

## Determining the Best-Fit Model for Oil Palm Yield and Planted Area in Malaysia

Abas Norzaida<sup>1</sup>, Aliman Kamariah<sup>2</sup> and Halim Shafrina<sup>3</sup>

<sup>1,2</sup>UTM Razak School of Engineering and Advanced Technology  
Universiti Teknologi Malaysia, K. Lumpur, Malaysia

<sup>3</sup>Department of Science in Engineering, Kuliyyah of Engineering,  
International Islamic University of Malaysia, Selangor, Malaysia

\*zaida.kl@utm.my, Kemm2009@live.com, syafrinaabdulhalim@iium.edu.my

### Abstract

Accurate information on agricultural production pattern is critical for foresighted and informed planning. Mathematical models have been successfully used to analyse and forecast a variety of agricultural data. In this study, time series models are employed for analysing and predicting the production and total planted area of oil palm, a significant agricultural crop export of Malaysia. Forecasting models, namely Linear Trend Model, Double Exponential Smoothing and Auto Regressive Integrated Moving Average (ARIMA) models are individually fitted to the 1974-2016 data to determine suitable models for forecasting purposes. Performances of the models are evaluated and compared using mean absolute percentage error (MAPE) and mean absolute deviation (MAD). The overall results demonstrated that ARIMA models are the best goodness-of-fit for both production and total planted area, indicating that there exists significant autocorrelation. In particular, ARIMA (2,2,3) is the best model to represent total planted area. Forecasted values of ten years show total planted land area will be gradually increasing, with an estimated increase of 0.14% to 1.98% per annum. Meanwhile, ARIMA (0,1,1) has the lowest MAPE and MAD value, suggesting that the model is most appropriate to represent oil palm production. The forecasted values show a consistent increase of about 0.12% per annum for the next ten years. The predicted results in this study could be used by relevant parties such as policy makers to foresee ahead of time the future requirement of import/export of palm production in Malaysia and enable them to do better strategic planning. This would indirectly contribute towards strengthening the position of oil palm industry in Malaysia and the world.

**Keywords:** Oil Palm, Time Series, Double Exponential Smoothing, ARIMA

### 1. Introduction

Identifying reliable mathematical models for agricultural data is crucial as the output of these models could be used for foresighted and informed planning. In particular, models based on time series approach are often employed to investigate the pattern of past data and subsequently this knowledge could be used as a basis to forecast future values. Analysing and forecasting agricultural data helps in explaining and predicting changes in crops production, hence help planners and policy makers in government and corporations to make appropriate decisions concerning import-export policies, monitoring of crop productions, procurement issues and propositions of improved crop techniques for higher yield.

This study focuses on the modelling of oil palm production, which is a very valuable agricultural crop export of Malaysia. Fresh fruit bunch (FFB) of the oil

---

Received (February 24, 2017), Review Result (August 10, 2017), Accepted (September 3, 2017)

palm tree produce kernel palm oil and palm oil [1]. Palm oil accounts for 35% of the world's vegetable oil market [2]. The crop is versatile as it could be harvested throughout the year and uses 10 times less land than other main oil producing crops such as soya, rapeseed and sunflower. In Malaysia, palm oil industry play a significant role in the country's economy by creating jobs opportunities and generation of income from exports. The industry is also the fourth largest contributor to Malaysia's Gross National income (GNI), with 36.6% shares of agriculture subsectors [3]. Malaysia is also the second palm oil producer in the world with 17,306 thousand tonne production in 2014 [4]. Oil palm plantation occupies the largest agriculture area in Malaysia with 48.5% planted in Peninsular Malaysia, 28.0% in Sabah and 23.5% in Sarawak [5].

Since oil palm is an important crop in Malaysia, various studies have been conducted. Talib and Darawi [6] analysed 1970 to 1999 data to develop a national model of the Malaysian palm oil market. Study by Abdullah [7] showed that the palm oil production in Malaysia depended on the maturity of oil palm tree, total area planted, replanting and yield. Khin *et al.*, [8] used vector error correction method (VECM), Multivariate Autoregressive Moving Average (MARMA) and ARIMA to forecast monthly price of palm oil of 1980 to 2011 data. The results showed that the MARMA model is more efficient and palm oil price was predicted to decrease. There are also studies focusing on the price of crude palm oil (CPO) using ARIMA-GARCH [9]. Gan and Li [4] studied on Malaysia's palm oil position in the world market in 2035, with results showing domestic palm oil production is projected to rise by about 50%, about 26.6 million tons in 2035 while domestic demand of palm oil is expected to increase by more than 200% to 1.4 million tons in 2035.

Realising the importance of oil palm to the Malaysia economy, this study aims to develop models on oil palm 1) production and 2) total planted area. Forecasting is then conducted based on the appropriate model for a ten year period.

## 2. Materials and Model Development

In this study, models on oil palm production and planted area are separately fitted and compared in order to determine the most suitable model for forecasting purposes. The data used are oil palm production in the form of yield per hectare and total planted area, and they are obtained from Malaysian Palm Oil Board (MPOB) portal. For the analysis of oil palm production, yearly yield per hectare of FFB production data for 1974-2016 are used while for planted area, data from the same interval (1974-2016) are used. Prior to selecting appropriate models, data are first analysed for trends and seasonality, subsequent, models are proposed based on the analysis results. Three forecasting models, linear trend model, double exponential smoothing and Auto Regressive Integrated Moving Average (ARIMA) model are fitted to determine the best model. However, before analysis is done, data are first transformed to natural log to stabilize the variance of time series.

In order to identify the most suitable model, performance of the models is evaluated by calculating mean absolute percentage error (MAPE) and mean absolute deviation (MAD) for each model. The values are then compared of which model with the lowest value is the better model.

### 2.1. Linear Trend Model

Trend model is a special case of the regression model in which the independent variable is just the time index variable.

$$\hat{y}_t = \beta_0 + \beta_1 t + \varepsilon \quad (1)$$

Where  $\hat{y}_t =$  the forecast at present time  $t$  and  $\varepsilon_t$  the error term

## 2.2. Double Exponential Smoothing

In Exponential smoothing [9] technique, the exponentially weighted average of past observations is used as a basis to make forecasting. The size of the weight assigned to each observation depends on the arrangement of data with regards to time. The biggest weight is assigned to the present observation while less weight is assigned to an older data.

Double exponential smoothing technique smooths out the data when a trend is present. Exponential smoothing with a pattern works much like simple smoothing apart from that two components must be updated every period - level and trend. The level is a smoothed estimate of the value of the data at the end of every period. The trend is a smoothed gauge of the estimation of average growth at the end of every period. The specific formula for simple exponential smoothing is:

$$S_t = \alpha * y_t + (1 - \alpha) * (S_{t-1} + b_{t-1}) \quad 0 < \alpha < 1 \quad (1)$$

$$b_t = \gamma * (S_t - S_{t-1}) + (1 - \gamma) * b_{t-1} \quad 0 < \gamma < 1 \quad (2)$$

Note that the present value of the series is used to calculate its smoothed value replacement in double exponential smoothing. There are numerous techniques to choose the initial values for  $S_t$  and  $b_t$ .  $S_1$  is in general set to  $y_1$ .

Three recommendations for  $b_1$

$$b_1 = y_2 - y_1$$

$$b_1 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3$$

$$b_1 = (y_n - y_1)/(n - 1)$$

## 2.3. Auto Regressive Integrated Moving Average (ARIMA)

Auto regression is a multiple regression technique in which the independent variables are time lagged version of the time dependent variable

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_n Y_{t-n} + \delta \quad (9)$$

where  $\delta$  - the error term of the model,

$b_1, b_2, \dots$  - coefficient of AR terms

MA term is the number of terms that describe the persistence of a random shock from one observation to the next. A model with two moving average terms ( $q=2$ ) is one in which an observation depends on two preceding random shocks.

$$\hat{Y}_t = \varepsilon_t + \theta_t \varepsilon_{t-1} + \theta_{t-1} \varepsilon_{t-2} + \dots + \theta_n \varepsilon_{t-n} \quad (10)$$

Where  $\theta$  is the coefficient of MA terms and  $\varepsilon$  is the residual or error term

$$\text{Then, } \hat{y}_n = \delta + \sum_{i=1}^p b_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (11)$$

If the time series data is not stationary, the data are differenced in order to transform them to become stationary. This results in an “integrated”, denoted by ARIMA( $p, d, q$ ), where  $d$  is the order of differencing. There are three major steps to build ARIMA model; 1) Model identification 2) Parameter estimation and 3) Diagnostic checking. In model identification, the possible model could be moving average, autoregressive model or the mix of this model. Autocorrelation function (ACF) and partial autocorrelation (PACF) can be applied to make the first guess. Once the model is identified, the parameter need to be estimated, and in principle the parameter must be significant and generate lowest residual. A common method is to test the randomness of the residuals using Ljung–Box Statistics, and non-significant p-values indicate that the residuals are uncorrelated and the proposed model is suitable for fitting the historical data.

### 3. Results and Discussions

#### 3.1. Total Planted Area

Figure 1 shows that total planted area from 1974-2016 has an increasing and upward trend. Hence time series forecasting models namely Linear Trend model, Double Exponential Smoothing and ARIMA model are therefore fitted to determine the best model. The graph comparing fitted and observed value of each model is shown in Figure 2 (a) – (c) while the comparison using MAPE and MAD is given in Table 1. Visual comparison in Figure 2 indicates that the performance of both Double Exponential Smoothing and ARIMA is similar. However numerical comparison results show that ARIMA (2,2,3) has the lowest MAPE and MAD value. This indicates that that ARIMA (2,2,3) is the best model to represent total planted area of oil palm in Malaysia from 1974 -2016. The forecast for total planted area using ARIMA (2,2,3) model for the period 2018 to 2027 is given in table 2. Based on the forecasted values, the total of planted land area with oil palm is increasing gradually. Total planted area for 2016 is 5,737,985 hectare, hence an estimated increase of 1.24% in 2018.

**Table 1. Numerical Comparisons of both Models**

	<b>Linear Trend Model</b>	<b>Double Exponential Smoothing</b>	<b>ARIMA (2,2,3)</b>
MAPE	0.6622	0.0943	<b>0.0745</b>
MAD	0.0418	0.0060	<b>0.0022</b>

**Table 2. Forecast Values for Total Planted Area**

<b>Year</b>	<b>Forecast Arima (2,2,3) (Hectare)</b>	<b>Increase (%)</b>
2018	5809383	1.24
2019	5883151	1.27
2020	5999706	1.98
2021	6042268	0.71
2022	6118571	1.26
2023	6169643	0.83
2024	6203117	0.54
2025	6240940	0.61
2026	6254319	0.21
2027	6263109	0.14

#### 3.2. Oil Palm Production

The second part of the study analyses oil palm production, measured in terms of yield of FFB per hectare. The first step is to investigate the time series data 1974 -2016 (see Figure 3) Visual inspection gives a strong indication of non-stationarity and no trend pattern for the yield of FBB production. Thus we consider Double Exponential Smoothing and ARIMA model. The graph comparing fitted value against observed data of each model is shown in Figure 4 (a) – (b) and the comparison using MAPE and MAD is given in Table 3. Results show ARIMA (0,1,1) has the lowest MAPE and MAD value, indicating that ARIMA (0,1,1) is the best model to represent yield of FFB in Malaysia from 1974 -2016. Forecast (Table 4) of yield per hectare from 2018 to 2027 show a gradual increase. The 2016 yield per hectare is 15.91 tonnes,

hence it is predicted that an increase of 18.22% in 2018 and a more gradual increase until 2027.

**Table 3. Numerical Comparisons of Both Models**

	Double Exponential Smoothing	ARIMA(0,1,1)
MAPE	1.68	1.57
MAD	0.02	0.0023

**Table 4. Forecast Values for Yield**

Year	Forecast Arima (0,1,1) (Tonne/Hectare)	Production Increase (%)
2018	18.81	18.22
2019	18.83	0.10
2020	18.85	0.11
2021	18.87	0.11
2022	18.89	0.11
2023	18.91	0.11
2024	18.93	0.11
2025	18.95	0.11
2026	18.97	0.11
2027	19.00	0.19

#### 4. Conclusion

The palm oil industry in Malaysia has a profound contributions towards the economy, hence studies related to oil palm would contribute towards strengthening the position of the industry in Malaysia and in the world. In this study, time series forecasting method is well implemented for forecasting yearly production of oil palm yield and planted area. In both cases, ARIMA models are found to be the best goodness-of-fit to historical data, indicating that there is significant autocorrelation. For total planted area, ARIMA (2,2,3) is the best model to represent the 1974 -2016 data. Forecasted values of ten years show total planted land area will be gradually increasing. It is estimated the increase is between 0.14% to 1.98%.

Again, for FFB oil palm yield, ARIMA (0,1,1) is found to be the best model to represent time series data 1974 -2016. The forecasts for period 2018 to 2027 show an increase in production. It is estimated an increase of 18.22% in 2018 and later, a more steady increase until 2027. The predicted result or forecast value could be used by relevant parties such as policy makers to foresee ahead of time the future requirement of import/export of palm production in Malaysia and enable them to make better strategic planning.

## Appendix

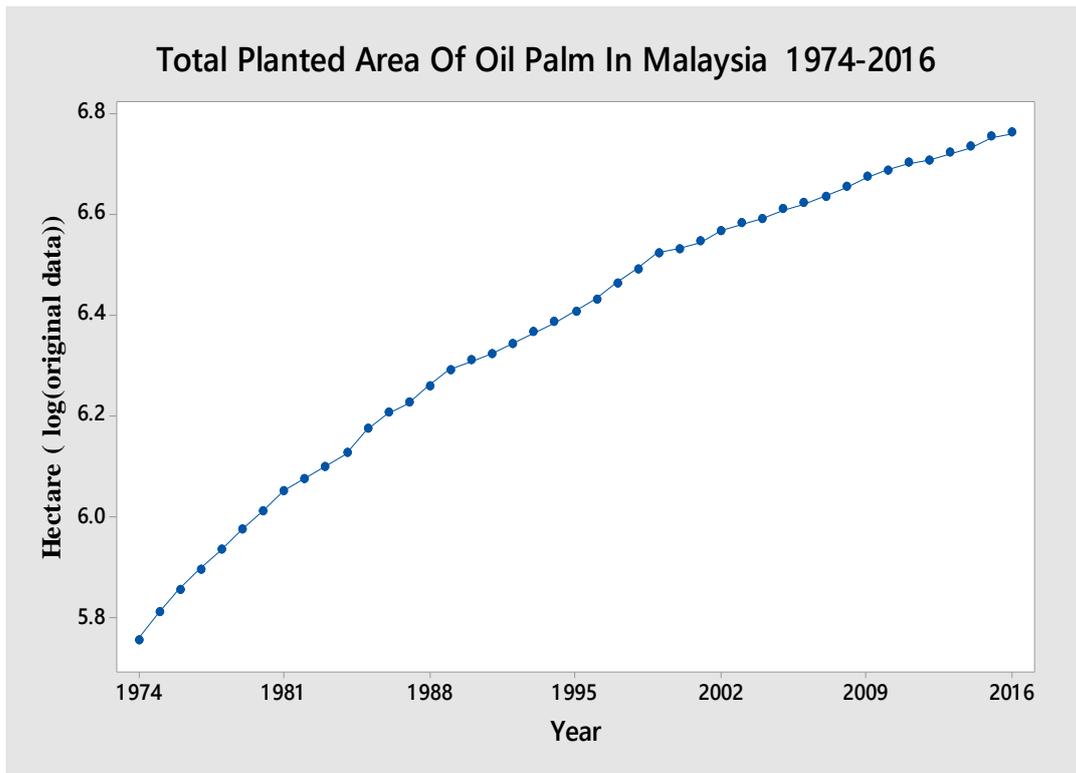
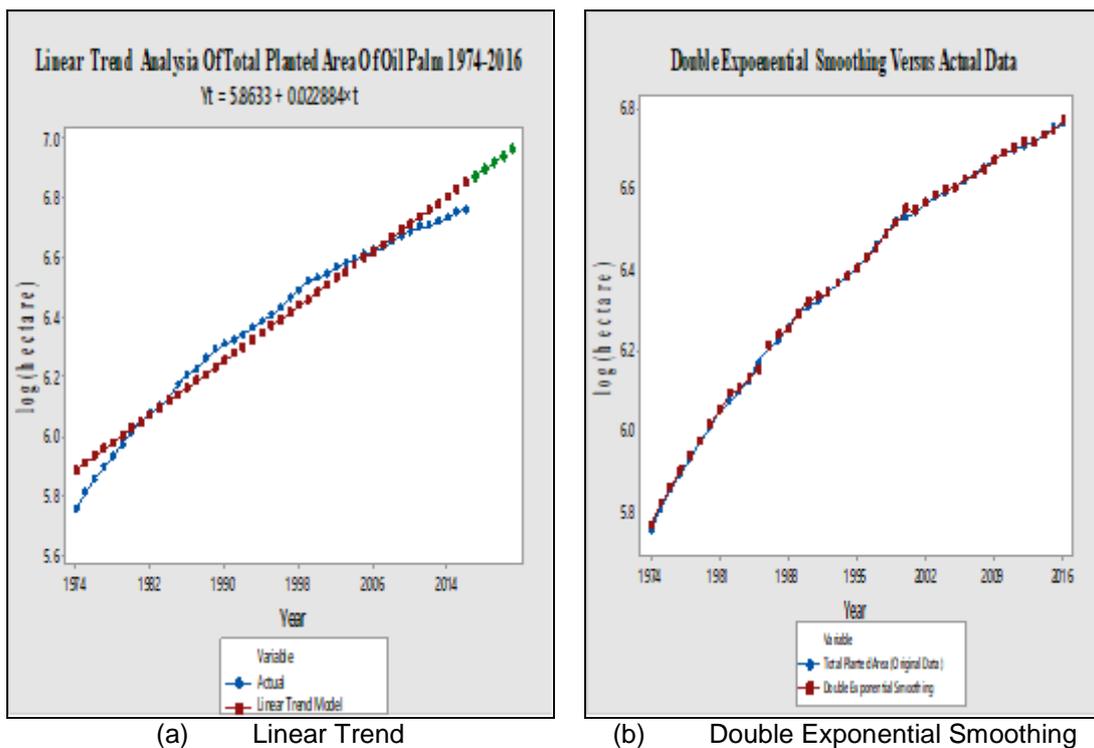
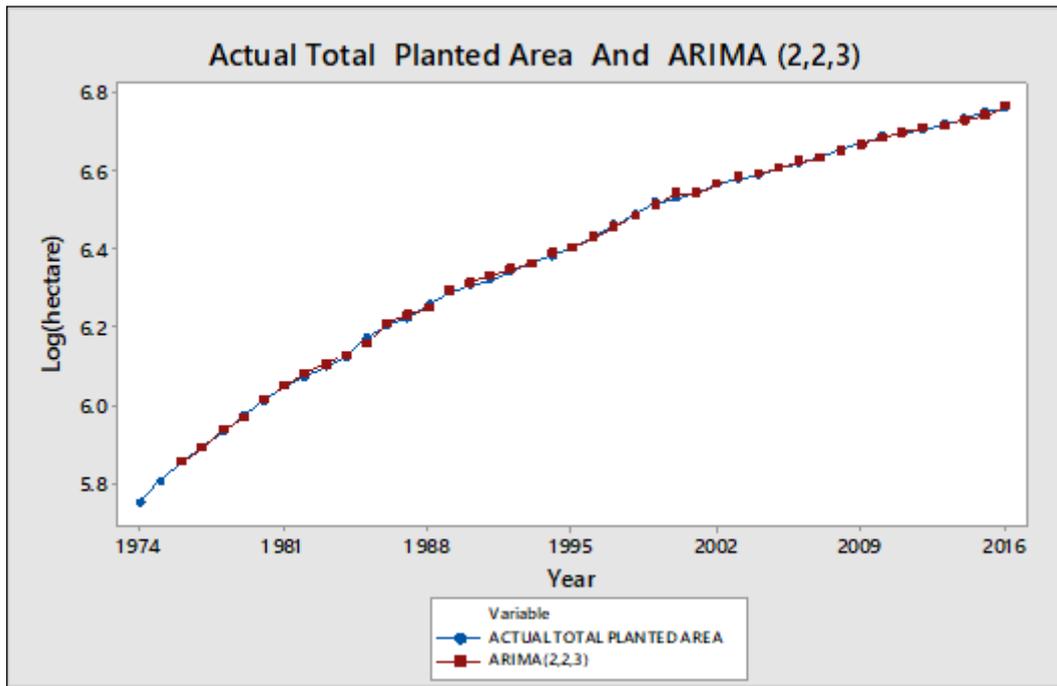


Figure 1. Total Planted Area of Oil Palm from 1974-2016





(c) ARIMA (2,2,3)

Figure 2. Graph of Observed and Fitted Values of Each Model for Total Planted Area

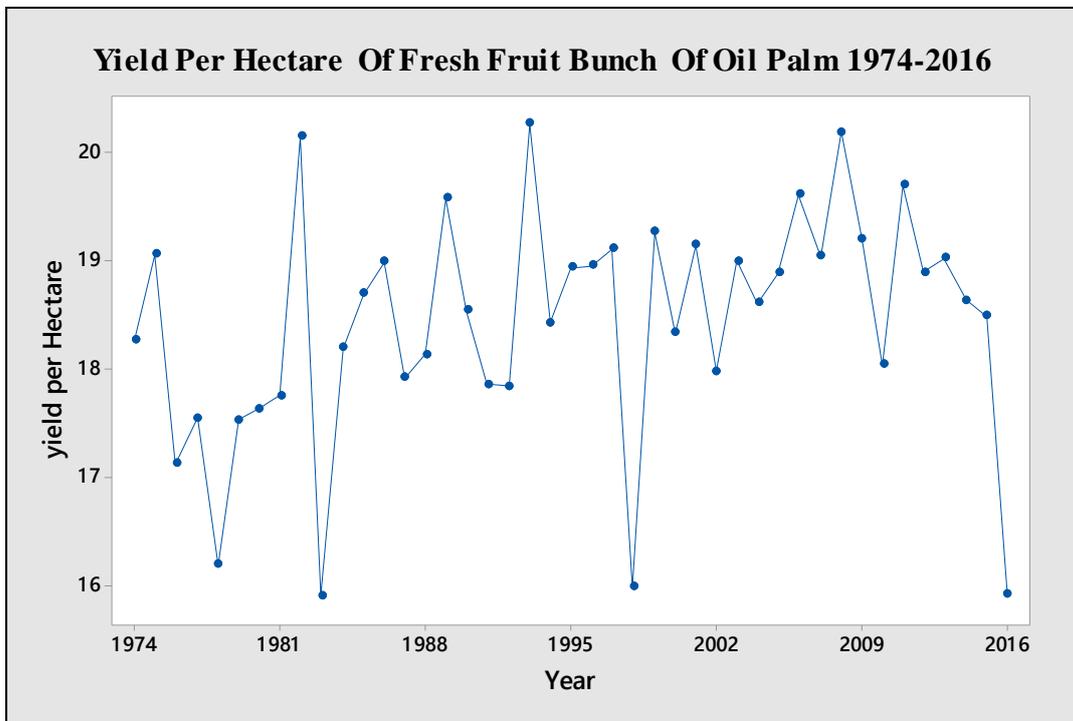
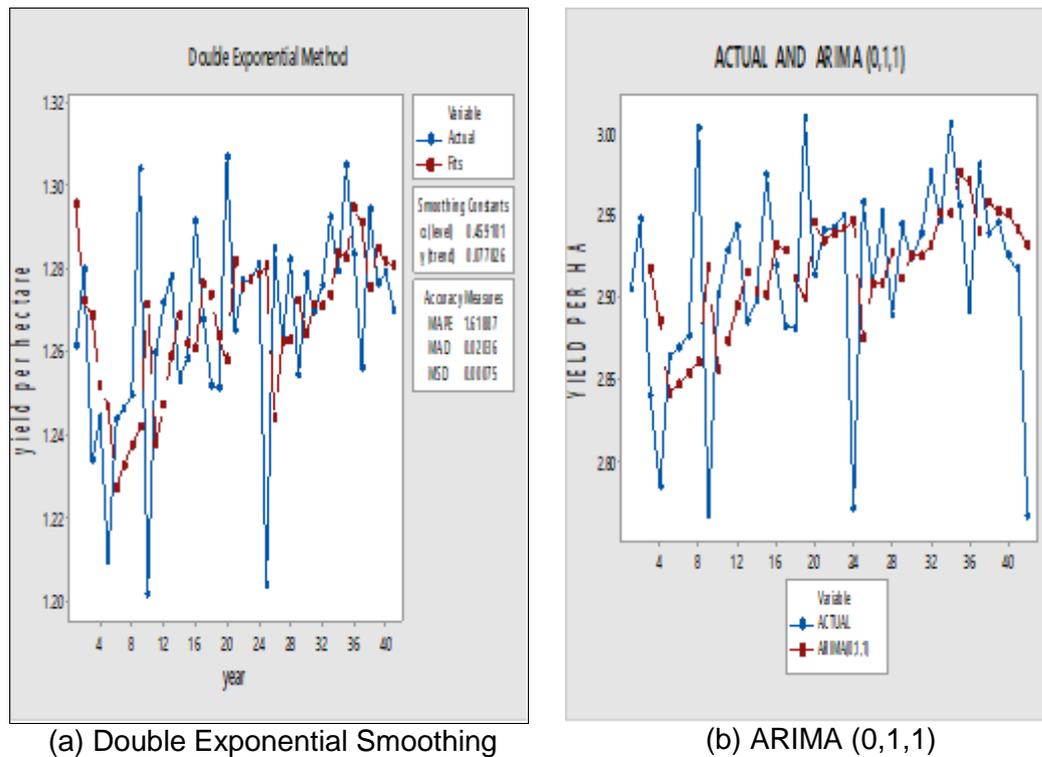


Figure 3. Oil Palm Yield per Hectare from 1974-2016



(a) Double Exponential Smoothing (b) ARIMA (0,1,1)  
**Figure 4. Graph of Observed and Fitted Values of Each Model for Yield per Hectare**

### Acknowledgements

We would like to thank Malaysian Palm Oil Board for the data used in this study. We also thank University Teknologi Malaysia for the support and funding of the study with Vote: Q.K130000.2540.15H74.

### References

- [1] Malaysian Palm Oil Industry, [www.palmoilworld.org/about\\_malaysian-industry.html](http://www.palmoilworld.org/about_malaysian-industry.html) (retrieved 2 March 2017)
- [2] Green Palm Sustainability, <http://www.greenpalm.org/> (retrieved 25 January 2017)
- [3] Palm Oil and Rubber, In: M.O.P.A. Industry, ed. ETP 2014. Malaysia: Ministry of Plantation and Industry, (2014), pp. 96-113.
- [4] G. P. Yan and L. Z. Dong, "A Study on Malaysia's Palm Oil Position in the World Market 2035", Ieej, Issue, (2012).
- [5] MPOB, 2016. Palm Oil, Malaysia: Department Of Statistics Malaysia.
- [6] B.A. Talib, and Z. Darawi, "An Economic Analysis of the Malaysian Palm Oil Market", Faculty of Economics, Universiti Kebangsaan Malaysia, Oil Palm Industry Economics J. vol. 2, no. 1, (2002), pp. 19-27.
- [7] Abdullah, Ramli . Short-Term and Long-Term Projection of Malaysian Palm Oil Production Malaysian Palm Oil Board, Malaysia (2003), 3(1): 32-36.
- [8] A. A. Khin, Z. Mohamed, C. Agamudai N.Malarvizhi and S. Thambiah, "Price Forecasting Methodology of the Malaysian Palm Oil Market", The International Journal of Applied Economics and Finance, vol. 7, (2013), pp. 23-36.
- [9] M. H. Ahmad, P. Y. Ping and N. Mahamed, "Volatility modelling and forecasting of Malaysian crude palm oil prices", Applied Mathematical Sciences, 6159-6169. DOI: 10.12988/ams.2014.48650, vol. 8, no. 121-124, (2014).

## **Authors**

**Norzaida Abas**, he is a Senior Lecturer at University Teknologi Malaysia in Mathematics and Statistics.

### Education

Degree : York University, Toronto, Canada

Masters : University of Wales, Cardiff, UK

Phd : University Teknologi Malaysia, Malaysia

### Area of Interest

Statistical Modeling

Modeling and Forecasting in Hydrology

Extreme Rainfall Analysis

Statistical Downscaling with respect to Climate Change

