMPTION OF ELECTRIC POWER IN THE BASRA CITY (2007-2016)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Jan.	261486	288488	281691	390764	451509	429593	807522	899989	838307	865526
Feb.	259998	285679	255394	344617	429083	484415	738966	726576	707156	780755
Mar.	245465	377126	251192	365066	437051	492188	721816	696173	690235	714554
Apr.	252826	370547	251535	424599	465765	474386	685595	709289	801766	872721
May	250000	370679	252166	493379	466934	473844	637647	927616	1011100	968996
June	260000	391280	294609	439938	486264	470051	726777	971336	1056622	977951
July	264379	420105	294480	484561	492394	467359	743681	1014811	1064439	1061552
Aug.	260934	424526	307352	500069	513132	565558	898910	950164	1097063	1109961
Sep.	252725	267127	272227	478386	523156	536642	719002	893891	900266	873789
Oct.	249128	268128	247681	391557	463160	544996	683716	756073	756589	823856
Nov.	248112	263011	260583	329518	488769	503104	632589	736583	757442	791445
Dec.	260000	280397	274029	347783	402923	488915	712643	774001	637434	714768

TABLE II . DESCRIBES OF DATA

No. of Observations	120	=12 months * 10 years
No. of Years	10	Between (2007-2016)
Mean	554,135	
Minimum	245,465	Recorded in 2007 (March)
Maximum	1,109,961	Recorded in 2016 (August)
Standard Deviation	250,846	
Monthly Growth	70,35%	
Rate		

The number of observations is sufficient to assume that the time series follows a natural distribution, and the model can be diagnosed in the best way.

#### B. Time Series Analysis

#### 1) Plotting time series

Before analysing the time series, the time series data in Table (I) were plotted, as shown in Fig. (2), using a statistical program (Minitab). Fig. (2) shows an increasing general trend with time as well as the presence of oscillations. These oscillations are repeated regularly at the same frequency each year, with the difference in frequency increasing yearly. These changes indicate the presence of a general-direction vehicle and its seasonal vehicle.



Fig. 2. Time series plot of monthly electric power consumption (2007-2016)

#### 2) Testing stability of time series

Data were processed by calculating the natural logarithm (Ln) once and calculating the square root again, as shown in Figs. (3) and (4), to obtain the stability of variance. From these figures, we note that stability was achieved to a certain extent by calculating the natural logarithm of data. Therefore, the time series was relied on when applying the model.





Fig. 4. Time series plot after calculating square root.

In Figs. (3) and (4), the general trend increases with time, indicating the instability of the data series. This behavior is confirmed by the values of the coefficients of the autocorrelation and partial autocorrelation, as shown in Fig. (5).



Lag	ACF	т	LBQ	Lag	ACF	т	LBQ	Lag	ACF	т	LBQ
1	0.95	10.43	111.56	11	0.69	2.04	883.94	21	0.50	1.21	1363.15
2	0.90	5.92	213.50	12	0.68	1.94	947.28	22	0.49	1.17	1400.32
3	0.86	4.49	307.25	13	0.65	1.80	1005.60	23	0.48	1.13	1435.77
4	0.83	3.75	395.11	14	0.61	1.66	1058.64	24	0.47	1.09	1469.70
5	0.80	3.25	477.66	15	0.59	1.57	1108.48	25	0.44	1.01	1499.89
6	0.76	2.87	553.68	16	0.58	1.50	1156.15	26	0.40	0.92	1525.64
7	0.74	2.59	625.01	17	0.56	1.43	1201.65	27	0.37	0.84	1547.52
8	0.71	2.36	691.50	18	0.54	1.36	1244.47	28	0.33	0.76	1565.76
9	0.69	2.21	755.71	19	0.53	1.30	1285.48	29	0.30	0.68	1581.01
10	0.69	2.11	819.57	20	0.51	1.25	1325.04	30	0.27	0.61	1593.34

Fig. 5. Coefficients of the autocorrelation and partial autocorrelation of the time series after calculating their natural logarithm

Fig. (5) shows that the autocorrelation coefficient values until the gap (30) are significantly different from zero, and the autocorrelation coefficients are beyond the limits of confidence. Entering all the parameters of the autocorrelation coefficients within the limits of confidence is necessary for the time series to be stable. Except for the first or second offsets, falling beyond the limits of confidence is possible. The confidence limits of the data are ( $-0.61 \le r_k \ge 0.61$ ). For the moral test, the total coefficients of the autocorrelation function using (Box & Jenkins) are evaluated as follows:

Q.state = 
$$LBQ = 1593.34 > X^2$$
 (30)

Therefore, the null hypothesis is rejected, meaning all coefficients of the autocorrelation are equal and equal to zero, as follows:

 $H_0 = \rho 1 = \rho 2 = \rho 3 = \Delta \Delta = \rho k = 0$ Thus, we accept the alternative hypothesis, which means that the time series is unstable.

#### 3) Removing general trend

The differences from the first grade were calculated, and we obtained the modified time series where  $\mathbf{\nabla} Z_t = z_t - z_{t-1}$  (Algorithm Series) to remove the general trend, as shown in Fig. (6). Fig. (6) shows the absence of the general trend in the time series with the existence of the seasonal vehicle. The instability of the time series is confirmed using statistics by (Box & Jenkins) where Q. state = LBQ =  $51.96 > X^2_{(30, 0.05)}$ .



Fig. 6. Logarithm time series plot after calculating its first difference

#### 4) Removing seasonality trend

The observations of the values of the autocorrelations of the modified time series after calculating the first difference, as shown in Fig. (7), show that these values are significant in periods 12–24. The series is seasonal, and it repeats every 12 months. We calculated the differences (seasonal) of grade 12 and obtained the modified series (C4) to remove the seasonality, as follows:

$$C4 = \nabla \nabla_{12} Z_t = Z_{t-1} - Z_{t-12}$$



Fig. 7. Coefficients of the autocorrelation of the time series after calculating its first difference.

Fig. (8) shows the modified time series after calculating the seasonal differences  $(\nabla \nabla_{12} Z_t)$ . The values of the autocorrelation and partial autocorrelation are shown in Figures (9), respectively.



Fig. 8. Time series plot after calculating its first and seasonal differences.



Fig. 9. Coefficients of the autocorrelation of the time series after calculating its first and seasonal differences.

Fig. (9) shows that the coefficients of the autocorrelation are within the limits of the confidence, and morality is observed only in period 12, indicating the stability of the time series.

#### 5. Diagnosis of the Model

Recognizing the model by specifying the grade of models, AR and MA, depends on the form of autocorrelation function (Conelogramme). The values of the coefficients of the autocorrelation and the partial autocorrelation of the time series are matched after calculating the first and seasonal differences, as shown in Fig.s (8) and (9), with the theoretical behaviour described in Table (1). The degree of autocorrelation and partial autocorrelation decrease gradually with increasing periods of displacement (K). We conclude that the model is a seasonal model multiplier of the class:

#### SARIMA (3, 1, 3) x (0, 1, 1)<sub>12</sub>.

#### 6. Model Estimation

After previewing possible models, we found that the appropriate model is: SARIMA  $(3, 1, 3)_X(0, 1, 1)_{12}$ , based on the standard AKIAKE, moral parameters and the homogeneity test of variance. Applying the method of non-linear smaller squares on the time series data under study and using the statistical program Minitab obtained the following results:

TABLE III . Final Estimates of Parameters

Туре	Coef	SE Coef	<b>T-Value</b>	<b>P-Value</b>
AR 1	1.181	0.109	10.81	0.000
AR 2	-1.2463	0.0882	-14.13	0.000
AR 3	0.540	0.114	4.72	0.000
MA 1	1.3868	0.0565	24.56	0.000
MA 2	-1.3964	0.0563	-24.80	0.000
MA 3	0.8786	0.0765	11.48	0.000
SMA 12	0.6993	0.0978	7.15	0.000

Differencing: 1 regular, 1 seasonal of order 12.

Number of observations: Original series 120, after differencing: 107

Residual Sums of Squares:

Standard error= 0.014 AIC= - 1.857372 MAIC= - 0.017358 BIC= - 1.913447

7. Checking the appropriate model

Ascertaining the significance, suitability and predictability of the model required the following steps:

 a) Perform autocorrelation error test: The autocorrelation and partial autocorrelation coefficients of the estimated model errors were extracted and plotted, as shown in Fig. (10). All values of the autocorrelation coefficients of errors fall within the limits of trust. Thus, the model used is good and appropriate.



Fig. 10. Autocorrelation and partial autocorrelation coefficients of the estimated model errors

- b) Test zero hypothesis: The mean average=zero, z=0.0004 and P-value=0.528. Therefore, zero hypothesis cannot be rejected.
- c) Draw histogram of errors: Fig. (11) is symmetrical and has an almost normal distribution form.



Fig. 11. Histogram of time series errors

All these procedures lead to the statistical acceptance of the model. Therefore, the model can be used for prediction.

#### 8- Prediction

Using the forecasting model selected in paragraphs (2) (4), the monthly electric power consumption of Basra City for 2017 and 2018 is predicted in Table (III). The time series for this prediction was plotted as in Figure (12), and it shows that the series for the predicted period follows the same behavior as the original sequence.



Fig. 12. The time series for this prediction

TABLE IV FORCATING MONTHLY ELECTRIC POWER CONSUMPTION OF BASRAH CITY FOR (2017–2018)

DASKAH CITY FOR (2017–2018)						
	2017	2018				
Jan.	1026804	1232918				
Feb.	994240	1163358				
Mar.	974279	1146559				
April	1096702	1223754				
May	1207539	1331394				
June	1238466	1365344				
July	1279979	1408117				
Aug.	1324948	1449092				
Sep.	1184205	1304580				
Oct.	1110024	1231290				
Nov.	1088332	1213671				
Dec.	1001990	1179583				

C. Use of Artificial Neural Networks Suggests the time series of photovoltaic electricity

The forecast plan monthly time series data of electric energy consumption of Basra City in table (I) based on neural networks consists of three requirements:Determine the predictive number of neural network

1) Determine the number of neurons in the hidden layer: Using the following formula Baum-Haussler known as a rule in determining the number of neurons in the hidden layer Input, that ia calculated by equation (4).

(4)

$$N_{Hidden} \leq \frac{N_{Train} E_{Tolerance}}{N_{Pts} + N_{Output}}$$

Where:

N train = Training Time = 10000 E Tolerance = Standard Error = 0.01 N  $_{pts}$  = Inputs Number = 9 N  $_{Output}$  = Outputs Number = 1

Data used for 120 views:

Training test =68.3% = 82 Validation test = 15.85% = 19 Training test = 15.85% = 19

TABLE (4) TABLE V .ILLUSTRATES THE ARCHITECTURE OF THE NETWORK USING NEURAL NETWORKS.

NET WORKS.						
Architect	Weights	Fitness	Test	Train	AIC	AIC <sub>C</sub>
ure			error	error		
[9-2-1]	23	0.011	490.55	319.25	0.00104	0.96442
[9-3-1]	34	0.011	628.36	321.54	0.00121	0.91237
[9-4-1]	45	0.011	723.68	325.97	0.00155	0.92540
[9-5-1]	56	0.012	859.37	332.93	0.00193	0.95501
[9-6-1]	67	0.012	822.91	427.74	0.00217	0.93122
[9-7-1]	78	0.012	801.42	337.68	0.00237	0.95411
[9-8-1]	89	0.013	865.07	245.78	0.00242	0.96587
[9-9-1]	100	0.013	761.62	145.04	0.00272	0.99131
[9-10-1]	111	0.014	730.53	776.95	0.00288	0.99775

#### 2) Build a neural network:

To obtain neural network prediction values for the months of 2017 and 2018, [9-2-1] the best network was trained and tested, which is the best network algorithm among the other networks using (Online) Incremental Back Propagation algorithm. This algorithm also installs the training parameters that represent the Learning Rate = 0.9 Learning Momentum Constant = 0.9 and the momentum constant from Iterations = 10000 and the number of repetitions as well as the number of retraining. Number of Retrains = 2, after training the network based on parameters, the following results were obtained:

 TABLE (VI)

 THE RESULTS AFTER TRAINING THE NETWORK.

Parameters						
	Training	Validation				
Absolute error:	522.44804	63048.326649				
Network error:	0.000003	0				
Error improvement:	2.42E-09					
Iteration:	10001					
Training speed, iter/sec:	33336.665342					
Architecture:	[9-2-1]					
Training algorithm:	Online Back Propagation					
Training stop reason:	All iterations done					

D. Comparison Between the Results of the Statistical Method and the Method of Artificial Neural Networks

The comparison between the results of the statistical method ( Box- Jenkins Method) and the method of artificial neural networks (Online Back-Propagation Algorithm), has been based on the statistical standards, the incremental RMSE, MSE, shown in the table VI:

TABLE (VII)
STATISTICALLY COMPARISON BETWEEN JENKINS' STYLE AND THE ANN

ical ards	Box-Jenkins	artificial neural networks [9-2-1]
Statist Standa	SARIMA(3 , 1 ,3)×(0 , 1 , 1)	Online Back propagation Algorithm
RMSE	0.010	0.9703
MSE	0.014	1.5321

As shown from the above table, the Box-Jenkins model is superior to the artificial neural network model (Online Back propagation Algorithm).

#### IV. CONCLUSION

Monthly electric energy consumption time series usually have two characteristics. First, the main trend of the said time series presents an approximately exponential increasing feature. Second, the period of the wave trends can only be 12, 6, 4, 3, or 2months. The purpose of this research was to find a suitable model to forecast the electric consumption in a Basra city, and to find the proper and efficient model for representing time series data. The study used the SARIMA model and forecasting electric power consumption in Basra city from January-2017 to December-2018. As for application, Box-Jenkins Method has given more appropriate forecasts than those given by Back propagation artificial neural network. Then, chose the suitable forecasting method and identified the most suitable forecasting period by considering the smallest values of MSE and RMSE, respectively.

#### REFERENCES

- Deng S. J., Oren S. and Gross G., Design and Valuation of Demand Response Mechanisms and Instruments for Integrating Renewable Generation Resources in a Smart Grid Environment, PSERC Report 12-24, September 2012.
- [2] Mancarella P. and Chicco G., Real-time demand response from energy shifting in Distributed Multi-Generation, IEEE Transactions on Smart Grid, vol. 4, no. 4, Dec.2013, pp. 1928-1938
- [3] Bruce L. Bowerman, Richard T. O' Connell, & Anne B. Koehler, "Forecasting, time series, and regression: an applied approach," 4th ed. The United States of America: Thomson Brooks, 2005.
- [4] Guarnaccia C., Quartieri J., Mastorakis N. E.and Tepedino C., Acoustic Noise Levels Predictive Model Based on Time SeriesAnalysis, in "Latest Trends in Circuits, Systems, Signal Processing and Automatic Control", Proc. of the 2nd Int. Conf. on Acoustics, Speech and Audio Processing (ASAP '14), Salerno, Italy, June 3-5, 2014, pp.140-147.
- [5] G.E.P. and G. Jenkins, BOX, "Times Series Analysis Forecasting and Control," Holden-Day, San Francisco, CA, 1976.
- [6] Guarnaccia C., Quartieri J., Rodrigues E. R.and Tepedino C., Time Series ModelApplication to ultiple Seasonality AcousticalNoise Levels Data Set, in "Latest Trends in Circuits, Systems, Signal Processing and Automatic Control", proceedings of the 2nd Int. Conf. on Acoustics, Speech and Audio Processing (ASAP '14), Salerno, Italy, June 3-5, 2014 , pp. 171-180
- [7] Tahmasebi, Pejman, and A. Hezarkhani. "A hybrid neural networksfuzzy logic-genetic algorithm for grade estimation." Computers & geosciences 42 (2012): 18-27.
- [8] Bisgaard S.and Kulahci M., Time Series Analysis and Forecasting by Example, Johin wiley & Sons. Inc, New Jersey, 2011.
- [9] Racelli, V., Bevilacqua, V., & Azzolini, M. (2011). An Artifical Neural Network Model to Forecasting Exchange Rates. Journal of Intelligent Learning Systems and Applications, 3, 57–69.

- [10] T., G. Sundar Mitra, R. Bhattacharyya, and S. Sarkar Mondal. "Artificial neural network based model for forecasting of inflation in India." Fuzzy Information and Engineering 8.1 (2016): 87-100.
- [11] H. Kocakl & Turgut Un1, Forecasting the Gold Returns with Artifice Neural Network and Time Series,1 Faculty of Economics, Department of Econometrics Goztepe, Marmara University, Istanbul, Turkey Correspondence: Habip Kocak, Faculty of Economics, Department of Econometrics Goztepe, Marmara University, Istanbul, Turkey.