Forecasting Annual Tea Production in Sri Lanka

H.P.A.S.S. Kumarasinghe[*] and B.L. Peiris[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Sri Lanka

ABSTRACT: *Tea production plays a significant role in the Sri Lankan economy. This study focused on modelling and forecasting the total annual tea production in Sri Lanka and also the annual tea productions in Sri Lankan major tea growing areas; low grown, mid grown, and high grown. The annual total tea production data from 1964 to 2015 and the annual tea production data of major tea growing areas from 1970 to 2015 were collected from the central bank reports of Sri Lanka. Time series models were fitted by the Box and Jenkins ARIMA model approach. Series were tested for stationary by the Augmented Dickey Fuller test. Differencing techniques were applied to transform non stationary series to stationary series. Model diagnostics were performed by Ljung Box test and autocorrelation function of residuals. ARIMA (2, 2, 1) was the best fitting model for the total annual tea production and also the annual tea production in the low grown areas. ARIMA (1,2,1) was the best fitting model for the annual tea production in mid grown areas and ARIMA (2,1,0) was the best fitting model for the annual tea production in high grown areas. Forecasting of annual production for respective series was made up to the year 2020. An increment of 4.08% of total annual tea production in 2020 is predicted by the fitted model and it is also expected that production increments of 5.15%, 2.6% and 4.5% in low, mid and high grown areas, respectively, compared to the average production from 2011 to 2015.*

Keywords: *ARIMA, forecasting, tea production, time series, stationary*

INTRODUCTION

Contribution of the agriculture sector for Sri Lankan economy is very high and the tea cultivation within the agricultural sector plays a significant role in socio-economic status of people living in the island. Tea cultivation could be considered as a contributing sector for formulating and exerting over one million various employment opportunities (Thushara, 2015) with uplifting the economic status of the country by the assistance of huge export earnings. In 2015, the cultivated land extend of tea in Sri Lanka was 203,000 ha and it is the secondly most cultivated crop (land extend wise) after coconut among plantation crops. At the same time the highest share for the GDP from plantation sector was also by tea cultivation which was 0.9% in 2013 (Central Bank, 2013). The majority of the cultivated lands (both small scale and large scale) are in rural areas and the rural population get benefits from the tea cultivation.

---

[1] Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka
* Corresponding Author : arunasanjeew@gmail.com

There are three major tea growing regions; high grown, medium grown and low grown. These three categories are based on the elevation where low grown is up to 600 m from the sea level, mid grown is from 600 to 1200 m and high grown is above 1200 m (Sri Lanka Export Development Board, 2016). Each category produces significantly different volumes of tea annually.

The productivity as well as the production of tea changes over the time, and is affected by several factors, such as climatic change, pest and disease problems. The knowledge on past, present and future production data are necessarily required for decision making. The past and present data could be measured or obtained by records although the future data are uncertain. Specific statistical approaches would be applicable for fitting appropriate models and forecasting (predicting) future production after capturing the existing patterns of current data series. Box and Jenkins ARIMA model fitting approach is commonly used in time series and most probably, understanding of patterns in existing data series is the basis of forecasting (assuming that, past patterns remain in the future as well) in this approach.

There are certain studies based on time series analysis techniques that have been carried out in the Sri Lankan context for the purpose of making forecasts in agriculture sector. Katangodage and Wijeratne (2016) have studied the overall plantation industry based on time series approach and Sivapathasundaram and Bogahawatte (2012) studied the forecasting of paddy production in Sri Lanka. However, studies for forecasting the tea production in individual growing areas along with the national tea production has not been studied and forecasted.

Therefore, this study mainly focused on investigating the behaviour of the annual national production of tea in Sri Lanka with the annual production of tea in major growing area categories; high grown, mid grown and low grown.

METHODOLOGY

Data collection and preliminary analysis

For the application of ARIMA models, time series should be formulated based on time resolution of historical data (Amini *et al*., 2016). Hence, the secondary data on annual national tea production from 1964 to 2015 and regional tea production from 1970 to 2015 were collected from the Central Bank Reports of Sri Lanka.

Initial time series plots were used for the preliminary understanding of existing patterns; trends, seasonality and the irregularity of the data series.

Checking for stationary condition

Stationary condition is necessarily required to apply this technique on data (Abdullah, 2012). Stationary refers that the properties of the time series are not affected by the change in time origin (Montgomery *et al*., 2008). By checking the presence of unit root, the stationary condition could be checked (Lawgali, 2008). Therefore, in this study Augmented Dickey Fuller (ADF) test was applied as a unit root test. With certain limitations, time series plot is used in identifying the underlying stationary condition.

There are different transformation techniques based on the characteristics of the series for converting non-stationary series in to stationary condition. Differencing technique is one of the commonly used techniques. This difference operator is incorporated to allow for the homogeneous non-stationarity series (Berthouex and Box, 1996).

A difference of order one means that each value of the time series is subtracted from the immediate previous value: (Paretkar, 2008),

$$y'_t = y_t - y_{t-1} \quad \text{……………………………………………………………}(1)$$

Second or higher order differencing is the applying same procedure for the previous order differenced time series.

$$y''_t = y'_t - y'_{t-1} \quad \text{………………………………………………………}(2)$$

First order differencing was used as the first approach and higher order differencing was also applied when it was confirmed by the ADF test that first order did not fulfil the stationary series for collected data.

Model identification, selection and fitting

Model identification and selection is the most important and crucial step in Box and Jenkins ARIMA approach. Autocorrelation and Partial Autocorrelation Function correlograms (ACF and PACF) are the main pathways which are used to identify possible components; Autoregressive (AR) order and Moving Average (MA) order, in ARIMA (p, d, q) models for the stationary series. In here, p is AR order obtained from PACF correlogram and q is MA order obtained from ACF correlogram while d is the order of differencing.

For a time series y_t, $t \geq 1$, the autocorrelation coefficient at lag k is,

$$\rho_k = \frac{Cov\,(y_t, y_{t+k})}{Var(y_t)} \quad \text{………………………………………………………………}..(3)$$

The sample k^{th} order autocorrelation is,

$$r_k = \frac{\sum_{t=1}^{n-k}(z_t - \bar{z})\,(z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(z_t - \bar{z}))^2} \quad \text{………………………………………………………}(4)$$

And the sample partial autocorrelation coefficients can be computed as,

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1}(r_{k-1,j}\,r_{k-j})}{1 - \sum_{j=1}^{k-1} r_{k-1,j}\,r_j}, \quad (k = 2,3, \dots\dots.) \quad \text{…………………………}(5)$$

For a single time series, there may be two or more possible ARIMA models, where the, Akaike Information Criterion (AIC) was used for selecting a well fitted model. In this case, AIC value for each of the possible ARIMA model was calculated and the model with minimum AIC was selected.

Model diagnostic and accuracy checking

After selecting a model, residual analysis is much important in order to check the model adequacy with independently and identically distributed residuals. Use of statistical test such as Ljung Box test and examination of residual autocorrelation function correlogram are commonly used as the model diagnostics.

The Ljung-Box test is as follows: H_0; The sequence data are iid, H_a; The sequence data are not iid with the test statistic: (Cui, 2011),

$$\hat{Q}(\hat{r}) = n(n + 2) \sum_{k=1}^{m}(n - k)^{-1}r_k^2 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (6)$$

where $\hat{r}_k = \left.\sum_{l=k+1}^{n} \hat{a}_l \hat{a}_{l-k} \middle/ \sum_{l=1}^{n} \hat{a}_l^2 \right.$, the estimated autocorrelation at lag k,

n = sample size
m = number of lags being tested

Forecasting accuracy was measured as to Mean Absolute Percentage Error (MAPE) of forecast series.

$$MAPE = \frac{1}{n} \sum \frac{|Actual - Forecast|}{|Actual|} * 100 \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots (7)$$

RESULTS AND DISCUSSION

The summary statistics of the time series are given in Table 1.

Table 1. Descriptive statistics of the tea production data

Time Series	Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
National	178.90	211.30	227.50	250.00	303.70	340.20
Low Grown	53.20	66.35	114.20	121.10	172.00	210.00
Mid Grown	37.90	50.68	53.85	55.69	56.88	76.00
High Grown	53.70	73.68	76.75	76.80	80.38	87.00

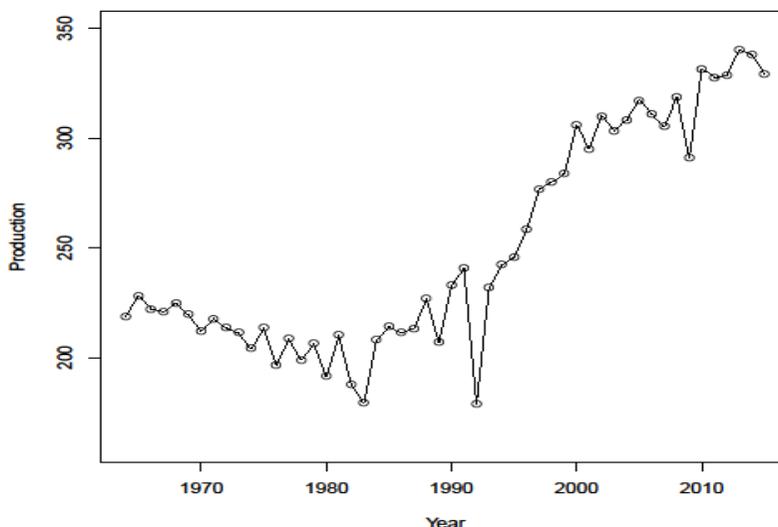The time series plot for annual total tea production is given in Figure 1.

Figure 1. Time series plot for national total tea production

According to the figure 1, there is no clear seasonal component seen in the series although a trend seems to be present. The other series of high grown, medium grown and low grown also showed the similar pattern as the national total tea production.

Any of the series was not confirmed with the stationary condition as per Augmented Dicky Fuller (ADF) test (not significant at 0.05 significance level). Therefore, differencing approach was applied for all the series as required and ADF test results for the series before and after applying the transformation are given in Table 2.

Table 2. Resulted Augmented Dicky Fuller (ADF) test statistics for first and second order differencing

Data Series	Resulted *P* Value for Original Data	Resulted *P* Value After First Order Differencing	Resulted *P* Value After Second Order Differencing
National	0.6456	0.3876	0.01[*]
Low Grown	0.5094	0.1852	0.01[*]
Mid Grown	0.5896	0.51	0.01022[*]
High Grown	0.4371	0.01823[*]	NA

*The P values marked with * are statistically significant at 0.05 level significance level.*

All the series except high grown were undertaken for the second order differencing as well but only first order differencing was adequate for high grown series.

The ACF and PACF correlograms were used for identifying possible models for all the four time series. The ACF and PACF correlograms for national total tea production are given in Figures 2 and 3.
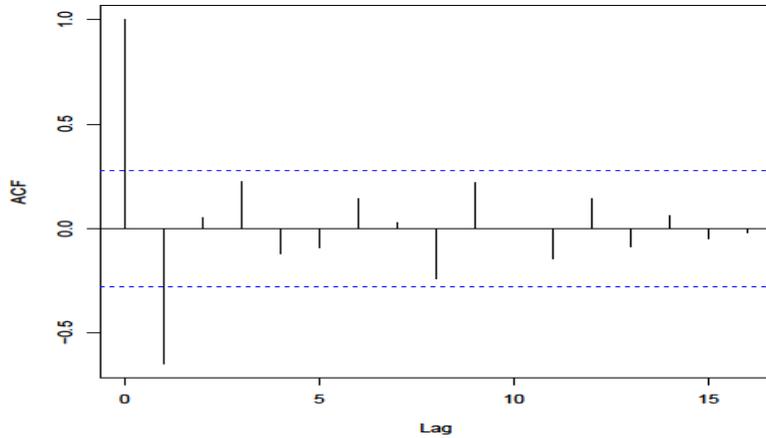
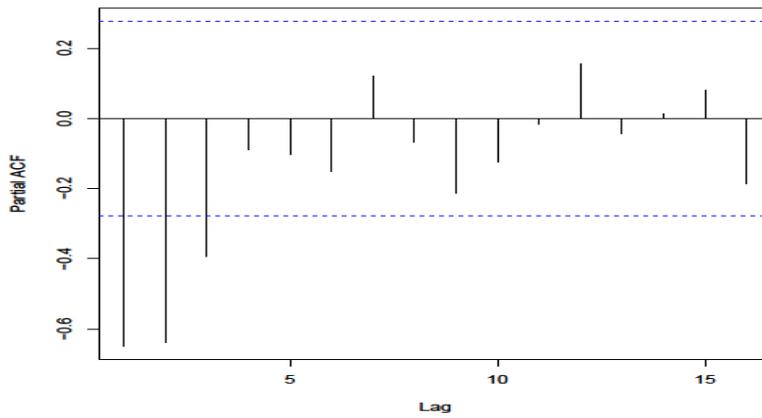**Figure 2. ACF for annual total tea production**

Figure 3. PACF for annual total tea production

According to the ACF and PACF correlograms of all the series possible best fitting models were identified and AIC values are given in Table 3.

Table 3.   Possible best fitting ARIMA models and respective AIC values for time series

Time Series	Possible Model	AIC
National	ARIMA (1,2,0)	404.64
	ARIMA (1,2,1)	377.68
	ARIMA (2,2,0)	382.73
	ARIMA (2,2,1)	372.41
	ARIMA (3,2,0)	375.26
	ARIMA (3,2,1)	374.33
Low Grown	ARIMA (1,2,0)	305.04
	ARIMA (1,2,1)	283.77
	ARIMA (2,2,0)	286.43
	ARIMA (2,2,1)	279.3
	ARIMA (3,2,0)	281.68
	ARIMA (3,2,1)	281.08

Table 3. Cont.

Mid Grown	ARIMA (1,2,0)	243.64
	ARIMA (1,2,1)	225.08
	ARIMA (2,2,0)	233.36
	ARIMA (2,2,1)	226.23
	ARIMA (3,2,0)	226.79
	ARIMA (3,2,1)	227.99
High Grown	ARIMA (1,1,0)	259.59
	ARIMA (1,1,1)	253.66
	ARIMA (2,1,0)	252.03
	ARIMA (2,1,1)	253.73

As to the minimum AIC, ARIMA (2,2,1) model was selected for national and low grown time series, while ARIMA (1,2,1) and ARIMA (2,1,0) models were selected for mid grown and high grown series, respectively. The model diagnostic checking was performed on each of the selected models.

For low grown and the national tea production series, a common model was well fitted indicating that there is a kind of common pattern in series. This similarity is more or less a result due to a significant portion of the national tea production which is shared by the low grown tea production. In 2015, the share of production by the low grown to the national production was 61.4% while it was 59.2% in 2010.

All the selected models confirmed that the independent and identical distribution of residuals with respect to the Ljung-Box test and ACF of residual. The plot of ACF of residuals of fitted model for national tea production series is given in Figure 4.
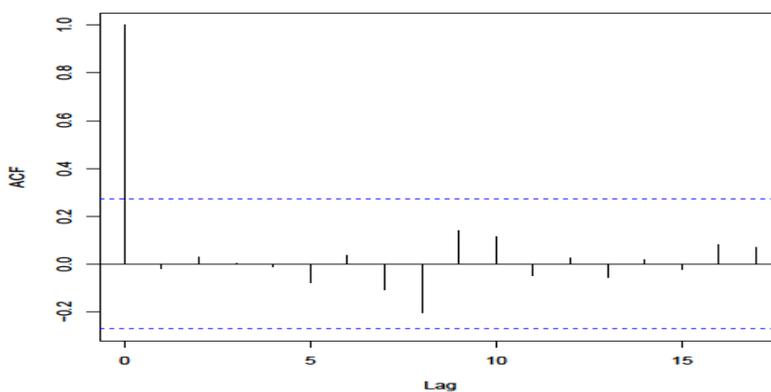

Figure 4. ACF of residuals of fitted model for national tea production series

Mean Absolute Percentage Error (MAPE) for all the forecast series are listed in Table 4.

Table 4.   Mean Absolute Percentage Error (MAPE) for selected models

Time Series	**MAPE %**
National	9.14
Low Grown	11.30
Mid Grown	17.21
High Grown	1.93

All the resulting MAPE values are less than 20% indicating that the selected models are accurate enough for the use of forecasting. Therefore, short term forecast up to 2020 is done by fitting the selected ARIMA models for each series.

The parameter estimates and their standard errors for the ARIMA (2,2,1) for national total tea production is given in Table 5.

Table 5. The parameter estimates and their standard errors for the best fitting model ARIMA (2,2,1) for national total tea production

	ar1	ar2	ma1
Coefficient	-0.7231	-0.4142	-0.8441
SE	0.1405	0.1369	0.1021

Forecast values for national total tea production is given in Table 6.

Table 6. Forecast values for national total tea production

Year	National Production (million Kg)	Standard Error (SE)
2016	342.9110	14.43552
2017	343.0709	15.72948
2018	343.6847	17.85081
2019	349.6662	21.79050
2020	351.5781	24.36733

By 2020, there will be 4.08% of increment in tea production at the national level in Sri Lanka compared to the average from 2011 to 2015. By looking at the basic time series plot also, this fact is confirmed by the clear positive trend starting from late 80's to the present.

By 2020, there will be 5.15% of increment in tea production of low grown region in Sri Lanka compared to the average from 2011 to 2015. Time series plot also confirmed that there was an increasing production (positive trend) resulting in the increment of tea production starting around 1990. The national tea production is comprised of a high proportion from the low grown tea production. So, this increment scenario of low grown region greatly affects the production at the national level and resulted in a positive trend. The increment of the production in mid grown area by 2020 is predicted as 2.6% compared to the average production from 2011 to 2015. By 2020, 4.5% of the production increment can be expected in high grown areas compared to the average production from 2011 to 2015.

Currently Sri Lanka is the second world exporter in the tea industry exporting 94% of the production in the country (Mujahid Hila, 2012).Therefore, it is important to study the future opportunities and constraints for the resilience of the tea industry in Sri Lanka. There have been some studies undertaken for fulfilling the above purpose.

Bordoloi (2012) has studied the trend of global tea production and exportation with reference to India. That study shows that there is a linear growth of tea production from India, Sri Lanka, Kenya, China and Vietnam from 1981 to 2010. Furthermore maximum growth was achieved by Vietnam followed by Kenya and China. The global tea production was more

than doubled during this time period. This study also indicated an increasing trend of the national tea production in Sri Lanka.

CONCLUSIONS

ARIMA approach had given four best fitted time series models; ARIMA (2,2,1) for the annual tea production in national total and low grown area, ARIMA (1,2,1) for mid grown area and ARIMA (2,1,0) for high grown area. Overall study resulted in short term forecasted values for four different data series. By the year 2020, an increment of tea production could be expected in all the three growing regions resulting in a production increment at national level as well. The percentage increments compared to the average of the preceding five years productions are 5.15%, 2.6% and 4.5% for low, mid and high grown regions, respectively and this will lead to 4.08% of total tea production increment in Sri Lanka by 2020 compared to the average production from 2011 to 2015.

REFERENCES

Abdullah, L. (2012). ARIMA model for gold bullion coin selling prices forecasting. Int. J. Adv. Appl. Sci., *1*, 153-158.

Amini, M.H., Kargarian, A. and Karabasoglu, O. (2016). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. Electr. Pow. Syst. Res., *140*, 378-390.

Annual report of the Central Bank of Sri Lanka (1964-2015). Colombo, Sri Lanka.

Berthouex, P.M. and Box, G.E. (1996). Time series model for forecasting wastewater treatment plant performance. Water Res., *30*, 1865-1875.

Cui, F. (2011). ARIMA models for bank failures: Prediction and comparison. UNLV Theses, Dissertations, Professional Papers, and Capstones, 1027.

Industry Capability Report (2016). Export Development Board, Sri Lanka.

Katangodage, B.T. and Wijeratne, A.W. (2016). Value-weighted price return index for plantation sector of Colombo stock exchange of Sri Lanka. International Journal of Agricultural Resources, Governance and Ecology (IJARGE), *12*, 27-52.

Lawgali, F.F. (2008). Forecasting water demand for agricultural, industrial and domestic use in Libya. International Review of Business Research Papers, *4*, 231-248.

Montgomery, D.C., Jennings, C.L. and Kulahci, M. (2008). Introduction to Time Series Analysis and Forecasting. John Wiley & Sons.

Mujahid Hilal, M.I. (2012). Export trend in global tea trade: A cross countries analysis with reference to Sri Lankan and Indian tea Industry. UMT 11th International Annual Symposium on Sustainability Science and Management 09th – 11th July 2012, Terengganu, Malaysia, 291-303.

Sivapathasundaram, V. and Bogahawatte, C. (2012). Forecasting of paddy production in Sri Lanka: A time series analysis using ARIMA model. Trop. Agric. Res., *24*, 21-30.

Thushara, S.C. (2015). Sri Lankan tea industry: prospects and challenges. Proceedings of the Second Middle East Conference on Global Business, Economics, Finance and Banking (ME15Dubai Conference), 22-24 May, 2015, Dubai-UAE. Paper ID: D533.