SARIMA Model Forecasting of Short-Term Electrical Load Data Augmented by Fast Fourier Transform Seasonality Detection

Hmeda Musbah
Department of Electrical and Computer Engineering
Dalhousie University
Halifax, Canada
Hm392855@dal.ca

Mo El-Hawary
Department of Electrical and Computer Engineering
Dalhousie University
Halifax, Canada
elhawary@dal.ca

Abstract- In this study, Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used to forecast short term electric load data. It is known that electrical load data is affected by weather conditions, therefore the electrical load data will have a seasonal component. For this reason, representing this data in the frequency spectrum domain will reflect the exact seasonal time. As such, Fast Fourier Transformation algorithm (FFT) has been used to detect the existence of the seasonal component in the time series of hourly electrical load data. The FFT technique gives a clear view of the behavior of the time series, which shows the three main components located at frequencies f_1 ***=3.17*** $e^{-8}\,Hz$***,*** f_2***=1.157***$e^{-5}\,Hz$***, and*** f_3***=2.315*** $e^{-5}\,Hz$***. The component that is located at*** f_2***= 1.157***e^{-5} ***Hz has a significant amplitude and tends to repeat itself every 24 hours. The model that has been selected to forecast one week ahead of the electrical load data is SARIMA (6,1,1) (4,1,1). Most of the predicted values fall into the 95% of confidence interval. The P-values for Ljung-Box statistic indicate that of P-values are higher than 5%. Unlike other techniques, FFT technique offers a quick view of the data behavior with high accuracy.***

Keywords- Stationary, Fast Fourier Transform, Frequency Spectrum, Seasonality

I. INTRODUCTION

Forecasting helps in the planning and decision-making process since it gives an insight of the future uncertainty using the past and current behavior of given observations. Load forecasting plays a vital role in the energy system, it has a direct effect on planning and management strategies, since inaccurate forecasting could cost power companies money and time. Therefore, it is quite important to choose a reliable and effective method of load forecasting. For this purpose, several approaches have been employed to study and forecast electrical load. These approaches can be divided into three groups [1]: traditional approaches, artificial intelligence approaches and support vector regression approaches.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model, based on Box-Jenkins approach, is one of the popular traditional approaches. In this paper, the SARIMA model has been applied on hourly electrical load data from January 2017 to December 2017, as shown in Fig. 1, to achieve a short-term prediction of electrical load data for the given data. Due to the reliability of this approach several researchers use this model, confirming its importance in the forecasting literature [2].

When studying and analyzing a time series it is important to identify its pattern, so the predication is more accurate. The seasonal component of the time series affects accuracy of the prediction. Usually the seasonal component is easily identified by visual inspection of graphing techniques such as run sequence plot, seasonal sub-series plot , multiple box plots or the autocorrelation plot [3]-[5]. But in some cases, when the time series is very long and there is a large concentration of observed data, the detection of seasonality is difficult. To the best of our knowledge most studies of electrical load forecasting use visual inspection to detect the seasonality. In this study the Fast Fourier Transform algorithm (FFT) is used to detect the seasonality in the electrical load data.

Two time series models were suggested in [6], the multiplicative decomposition model and the SARIMA model to forecast short term electricity demand in Singapore. The authors stated that the multiplicative decomposition model gave minor improvements of accuracy over the SARIMA model. The autoregressive integrated moving average ARIMA model has been applied to forecast seven years of domestic, commercial and industrial electricity demand in Tamale, Ghana[7]. The selected models showed that the industrial electricity demand was not rising faster than the domestic and commercial electricity demand. For forecasting load demand in distribution substations, [8] conducted a study to compare - ARIMA model, and Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy System techniques. Their study shows that ANN outperforms ARIMA model and Adaptive Neuro-Fuzzy System techniques. In [1] energy consumption was predicted using two methods, ARIMA model and a non-linear autoregressive neural network (NAR) model. Even though both models are considered effective, the comparison of the predictive error value favors the performance of the ARIMA model. The study found that the two methods performed well, but they preferred the ARIMA model because of its simple structure.

II. METHODOLOGY AND MATERIAL

A. *The Autoregressive Integrated Moving Average (ARIMA) Model*

ARIMA model consists of two parts, Autoregression (AR) and Moving Average (MA) which form an ARMA model, in addition to the possibility of converting the time series from non-stationary to stationary time series by the integrated term I. ARIMA model has been expanded to

978-1-7281-0319-8/19/$31.00 ©2019 IEEE

seasonal autoregressive integrated moving average SARIMA model, thereby increasing the ability to deal with seasonal time series. SARIMA model could be written as:

$$\emptyset_p(\boldsymbol{B})\phi_p(\boldsymbol{B}^s)\,\boldsymbol{X}_t = \boldsymbol{\Theta}_q(\boldsymbol{B})\boldsymbol{\Theta}_Q(\boldsymbol{B}^s)\boldsymbol{\epsilon}_t \qquad \textbf{(1)}$$

where:

$\boldsymbol{B}$ is the backward shift operator,

$\emptyset_p, \phi_p, \boldsymbol{\Theta}_q, \boldsymbol{\Theta}_Q$ are the polynomials of p, q, P, Q.

B. Box and Jenkins Approach

In 1970, Box and Jenkins proposed a new celebrated approach to forecast the data from historical data. The Box and Jenkins Approach is able to analyze and predict several types of time series. The Box and Jenkins approach is indicated in [9] as suitable for short-term and medium-term prediction. This approach is based on four basic steps:

- Model Identification: The right model structure of the AR, MA, or ARIMA and its order (p, d, q) are identified using the autocorrelation function (ACF) and partial autocorrelation (PACF) plots.
- Estimation of model parameters: Determining which ARIMA model to choose depends on the Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) value that make a tradeoff between the fit statistics of the model and its complexity. Therefore, the one with the minimum AIC value should be chosen as the appropriate model.
- Diagnostic of the model: The final model is chosen using two factors, residuals and estimated parameters.
- Forecasting: The tested model that has been chosen is ready for forecasting.

C. Stationary and Non-stationary Data

When the mean, covariance and autocorrelation are constant over time the data becomes stationary. Non-stationary data can never be modeled and forecasted due to its unpredictable nature. This indicates that all the non-stationary data should be converted to a stationary form before performing any forecasting.

D. Fast Fourier Transforms

The time series consists of several signals that have different amplitudes and periods. FFT converts these time series from time domain to frequency domain, and clearly shows for each signal the individual frequencies and the dominant frequency. This can help in identifying the seasonality in the time series, specifically when the seasonality is not show in the time series.

III. RESULTS AND DISCUSSION

Before applying the Box and Jenkins approach, the stationarity of the data must first be examined. It is very clear from the visual inspection of Fig. 1 that the mean and covariance of the data do not look constant over time which results in the non-stationarity of the data. Also, the Augmented Dicker-Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test with p-values of 0.6522 and 0.01 respectively confirm that the data is not stationary. Therefore, differencing is mandatory to make the series stationary [5]. Since the time series of electrical load data might be affected by different factors such as weather conditions, we can assume that the time series of electrical load data is seasonal. Hence the time series has been converted from the time domain to the frequency domain by Fourier Transformation to check the seasonality.

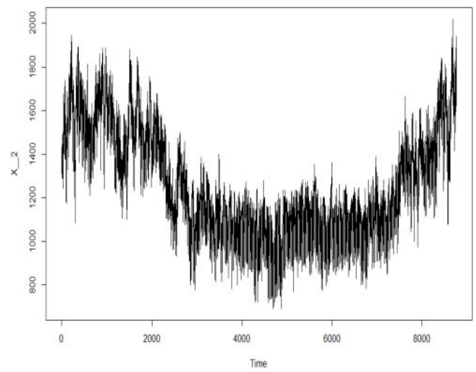

Fig. 1. Hourly electrical load data from January 2017 to December 2017

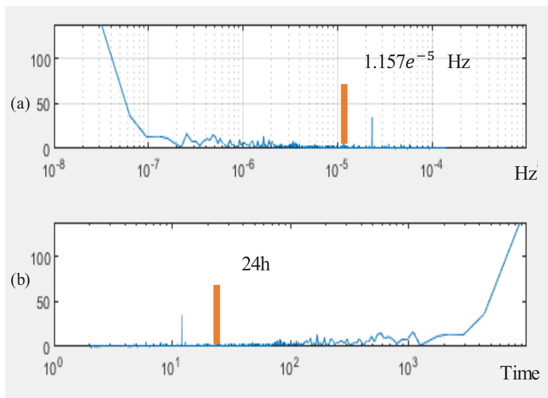

Fig. 2a and 2b. Time series of electric load data in frequency domain

The frequency spectrum of the time series of electric load data of the Fig. 2a shows there are three components at frequencies f_1=3.17e^{-8} Hz, f_2= 1.157 e^{-5} and f_3= 2.315 e^{-5} . Fig. 2b corresponds to time $t_1 = 8760$h , $t_2 = 24$h and $t_3 =$ 12 h. These components form the time series of electric load data. The component at frequency 1.157 e^{-05} Hz represents the day pattern, while the component at frequency 3.171e^{-08} Hz represents the whole pattern of time series of electric load data. The component at frequency f_2= 1.157e^{-5} Hz, that has the strongest significance in its amplitude, represents the pattern of two days and it tends to repeat itself every 48 hours. The time series of electric load data would need to be different more than once. Fig. -3- shows ACF and PACF after one non-seasonal and one seasonal difference.

The ACF plot of Fig. 3 shows two significant spikes at low lags that exceed the confident limits and one significant spike at lag 48, while the PACF plot in Fig. 3 shows a number of significant spikes that exceed the confident limits. Therefore, the suggested model is SARIMA (7,1,2) (1,1,1).

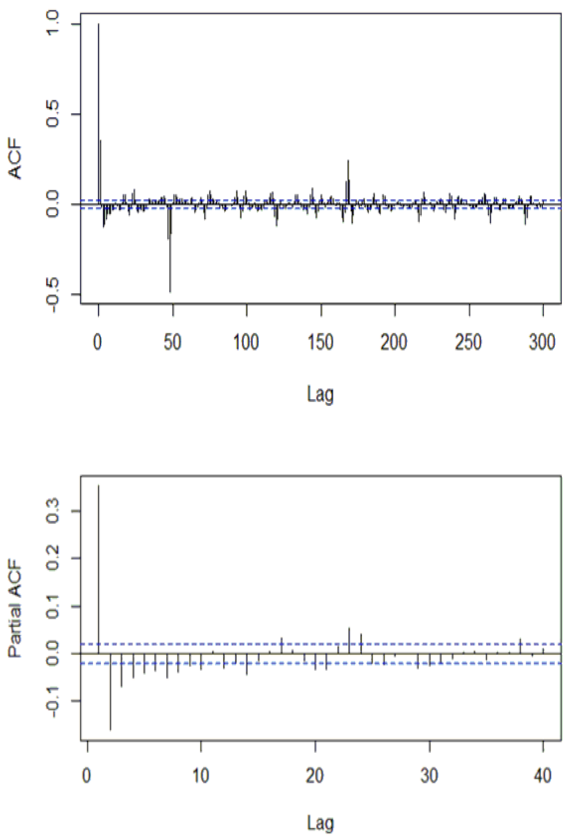

Fig. 3 ACF and PACF plots after one non-seasonal and seasonal differencing.

Table 1 Suggested models and their AIC and BIC

Model	AIC	BIC
SARIMA (1,1,1) (1,1,1)	88587.55	88622.94
SARIMA (4,1,1) (2,1,2)	85364.84	85435.62
SARIMA (6,1,1) (4,1,1)	**85306.71**	**85398.72**
SARIMA (7,1,2) (1,1,1)	85345.51	85423.36

- Estimation of model parameter

Table 1 shows different models that have been suggested. The model that has been chosen based on the minimum AIC and BIC values is SARIMA (6,1,1) (4,1,1). Table 2 shows the fitted model accuracy statistic.

Table 2 Accuracy tests of the fitted model

ME	RMSE	MAE	MPE	MAPE	MASE
0.064	31.46	23.91	-0.053	1.94	0.69

- Diagnostic of the model:

Fig. 4 shows that the standardized residual has no specific pattern with zero mean and constant variance between -4 to 4. ACF of residual shows that there are a few spikes at different lags that exceed the confidence limits. Also, P-values for Ljung-Box test show that the P-value that is located at lag 1 to lag 7 is higher than 5%.

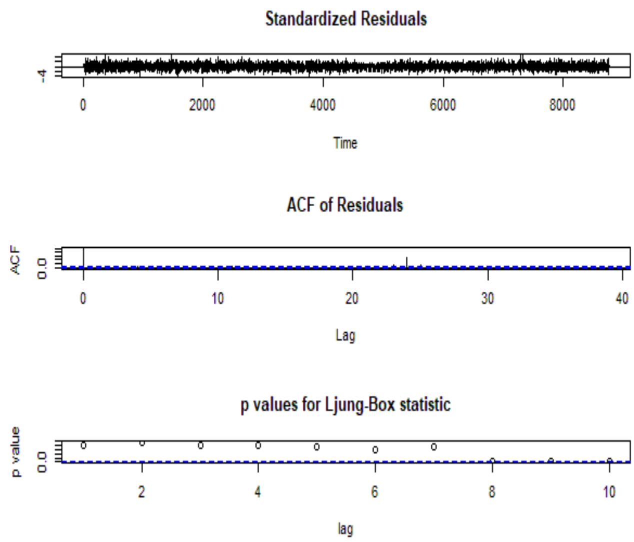

Fig. 4 Shows the residual plots of the electrical load data

- Forecasting

The model SARIMA (6,1,1) (4,1,1). has been used to forecast a week ahead for the electrical load data. Table 3 shows the first 5 hours of the forecasted values and its 95% confidence intervals.

Table 3 First 5 hours of the forecasted values

Date & Time	Observation	Forecasted	Lo 95	Hi 95
01/01/2018 01:00	1573.78	1600.821	1539.13	1662.50
01/01/2018 02:00	1537.95	1603.564	1483.99	1723.13
01/01/2018 03:00	1509.02	1623.091	1456.92	1789.26
01/01/2018 04:00	1495.29	1645.049	1445.05	1889.51

IV. CONCLUSION

The time series of electrical load data has been studied and its stationarity has been checked using ADF and KPSS tests.

These tests show that the series is not stationary. In addition, the seasonality of the series has been determined by converting the time series of electrical load from the time domain to the frequency domain by using FFT algorithm. The algorithm shows the main components that form the series and shows the significant component that tends to repeat itself. The best model was selected based on AIC and BIC and has high order of p and q which increased the complexity. The model was used to forecast a week ahead to the time series of electrical load data. Most of the forecasted data fell into the 95% confidence interval. We can determine from this that it is possible to employ the SARIMA models to forecast the hourly electrical load data, but it is better to be used with monthly electrical load data.

REFERENCES

[1] C. Nichiforov et al., "*Energy consumption forecasting using ARIMA and neural network models*," in Electrical and Electronics Engineering (ISEEE), 2017 5th International Symposium *On,* 2017.

[2] E. S. El-Mallah1 and S. G. Elsharkawy, "*Time-Series Modeling and Short-Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model*" 2016.

[3] J. M. Chambers *et al.*, "*Graphical methods for data analysis. 1983*" *Wadsworth, Belmont, CA,* vol. 35.

[4] W. Cleveland, *Visualizing Data, Murray Hill: AT&T Bell Laboratories.360 S,* 1993.

[5] Box, George EP, et al., *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[6] J. Deng and P. Jirutitijaroen, "*Short-term load forecasting using time series analysis: A case study for singapore*" in *Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference On,* 2010.

[7] S. Katara, A. Faisal and G. M. Engmann, "*A time series analysis of electricity demand in Tamale, Ghana*" *International Journal of Statistics and Applications,* vol. 4, *(6),* pp. 269-275, 2014.

[8] de Andrade, Luciano Carli Moreira and I. N. da Silva, "*Very short-term load forecasting based on ARIMA model and intelligent systems*" in *Intelligent System Applications to Power Systems, 2009. ISAP'09. 15th International Conference On,* 2009.

[9] E. Afrifa-Yamoah, "*Application of ARIMA models in forecasting monthly average surface temperature of Brong Ahafo Region of Ghana*" *International Journal of Statistics and Applications,* vol. 5, *(5),* pp. 237-246, 2015.