B. Yogarajah[(1)], C. Elankumaran[(2)] and R. Vigneswaran[(3)]

Application of ARIMAX Model for Forecasting Paddy Production in Trincomalee District in Sri Lanka

(1) Department of Physical Science, Vavuniya Campus, University of Jaffna, Sri Lanka.
(2) Department of Economics, University of Jaffna , Jaffna, Sri Lanka.
(3) Department of Mathematics and Statistics, University of Jaffna, Sri Lanka.
(e-mail: yoganbala@vau.jfn.ac.lk)

Abstract: In the post-war climes, the government's mandates focus on reviving one of the paddy production region Trincomalee district in SriLanka to meet the growing demands of the nation. Such a rehabilitation program requires the understanding of how the paddy producing industry has fared along the historical time-lines. This understanding is essential for developing the necessary development plans for the Rice sector.

When an ARIMA model includes other time series as input variables, the model is referred to as an ARIMAX model. Pankratz (1991). In this paper, ARIMAX model has been applied to forecast annual paddy production with includes rainfall time series as input variable for both seasons in this district. The validity of the model is verified with various model selection criteria such as Adj R2, minimum of AIC and SBC lowest MAPE values.

Keywords: ARIMAX model; Forecasting; AIC; SBC; MAPE

Introduction

Rice is the most extensively cultivated crop in Trincomalee district in SriLanka. Due to the unsettled situation for last two decades in this district the rice sector entered the difficult stage of development and faces adjustment problems(IRI). Attention should be drawn to the fact that there is room for further improvement by finding ways and means of fully utilizing the general cultivable area in this district.

In this paper rainfall and annual paddy production are time series event processes. There are two cropping seasons in the district corresponding with the northeast monsoon, or Maha season, and the south-west monsoon, or Yala season. This dry-zone district show a significant relationship between annual Paddy production and rainfall in both Maha and Yala seasons(Table3) (Yoshino, 1984b: 95). In regard to Yoshino's Sri Lankan data, he confirmed the generality of the relationship between rainfall and the planted area.

Applications of time series technique ARIMA have been used to model for forecasting agriculture product. Some of research papers are Applying ARIMA models are Hossian et al. (2006) forecasted three different varieties of pulse prices namely motor, mash and mung in Bangladesh with monthly data from Jan 1998 to Dec 2000; Wankhade et al. (2010) Forecasted pigeon pea production in India with annual data from 1950-1951 to 2007-2008; Mandal (2005) forecasted sugarcane production in India; Iqbal et al. (2005) forecasted area and production of wheat in Pakistan; Masuda and Goldsmith (2009) forecasted world Soybean productions; Cooray (2006) forecasted Sri Lanka's monthly total production of tea beyond Sept 1988 using monthly data from January 1988 to September 2004. With these exceptions, there is paucity of studies regarding applications of ARIMA model for forecasting agricultural products

When an ARIMA model includes other time series as input variables, the model is referred to as an ARIMAX model. Pankratz (1991). In this paper, ARIMAX model has been applied to forecast annual paddy production with includes rainfall time series as input variable for both seasons in this district.This paper applies Autoregressive Integrated Moving Average (ARIMAX) forecasting model, the most popular and widely used forecasting models for uni-variate time series data. Although it is applied across various functional areas, it's application is very limited in agriculture, mainly due to unavailability of required data and also due to the fact that agricultural product depends typically on monsoonal rain and other factors, which the ARIMA models failed to incorporate

Materials and Methods

The existing study applies Box-Jenkins (1970) forecasting model popularly known as ARIMA model. The ARIMA is an extrapolation method, which requires historical time series data of underlying variable. The ARIMA approach was first popularized by Box and Jenkins, and ARIMA models are often referred to as Box-Jenkins models. The general transfer function model employed by the ARIMA procedure was discussed by Box and Tiao (1975). When an ARIMA model includes other time series as input variables, the model is sometimes referred to as an ARIMAX(p,d,q) model. Pankratz (1991) refers to the ARIMAX model as *dynamic regression*.

$$W_t = (1-B)^d Y_t = \mu + w_1\vartheta_1(B)X_{1t} + w_2\vartheta_2(B)X_{2t} + \frac{\theta_1(B)\theta_2(B)}{\varphi_1(B)\varphi_2(B)}a_t$$

Where t indexes time w_t ; μ is the mean term; B is the backshift operator; that is, $BX_t = X_{t-1}$

φB=1− $\varphi 1B$−…− φpBq is the autoregressive operator, represented as a polynomial in the back shift operator: θB=1−$\theta 1B$−…−θqBq is the moving-average operator, represented as a polynomial in the back shift operator: a_t is the independent disturbance, also called the random

ESACF and SCAN Methods

The Extended Sample Autocorrelation Function (ESACF) The Smallest CANonical (SCAN) methods can tentatively identify the orders of a *stationary or nonstationary* ARMA process based on iterated least squares estimates of the autoregressive parameters. Tsay and Tiao (1984) proposed the technique, and Choi (1990) provides useful descriptions of the algorithm.

Data Collection and Arrangement

The monthly rainfall and the yearly paddy production data from the period of 1970 to 2010 collected by the Trincomalee meteorological station and the Statistical Abstract release 1992, 1996, 2000, 2004, 2008 and 2010 by the Department of Census and Statistics are used. These data are used to explore annual climatic trend of paddy production and seasonal variation of Rainfall of the district. These data are used for driving the Auto Regressive Integrated Moving Average (ARIMAX) models with rainfall as input variables for both seasons. The collected data will be divided into two sets, calibration data and validating data, in order to testify the performance of the suggested model.

Result and Discussion

From the Table1 the Correlation between production in maha season(Mprd) and rainfall of the season (MRfl) is 0.48207. The Correlation between production in yala season(Yprd) and rainfall of the season (YRfl) is 0.42209. Therefore this auxiliary information of rainfall can be used for this model building process.

The Figure 1 and Figure 2 shows the relationship between the paddy production and rainfall season wise in this district.

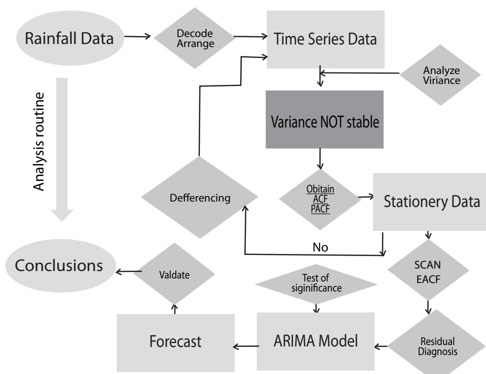
Table 1

Pearson Correlation Coefficients, N = 30 Prob > |r| under Ho: Rho=0					
	TotalPrd	Mprd	Yprd	MRfl	YRfl
TotalPrd	1.000	0.926	0.809	0.377	0.160
Mprd	0.926	1.000	0.612	0.482	0.101
Yprd	0.809	0.612	1.000	0.091	0.423
MRfl	0.377	0.482	0.091	1.000	0.011
YRfl	0.160	0.101	0.423	0.011	1.000

Descriptive statistics

The preliminary understating about the nature of data (Figure1 and Figure2) the output have shown highly volatile pattern, showing ups and downs over a period of time; some of their trends may be non-linear and non normal. In time series language the variables are non-stationary in nature; hence their mean and variance are not-constant and time variant which means output of these series are less and highly dispersed from the mean values. This is also reflected in Conclusion.

Model Building and Analysis



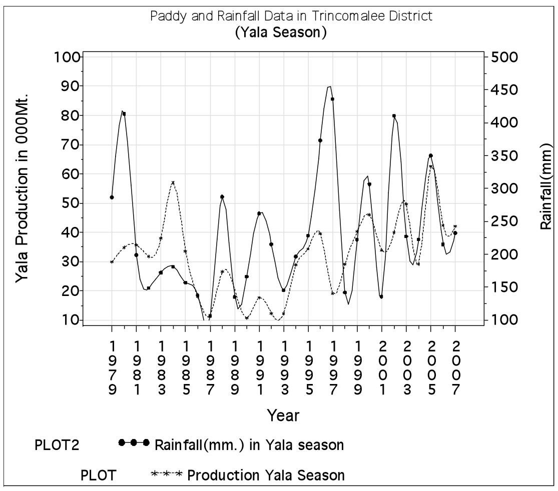


Figure 1: Plot of paddy and Rainfall data in the Yala Season

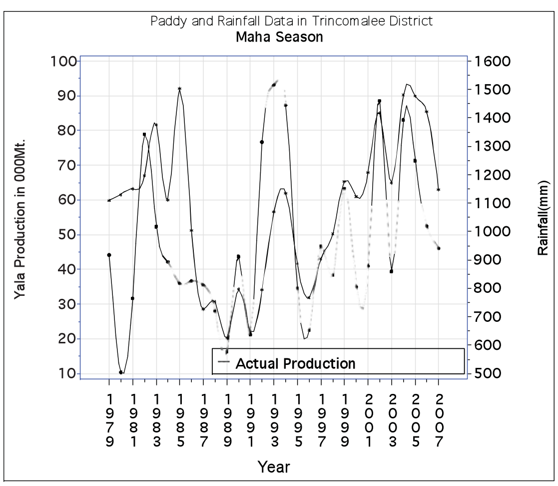


Figure 2: Plot of Paddy and Rainfall data in the Maha season

Figure 3 gives scatter plot ACF, PACF and IACF graph at autoregressive order p=1. This shows the stationarity of rainfall in the *maha* season.

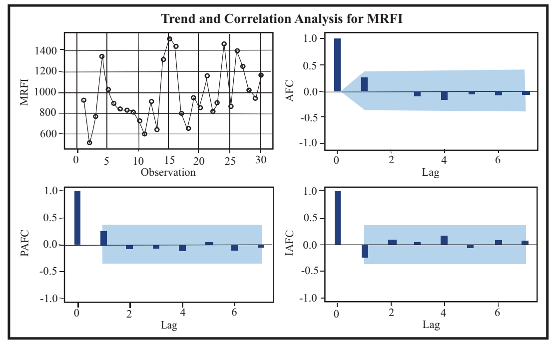


Figure 3: Trend and Correlation Analysis

Stationary vs. non-stationary

ARIMAX model is generally applied for stationary time series data. The time series properties of stationary and non-stationary are checked applying Augmented Dickey Fuller Test (Dickey-Fuller, 1979). The Augmented Dickey Fuller (ADF) tests results are estimated with level of first differences both seasons. The result shows that, the variables production and rainfall are non-stationary at level but stationary at first difference. The null hypothesis of non-stationary at level data is rejected at first difference data.

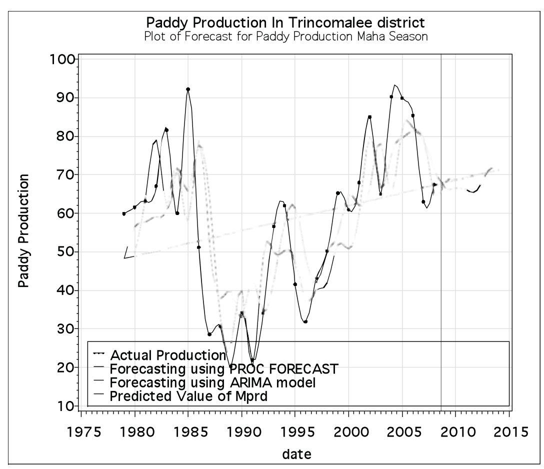


Figure 4: Plot of Forecast for Paddy Production in the Maha Season'

Now the question may arise, how do we know whether the identified model is appropriate or not? One simple way to answer is diagnostic checking on residual term obtained from ARIMA model applying the same ACF and PACF functions. Obtain ACF and PACF of residual term up to certain lags of the estimated ARIMAX model and then check whether the coefficients are statistically significant or not with Box-Pierce Q and LjungBox LB statistics, respectively(Figure4). If the result obtains from the model is purely random, then estimated ARIMA model is correct or else we have to look for alternative specification of the model. Similarly, diagnostic checking can also be done through Adjusted R2, minimum of Akaike Information Criteria (AIC) and Schwarz Bayesian Criteria (SBC) Table III reports the estimated results.

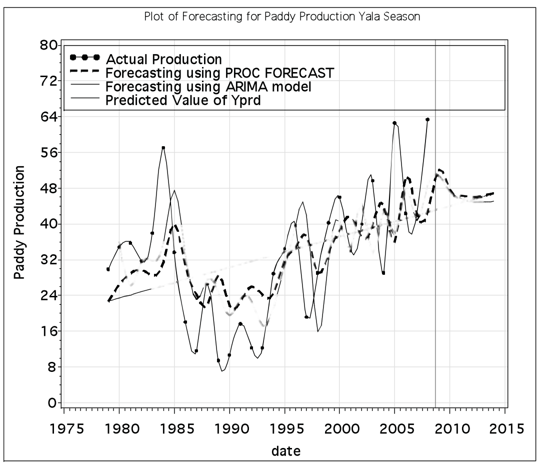


Figure 5: Plot of Forecasting for Paddy Production in the Yala Season

As per findings, the best ARIMA model for Trincomalee District

Yala Season is ARIMA(1,1,1)

$$(1-B)Yprd_t = 0.422071 + 0.04412(1-B)Yrfl_t + \frac{(1 - 0.99997\,B**(1))}{(1 - 0.53037\,B**(1))}a_t$$

Maha season is ARIMA(1,1,0*)*

$$(1-B)Mprd_t = 0.085781 + 0.01164(1-B)Mrfl_t + \frac{1}{(1 + 0.38282\,B**(1))}a_t$$

Diagnostic checking:

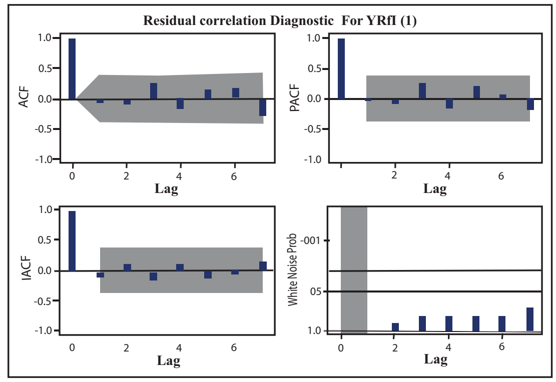


Figure 6: Residual Correlation diagnostic

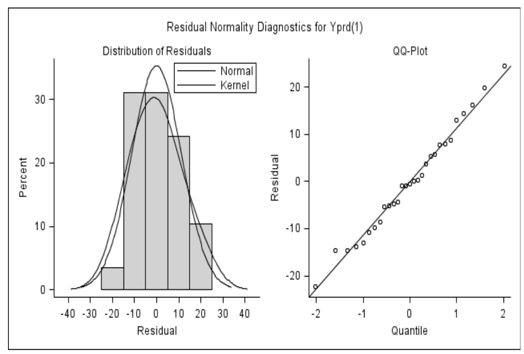


Figure 7: Residual Normality Diagnostic

Forecasting: Once the three previous steps of ARIMAX model is over, then we can obtained forecasted values by estimating appropriate model, which are free from problems. The forecasted values obtained from ARIMA model are reported in Table 3. The forecasted values are reported for a maximum 5 years as too much long term forecasting might not be appropriate

Conclusion

While applying various quantitative and qualitative models for forecasting, it is essential to understand the productivity is not an exception to it. In this paper ARIMA model has been applied on few selected agricultural products in India. As the model requires large data points, considering the availability of required annual data, 34 different agricultural products has been selected. Annual data from 1950 and 1957 onwards to 2010 as the case may be have been used. All the necessary steps of ARIMA model have been applied systematically for forecasting 5 periods ahead from 2011 onwards. Among these items, tea provides lowest MAPE value, whereas cardamom provides lowest AIC value. Similarly, highest MAPE is obtained for papaya and highest AIC value is for sugarcane. Now the question may arise is since agricultural productivity depend upon many factors such as rainfall, irrigation facility, monsoon, climate, soil, fertilizer etc., forecasted values might be more accurate only with ceteris paribus assumption. However, generally all the factors do not go well every time and in right direction; therefore reliability of these forecasted values might be questionable. In this context one need to rethink about other forecasting model, which could incorporate more information for forecasting the agricultural products. This could be one of the limitations of the paper.

underlying factors affecting it. Thus, forecasting agricultural

In this model fitting process for the monthly rainfall data from 1952 to 2010 in Trincomalee district the more appropriate model than other models is ARIMA$(0,0,1)(0, 0,1)_{12}$

The seasonal ARIMA model for monthly data with the following mathematical form:

$(1- B)(1- B^{12})$***Monthly Rainfall***$= (1 - 0.903 B)(1 - 0.999 B^{**}(12))a_t$

Where B is the backshift operator; that is and a_t is the independent disturbance, also called the random error.

Key findings indicate that the rainfall patterns in the study area are non auto-regressive, as such - they do not depend on the past history of rainfall; but are predominantly depending on nonlinear trend and a seasonal pattern of order 12. This indicates that in-order to arrive at a comprehensive forecasting model for rainfall in Trincomalee, the need to focus on the influences of non-endemic and regional-to-global climatic phenomena is apparent.

However, the model does not account for variations in precipitation due to cyclonic activity, which is a significant factor of variation in the local climate. This shortcoming should be addressed in the future research projects of this nature to enable a much clearer picture of climatic changes in the district.

The findings of this work can be applied in decision making for paddy cultivation and related water resources management.

Future work

Modeling Paddy production in the district and determine whether the annual rainfall of the district impacts on Paddy production in the district.

References

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. "Time Series Analysis: Forecasting and Control". Third Edition, Englewood Cliffs, NJ: Prentice Hall, pp.197-199 1994.

Box, G.E.P. and Tiao, G.C. "Intervention Analysis with Applications to Economic and Environmental Problems". JASA, pp.70, 70-79 1975.

B.Yogarajah. "A time-series modeling approach for understanding the behavior of rainfall patterns in the Anuradhapura District". Journal of the SrLanka Association of Geographers.pp.82-92 2010.

Brocklebank, J.C. and Dickey, D.A. "SAS System for Forecasting Time Series". Edition, Cary, North Carolina: SAS Institute Inc.1986.

Chatfild, C. "The Analysis of Time Series-an introduction." 5th Edn., Chapman and Hall, UK. 1996

Demroes, M. "Aspects of aridity and drought in the monsoon climate of Sri Lanka". Indian Journal. Hydrology and geophysics. 21(1and2): 384-394. 1978.

Manfred Domroes, Edmound Ranatunge "A statistical approach towards a regionalization of daily rainfall in Sri Lanka" Article first published online: DOI: 10.1002/joc.3370130704. 2006

Manobavan, M., "The Response Of Terrestrial Vegetation To El Niño Southern Oscillation". Unpublished PhD Thesis. Faculty of Science, Kingston University, Surry, United Kingdom. (2003).

Meyler, A., Kenny, G., and Quinn, T. "Forecasting Irish Inflation Using ARIMA Models". Technical Paper (3/RT/98). Economic Analysis, Research and Publication Department, Central bank of Ireland. 1998.

Sabita Madhvi Singh, "Statistical Modeling of Climate Parameters". Asian Journal of Current Engineering and Maths1: 2 March - April (2012) 29.

Thambyahpillay, G. "Climatic changes in Ceylon". MA. Thesis, University of Cambridge.1958.