

Application of ARIMAX Model to Forecast Weekly Cocoa Black Pod Disease Incidence

Ling, A. S. C.^{1,*}, Darnesah, G.², Chong, K. P.², Ho, C. M.²

¹Malaysian Cocoa Board, Wisma SEDCO, Locked Bag 211, 88999 Kota Kinabalu, Sabah, Malaysia

²Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Received July 1, 2019; Revised September 8, 2019; Accepted September 23, 2019

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The losses caused by cocoa black pod disease around the world exceeded \$400 million due to inaccurate forecasting of cocoa black pod disease incidence which leads to inappropriate spraying timing. The weekly cocoa black pod disease incidence is affected by external factors, such as climatic variables. In order to overcome this inaccuracy of spraying timing, the forecasting disease incidence should consider the influencing external factors such as temperature, rainfall and relative humidity. The objective of this study is to develop a Autoregressive Integrated Moving Average with external variables (ARIMAX) model which tries to account the effects due to the climatic influencing factors, to forecast the weekly cocoa black pod disease incidence. With respect to performance measures, it is found that the proposed ARIMAX model improves the traditional Autoregressive Integrated Moving Average (ARIMA) model. The results of this forecasting can provide benefits especially for the development of decision support system in determine the right timing of action to be taken in controlling the cocoa black pod disease.

Keywords ARIMAX, Black Pod, Climate, Cocoa, Forecasting

1. Introduction

Black pod or *Phytophthora* pod rot is the most economically important and widespread disease of cocoa, *Theobroma cacao* L. in Malaysia. The losses due to *Phytophthora* exceed \$400 million worldwide [1]. Among the *Phytophthora* species which attacked the cocoa, *Phytophthora palmivora* is the most widely distributed in the world and it caused global yield loss of 20-30% and tree deaths of 10% annually [1]. The intensity of black pod disease in cocoa was influenced by numerous climatic

parameters such as rainfall, temperature and high humidity as reported by Thorold [2,3], Dakwa [4], Wood [5] and Mpika [6]. Both rainfall and relative humidity has strong correlation which encourages pathogen sporulation by reproduction of zoospores to infect the cocoa pods while the optimum temperature is propitious to black pod symptom development [2-6]. It is important to quantify the black pod disease fluctuations due to the real effect of climatic parameters. Understanding the effect of the climatic variables on the cocoa black pod incidence can identifying the suitable management options such as fungicide spraying and culture practices to control the disease incidence under projected climate change scenario. The time-fractional partial differential equation was widely used in study the mathematical biology including the description of plant disease epidemic [7] but need the numerical approach to solve the equation. Another approach was developing the time series model especially auto-regressive moving average (ARIMA) described by Box and Jenkins [8] that widely used in forecasting where it used the historical sequences of observations to do the forecasting. In agriculture, ARIMA model used to forecast the annual production of several crop in countries that relied the crop for daily life or economy, for example production of rice [9], wheat [10], coffee [11] and cocoa [12].

As for the disease monitoring program, ARIMA model was used to predict *Botrytis cinerea* spore concentrations that caused grey mould in Spain and assist in deciding number of treatments needed [13]. In cocoa, it is very useful in forecasting the cocoa black pod incidence to understand the effect of previous incidence on the current incidence. However, ARIMA model along can't quantify the effect of climate variables on the cocoa black pod disease incidence and help in decision making process. The key problem is how to incorporate the climate information into the forecasting process and subsequently into the decision making process. When an ARIMA model

includes other time series as input variables, the model is referred to as an ARIMAX [14]. The application of ARIMAX model has been used in agriculture, economics and engineering as ARIMAX model provides better forecast performance than ARIMA model [15-17].

Therefore, the present study was undertaken with the following objectives (i) Developing the univariate ARIMA cocoa black pod disease forecasting models and (ii) Fitting ARIMAX (climate parameters as regressors) models and testing the validity of the developed models.

2. Methodology

2.1. Theory of Methods

2.1.1. ARIMA Models

ARIMA models are also known as Box-Jenkins models which require historical time series data of underlying variables. There are three stages involved in time series approach, namely process of model identification, parameter estimation and model checking. In the stage of model identification, the data series is determined whether the series is stationary before the Box-Jenkins model or ARIMA model is developed. A stationary series in Box-Jenkins model will have constant mean, constant variance and constant autocorrelation. For a nonstationary series, a differencing on the non-stationary series one or more times will be carried out to achieve stationary. For example, when we applied a first difference $\Delta y_t = y_t - y_{t-1}$ on non-stationary series to achieved stationary, so y is called as integrated of order 1 or $y \sim (1)$.

2.1.2. Test of Stationary Series of Data

In order to test for non-stationary, the Augmented Dickey-Fuller (ADF) test [18] is used where it test for a unit root in a time series sample. Given

$$y_t = a + by_{t-1} + \mu_t \quad (1)$$

where μ_t would be a non-stationary random walk if $b = 1$. To test whether y has a unit root or non-stationary, we regress

$$\Delta y_t = a + cy_{t-1} + \mu_t \quad (2)$$

where $c = b-1$ and test hypothesis that $c = 0$ against $c < 0$. For non-stationary series data, the differencing function will be applied to the data sequence to transform into stationary series.

2.1.3. Model Identification

When the stationarity has been addressed, then we need to identify the order (the p and q) of the autoregressive and moving average terms. The tools used are the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The formula for both ACF and PACF are available in most of time series book

[8]. The protocol used in identifying what autoregressive or moving average models terms are given below;

- An ACF with steadily decline and a PACF will cut off suddenly after p lags indicates an autoregressive process at lag p , AR(p).
- An ACF cuts off suddenly after q lags and a PACF with steadily decline indicates a moving average process at lag q , MA(q).

The ACF and the PACF both exhibiting large spikes that gradually die out indicates that there are both autoregressive and moving averages processes.

2.1.3. Estimation of Model Parameters

Given a tentative ARIMA (p, d, q) model as follows;

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (3)$$

which can also be written as

$$\alpha(L)Y_t = \beta(L)\varepsilon_t \quad (4)$$

where L is the lag operator, $LY_t = Y_{t-1}$ and $L^2 Y_t = Y_{t-2}$

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p \quad (5)$$

$$\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q \quad (6)$$

The degree of homogeneity d is determined in the identification, and Y_t is a dependent variable representing stationary series. This stage is to estimate all unknown population parameters, which are $\alpha_1, \alpha_2, \alpha_p$, and $\beta_1, \beta_2, \dots, \beta_q$ in equation (3). The least squares method is used in estimating the model parameters of $\alpha_1, \alpha_2, \dots, \alpha_p$, and $\beta_1, \beta_2, \dots, \beta_q$. The estimation procedure used in the least squares method involved minimizing the sum of squared residuals,

$$\sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (\varepsilon_t - \beta(L)^{-1} \alpha(L)Y_t)^2 \quad \text{where } \varepsilon_t =$$

2.1.4. Diagnostic Checking

Diagnostic checking in time series model is similar to the regression analysis which included testing the parameters and residual tests. Parameters testing by using the t -test is to check and retain only those estimate parameters $\hat{\alpha}(L)$, $\hat{\alpha}(L)$ and $\hat{\beta}(L)$ whose t -ratios are significantly greater than a predetermined critical value (that is, $|t| > 2$ at 5% significance level). Then, the residual tests are carried out using the Akaike Information Criterion (AIC) test and the Ljung-Box test or also known as Q statistics.

The AIC test formula is calculated as:

$$AIC(p, q) = n \log(RSS/n) + 2(p + q) \quad (7)$$

where n is the number of data points (observations) and RSS is the residual sums of squares. The model with smaller AIC values is considered to be the best model.

Meanwhile the Ljung-Box test the magnitudes of the

residuals autocorrelation for significance:

H₀: The data are independently distributed with no serial correlation

H_a: The data are not independently distributed which they exhibit serial correlation.

The test statistic is:

$$Q = n(n+2) \sum_{k=1}^n \frac{\hat{\rho}_k^2}{n-k} \tag{8}$$

where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k , and h is the number of lags being tested. Under H_0 , the statistic Q follows a chi-square distribution, $\chi^2(k-p-q)$. For significance level α , the critical region for rejection of the hypothesis of randomness is $Q > \chi^2(\alpha, k-p-q)$.

2.1.5. ARIMAX Model

An ARIMAX model is viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. Autoregressive terms for a dependent variable are merely lagged values of that dependent variable that have a statistically significant relationship with its most recent value. Moving average terms is nothing more than residuals (i.e., lagged errors) resulting from previously made estimates. The general ARIMAX models are as follows:

Autoregressive model with exogenous variables (ARX):

$$y_t = \phi(L)y_t + \beta x_t + \varepsilon_t \tag{9}$$

Moving average model with exogenous variables (MAX):

$$y_t = \beta x_t + \theta(L)\varepsilon_t \tag{10}$$

Autoregressive Moving Average with exogenous variables model (ARMAX):

$$\phi(L)y_t = \beta x_t + \theta(L)\varepsilon_t \tag{11}$$

where x_t represents exogenous variables, β their coefficients, $\phi(L)y_t$ is an AR model ($\phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p}$) and $\theta(L)\varepsilon_t$ is the MA model ($\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \dots + \theta_q \varepsilon_{t-q}$).

The approach used in ARIMAX model was based on the steps described by Andrews [19] with flowchart shown in Fig. 1. The approach involved two phases where the first phase deals with a linear regression model and the second phase deals with integration of AR and MA terms into a multiple regression model. In the first phase, the linear regression is used to identify the exogenous variables that are significant. Meanwhile, the second phase deals with iterative searching process to search the order of ARIMA part of the model. When there is large number of exogenous variables to be screened, the stepwise regression process is used in selecting and introducing a new exogenous variable into the ARIMAX model. This is followed by the iteration process of finding newly AR or MA or both terms to re-establish the random pattern of residuals in the ARIMAX models.

2.2. Data Collection and Statistical Analysis

The study is conducted in the cocoa mature areas in the Cocoa Research and Development Madai, Sabah with the coordinates of latitude and longitude as 4° 47' 10" N and 117° 57' 54" E. The plot's size of 1 hectare was planted with cocoa clone of PBC 123. The study plot was divided into two treatments, i.e. first treatment was the combination of fungicide, pruning and phytosanitary of diseased pods removed from the field while second treatment was the combination between fungicide and phytosanitary of diseased pods removed from the field. Each treatment has three replications with 30 trees per replication. The random sampling method was used to select 10 trees in each replication for the black pod incidence studied meanwhile the assessment of the black pod incidence was adapted from Ndoumbe-Nkeng [20] and Ngoh Dooh [21] as follows:

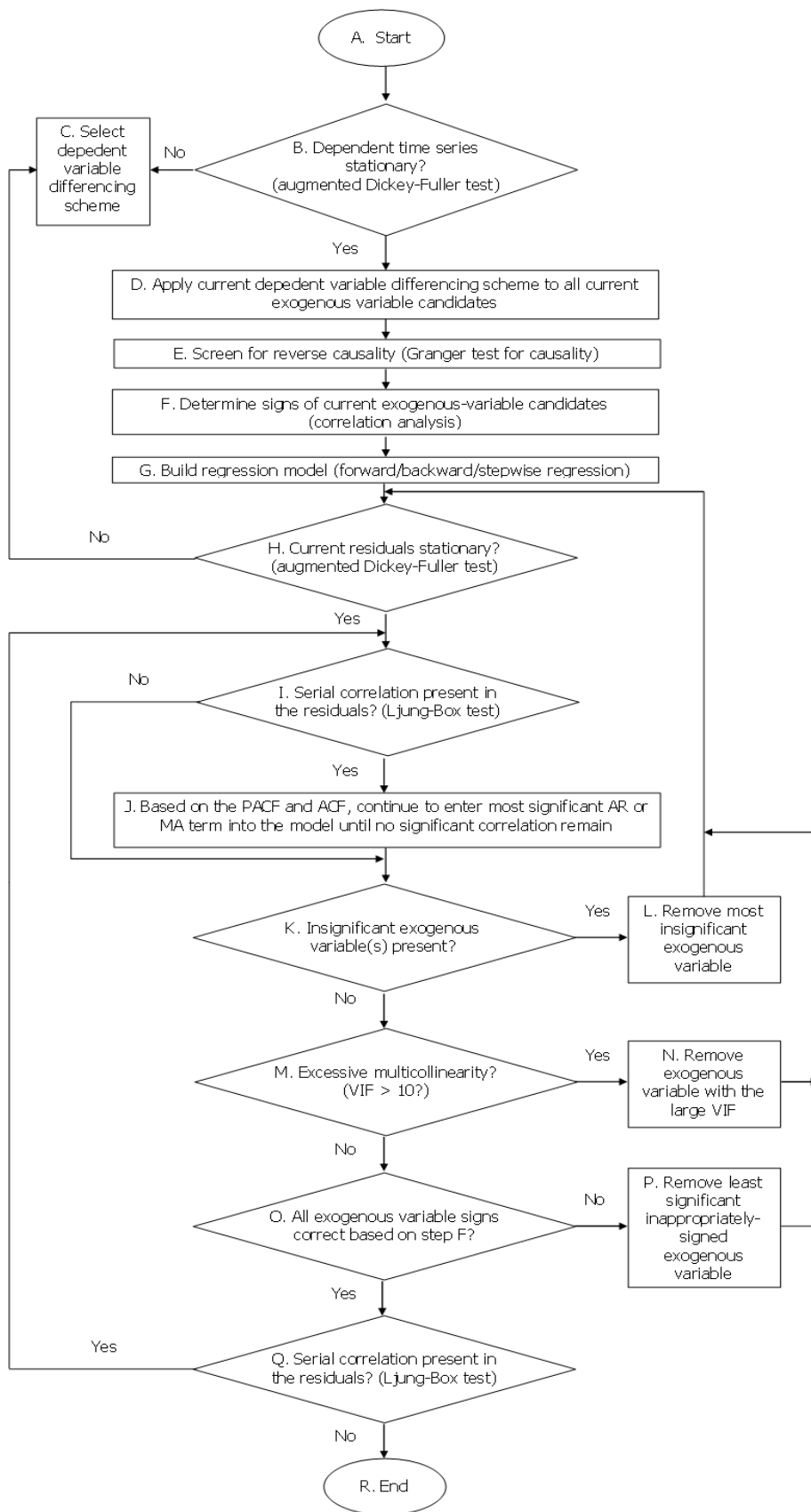


Figure 1. ARIMAX model building procedure

Black pod rate in treatment n (BPTRT n) was calculated each week i ,

$$BPTRTn_i = \frac{BP_i * 100}{(YP_i + AP_i)} (BP_i \times 100) / (YP_i \frac{BP_i * 100}{(YP_i + AP_i)} + AP_i) \tag{12}$$

$(BP_i \times 100) / (YP_i + AP_i) BP_i * 100$ where BP_i is the number of black pods observed over 30 selected trees in week i , YP_i is the number of young pods over 10 selected trees in week i and AP_i is the number of adult pods over 30 selected trees in week i .

The daily weather variables (rainfall, relative humidity and temperature) and the weekly black pod incidence data from the period of February 2015 to June 2016 collected. The weather data were recorded using WatchDog data loggers (1000 Series Micro Station, Spectrum Technologies, Aurora, IL, USA). The statistical tool used in this paper is EViews version 6.0 to analyse the time series data and building the ARIMA and ARIMAX model. The forecasting performance of the best-fitting between ARIMA and ARIMAX models as plant disease forecasting model are compared using three criteria, namely mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE). The dataset for model

validation purpose is 23 observations (or weeks) that covered observations from July 2016 to December 2016.

3. Results and Discussion

3.1. Descriptive Statistics

From Fig. 1 to 4, the black pod incidence and weather data have shown volatile pattern, showing ups and downs over a period of time except rainfall shows less volatile pattern as less rain occurred during that period. The high black pod incidence is recorded to be high especially during the rainfall and relative humidity were high during earlier days which is known as lag effect. The high relative humidity was also associated with a high frequency of rainfall often favorable for the occurrence of the disease as the presence of moisture on cocoa trees from rain or dew provided ideal conditions for *P. palmivora* infection [4]. The lag effect of the climate change could be due to the incubation period which is the period from fruit infection to the first disease symptoms [22].

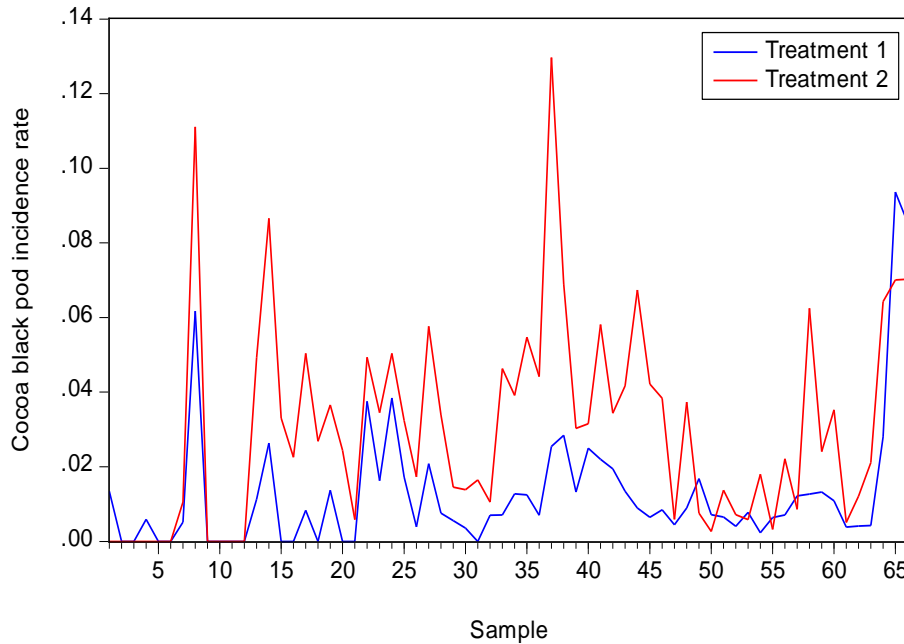


Figure 1. The black pod disease incidence

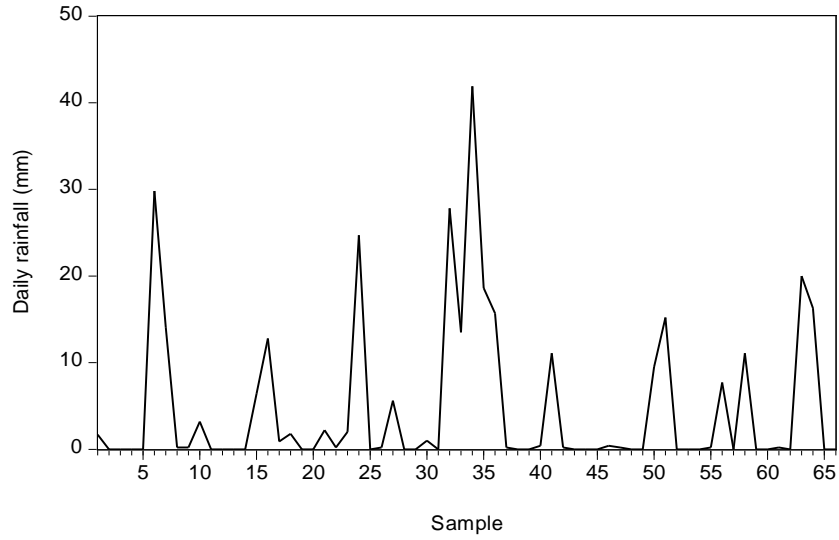


Figure 2. The daily rainfall (mm) recorded

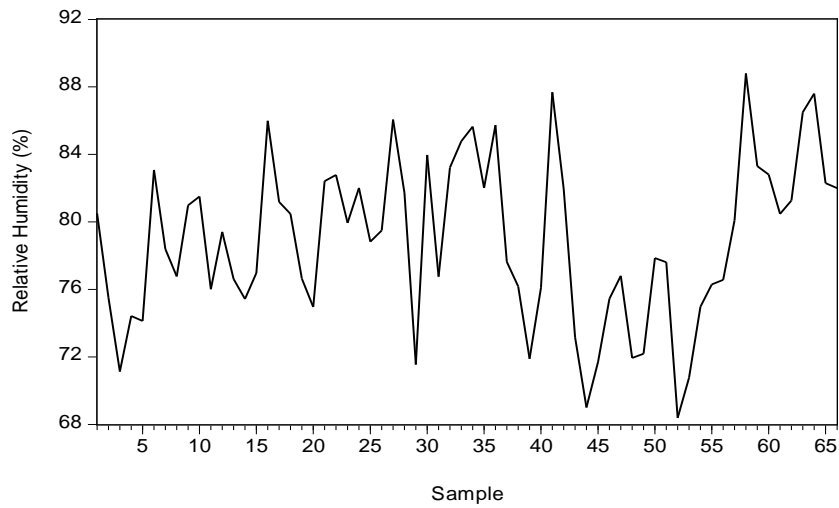


Figure 3. The relative humidity (%) recorded

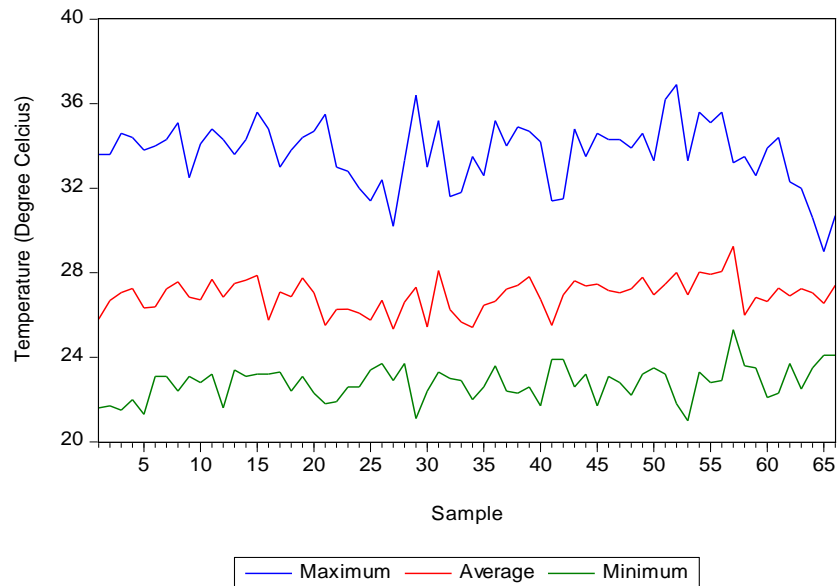


Figure 4. The temperature (°C) recorded

3.2. Testing the Stationarity of Black Pod Incidence

The cocoa black pod incidence data in both treatments as shown in Fig. 1 were found to be stationary for both treatments based on the ADF test or unit root test given in Table 1. ADF test successfully rejected hypothesis null of unit root at the 5% significant level in both treatments.

Table 1. Results of ADF Test on disease incidence data

Test equation	Treatment 1		Treatment 2	
	t_{ADF}	p value	t_{ADF}	p value
Without intercept and trend	-2.255	0.024	-1.943	0.050
Intercept without Trend	-3.427	0.013	-5.572	<0.001
Intercept With Trend	-3.738	0.027	-5.614	<0.001

3.3. ARIMA Modeling

Fig. 5 shows the ACF and PACF plots were cut-off after one lag in treatment 1. This suggested the presence of autoregressive models AR(1) through PACF plot and also presence of moving average models MA(1) through ACF plot. So, possible ARIMA models to be considered are ARIMA(1,0,0), ARIMA(1,0,1) and ARIMA(0,0,1). Fig. 6 shows the ACF and PACF plots with the partial autocorrelation coefficients cut-off after one lag and the autocorrelation coefficients cut-off after two lags in treatment 2.

This suggested the presence of autoregressive model AR(1) through PACF plot and also presence of moving average models MA(1) and MA(2) through ACF plot. So, possible ARIMA models to be considered are ARIMA(1,0,2), ARIMA(1,0,1), ARIMA(0,0,1), ARIMA(0,0,2), and ARIMA(1,0,0).

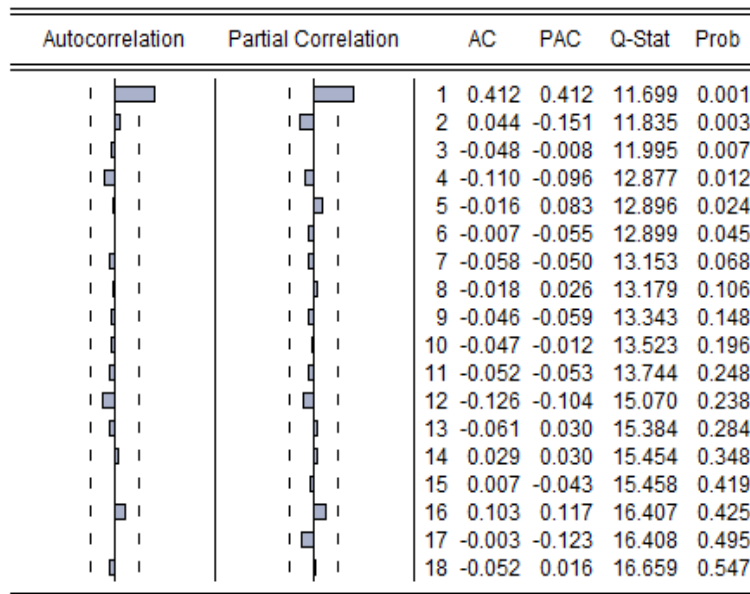


Figure 5. The correlogram of black pod disease incidence in treatment 1

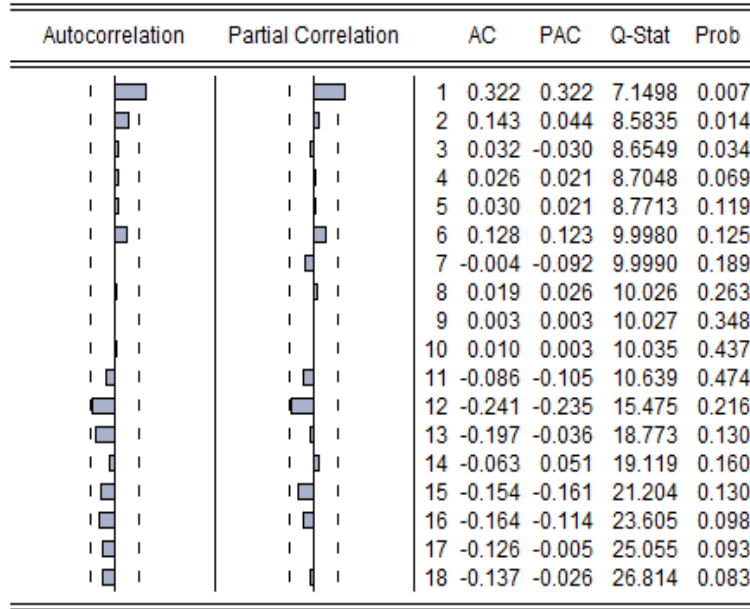


Figure 6. The correlogram of black pod disease incidence in treatment 2

Table 2. Estimated parameters of ARIMA models in Treatment 1

Variable	Coefficient	Standard Error	t	p value
Constant	0.00726	0.00448	1.619	0.111
AR(1)	0.48072	0.32276	1.489	0.141
MA(1)	0.08530	0.31566	0.270	0.788
Constant	0.00632	0.00247	2.558	0.013
AR(1)	0.55861	0.12880	4.337	<0.001
Constant	0.01321	0.00288	4.590	<0.001
MA(1)	0.48622	0.10938	4.445	<0.001

Table 3. Estimated parameters of ARIMA models in Treatment 2

Variable	Coefficient	Standard Error	t	p value
Constant	0.00606	0.00516	1.174	0.245
AR(1)	0.82492	0.15791	5.224	<0.001
MA(1)	-0.58850	0.20200	-2.914	0.005
MA(2)	-0.18836	0.13235	-1.423	0.160
Constant	0.00941	0.00627	1.502	0.138
AR(1)	0.71973	0.19560	3.680	<0.001
MA(1)	-0.51132	0.24698	-2.070	0.043
Constant	0.02101	0.0048	4.340	<0.001
AR(1)	0.33293	0.11972	2.781	0.007
Constant	0.02052	0.00560	3.665	0.001
AR(1)	0.30074	0.12818	2.346	0.022
AR(2)	0.06131	0.12918	0.474	0.637
Constant	0.03055	0.00418	7.301	<0.001
MA(1)	0.28092	0.12070	2.327	0.023

Table 2 shows the parameters estimation for three ARIMA models that have potentially fitted to the dataset in treatment 1. Table 3 shows the parameters estimation for

five ARIMA models that have potentially fitted to dataset in treatment 2. Both treatments have two ARIMA models, i.e. ARIMA(1,0,0) and ARIMA(0,0,1) with all the parameters significant at 5% level and need validated for the best fitted model.

The Ljung-Box Q test in treatment 1 gave the Ljung-Box Q statistics of ARIMA(1,0,0) and ARIMA(0,0,1) at lag 24 as 10.750 and 10.248 with p values greater than 5% significant level. This proved residuals of fitted ARIMA(1,0,0) and ARIMA(0,0,1) in treatment 1 are independently distributed with no serial correlation (Table 4). Meanwhile Ljung-Box Q test in treatment 2 gave the Ljung-Box Q statistics of ARIMA(1,0,0) and ARIMA(0,0,1) at lag 24 as 14.091 and 17.504 with p values greater than 5% significant level to show there was no autocorrelation among the residuals of fitted ARIMA(1,0,0) and ARIMA(0,0,1) (Table 4).

Table 4. Results of the Ljung-Box Q test on residuals on ARIMA models

Treatment	Time series model	Ljung-Box Q statistic	p value
1	ARIMA(1,0,0)	10.750	0.986
	ARIMA(0,0,1)	10.248	0.990
2	ARIMA(1,0,0)	14.091	0.924
	ARIMA(0,0,1)	17.504	0.784

The best fitted ARIMA model based on the AIC test and BIC test in Table 5 show ARIMA(1,0,0) in treatment 1 has the smallest value of AIC (-5.440) and BIC (-5.373) compared to ARIMA(0,0,1) with AIC value of -5.430 and BIC value of -5.363. Meanwhile ARIMA(1,0,0) in treatment 2 has the smallest value of AIC (-4.409) and BIC (-4.342) compared to ARIMA(0,0,1) with AIC value of -4.385 and BIC value of -4.3193 (Table 5).

Table 5. Results of the AIC test and BIC test

Treatment	Time series model	AIC value	BIC value	Decision
1	ARIMA(1,0,0)	-5.440	-5.373	Selected
	ARIMA(0,0,1)	-5.430	-5.363	
2	ARIMA(1,0,0)	-4.409	-4.342	Selected
	ARIMA(0,0,1)	-4.385	-4.319	

The selected ARIMA(1,0,0) model for both treatments explained that the forecast cocoa black pod disease incidence at time t was depend on the incidence at time $t - 1$. This was because the cocoa black pod disease incidence existed in treatments 1 and 2 mostly caused by the environment as study by Guest [23] that showed under favorable condition will caused the sporangia from diseased pod to develop within 48 hours of infection.

3.4. ARIMAX Modeling

The twelve steps given in the flowchart shown in Fig. 1 was used to build the ARIMA model. Part of the tests have been carried out in ARIMA modeling. The pair-wise Granger Causality tests in Table 6 for treatments 1 and 2 showed three null hypotheses of Granger Causality tests being rejected at 5% significant level to indicate one-way causal relationships from the exogenous variables to a dependent variable. The exogenous variables in treatment 1 were the relative humidity at time t (MeanRH _{t}), average temperature at time t (Tmean _{t}) and maximum temperature at time t (Tmax _{t}). As for treatment 2, the exogenous variables were the total rainfall at time t (TRainF _{t}), average relative humidity at time t (MeanRH _{t}) and maximum temperature at time t (Tmax _{t}).

Table 6. Granger Causality test involved climate variables

Null Hypothesis	p value	
	Treatment 1	Treatment 2
TRainF _{t} , does not granger cause BPTRT _{t}	0.062	0.007
BPTRT _{t} , does not granger cause TRainF _{t}	0.346	0.407
MeanRH _{t} , does not granger cause BPTRT _{t}	0.050	0.026
BPTRT _{t} , does not granger cause MeanRH _{t}	0.944	0.917
Tmean _{t} , does not granger cause BPTRT _{t}	0.042	0.702
BPTRT _{t} , does not granger cause Tmean _{t}	0.345	0.578
Tmin _{t} , does not granger cause BPTRT _{t}	0.474	0.142
BPTRT _{t} , does not granger cause Tmin _{t}	0.094	0.446
Tmax _{t} , does not granger cause BPTRT _{t}	<0.001	0.021
BPTRT _{t} , does not granger cause Tmax _{t}	0.686	0.832

It is important to ensure the exogenous variables have correct correlation sign to the disease incidence. The Spearman correlation analysis in treatment 1 determined four climate variables have correct correlation at 5% significant level (Table 7). The climate variables were the average relative humidity at time $t - 6$ (MeanRH _{$t-6$}),

average temperature at time $t - 3$ (Tmean _{$t-3$}), maximum temperature at time t (Tmax _{t}) and $t - 2$ (Tmax _{$t-2$}). There were nine climate variables in treatment 2 have correct correlation at 5% significant level (Table 8). The climate variables were the total rainfall at time $t - 4$ (TRainF _{$t-4$}) and $t - 5$ (TRainF _{$t-5$}), average relative humidity at time $t - 3$ (MeanRH _{$t-3$}), $t - 4$ (MeanRH _{$t-4$}), $t - 5$ (MeanRH _{$t-5$}), $t - 6$ (MeanRH _{$t-6$}) and $t - 7$ (MeanRH _{$t-7$}), maximum temperature at time t (Tmax _{t}) and $t - 5$ (Tmax _{$t-5$}).

Table 7. Correlation analysis on climate variables in treatment 1

Climate variable	Correlation coefficient	p value
MeanRH _{$t-6$}	0.293	0.017
Tmean _{$t-3$}	0.243	0.050
Tmax _{t}	-0.392	0.001
Tmax _{$t-2$}	-0.256	0.038

Table 8. Correlation analysis on climate variables in treatment 2

Climate variable	Correlation coefficient	p value
TRainF _{$t-4$}	0.265	0.031
TRainF _{$t-5$}	0.283	0.021
MeanRH _{$t-3$}	0.285	0.020
MeanRH _{$t-4$}	0.374	0.002
MeanRH _{$t-5$}	0.416	<0.001
MeanRH _{$t-6$}	0.438	<0.001
MeanRH _{$t-7$}	0.355	0.004
Tmax _{t}	-0.268	0.660
Tmax _{$t-5$}	-0.266	0.069

The stepwise regression that develop a linear regression model between the climate variables and the cocoa black pod disease incidence showed both Tmean _{$t-3$} and Tmax _{t} variables have significant F values at 5% level in treatment 1 while only MeanRH _{$t-4$} in treatment 2 has significant F values at 5% level (Table 9).

Table 9. Stepwise regression analysis

Variable	Parameter Estimate	Standard Error	F	p value	R Squared
Treatment 1					
Constant	0.07897	0.07109	1.23	0.271	0.2892
Tmean _{$t-3$}	0.00492	0.00232	4.50	0.038	
Tmax _{t}	-0.00588	0.00123	22.93	<0.001	
Treatment 2					
Constant	-0.16575	0.05086	10.62	0.002	0.1892
MeanRH _{$t-4$}	0.00245	0.00063	14.94	<0.001	

Table 10 shows the residuals from stepwise regression model in treatments 1 and 2 rejected the hypothesis null of unit root at the 5% significant level in ADF test. This indicated the residuals from the stepwise regression model in treatments 1 and 2 were stationary.

Table 10. Results of ADF Test on residuals from stepwise regression

Test equation	Treatment 1		Treatment 2	
	t_{ADF}	p value	t_{ADF}	p value
Without intercept and trend	0.656	0.855	-6.120	<0.001
Intercept without Trend	-5.317	<0.001	-6.070	<0.001
Intercept With Trend	-5.756	<0.001	-6.039	<0.001

Both the Ljung-Box Q statistics in treatments 1 and 2 (Table 11), 12.289 (lag 5) and 4.218 (lag 1) have p value greater than 0.05 which indicated autocorrelation exist in the residuals at 5% significant level. This provides an indication that AR and/or MA terms must be added into the model to remove the serial correlation.

Table 11. Results of the Ljung-Box Q test on residuals from stepwise regression

Treatment	Lags	Ljung-Box Q statistic	p value
1	5	12.289	0.0153
2	1	4.218	0.0400

Table 12 shows the ARIMAX model in treatment 1 was established using the AR(1) term and variable $Tmax_t$ as both were significant at 5% level while ARIMAX model in treatment 2 involved the AR(1) term and variable $MeanRH_{t-4}$ as both were significant at 5% level. The sign check on the coefficient of the exogenous variable as given in Table 12 correctly correlated with the disease incidence.

Table 12. Fitted ARIMAX models

Treatment	Variable	Coefficient	Standard Error	t	p value
1	Constant	0.14900	0.04316	3.452	0.0010
	AR(1)	0.40777	0.12807	3.184	0.0023
	$Tmax_t$	-0.00418	0.00126	-3.310	0.0016
2	Constant	-0.1563	0.04948	-3.159	0.0024
	AR(1)	0.29427	0.11029	2.668	0.0097
	$MeanRH_{t-4}$	0.00222	0.00062	3.598	0.0006

The selection of maximum temperature as one of the predictors in ARIMAX model in treatment 1 was due to its treatment applied the pruning as one of its control measure in managing the cocoa black pod disease incidence. The pruning provided good ventilation around and within the cocoa tree which can reduced the relative humidity to less volatile compared to temperature [24]. There were study

proved that the difference in maximum temperature measured outside and inside canopy layers of cocoa trees caused the maximum temperature to be an important factor in effecting the cocoa production and also the disease infection [25-26]. Meanwhile the selection of average relative humidity as one of the predictors in ARIMAX model in the treatments 2 due to the control measure applied to the cocoa trees were free from pruning with fungicide application only. The cocoa tree without pruning have dense canopy and caused little sunshine penetration that will create highly relative humidity to favor the *Phytophthora*'s inoculums to infect the cocoa pods [5, 27].

3.5. Model Validation

The best cocoa black pod disease forecasting model between ARIMA and ARIMAX model was selected based on its forecasting performance using three criteria, namely mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE). Table 13 shows ARIMAX model in treatment 1 has the lowest values of MSE (i.e. 0.00057), RMSE (i.e. 0.02389) and MAE (i.e. 0.01730) as compared to ARIMA model with MSE value of 0.00061, RMSE value of 0.02467 and MAE value of 0.01773. Meanwhile ARIMAX model in treatment 2 has the lowest values of MSE (i.e. 0.00955), RMSE (i.e. 0.09773) and MAE (i.e. 0.07765) as compared to ARIMA model with MSE value of 0.01003, RMSE value of 0.10013 and MAE value of 0.08139. This showed ARIMAX model has performed better compared to ARIMA model and therefore, ARIMAX model was selected as the best disease forecasting model in this study. The gap between the actual and ARIMAX predicted values seen to be far especially on sample with higher black pod incidence rates as seen in Fig.7 and Fig.8. The reason probable due to nonlinear effects in cocoa black pod incidence rate and also influencing external factors that not being considered in the study, i.e. duration of pods being exposed to wetness as direct impact to the disease.

Table 13. Actual versus forecasted values

Validation test	Treatment 1		Treatment 2	
	ARIMA	ARIMAX	ARIMA	ARIMAX
MSE	0.00061	0.00057	0.01003	0.00955
RMSE	0.02467	0.02389	0.10013	0.09773
MAE	0.01773	0.01730	0.08139	0.07765

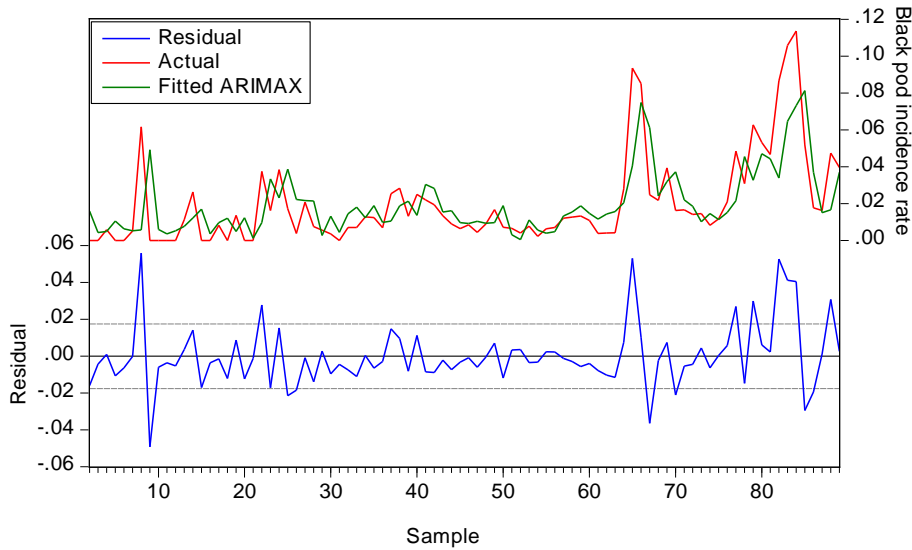


Figure 7. Plot of forecasted value of ARIMAX model in treatment 1

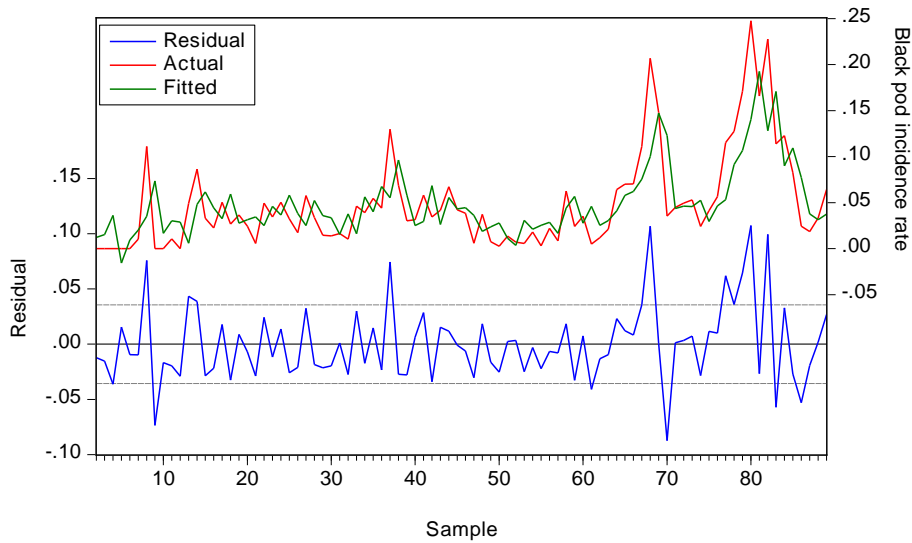


Figure 8. Plot of forecasted value of ARIMAX model in treatment 2

4. Conclusions

This study shows that there were an effect of climate variables to the cocoa black pod incidence. Key findings indicate that maximum temperature and relative humidity have significant correlation with black pod incidence and suggested as indicator in forecasting the cocoa black pod incidence. The results in this paper has found that ARIMA model with climate variables as input series *i.e.* ARIMAX consistently showed the superiority over ARIMA models in capturing the percent relative deviations pertaining to cocoa black pod incidence forecasts in two different treatment plots. The ARIMAX models performed well with lower error as compared to the ARIMA models. The plant disease forecasting models developed from ARIMAX model able to provide benefits especially for the development of decision support system in determine the right timing of action to be taken in controlling the cocoa

black pod disease.

Acknowledgements

The author also would like to thank the Director-General, Malaysian Cocoa Board (MCB) and Director of Upstream Technology Cocoa for permission to publish and reviewing this paper.

REFERENCES

- [1] International Cocoa Organization, Online available from <http://www.icco.org/about-cocoa/pest-a-diseases.html>
- [2] C. A. Thorold. The epiphytes of *Theobroma cacao* in Nigeria in relation to the incidence of black pod disease

- (*Phytophthora palmivora*), Journal of Ecology, Vol.40, No.1, 125-142, 1952.
- [3] C. A. Thorold. Observations on black pod disease (*Phytophthora palmivora*) of cacao in Nigeria, Transactions of the British Mycological Society, Vol.38, 435-452, 1955.
- [4] J. T. Dakwa. The relationship between black pod incidence and the weather in Ghana, Ghana Journal of Agricultural Science, Vol.6, 93-102, 1973.
- [5] G. A. R. Wokod. Black pod: Meteorological factors, In: Gregory PH (ed) *Phytophthora diseases of cocoa*, Longman, London, 1974.
- [6] J. Mpika, E. I. Auguste, B. K. Ismael, K. Brou, L. K. Jean, A. Kouassi, A. Z. Nicodeme, A. Severin. Effects of climatic parameters on the expression of the black pod disease on *Theobroma cacao* in Cote d'Ivoire, Journal of Applied Biosciences, Vol.20, 1183-1193, 2009.
- [7] O. A. Arqub, Z. Odibat, M. Al-Smadi. Numerical solutions of time-fractional partial integrodifferential equations of Robin functions types in Hilbert space with error bounds and error estimates, International Journal of Numerical Methods for Heat and Fluid Flow, Vol.28, No.4, 828 – 856, 2018.
- [8] G. E. P. Box, G. M. Jenkins. Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976.
- [9] M. Hemavathi, K. Prabakaran. ARIMA Model for Forecasting of Area, Production and Productivity of Rice and Its Growth Status in Thanjavur District of Tamil Nadu, India, International Journal of Current Microbiology and Applied Sciences, Vol.7, No.2, 149-156, 2018.
- [10] M. Amin, M. Amanullah, A. Akbar. Time series modeling for forecasting wheat production of Pakistan, Journal of Animal and Plant Sciences, Vol.24, No.5, 1444-1451, 2014.
- [11] E. Harris, A. R. Abdul-Aziz, R.K. Avuglah. Modeling annual Coffee production in Ghana using ARIMA time series Model, International Journal of Business, Social and Scientific Research, Vol.2, No.7, 175-186, 2012.
- [12] S. Ankrah, K. A. Nyantakyi, E. Dadey. Modeling the Causal Effect of World Cocoa Price on Production of Cocoa in Ghana, Universal Journal of Agricultural Research, Vol.2, No.7, 264-271, 2014.
- [13] F. J. Rodríguez-Rajo, V. Jato, M. Fernández-González, M. J. Aira. The use of aerobiological methods for forecasting *Botrytis* spore concentrations in a vineyard. Grana, Vol.49, No.1, 56-65, 2010.
- [14] A. Pankratz. Forecasting with Dynamic Regression Models, Wiley-Interscience, New York, 1991.
- [15] B. Yogarajah, C. Elankumaran, R. Vigneswaran. Application of ARIMAX Model for Forecasting Paddy Production in Trincomalee District in Sri Lanka, In: 3rd International Symposium, South Eastern University of Sri Lanka, 21-25, 2013.
- [16] C. Kongcharoen, T. Kruangpradit. Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export, In: 33rd International Symposium on Forecasting, South Korea, South Korea, 1-8, 2013.
- [17] H. Cui, X. Peng. Short-Term City Electric Load Forecasting with Considering Temperature Effects: An Improved ARIMAX Model, Mathematical Problems in Engineering Vol.2015, 2015.
- [18] E. Graham, J. R. Thomas, H. S. James. Efficient tests for an autoregressive unit root, Econometrica, Vol.64, No.4, 813-836, 1996.
- [19] B. Andrews, M. Dean, R. Swain, C. Cole Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors, Society of Actuaries, University of Southern Maine, 2013.
- [20] M. Ndoumbe-Nkenga, C. Cilasb, E. Nyemba, S. Nyassea, D. Bieysseb, A. Florib, I. Sachec. Impact of removing diseased pods on cocoa black pod caused by *Phytophthora megakarya* and on cocoa production in Cameroon, Crop Protection, Vol.23, 415-424, 2004.
- [21] J. P. Ngoh Dooh, B. N. Zachée Ambang, W. N. Kuate Tueguem, H. Alain, N. G. Ntsomboh. Development of Cocoa Black Pod Disease (Caused by *Phytophthora megakarya*) in Cameroon when Treated with Extracts of *Thevetia peruviana* or Ridomil, International Journal of Current Research in Biosciences and Plant Biology, Vol.2, No.3, 47-59, 2015.
- [22] M. Ndoumbè-Nkeng, M. I. B. Efombagn, S. Nyassé, E. Nyemb, I. Sache, C. Cilas Relationships between cocoa Phytophthora pod rot disease and climatic variables in Cameroon, Canadian Journal of Plant Pathology, Vol.3, 309-320, 2009.
- [23] D. I. Guest. Black pod: diverse pathogens and the greatest global impact on cocoa yield, Phytopathology, Vol.97, No.12, 1650-1653, 2007.
- [24] C. H. Lee. Planting cocoa challenges and reality in Malaysia, Online available from <http://www.iipm.com.my/ipicex2014/docs/2012/oral/PLANTING%20COCOA-%20CHALLENGES%20AND%20REALITY%20IN%20MALAYSIA.pdf>.
- [25] R. A. C. Miranda, L. C. E. Milde, A. L. Bichara, S. Cornell Daily characterisation of air temperature and relative humidity profiles in a cocoa plantation, Pesquisa Agropecuária Brasileira, Vol.29, 345-353, 1994.
- [26] A. J. Daymond, P. Hadley. The effects of temperature and light integral on early vegetative growth and chlorophyll fluorescence of four contrasting genotypes of cacao (*Theobroma cacao*), Annals of Applied Biology, Vol.145, 257-262, 2004.
- [27] J. Flood, D. Guest, K. A. Holmes, P. Keane, B. Padi, E. Sulistyowati. Cocoa under attack, In: J. Flood, R. Murphy (eds) Cocoa futures. Chinchina, CO: CABIFEDERACAFE, 2004.