

Time Series Analysis and Prediction of Electricity Consumption of Health Care Institution Using ARIMA Model

Harveen Kaur¹(✉) and Sachin Ahuja²

¹ School of Computer Sciences, Chitkara University, Rajpura, India
harveen.kaur@chitkara.edu.in

² Chitkara University Research and Innovation Network, Rajpura, India
sachin.ahuja@chitkara.edu.in

Abstract. The purpose of this research is to find a best fitting model to predict the electricity consumption in a health care institution and to find the most suitable forecasting period in terms of monthly, bimonthly, or quarterly time series. The time series data used in this study has been collected from a health care institution Apollo Hospital, Ludhiana for the time period of April 2005 to February 2016. The analysis of the time series data and prediction of electricity consumption have been performed using ARIMA (Autoregressive Integrated Moving Average) model. The most suitable candidate model for the three time series is selected by considering the lowest value of two relative quality measures i.e. AIC (Akaike Information Criterion) and SBC (Schwarz Bayesian Criterion). The appropriate forecasting period is selected by considering the lowest value of RMSE (Root Mean Square Error) and MPE (Mean Percentage Error). After building the final model a two-year prediction of electricity consumption of the health care institution is performed.

Keywords: Time series · Akaike Information Criterion · Schwarz Bayesian Criterion · Autoregressive integrated moving average model · SAS University Edition

1 Introduction

It is critical for electricity traders [1] to predict electricity consumption for balancing the electricity purchase and sales portfolios. Electricity consumption is a significant commodity to promote society's economic development and raise people's standard of living [2]. Difficulties associated with electricity storage necessitate accurate prediction of its consumption. This further emphasizes the use of apt and correct approaches for electricity consumption prediction. Amongst the methods for performing the prediction, ARIMA is a statistical method for analyzing and building the forecasting model which best represents a time series by modeling the correlations in the Data.

1.1 Autocorrelation Functions

- Autocorrelation: Autocorrelation, also called a lagged correlation or serial correlation is the way in which the observations in time series are related to each other and is calculated as the simple correlation between current observation y_t and the observation from p periods before the current one i.e. y_{t-p} for a given interval.
- Partial Autocorrelation: Partial autocorrelation is the correlation calculated after removing the linear relationship between the two consecutive observations. Basically, it is the correlation between y_t and y_{t-p} when the effect of y at other time lags 1, 2, 3, ..., $p-1$ is removed.
- Both autocorrelations and partial autocorrelations are computed for lags in sequence for the given time series. The autocorrelation at lag 1 is between Y_{t-1} and Y_t , the autocorrelation and partial autocorrelation at lag 2 is between Y_{t-2} and Y_t and so on for n lags.
- Autocorrelation function (ACF) and Partial autocorrelation function (PACF): The identification of the basic model is done by observing the patterns of ACFs and PACFs (Autocorrelations versus lags). ACFs and PACFs act as a tool for finding the values of the orders of autoregressive and moving average components i.e. p and q of ARIMA(p, d, q) respectively.

1.2 Autoregressive Integrated Moving Average Model

The notation for ARIMA model is ARIMA(p, d, q) where p, q, d denotes the order of autoregression, the order of moving average and the degree of differencing respectively. The final prediction depends on these parameters. Hence, the prediction equation for ARIMA (p, d, q) process is denoted by:

$$\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + \varepsilon_t + \lambda_1 \varepsilon_{t-1} \quad (1)$$

Here β_0, β_1 and λ_1 are the parameters of the model and ε_t is the white noise error term (the unidentifiable form of data which doesn't exhibit any patterns).

2 Literature Review

Currently, ARIMA model is avidly used for building a forecasting model for the time series of the electricity consumption. Katara et al. [3] used this method to forecast electricity demand in Tamale, Ghana using data from the Northern Electricity Department Tamale during 1990 to 2013. The model provides a seven-year forecast of the electricity demand in the city. Kandanand [4] implemented three forecasting techniques— autoregressive integrated moving average (ARIMA), multiple linear regression (MLR) and artificial neural network (ANN) to built forecasting model of the electricity demand in Thailand. Yasmeeen and Sharif [5] studied, (the least out of sample forecast performance) i.e. the minimum forecast standard deviation value and Mean Absolute Percentage Error (MAPE) value of the four competing time series

models. On the basis of these accuracy measures, the most suitable model to predict electricity consumption in Pakistan was built. Erdogdu [6] used co-integration analysis and implemented ARIMA modeling to estimate electricity demand and prediction respectively. Bianco et al. [7] evolved a long-term forecasting time series model by developing different regression models using historical electricity consumption, gross domestic product per capita (GDP per capita), gross domestic product (GDP) and population for the time period 1970 to 2007.

3 Methodology

The objective of this research is to develop a set of programs using SAS University Edition [8, 9] to find the best fitted model out of monthly, bimonthly and quarterly time series data [10–12] to predict the Electricity Consumption using ARIMA [13–16] model by considering the lowest value of Root Mean Square Error and Mean Percentage error for all three models. The selected forecasting model should have the lowest prediction error.

This work is purely based on ARIMA Model. The time series variable being used is the electricity consumption of the health care institution.

3.1 Data Preprocessing

Monthly data over a long period of time from April 2005 to February 2016 has been used for this research. SAS University Edition is used for building the model. Initially, data was not in a proper format and consisted of some missing values and so wasn't ready for building the forecasting model.

The data is monthly, but we have to find the suitable time period to predict the future values among three time periods i.e. monthly, bimonthly and quarterly time series.

The following pre-processing steps has been performed:

3.1.1 Converting the Data in Proper Format

By default, SAS displays the date values in a numeric format which is difficult to recognize.

Table 1 represents the sample of the raw data displaying the date values in numeric format, when it is imported to SAS. Hence, a proper date format is assigned to the variable. Table 2 represents a sample of the final data ready to be analyzed on the SAS.

3.1.2 Filling the Missing Information

In a particular observation if no value exists for a particular variable, then a missing value is said to occur. Missing data are very common and can have a huge effect on the data analysis. If not handled properly, these can lead to unstable or inaccurate results. There are various methods to fill the missing data with mean, median or previous value. Table 3 represents the sample of the current data set with some missing data. In this paper, previous values have been used to fill the current value by assuming that current value will be similar to the previous value.

Table 1. Sample of raw data displaying the date values in numeric format

Date	Electricity consumption
42465	137420
42495	240170
42526	287080
42556	295480
42587	304430
42618	285850

Table 2. Sample of final data ready to be analyzed

Date	Electricity consumption
Apr-05	137420
May-05	240170
Jun-05	287080
Jul-05	295480
Aug-05	304430
Sep-05	285850

Table 3. Sample of data with missing values

Date	Electricity consumption
Jan-06	170600
Feb-06	
Mar-06	204860
Apr-06	283940
May-06	
Jun-06	366000

3.2 Seasonal Adjustment of Data

For the seasonal adjustment of the data, a time series consists of three components: trend-cycle, combined seasonal effects, and irregular component. Here, Fig. 1 shows the seasonally adjusted data, while Figs. 2 and 3 depict the trend-cycle and irregular data in the time series respectively. The main purpose of the seasonal adjustment of the data is to identify the seasonal effects and remove them from the time series data.

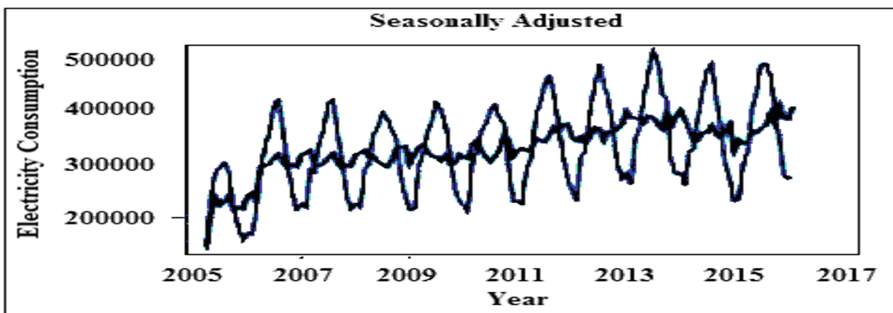


Fig. 1. Seasonally adjusted data

The time series is then called seasonally adjusted series and constitutes the trend-cycle component and irregular component.

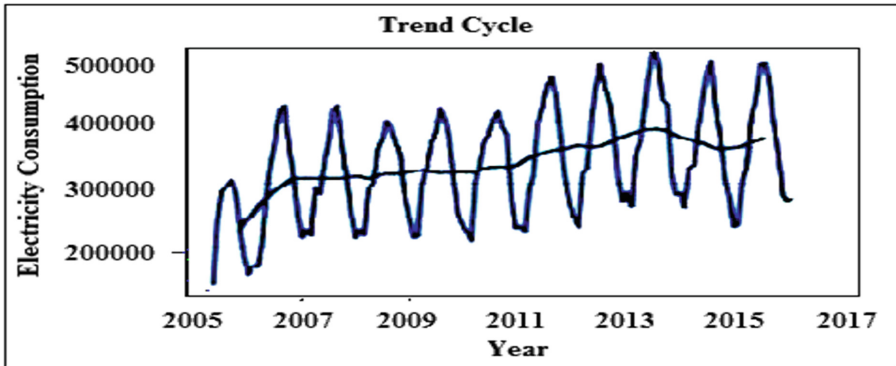


Fig. 2. Trend cycle of data

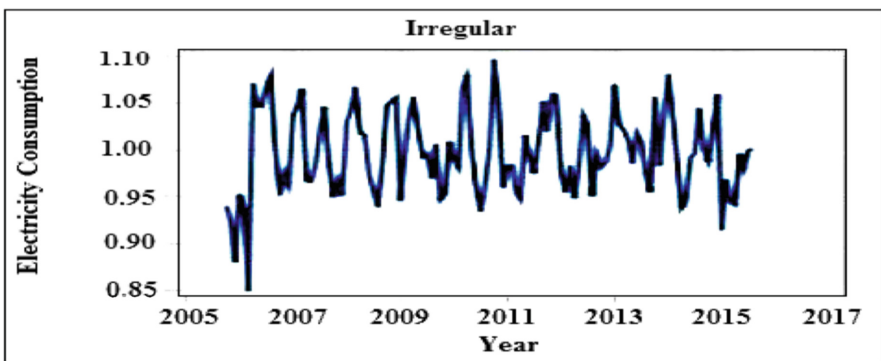


Fig. 3. Irregular form of data

3.3 Steps for Building the Model

Following are the steps for building the forecasting model using ARIMA:

3.3.1 Stationarity Checking

The Electricity Consumption of the health care institution is modeled variable. The time series is monthly time series from April 2005 to February 2016. The Fig. 4 represents the property of stationarity of the monthly data i.e. stationary or nonstationary times series. After the visual inspection of the graphs displayed in the figure, we can see that the time series is nonstationary and needs to be transformed into stationary time series; which is the first condition for applying ARIMA model. Here dB is the stationary time series. This is achieved by taking the first difference of the time series variable in the data. The blue line represents the nonstationary time series and redline represents the stationary time series. Data is prepared into two more formats i.e. bimonthly and quarterly time series.

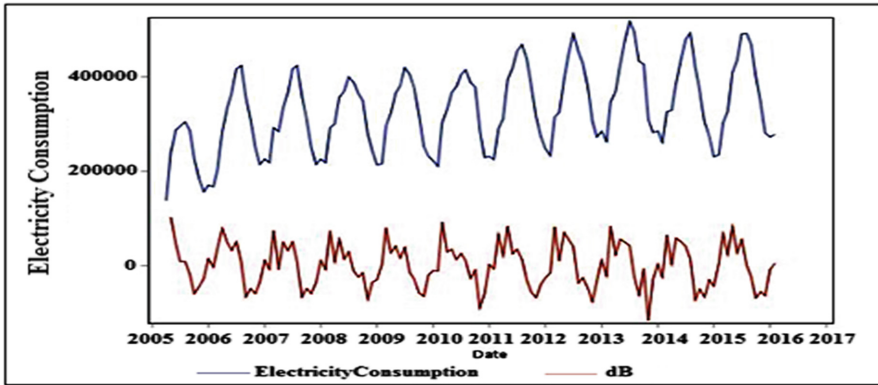


Fig. 4. Stationary vs nonstationary monthly time series.

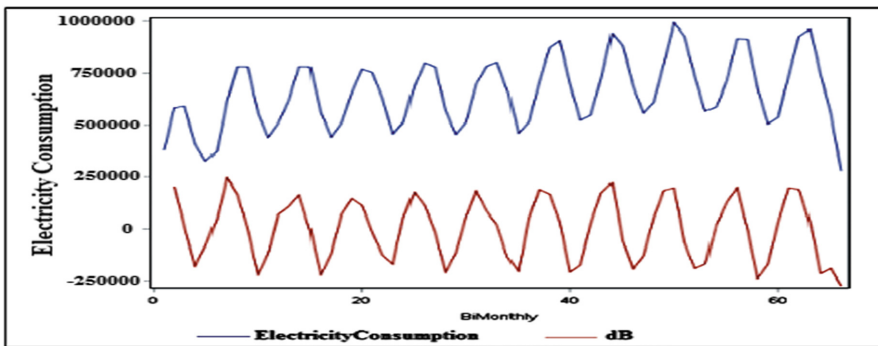


Fig. 5. Stationary vs nonstationary bimonthly time series

The Fig. 5 given below represents the property of stationarity of the bimonthly data while Fig. 6 represents the property of stationarity of the quarterly data.

3.3.2 Model Identification and Model Estimation

By examining the Autocorrelation Function (ACF) plot and Partial Autocorrelation Function (PACF) plot of the stationary time series, the basic ARIMA(p, d, q) model can be identified, but final models can be estimated by conditional least squares estimation method. In this estimation method if there is unit root in any of the AR (autoregressive) or MA (Moving Average) terms i.e. sum of parameter estimates is 1 or close to 1 then the current model is stabilized by adding or removing the number of AR or MA terms so that the unit roots can be removed. By considering the lowest value of AIC and SBC of the candidate models the most suitable model for the three time series can be decided. PACF and ACF plot of stationary monthly time series have been shown in Fig. 7. By examining the PACF plot the value of “p” - the order of autoregressive component - can be estimated. As PACF function plot drops to zero after first two lags, the value of “p” is 2. By examining the ACF plot we can estimate

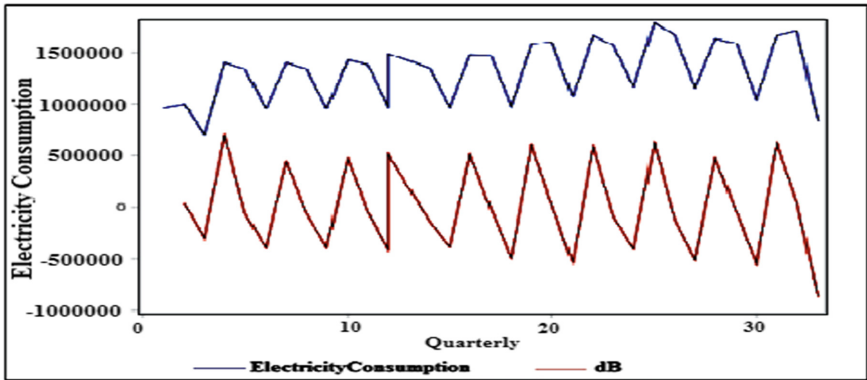


Fig. 6. Stationary vs nonstationary quarterly time series

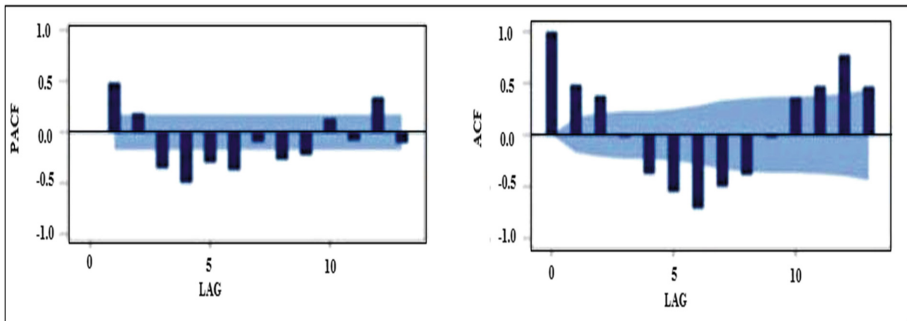


Fig. 7. PACF and ACF plot for differenced monthly time series.

the value of “ q ” - the order of moving average component. ACF function plot drops to zero after first three lags and hence, the value of “ q ” is 3. Since, the time series is differenced once, hence the value of “ d ” is 1. Basic model for the monthly time series of electricity consumption is ARIMA (2, 1, 3).

Similarly, referring to Fig. 8 the values of “ p ” and “ q ” can be estimated for differenced bimonthly time series data as 1 and 2 respectively. As illustrated in Fig. 9 “ p ” and “ q ” values for differenced quarterly time series data can be ascertained as 2 and 1 respectively.

The basic models are estimated by examining the ACF and PACF plots for all the three models. However, for bringing stability in the models, the existence of the unit roots is checked by conditional least squares estimation method.

If any unit root is found, then the current model is stabilized by removing or adding the number of AR or MA terms. The estimates for the parameters of all three models are calculated using the conditional least squares estimation method.

Tables 4, 5 and 6 display the parameter estimates for monthly, bimonthly and quarterly series respectively. It is worthwhile to note that in SAS, the estimate for moving average is always taken with opposite sign as displayed in table. Hence, after

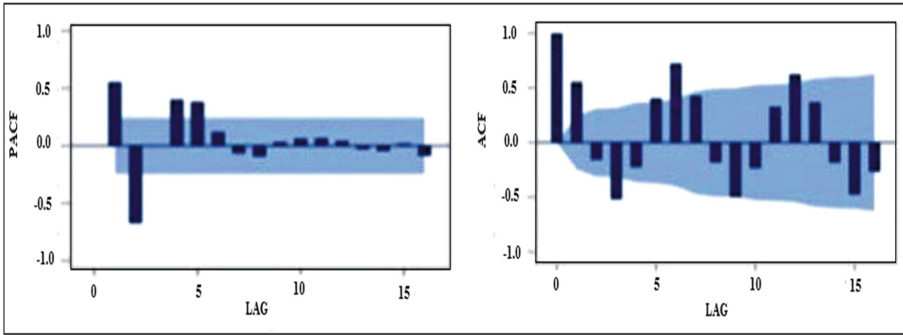


Fig. 8. PACF and ACF plot for differenced bimonthly time series

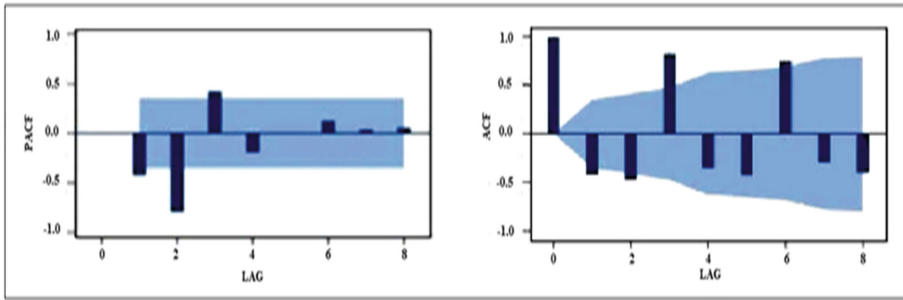


Fig. 9. PACF and ACF plot for differenced quarterly time series

Table 4. Conditional least squares estimation for monthly time series.

Parameter	Estimate	Standard error	t value	Approx Pr > t	Lag
MU	1588.1	898.69711	1.77	0.0797	0
MA1,1	1.99693	0.08688	22.98	<0.0001	1
MA1,2	-1.42805	0.15598	-9.16	<0.0001	2
MA1,3	0.32658	0.08748	3.73	0.0003	3
AR1,1	1.72513	0.01148	151.09	<0.0001	1
AR1,2	-1	0.01144	-87.42	<0.0001	2

Table 5. Conditional least squares estimation for bimonthly time series.

Parameter	Estimate	Standard error	t value	Approx Pr > t	Lag
MU	3638.5	2935.1	1.24	0.2199	0
MA1,1	0.68181	0.12167	5.6	<0.0001	1
AR1,1	1.00597	0.06042	16.65	<0.0001	1
AR1,2	-0.95325	0.05881	-1621	<0.0001	2

Table 6. Conditional least squares estimation for quarterly time series.

Parameter	Estimate	Standard error	t value	Approx Pr > t	Lag
MU	14328.6	10062	1.42	0.1655	0
MA1,1	-0.17119	0.21144	-0.81	0.4231	1
AR1,1	-0.99001	0.07057	-14.03	<0.0001	1
AR1,2	-1	0.06747	-14.82	<0.0001	2

Table 7. Possible models for monthly series.

Model	AIC	SBC
ARIMA(1,1,2)	3106.59	3118.057
ARIMA(2,1,2)	3098.66	3112.995
ARIMA(2,1,3)	3014.03	3031.238

Table 8. Possible models for bimonthly series.

Model	AIC	SBC
ARIMA(1,1,1)	1696.89	1703.413
ARIMA(1,1,2)	1695.576	1704.274
ARIMA(2,1,1)	1630.904	1639.602

Table 9. Possible models for quarterly series.

Model	AIC	SBC
ARIMA(1,1,0)	917.7574	920.6889
ARIMA(1,1,1)	922.5976	926.9948
ARIMA(2,1,1)	853.8797	859.7425

analysis of time series and performing the conditional least squares estimation methods, some possible models for the three series has been found. The possible candidate models for monthly, bimonthly and quarterly time series are depicted below in Tables 7, 8, and 9, respectively:

The values of the relative quality measures AIC and SBC are also mentioned. The best model is selected by considering the lowest value of AIC and SBC for all the time series. Hence the final model selected for monthly series is ARIMA(2, 1, 3)

$$\Delta y_t = 1588.1 + 1.72513\Delta y_{t-1} - \Delta y_{t-2} + \varepsilon_t - 1.99693\varepsilon_{t-1} + 1.42805\varepsilon_{t-2} - 0.32658\varepsilon_{t-3} \tag{2}$$

The final model selected for bimonthly series is ARIMA(2, 1, 1)

$$\Delta y_t = 3638.5 + 1.00597\Delta y_{t-1} - 0.95325\Delta y_{t-2} + \varepsilon_t - 0.68181\varepsilon_{t-1} \tag{3}$$

The final model selected for quarterly series is ARIMA(2, 1, 1)

$$\Delta y_t = 14328.6 - 0.99001\Delta y_{t-1} - \Delta y_{t-2} + \varepsilon_t + 0.17119\varepsilon_{t-1} \tag{4}$$

4 Results and Discussion

The actual electricity consumption and that forecasted by ARIMA method are compared from April 2005 to February 2016. An ARIMA based forecasting of electricity consumption has been performed and applied to the practical power system of the health care institution. ARIMA based method is more reliable and better than the traditional forecasting methods.

The data has been analyzed for all the three series. Table 10 shows the calculated RMSE (Root Mean Square Error) and MPE (Mean Percentage Error) of the best ARIMA models selected for monthly, bimonthly and quarterly time series

Table 10. RMSE and MPE for monthly, bimonthly and quarterly time series.

ARIMA	RMSE	MPE
Monthly	40021.25	0.22
Bimonthly	63891	0.39
Quarterly	132976.98	1.22

As depicted in Table 10, the most suitable period for forecasting is the monthly time series model with the lowest value of RMSE and MPE. On the basis of these forecasting accuracy measures of RMSE and MPE, the monthly forecasting model will be used to perform the final prediction of the electricity consumption of the health care institution.

Figure 10 shows the comparison of the actual and forecasted observations for the time period of April 2005 to February 2016, while, Fig. 11 illustrates the prediction of electricity consumption for 2 years ahead. Hence, the resultant prediction of electricity

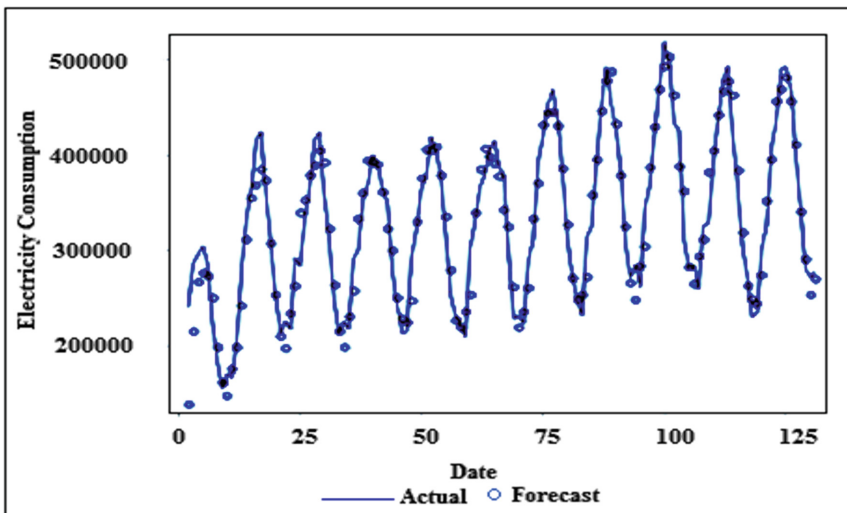


Fig. 10. The actual and forecasted values of monthly time series data

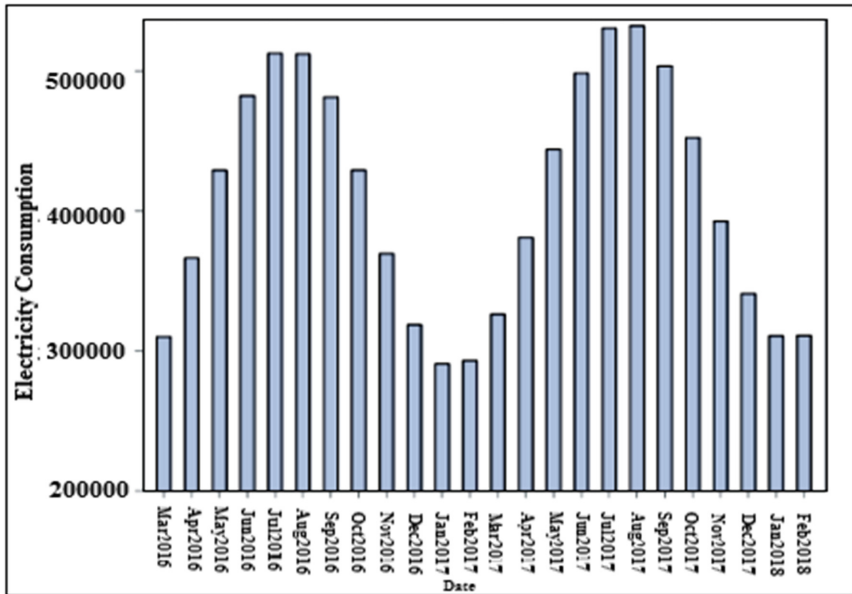


Fig. 11. Prediction of electricity consumption for two years ahead

consumption of the health care institution has been performed for the time period of March 2016 to February 2018.

5 Conclusion

The ARIMA based forecasting of electricity consumption has been performed and applied to the practical power system of the health care institution. From the results it can be concluded that this methodology is very efficient, and is more accurate and reliable than any other single forecasting methods which bring out imprecise predictions. Electricity being a crucial requirement of society needs to be predicted in advance as it is difficult to store. This forecasting will benefit the health care institution in many ways. It will help the management to anticipate the increase in capacity of the power system to cater to the futuristic needs, in addition to allocating the budget for planning and updating of their power system.

References

1. Sauhats, A., Varfolomejeva, R., Lmkevics, O., Petrecenko, R., Kunickis, M., Balodis, M.: Analysis and prediction of electricity consumption using smart meter data. In: IEEE 5th International Conference on Power Engineering, Energy and Electrical Drives (POWER-ENG) (2015)

2. Ahuja, D., Tatsutani, M., Schaffer, D.: Sustainable energy for developing countries. S.A.P.I. EN. S. (2.1) (2009)
3. Katara, S., Faisal, A., Engmann, G.: A time series analysis of electricity demand in Tamale, Ghana. *Int. J. Stat. Appl.* **4**(6), 269–275 (2014). Scientific & Academic Publishing
4. Kandananond, K.: Forecasting electricity demand in Thailand with an artificial neural network approach. *Energies* **4**(12), 1246–1257 (2011)
5. Yasmeen, F., Sharif, M.: Forecasting electricity consumption for Pakistan. *Int. J. Emerg. Technol. Adv. Eng.* **4**(4), 496–503 (2015)
6. Erdogdu, E.: Electricity demand analysis using cointegration and ARIMA modeling: a case study of Turkey. *Energ. Policy* **35**(2), 1129–1146 (2007)
7. Bianco, V., Manca, O., Nardini, S.: Electricity consumption forecasting in Italy using linear regression models. *Energy* **34**(9), 1413–1421 (2009)
8. Analytics, Business Intelligence and Data Management (n.d.). http://www.sas.com/en_us/software/university-edition.html.
9. Discover SAS® Customer Resources (n.d.). <https://support.sas.com>
10. Ghosh, S.: Univariate time-series forecasting of monthly peak demand of electricity in Northern India. *Int. J. Indian Cult. Bus. Manag.* **1**(4), 466 (2008)
11. Janacek, G.: Time series analysis forecasting and control. *J. Time Ser. Anal.* **31**(4), 303 (2010)
12. Dordonnat, V., Pichavant, A., Pierrot, A.: GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models. *Int. J. Forecast.* **2**(3), 1005–1011 (2016)
13. Camara, A., Feixing, W., Xiuqin, L.: Energy consumption forecasting using seasonal ARIMA with artificial neural networks models. *Int. J. Bus. Manag.* **11**(5), 231 (2016)
14. Razak, A.F., Shitan, M., Hashim, A.H., Abidin, I.: Load forecasting using time series models. *Jurnal Kejuruteraan* **21**(1), 53–62 (2009)
15. Nie, H., Liu, G., Liu, X., Wang, Y.: Hybrid of ARIMA and SVMs for short-term load forecasting. *Energ. Procedia* **16**, 1455–1460 (2012)
16. Lee, C.-M., Ko, C.-N.: Short-term load forecasting using lifting scheme and ARIMA models. *Expert Syst. Appl.* **38**(5), 5902–5911 (2011)