

RESEARCH ARTICLE

Modelling electricity sales in Sri Lanka and Colombo city using different statistical methods[†]

H.A. Pathberiya* and P. Dias

Department of Statistics and Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura, Gangodawila, Nugegoda.

Revised: 29 May 2012 ; Accepted: 19 November 2012

Abstract: Electricity is an essential form of energy used all over the world and the sales are growing each year owing to population growth, industrialization, etc. Forecasts of electricity sales would be of great importance to electric utilities when deciding on construction and investments. The objective of this study was to develop models suitable for forecasting monthly electricity sales for Sri Lanka and Colombo city. The monthly electricity sales data (in GWh) for Sri Lanka and for Colombo city from 2001 to 2009 were obtained from the Ceylon Electricity Board for this purpose. Three models for forecasting electricity sales were studied using three different approaches, namely, classical decomposition approach, stochastic approach and exponential smoothing approach. The estimated models were compared and the best model for forecasting electricity sales was selected. It is shown that the stochastic model given by ARIMA (0,1,1)(0,1,1)₁₂ generates more accurate forecasts for Colombo city and for Sri Lanka, compared with the other two models. The corresponding mean square deviations are 181.658 and 6.703 for Sri Lanka and Colombo city, respectively.

Keywords: Classical approach, exponential smoothing approach, forecasting, stochastic approach, time series.

INTRODUCTION

The need for electric energy is growing each year all over the world with the growth of the population and due to the enhancement of their requirements. When electricity sales to household, commercial and industrial customers were observed it was clear that household electrification has risen from 47 % to 80 % from 1996 to 2008 (Siyambalapitiya & Wickramasinghe, 2009). Electricity has become an essential form of energy at present and the society has become dependent on it.

The demand for electricity depends on factors such as population growth, economic activities and electricity price. A study by Ruwanthi and Wickremasinghe (1999) has shown that forecasts of electricity demand in Sri Lanka in any given period depends on the price of electricity and the GDP pertaining to that period. The demand for electricity is increasing over time, and an energy crisis is one of the most critical problems that may be faced by future generations. Hence, the concept of sustainable development has become a widely discussed topic at present and according to this concept each person has a responsibility for saving energy for future generations.

Moreover, it is vital to be knowledgeable about the future demand for energy, including electricity. Forecasts on electricity sales will determine the amount of electric power that should be generated, transmitted and supplied in the immediate future. Some past studies (Ruwanthi & Wickremasinghe, 1999; Zuhaimy & Khairil, 2005; Kareem & Majeed, 2006) have investigated various techniques to obtain accurate forecasts for electricity sales. Forecasts are crucial for electricity utilities to establish sustainable power development plans in order to achieve their objective of providing a good service to meet consumers' demand. Similarly, forecasts on electricity sales are very important for government authorities in making decisions on investments in the sector. If the existing power generating system seems insufficient to fulfill the future demand on electricity, then the government will have to consider new power sources and construct new power plants.

* Corresponding author (hasanthi@sjp.ac.lk)

[†] An abstract of this paper was presented and published in the Proceedings of the International Statistics Conference 2011 organized by the Institute of Applied Statistics of Sri Lanka jointly with the School of Mathematics and Statistics, University of Sydney, Australia.

The main objective of this study was to forecast electricity sales in Sri Lanka and in Colombo city based on the historical data. The Colombo city itself consumes around 14 % of the total electricity sold by the Ceylon Electricity Board (CEB). In terms of electricity consumption per account holder, the highest consumption in year 2000 has been in the Colombo city (Samaranayaka, 2008). As such, the prediction of electricity sales in Colombo would be of great importance to electricity utilities, to the government and to the entire community.

This study was carried out with the intention of investigating the best model out of three statistical methods, namely, classical decomposition approach, stochastic approach and exponential smoothing approach in order to forecast electricity sales in Sri Lanka and in the Colombo city.

METHODS AND MATERIALS

Monthly electricity sales data (in GWh) in Sri Lanka and in the Colombo city from 2001 to 2009 were obtained from the Ceylon Electricity Board. Data from 2001 to 2008 were used as the identification set for estimating models, whereas the monthly sales data of 2009 were used to verify the models.

Graphical methods were used to identify the behaviour of the data in the initial phase of analysis. Presence of long term trends, seasonal patterns and cyclical variations were examined using time series plots and plots of autocorrelations.

Thereafter, (i) classical decomposition approach, (ii) stochastic approach, and (iii) exponential smoothing approach were used to obtain forecasting models. In the stochastic approach, "Box-Jenkins" methodology was used to determine the most suitable stochastic model, which can be used to develop more reliable forecasts. Augmented Dickey-Fuller test was used to examine the stationarity of the series. Appropriate stochastic models were identified by a methodical examination of the behaviour of the auto correlation function (ACF) and partial auto correlation function (PACF). Adequacy of the stochastic model was tested with respect to the normality and the randomness of the residuals. Ljung-Box Q statistic was used to test the randomness of residuals. When developing an exponential smoothing model, Holt-Winters' exponential smoothing method was used, since the data contains both trend and seasonal components. Ultimately, the best approach out of these three approaches was selected to predict future

electricity sales in Sri Lanka and in the Colombo city by considering the mean squared deviation (MSD), correlation coefficient between the observed values and the fitted values of the models and Akaike Information Criteria (AIC) as the model selection criteria (Makridakis *et al.*, 1998).

RESULTS

Classical approach

Electricity sales data for Sri Lanka exhibited a long term upward trend in the time series plot. The ACF of the first differenced series was examined to identify the existence of seasonality. It indicates a significant autocorrelation at lag 1. It was recognized that the ACF of detrended series contains relatively large autocorrelation coefficients at lag 12 and lag 24. Thus, it was identified that the data contain a slight seasonal pattern with a seasonal length of 12 months.

Time series plot of the Colombo city electricity sales data clearly indicated the long term upward trend in monthly electricity sales. Moreover, a significant autocorrelation at lag 1 and a marginally significant autocorrelation at lag 12 were identified in the ACF of the detrended series. Hence, the Colombo city data were considered to have a slight seasonal pattern with a seasonal length of 12 months.

Long term trend in both series were estimated, using the method of least squares. Normality and homoscedasticity of residuals of the fitted models were examined with the use of Anderson-Darling test and the plot of residuals against the fitted values, respectively (Figure 01). Normality is significant in both Sri Lanka ($p = 0.198$) and in the Colombo city ($p = 0.653$) models. According to the residual plots of both models, the residual variance seemed to remain constant around zero. The estimated trend equations for all Island and the Colombo city are given by equations 1 and 2, respectively,

$$Y_t = 421.4 + 3.274 * t \quad (R^2 = 96.4 \%) \quad \dots(1)$$

$$Y_t = 60.2 + 0.406 * t \quad (R^2 = 90.5 \%) \quad \dots(2)$$

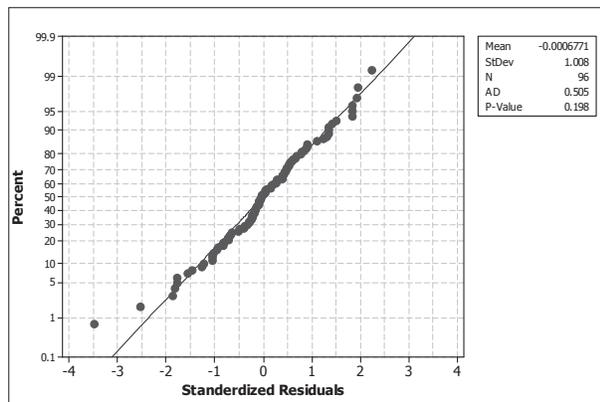
Where, Y_t = Electricity sales in GWh at time t
 $t = 1, 2, 3, \dots$

January 2011 was set as the origin of 't' and t increase by one unit for each month.

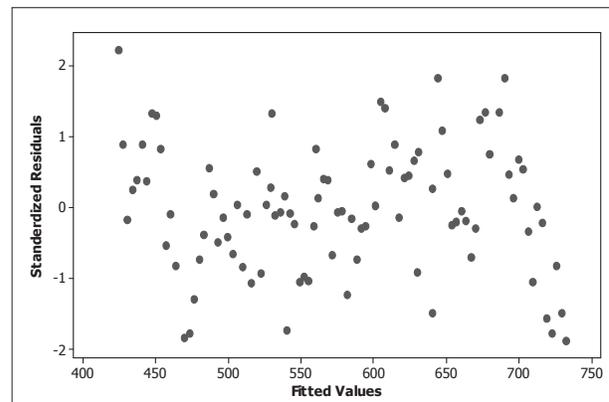
Table 1: Parameter estimates in linear trend models

	coefficient	All Island Standard Error	p value ^a	coefficient	Colombo city Standard Error	p value ^a
Constant	421.4	3.603	0.000	60.2	0.7545	0.000
t	3.274	0.0645	0.000	0.406	0.0135	0.000

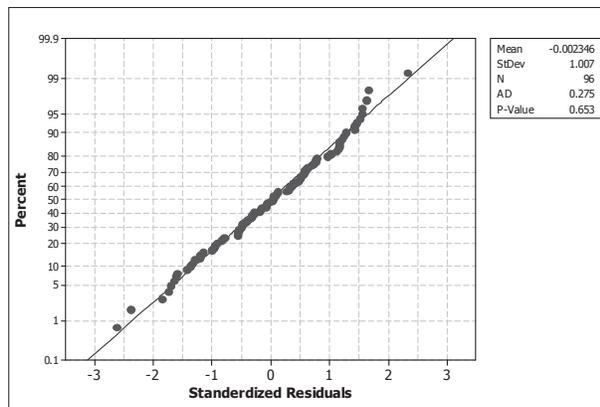
^a Null hypothesis: parameter is not significant



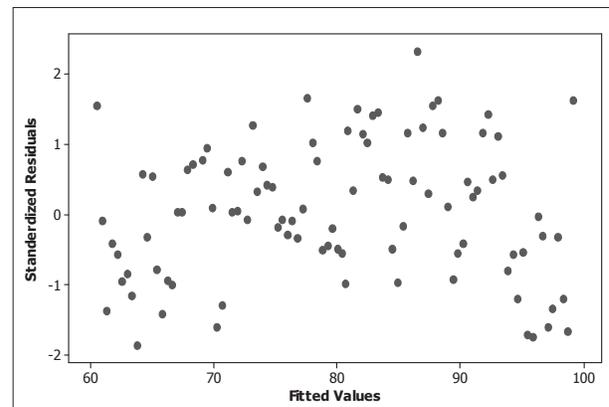
(a)



(b)



(c)



(d)

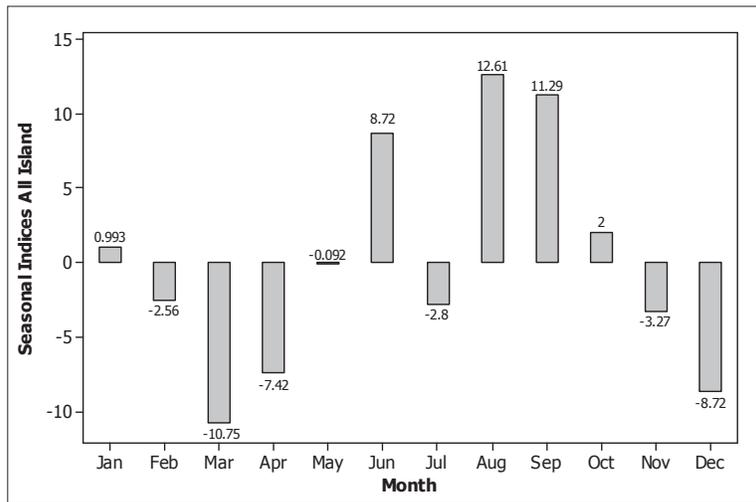
Figure 1: Plots for model adequacy of Stochastic model; (a) Normal probability plot of residuals – all island model; (b) Plot of residuals versus fitted values – all island model; (c) Normal probability plot of residuals – Colombo city model; (d) Plot of residuals versus fitted values – Colombo city model

An additive model was considered to be more suitable to represent the monthly electricity sales in Sri Lanka since the variation of electricity sales remains approximately the same over the period, whereas a multiplicative model was considered to be more suitable to represent the Colombo city data since the variation of the series seemed to be slightly changing over time. The means of the detrended values of corresponding months in different years were obtained as seasonal indices. A seasonal index represents how higher or lower the electricity sales have been for a given month, when compared with the average monthly electricity sales. Arithmetic mean of the seasonal indices

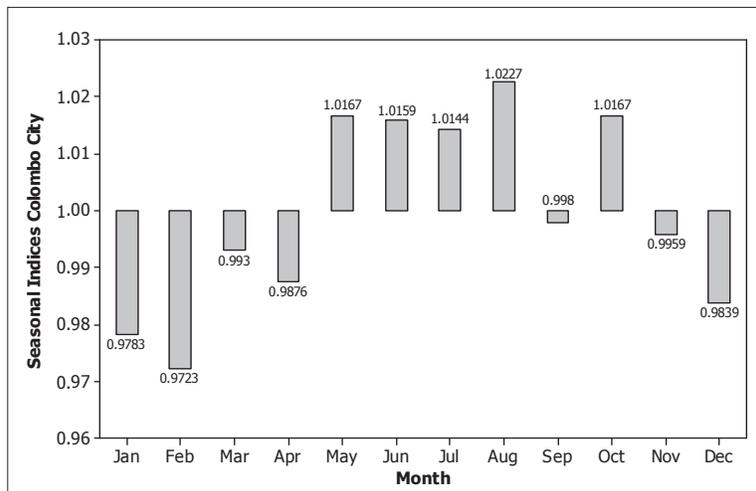
Table 2: Accuracy measures for identification and verification periods for the classical model

	MSD		r^a	
	All Island	Colombo city	All Island	Colombo city
Identification period	247.138	11.198	0.985	0.959
Verification period	186.476	10.050	0.906	0.547

^a r - Correlation coefficient between observed and fitted values



(a)



(b)

Figure 2: Seasonal indices for (a) All Island data; (b) Colombo city data

was considered as the average monthly electricity sales. For an additive model, this average equals to zero, whereas it equals to one for a multiplicative model.

Figure 02 (a) shows that the monthly electricity sales in Sri Lanka in June, August and September are higher than the average monthly electricity sales, whereas in other months it is lower than the average monthly electricity sales. Electricity sales in January and October are substantially closer to the average monthly electricity sales.

Figure 02 (b) shows that the monthly electricity sales in Colombo city for May, June, July, August and October are higher than the average monthly electricity sales, whereas it is lower than the average in January, February, March, April and December. Electricity sales in September and November are substantially closer to the average monthly electricity sales.

Seasonal changes in the electricity sales may be due to the climatic changes during the year. According to the climate of Sri Lanka, the night times of the months from April to September are warmer than the other months (*unpublished data*). Use of electrical appliances such as air conditioning equipment may increase with the increase of the temperature. This may account to the higher electricity sales than the average during the months of June to September when compared with the other months.

The classical model was verified by comparing the predicted monthly electricity sales for the year 2009 with the corresponding actual electricity sales. MSD and the correlation coefficient between the observed values and the fitted values for the identification period and the verification period were taken into account (Table 2).

Table 3: Results of the augmented Dickey-Fuller test

	p-value ^a	
	All Island	Colombo city
Original series	0.7669	0.5542
1 st differenced series	0.0001	0.000

^a Null hypothesis: Series has a unit root

Stochastic approach

The behaviour of the ACF and PACF was examined methodically in order to identify the potential stochastic models. Since the variation of the Colombo city data seems to slightly increase over time, the natural log transformation was used to stabilize the variance. In order to examine the stationarity of the series, unit root tests were performed. According to the Augmented Dickey-Fuller test, it was concluded that the stationarity of both the All Island and the Colombo city series can be obtained by applying the first order differencing (Table 3).

ACF of the first differenced series exhibits marginally significant auto correlation coefficients at lag 12 and 24, which indicate slight seasonality in the data. Hence, the first order seasonal differencing was applied to the first differenced series in order to obtain the deseasonalized series. Further, ACF and PACF were examined to identify the potential stochastic models. Out of several possible models ARIMA (0,1,1) (0,1,1)₁₂ model was selected as the most appropriate model for both Sri Lanka and for the Colombo city since those models produced the lowest MSE and the highest correlation coefficient between the observed values and the fitted values (Table 4) .

Further, it was concluded that the residuals related to these models satisfied the normality assumption with p values of 0.394 in the Colombo city model and 0.146 in Sri Lanka model (Figure 03). Moreover, randomness of the residuals of both models were significant according to Ljung-Box Q statistic (Table 5). The models for Sri Lanka and for the Colombo city are given by equations 3 and 4 respectively, where electricity sales at time t is denoted by Y_t and residual at time t is denoted by Z_t ;

$$\begin{aligned}
 Y_t = & Y_{t-1} + Y_{t-12} - Y_{t-13} + Z_t - 0.3747Z_{t-1} \\
 & - 0.8318Z_{t-12} + 0.3117Z_{t-13} \\
 & \dots(03)
 \end{aligned}$$

$$\begin{aligned}
 \ln Y_t = & \ln Y_{t-1} + \ln Y_{t-12} - \ln Y_{t-13} + Z_t - \\
 & - 0.7386Z_{t-1} - 0.5201Z_{t-12} + 0.3841Z_{t-13} \\
 & \dots(04)
 \end{aligned}$$

Note that Y is in GWh and t is the month, where in January 2001, t = 1

With the use of the above models, the monthly electricity sales in 2009 were forecasted. Model verification was done by comparing the predicted monthly electricity

sales for 2009 with the corresponding actual figures. Table 6 describes the performance of both the Sri Lanka model and the Colombo city model, using MSD and correlation coefficient between the observed values and the fitted values, both for the identification period and the verification period.

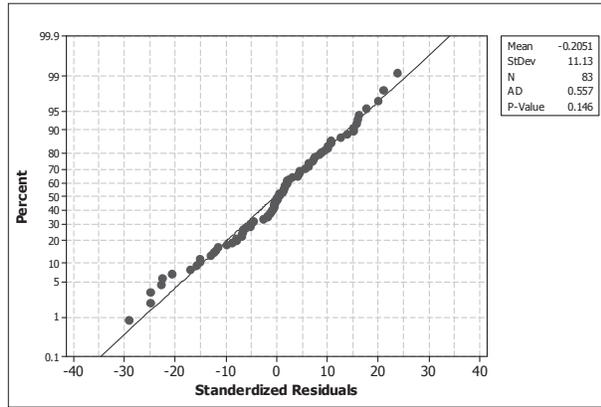
Table 4: Accuracy measures for possible stochastic models

	MSE		r ^a	
	All Island	Colombo city	All Island	Colombo city
ARIMA (0,1,1) (0,1,1) ₁₂	125.4	0.0018	0.992	0.951
ARIMA (1,1,0) (1,1,0) ₁₂	142.4	0.0024	0.991	0.934
ARIMA (1,1,0) (0,1,1) ₁₂	125.6	0.0021	0.992	0.940
ARIMA (0,1,1) (1,1,0) ₁₂	144.7	0.0019	0.991	0.949

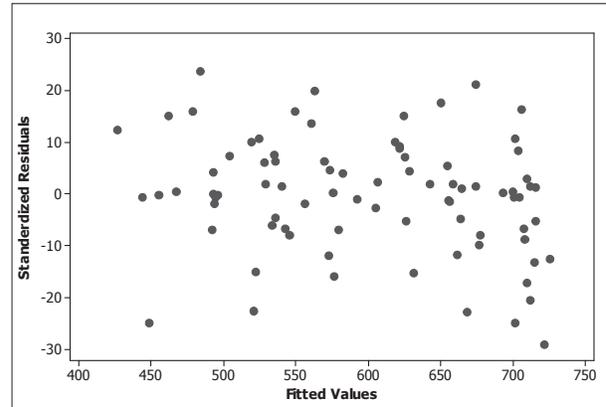
^a r - Correlation coefficient between observed and fitted values

Exponential smoothing approach

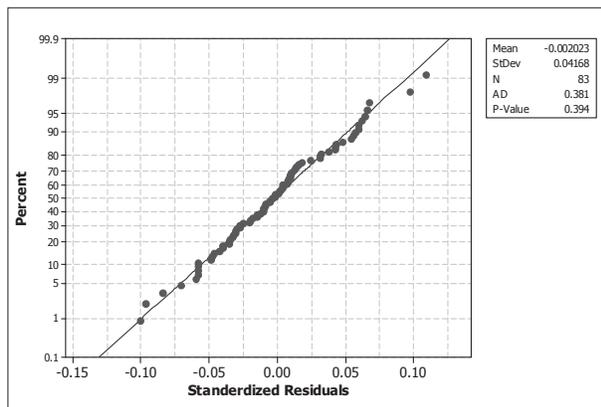
In the preliminary analysis, it was identified that the electricity sales both in Sri Lanka and the Colombo city contain long term trend and seasonal variations. Hence, Holt-Winters' method was used to develop an appropriate exponential smoothing model. Suitable smoothing constants were determined using a programme written in R (R Development Core Team, 2010) by minimizing the MAPE.



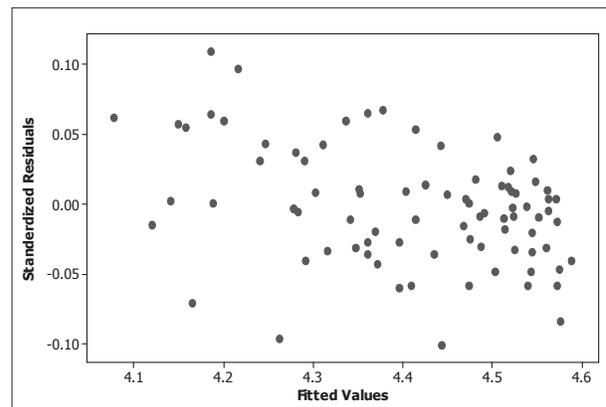
(a)



(b)



(c)



(d)

Figure 3: Plots for model adequacy of Stochastic model; (a) Normal probability plot of residuals – all island model; (b) Plot of residuals versus fitted values – all island model; (c) Normal probability plot of residuals – Colombo city model; (d) Plot of residuals versus fitted values – Colombo city model

Table 5: Results of the Ljung-Box Q test and corresponding p-values

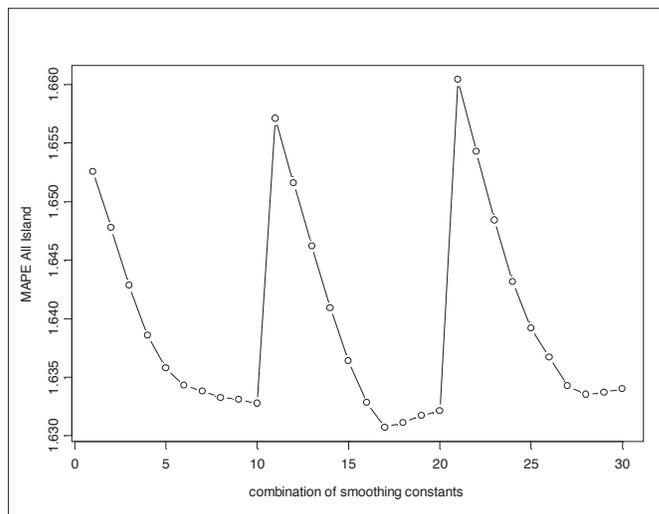
	All Island	Colombo city
Lag	12	12
Chi-Square statistic	13.4	10.2
Degree of freedom	10	10
p-value ^a	0.201	0.426

^a Null hypothesis: residuals are uncorrelated

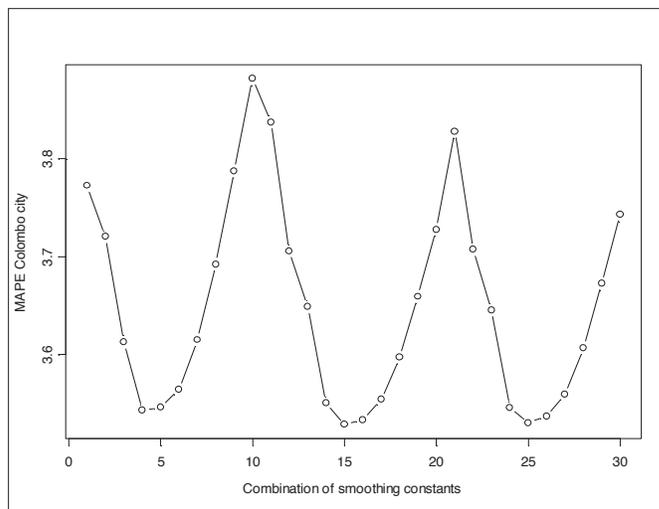
Table 6: Accuracy measures for identification and verification period for stochastic model

	MSD		r ^a	
	All Island	Colombo city	All Island	Colombo city
Identification period	122.330	10.602	0.992	0.953
Verification period	181.658	6.703	0.909	0.724

^a r - Correlation coefficient between observed and fitted values

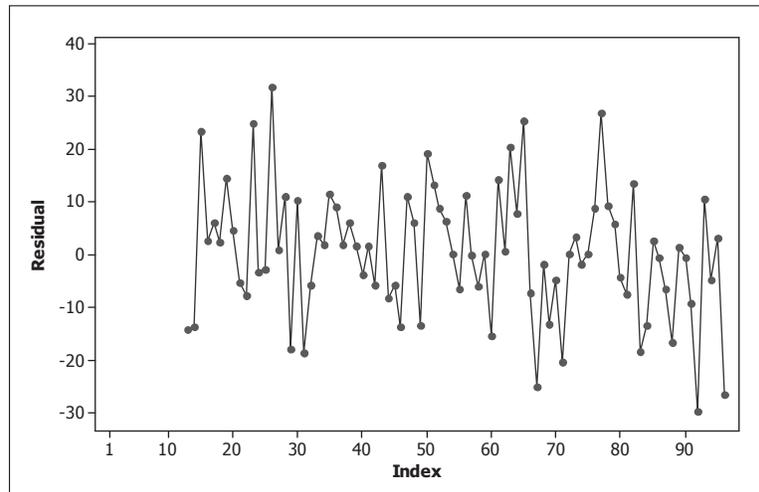


(a)

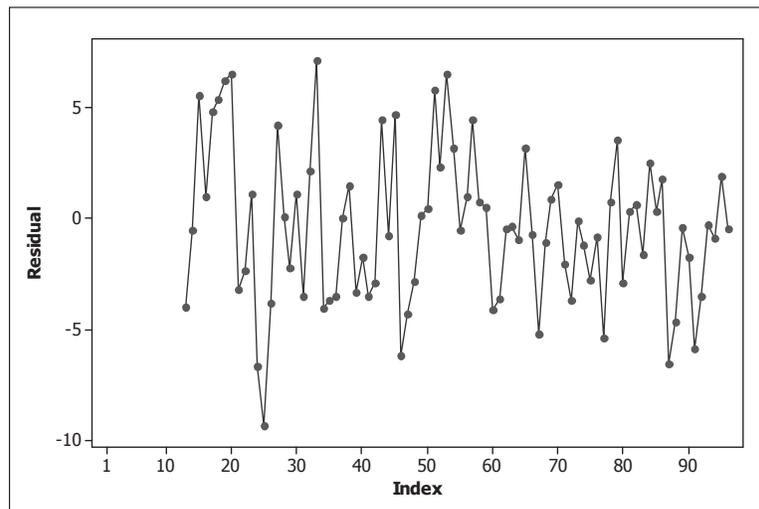


(b)

Figure 4: MAPE for different combinations of smoothing constants in (a) Sri Lanka model ; (b) Colombo city model



(a)



(b)

Figure 5: Time series plot of residuals of the exponential smoothing model
(a) Sri Lanka model; (b) Colombo city model

According to Figure 4, in the Sri Lanka model, the lowest MAPE was produced at $\alpha = 0.57$, $\beta = 0.03$ and $\gamma = 0.83$, and for the Colombo city model, it was at $\alpha = 0.05$, $\beta = 0.41$ and $\gamma = 0.59$, where α , β and γ represent the smoothing constant for the data, the smoothing constant for trend estimate and the smoothing constant for seasonality estimate, respectively.

Figure 5 exhibits that the residuals of the exponential smoothing models appear to be uncorrelated.

Monthly electricity sales forecasts were obtained for 2009 from the above exponential smoothing models. Those estimated values were compared with the actual figures in order to verify the models. For the verification of the model, MSD the correlation coefficient between the observed values and the fitted values for the identification period and the verification period were taken into account (Table 7).

Table 7: Accuracy measures for identification and verification periods for exponential smoothing model

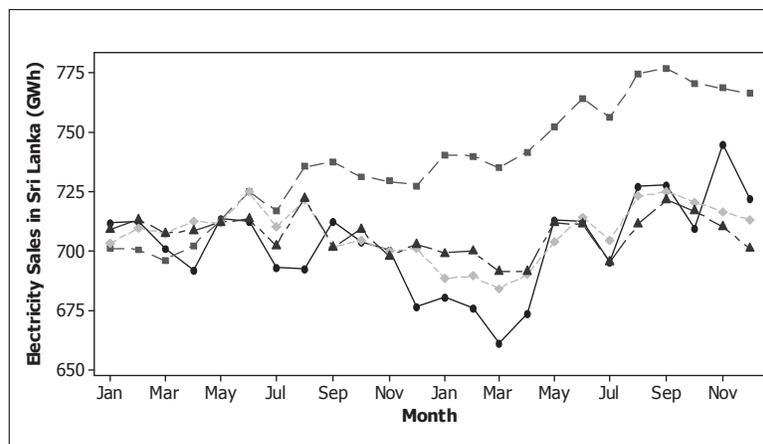
	MSD		r ^a	
	All island	Colombo city	All island	Colombo city
Identification period	151.717	11.972	0.990	0.945
Verification period	344.937	9.553	0.771	0.563

^a r - Correlation coefficient between observed and fitted values

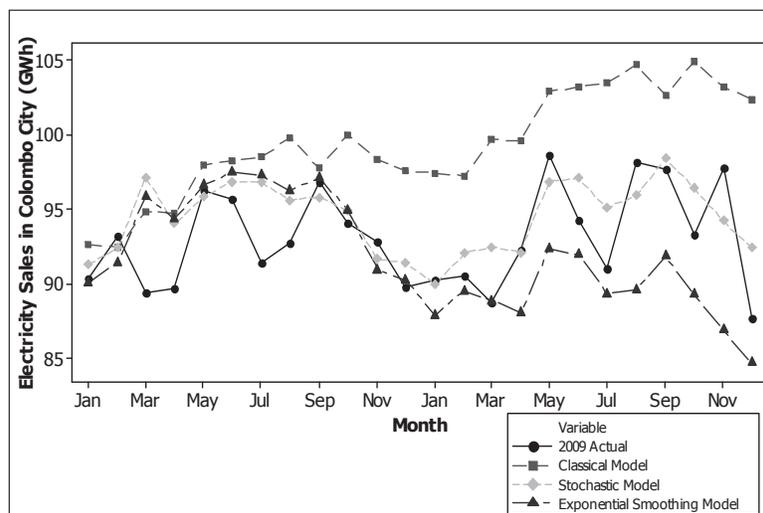
Table 8: Accuracy measures of the models for verification period

	MSD		r ^a	
	All island	Colombo city	All island	Colombo city
Classical method	186.476	10.050	0.906	0.547
Stochastic method	181.658	6.703	0.909	0.724
Exponential smoothing method	344.937	9.553	0.771	0.563

^a r - Correlation coefficient between observed and fitted values



(a)



(b)

Figure 6: Plot of predicted values and actual values for 2008 and 2009
 (a) Sri Lanka model; (b) Colombo city model

Comparison of models

The ultimate objective of this study was to identify the best forecasting models, which can be used for the purpose of forecasting monthly electricity sales in Sri Lanka and in the Colombo city. Models with the lowest MSD and the highest correlation coefficient between the observed values and the fitted values for verification period were selected as the best forecasting models.

According to Figure 6, the predicted electricity sales of the stochastic models for both Sri Lanka and the Colombo city were closer to the actual values than that of the classical model and the exponential smoothing model. The predicted values of the classical model for both Sri Lanka and the Colombo city data were slightly higher than the actual electricity sales data. The electricity sales continuously declined from September 2008 to March 2009 and the classical model was not capable of capturing this change in the trend direction. This may be the reason for the difference in the predictions of the classical model as exhibited in Figure 6.

Accuracy of the forecasts was checked with respect to the MSD and correlation coefficient between the observed values and the fitted values. According to Table 8, the lowest MSD and the highest correlation coefficient between the observed values and the fitted values for verification period was produced by the stochastic model for both Sri Lanka and the Colombo city. When considering the models for Sri Lanka data, the highest MSD was produced by the exponential smoothing model whereas for the Colombo city data the highest MSD was produced by the classical model. This indicates that more accurate forecasts for electricity sales can be obtained by the stochastic approach when compared with the classical decomposition approach and the exponential smoothing approach.

Hence, for the purpose of forecasting monthly electricity sales in Sri Lanka and in the Colombo city ARIMA (0,1,1) (0,1,1)₁₂ model was selected as the most appropriate model.

DISCUSSION

This study was focused on predicting electricity sales at the country-level and within the Colombo city. Three models, namely, the classical model, Winters' exponential smoothing model and the stochastic model were used for forecasting electricity sales on a monthly basis in Sri Lanka and in the Colombo city. Comparison of the forecasts made with these three models showed that the

forecasts generated from the stochastic model were more accurate than of the other two models. Therefore, it was concluded that the ARIMA (0,1,1) (0,1,1)₁₂ model is the most appropriate model to forecast electricity sales in Sri Lanka and in Colombo city. The corresponding mean squared deviations were 181.658 and 6.703 for Sri Lanka and for the Colombo city, respectively.

These conclusions are based on the monthly electricity sales data in Sri Lanka and in the Colombo city for the period 2001 to 2009 and the results may differ according to the study period.

According to the Ceylon Electricity Board the distribution of electricity is handled by four distribution divisions. Hence, the study can be further extended to forecast electricity sales in other provinces and in the distribution divisions as well.

In this study, only the classical approach, stochastic approach and the exponential smoothing approach were used. However, there are many other techniques, which can be used for the purpose of predicting electricity sales such as neural network models, data mining techniques, etc. An investigation of such techniques with a view to identifying methods for more accurate forecasts of electricity sales may be useful.

Acknowledgement

The authors appreciate the support given by Mr. L. Senaratne, Mr. K.G. Wijeratne, Ms. L.L. Hemamali and all the staff of the Statistical Unit of the Ceylon Electricity Board for providing the necessary data for this study.

REFERENCES

1. Ceylon Electricity Board (2005). *Annual Report and Accounts*, pp. 15 – 19, 26. Ceylon Electricity Board, Colombo 02.
2. Ceylon Electricity Board (2001 – 2007). Sales and generation. *Annual Reports*. Ceylon Electricity Board, Colombo 02.
3. Kareem Y.H. & Majeed A.R. (2006). Monthly peak-load demand forecasting for Sulaimany Governorate using SARIMA. *Transmission & Distribution Conference and Exposition: Latin America*, 15 – 18 August, pp. 1 – 6.
4. Makridakis S., Wheelwright S.C. & Hyndman R.J. (1998). *Forecasting Methods and Applications*, 3rd edition, pp. 111 – 387. John Wiley & Sons, Inc., USA.
5. Ruwanthi K.D.R. & Wickremasinghe W.N. (1999). Modelling sector-wise demand for electricity in Sri Lanka using a multivariate regression approach. *Journal of the National Science Foundation of Sri Lanka* 27(1): 55 – 64.

6. Samaranayaka D.N.R. (2008). Power crisis: is Upper Kotmale the answer? *Financial Times*. Available at <http://sundaytimes.lk/021208/ft/feature.html>, Accessed 17 January 2012.
7. Siyambalapatiya T. & Wickramasinghe A. (2009). *The Electricity Act of 2009 and the Development of the Sector*. Available at http://sanvada.org/index.php?option=com_content&task=view&id=87&Itemid=32, Accessed 17 January 2012.
8. Zuhaimy I. & Khairil A.M. (2005). SARIMA model to forecast Malaysian electricity generated. *Matematika* **21**(2): 143 – 152.