Electricity Consumption Forecasting Using Time Series Analysis

Praphula Kumar Jain(✉), Waris Quamer, and Rajendra Pamula

Department of Computer Science and Engineering,
Indian Institute of Technology (Indian School of Mines),
Dhanbad 826004, Jharkhand, India
praphulajnl@gmail.com, mr.warisquamer@gmail.com,
rajendrapamula@gmail.com

Abstract. The demand for electricity has been continuously increasing over the years. To understand the future consumption, a good predictive model is entailed. The ARIMA models have been extensively used for time series prediction showing encouraging results. In this paper, an attempt is made on forecasting the electricity consumption using the ARIMA model. Using the mean absolute percentage error (MAPE) to measure forecast accuracy, the model was able to forecast with an error of 6.63%. Results shows that the ARIMA model has a potential to compete with existing techniques for electricity consumption forecast.

Keywords: Electricity consumption · Forecast · ARIMA · Time series

1 Introduction

The demand for electricity has been continuously increasing in every sector [1]. The increased dependency on the electronic and electrical appliances necessitates the need for future demand forecast. Electricity consumption forecast plays an integral role in planning the future in terms of the size, location and type of the future generating plants as well as in deciding and planning for maintenance of the existing power systems [2].

The ARIMA model has been extensively used in forecasting economic, stock prices, marketing, social problems, industrial production etc. It is a statistical analysis model known to be efficient and robust for short-term forecast and requires at least 40 past data point's values.

In this paper, the electricity consumption in IIT(ISM) Dhanbad for the year 2008–09 is forecasted based on data from the year 2004 to 2008 using the ARIMA models, and then root mean square error (RMSE) and mean absolute percentage error (MAPE) is used to select the best model as the basis of model performance [24, 25].

Rest of the paper is organized as follows. In Sect. 2, the approaches used in the earlier research paper in the forecast of electricity consumption are reviewed and Sect. 3 presents a brief overview of ARIMA model. Section 4, then lays out the dataset used. Section 4 describes the methodology employed while in Sect. 5, the experimental results obtained are presented and analyzed. Last of all, Sect. 6 finalizes this paper with a conclusion and future research potentials.

M. Singh et al. (Eds.): ICACDS 2018, CCIS 906, pp. 327–335, 2018.
https://doi.org/10.1007/978-981-13-1813-9_33

2 Related Work

For forecasting the electricity consumption, there are different methods deployed by researchers. Author [2] presented an integrated framework based on Artificial Neural Network, Multilayer Perceptron, conventional regression and design of experiment for forecasting household electricity consumption using five input variables viz. electricity price, urban house income, urban household size, refrigerator price index and TV price index.

Different variations are observed in the load profile of consumers depending on income level, residence type and locality as well as environmental factors [3, 4]. Author [5] adopted data mining techniques for analysis of electricity consumption in order to extract information using the K-means clustering algorithm.

Authors [6, 7] carried out an analysis of seasonal electricity consumption and made an attempt in recognizing environmental effects on the consumption.

Author [8] carried out a comparison of different models for electricity consumption forecasting, like regression, neural network, and least square support vector machine. Author [9] presented a approach for forecasting of daily electricity consumption in the administrating buildings. In the paper [10] author's indication is to a warning in the shortfall of electricity if the same situation exists.

3 ARIMA Model - an Overview

The ARIMA model, also known as Box-Jenkins has been widely used for short-term forecast. The Autoregressive (AR) part of the ARIMA indicates the regression of the time series over its own lagged values. Integrated (I) indicates that the values have undergone differencing and the Moving Average (MA) Indicates weighted moving average over regression errors [11]. A Non-seasonal ARIMA model is represented as ARIMA (p, d, q) where, p is the order (number of time lags) of AR model, d represents the degree of integration (differencing), and q is the order of the MA [12].

Seasonal ARIMA model is represented as ARIMA(p, d, q) $(P, D, Q)_m$, where the P, D, Q depict the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model and m is the number of periods in each season [13, 14].

In order to estimate the values for the various terms of ARIMA model, the steps involving in finding autocorrelation and partial autocorrelation between the values of the data [15, 16]. Autocorrelation is the correlation of a time series with a delayed copy of itself and is defined as ACF = corr(X_t, X_{t+k}). Here X_t and X_{t+k} are the current observation and the observation after k period respectively.

Partial Auto-Correlation (PACF) [17] is the partial correlation of X_{t+k} with X_t i.e. it controls the values of the time series at all shorter lags which ACF does not. It is defined for positive lag only with values lying between −1 and +1. Table 1 gives the idea as how to make the estimation for initial values of ARIMA (p, d, q) [18].

Table 1. Characteristics of ACF and PACF graph for AR, MA and ARMA.

Characteristics	AR (p)	MA (q)	ARMA (p, q)
ACF	Decays	Cuts after q lags	Decays
PACF	Cuts after p lags	Decays	Decays

4 Dataset Description

Day by day demand of electricity is increasing because of uses of electrical and electronics instruments. To forecasting the electricity consumption, the electricity consumption data of Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, India, collected from the electricity distribution unit of it. This data set contains the electricity consumption in unit (in kWh) from July 2004 to June 2009. The data specified the unit consumed every month between the mentioned periods [19–23].

5 Methodology Used

In order to build an ARIMA model, the steps used are as follow:

Step 1. Data Visualization
The data is visualized and it is determined whether the data shows any overall trend or seasonal trend. The time series data was decomposed into constituent's viz. Trend, Seasonality and Residual Values. The trend would represent the optional and often linear increasing or decreasing behavior of the series over time, whereas the seasonality would depict its optional repeating patterns or cycles of behavior over time. The residual values essentially take out the trend and seasonality of the data, making the values independent of time. The seasonal_decompose function in stats models was used for the same.

Step 2. Stationarity Testing
A time series is stationary if its statistical properties such as mean, variance are constant over time. In a time series, observations are dependent on time, but a linear regression assumes all the observations to be independent of each other. So stationarising the data could enable us to apply regression techniques to time dependent variables. The series is made stationary by estimating the trend and seasonality and eliminating them from the series. For this purpose the logarithmic transformation and differencing methods are applied.

In order to test the Stationarity of data following methods were used:

1. **Plotting Rolling Statistics:** A plot depicting moving average and moving standard deviation is drawn and it is observed if it varies with time. A moving average/standard deviation means that at any instant 't', the average/standard deviation of the last year, i.e. last 12 months is taken.
2. **Dickey-Fuller Test:** It is a statistical test for testing stationarity. Here, the null hypothesis is that the time series is non-stationary. The test results consists of a Test Statistic and some Critical Values for different confident levels. If the 'Test Statistic'

is less than the ‘Critical Value’, the null hypothesis can be rejected and it can be said that the series is stationary.

The two major reasons for the time series to be non-stationary is the trend and the seasonality. The series is made stationary by estimating the trend and seasonality and eliminating them from the series. For this purpose the logarithmic transformation and differencing methods are applied.

Step 3. Deduction of Optimal Parameters
ACF and PACF are used to determine the suitable model parameters.

Step 4. Model Validation
This step involved validating the model using statistics and confidence intervals and tracking of model performances.

Step 5. Forecast
The best model obtained is implemented on the series and used to forecast the future values. The values are reverted back to the original scale.

6 Result Analysis

The electricity consumption in IIT(ISM) during the period July 2004 to June 2009 is depicted in Figs. 1 and 2 represents the constituents of the time series viz. are trend, seasonality and residual values. It can be observed that the electricity consumption data contains both an overall upward trend and has a seasonality to it and thus, seasonal ARIMA was used for forecasting.

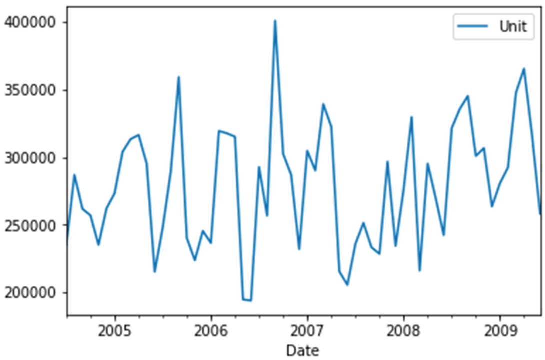

Fig. 1. Electricity consumption data for period 2004–09.

The presence of trend and seasonality makes the data non-stationary and the same can be confirmed by the rolling statistics and Dickey-Fuller test on the electricity consumption data as illustrated in Fig. 3 and Table 2 respectively. Although slight change in standard deviation is seen, but it can be clearly observed that the mean is varying with time. Also, the test statistics confirm the same since it is greater than the the critical values.

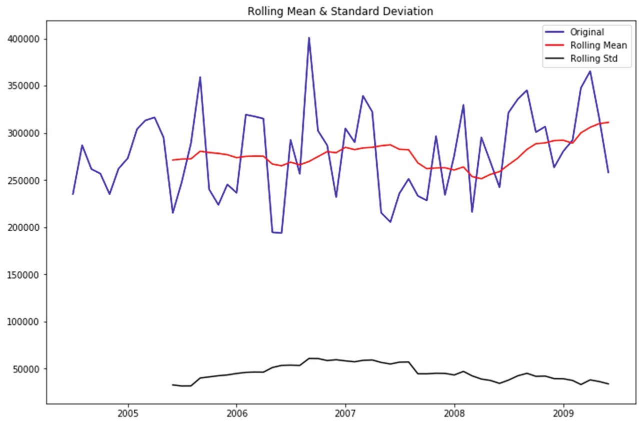

Fig. 2. Constituents of the electricity data.

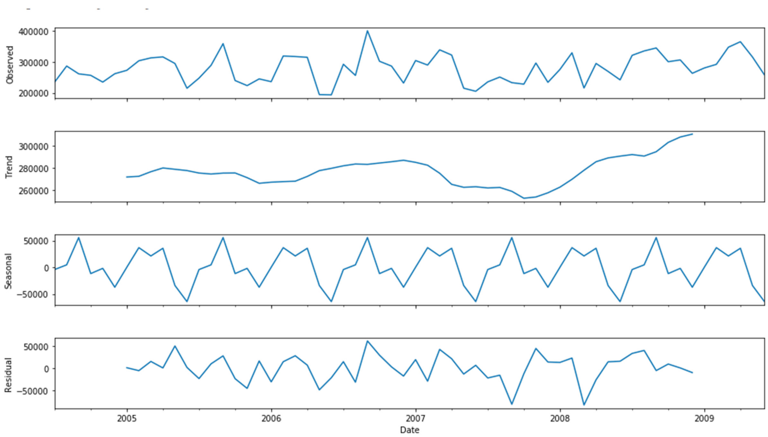

Fig. 3. Rolling statistics for the electricity data.

Table 2. Results of dickey-fuller test on the electricity data.

Statistics	Value
Test statistics	−2.144462
p-value	0.227016
#Lags used	7
Number of observations	52
Critical value (1%)	−3.562879
Critical value (5%)	−2.918973
Critical value (10%)	−2.597393

The series was made stationary using logarithmic transformation and differencing methods. It became stationary after the seasonal first difference was taken. The ACF and PACF correlogram was plotted as depicted in Fig. 4, to select the suitable AR, MA, SAR, and MAR terms for the model.

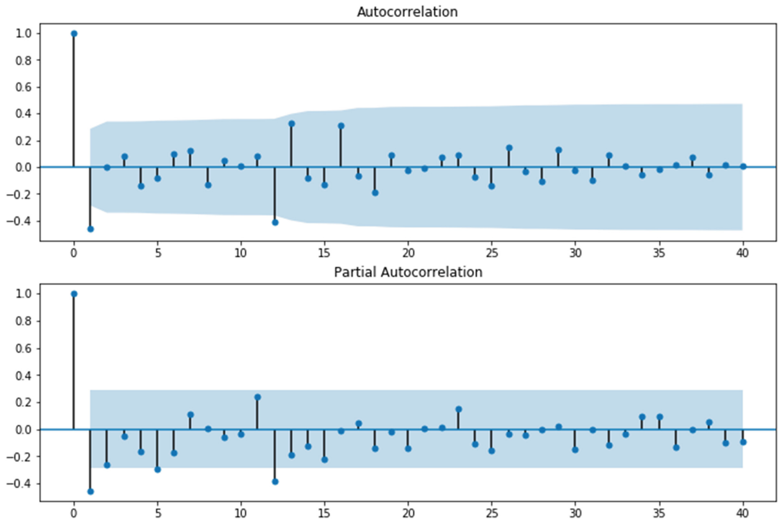

Fig. 4. The ACF and PACF graph for the first seasonal difference.

From the correlograms, it was observed that both the ACF and PACF cuts the upper confidence level for the first time at lag value 0 and hence, the coefficients of both AR and MA terms would zero i.e. p = 0 and q = 0. Since the ACF and PACF plot is negative at lag 12, there should be a SMA and SAR term to the model. A function was created using all possible combinations of parameters for fitting the models, the outcome was predicted using the models, and the model with the smallest MAPE was selected. The best model was found to be seasonal ARIMA(0, 1, 0)$\times$(2, 0, 1, 12) model which was used to forecast the future electricity consumption.

Figure 5 shows the forecasted electricity consumption (yellow) for the academic year 2008–09 and the actual data (blue), also tabulated in Table 3. The best model was able to forecast the consumption with a MAPE of 6.63%.

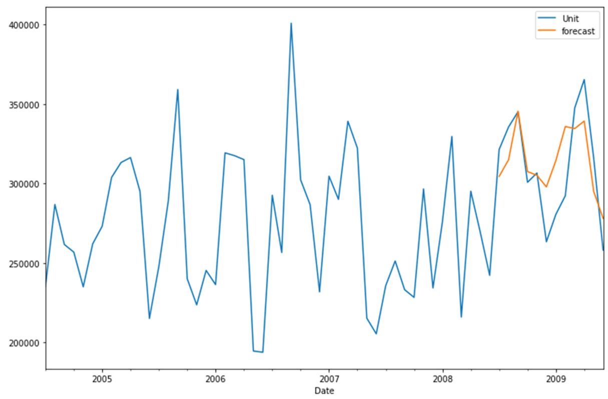

Fig. 5. Actual data Vs forecasted data.

Table 3. Actual data of electricity consumption Vs the forecasted data.

Month	JAN	FEB	MARCH	APRIL	MAY	JUNE	JULY	AUG	SEP	OCT	NOV	DEC
Actual	280572	292272	347604	365376	316116	258036	321408	335580	345156	300744	306600	263340
Predicted	314373	335838	334558	339294	295065	278028	304447	314965	345496	307510	305122	297856

7 Conclusions

Electricity demand forecasting plays an integral role in planning for the electricity production and determine the resources needed to operate the plants such as fuels. Furthermore, it helps in planning for future electricity needs and thus establishing new plants and networks.

The analysis of the electricity consumption in IIT(ISM) for the period 2004–2008 gave us a seasonal ARIMA (0, 1, 0)×(2, 0, 1, 12) model as the best model and it was able to forecast the consumption for year 2008–2009 with a MAPE of 6.63%.

References

1. Navani, J.P., Sharma, N.K., Sapra, S.: Technical and non-technical losses in power system and its economic consequence in Indian economy. Int. J. Electron. Comput. Sci. Eng. **1**(2), 757–761 (2016)
2. Azadeh, A., Faiz, Z.S.: A meta-heuristic framework for forecasting household electricity consumption. Appl. Soft Comput. **11**(1), 614–620 (2011)
3. Dzobo, O., et al.: Multi-dimensional customer segmentation model for power system reliability-worth analysis. Int. J. Electr. Power Energy Syst. **62**, 532–539 (2014)
4. Min, B., Golden, M.: Electoral cycles in electricity losses in India. Energy Policy **65**, 619–625 (2014)

5. Rathod, R.R., Garg, R.D.: Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. Int. J. Electr. Power Energy Syst. **78**, 368–374 (2016)
6. Chen, C.S., Hwang, J.C., Huang, C.W.: Application of load survey systems to proper tariff design. IEEE Trans. Power Syst. **12**(4), 1746–1751 (1997)
7. Benítez, I., et al.: Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. Int. J. Electr. Power Energy Syst. **55**, 437–448 (2014)
8. Kaytez, F., et al.: Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. Int. J. Electr. Power Energy Syst. **67**, 431–438 (2015)
9. Dong, B., et al.: A hybrid model approach for forecasting future residential electricity consumption. Energy Build. **117**, 341–351 (2016)
10. Hussain, A., Rahman, M., Memon, J.A.: Forecasting electricity consumption in Pakistan: the way forward. Energy Policy **90**, 73–80 (2016)
11. Nguyen, H., Hansen, C.K: Short-term electricity load forecasting with time series analysis. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, pp. 214–221 (2017)
12. Nichiforov, C., Stamatescu, I., Făgărăşan, I., Stamatescu, G.: Energy consumption forecasting using ARIMA and neural network models. In: 2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE), Galati, pp. 1–4 (2017)
13. Mahalakshmi, G., Sridevi, S., Rajaram, S.: A survey on forecasting of time series data. In: 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE 2016), Kovilpatti, pp. 1–8 (2016)
14. Papadopoulos, S., Karakatsanis, I.: Short-term electricity load forecasting using time series and ensemble learning methods. In: 2015 IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL, pp. 1–6 (2015)
15. Zhou, R., Pan, Y., Huang, Z., Wang, Q.: Building energy use prediction using time series analysis. In: 2013 IEEE 6th International Conference on Service-oriented Computing and Applications, Koloa, HI, pp. 309–313 (2013)
16. Fahmi, F, Sofyan, H.: Forecasting household electricity consumption in the province of Aceh using combination time series model. In: 2017 International Conference on Electrical Engineering and Informatics (ICELTICs), Banda Aceh, pp. 97–102 (2017)
17. Deng, J., Jirutitijaroen, P.: Short-term load forecasting using time series analysis: a case study for Singapore. In: 2010 IEEE Conference on Cybernetics and Intelligent Systems, Singapore, pp. 231–236 (2010)
18. Haydari, Z., Kavehnia, F., Askari, M., Ganbariyan, M.: Time-series load modelling and load forecasting using neuro-fuzzy techniques. In: 2007 9th International Conference on Electrical Power Quality and Utilisation, Barcelona, pp. 1–6 (2007)
19. Jifri, M.H., Hassan, E.E., Miswan, N.H.: Forecasting performance of time series and regression in modeling electricity load demand. In: 2017 7th IEEE International Conference on System Engineering and Technology (ICSET), Shah Alam, pp. 12–16 (2017)
20. Ridzuan, M.R.M., Hassan, E.E., Abdullah, A.R., Bahaman, N., Kadir, A.F.A.: A new meta heuristic evolutionary programming (NMEP) in optimizing economic energy dispatch. J. Telecommun. Electron. Comput. Eng. **8**(2), 35–40 (2016)
21. Miswan, N.H., Said, R.M., Anuar, S.H.H.: ARIMA with regression model in modelling electricity load demand. J. Telecommun. Electron. Comput. Eng. **8**(12), 113–116 (2016)
22. Ferhatosmanoglu, N., Macit, B.: Incorporating explanatory effects of neighbour airports in forecasting models for airline passenger volumes. In: Proceedings of 5th the International Conference on Operations Research and Enterprise Systems, ICORES, pp. 178–185 (2016)

23. Miswan, N.H., Ping, P.Y., Ahmad, M.H.: On parameter estimation for malaysian gold prices modelling and forecasting. Int. J. Math. Anal. **7**(21–24), 1059–1068 (2013)
24. Usha, T., Balamurugan, S.: Seasonal based electricity demand forecasting using time series analysis. Circuits Syst. **7**, 3320–3328 (2016)
25. Espinoza, M., Joye, C., Belmans, R., Moor, B.D.: Short-term load forecasting profile identification and customer segmentation: a methodology based on periodic time series. IEEE Trans. Power Syst. **20**(3), 443 (2005)