*The Andhra Agric. J* 61(2): 408-412, 2014

# Comparative Performance of Time Series Models on Maize in North Coastal Zone of Andhra Pradesh

**G Srinivas, V Srinivasa Rao, D Vishnu Sankara Rao, and G Raghunadha Reddy**
Department of Agricultural Statistics, Agricultural College, Bapatla -522 101

**ABSTRACT**

The present study was carried out to the "*Comparative performance of time series models of area, production and productivity of maize in North coastal zone of Andhra Pradesh*". It has been undertaken to estimate the future trends and to fit the adequate model for the future projections by 2020 AD. The study was carried for North coastal zone of Andhra Pradesh using time series data from 1971 to 2011. Different linear, non linear and time series models were fitted to the area, production and productivity of maize and the best-fitted model was chosen based on least Mean Absolute Percent Error (MAPE) value and highest $R^2$ value for future projections. The result also shows that maize area, production and productivity would be 34.75 thousand hectares, 140.75 thousand tonnes and 4048 kg $ha^{-1}$ respectively by 2020 AD. The conclusion from the study is that, ARIMA model indicated that there would be substantial increasing in the area, production and productivity of the maize crop in the future.

**Key words :** ARIMA, ARCH, GARCH, Exponential smoothing, Linear trend, MAPE, $R^2$.

Maize (*Zea mays*) is the world's leading cereal crop. It is being cultivated globally over an area of 147.26 M ha with a production of 724.6 Mt of grain. Maize is emerging as third most important crop in India, after rice and wheat. Maize has its significance as a source of a large number of industrial products besides its uses as human food and animal feed. Maize is being grown in all the regions of Andhra Pradesh. In this study an attempt has been made to assess the growth rates in area, production and productivity of Maize crop in North coastal zone of Andhra Pradesh by using 41 years of data from 1971 to 2011. Besides, the projections were also estimated up to 2020 AD. The data of the study for a period of 41 years (1971 to 2011) in North coastal zone of Andhra Pradesh pertaining to area, production and productivity of Maize were collected from the statistical abstracts of Andhra Pradesh and web resources http:// www.Indiastat.com.

The future projections of area, production and productivity of maize crop in North coastal zone up to 2020 AD were estimated upon the best fitted time series model used for fitting the trend equations. The trend equations were fitted by using different linear, non linear and time series models. Among these models the model with least MAPE and highest $R^2$ was considered as the best fitted model for the projection purpose.

## MATERIAL AND METHODS

### Linear model:

Most naturally-occurring time series data are not at all stationary, they exhibit various kinds of trends i.e, cycles, seasonal and irregular patterns. Linear trend model is one of simplest method used to fit the time series data. The linear trend model is:

$$Y(t) = \alpha + \beta t$$

Where't' is the time index. The parameters alpha and beta (the "intercept" and "slope" of the trend line) are usually estimated via a simple regression in which Y is the dependent variable and the time index t is the independent variable.

### Exponential smoothing:

The time series data $Y_1$, $Y_2$….. $Y_t$. is to forecast the next value of time series $Y_{t+1}$ that is yet to be observed with forecast for $Y_t$ denoted by

$F_t$, then the forecast $F_{t+1}$ is based on weighting the most recent observation $Y_t$ with a weight value $\alpha$ and weighting the most recent forecast $F_t$ with a weight of $(1-\alpha)$. The weight is between 0 and 1. Thus the forecast for the period t+1 is given by

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

Where $\alpha$ is a smoothing constant.

**TIME SERIES MODELS:**

Autoregressive Process of order (p) is:

$$Y_t = \mu + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots\ldots\ldots + \emptyset_p Y_{t-p} + \varepsilon_t$$

Moving Average process of order (q) is:

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

And the general form of ARIMA model of order (p, d, q) is

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots\ldots\ldots + \emptyset_p Y_{t-p} - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots\ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where, $Y_t$ is the dependant variable, $\varepsilon_t$ s are independently and normally distributed with zero mean and constant variance for t = 1, 2…..n; d is the fraction differenced, Ø's and è's are coefficients to be estimated.

**ARIMA: Four steps**

**Identification:**

This step involves to determine the values of p, d and q. When a time series data is non-stationary, it can be often made by stationary by taking first differences of the series i.e., creating a new time series of successive differences i.e., $(Y_t - Y_{t-1})$, this is first order differentiation (d=1). If first difference does not convert the series to stationary, then go for second order differentiation (d=2). Values of p and q are determined by Autocorrelation function (ACF) and Partial Autocorrelation function (PACF), when the data is stationary only.

**Estimation of parameters**:

In this step, the precise estimates of parameters of the model are obtained by least-squares. Here, Standard Statistical package SAS was used for finding the estimates of relevant parameters.

**Diagnostic checking:**

The estimated model must be checked to verify if it adequately represents the series or not. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics are available, but Mean Absolute Percentage Error (MAPE) and $R^2$ are chosen as diagnostic checks in the present study.

**Auto Regressive Conditionally Heteroscedastic (ARCH) model:**

If the series is being high volatility in the data to capture the volatility ARCH model can be used. In this model, $\sum_t$ denote the error terms (return residuals, with respect to a mean process) i.e. the series terms. These $\sum_t$ are split into a stochastic piece $Z_t$ and a time-dependent standard deviation $ó_t$ characterizing the typical size of the terms so that $\sum_t = \alpha_t Z_t$ The random variable $Z_t$ is a strong process. The series $\sigma_t^2$ is modeled by Engle R.F (1982).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

Where $\alpha_0 > 0$ and $\alpha_i \geq 0,\ i > 0$.

**Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Process:**

Bollerslev (1986) developed a GARCH (p, q) model which takes both the components of autoregressive and moving average in the form of the heteroscedasticity variance, GARCH model was the extension of the ARCH model which was proposed by Engle (1982).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

Table 1. Different models on area, production and productivity of maize in North coastal zone of Andhra Pradesh.

| | | MAIZE | | | | |
|---|---|---|---|---|---|---|
| | CRITERIA | LINEAR | EXP. SMOOTHING | ARIMA | ARCH | GARCH |
| AREA | $R^2$ | 0.76 | 0.91 | 0.92 | 0.90 | 0.92 |
| | MAPE | 34.47 | 12.42 | 12.05 | 14.21 | 13.69 |
| PRODUCTION | $R^2$ | 0.65 | 0.93 | 0.96 | 0.93 | 0.95 |
| | MAPE | 128.46 | 29.67 | 28.98 | 32.77 | 30.76 |
| PRODUCTIVITY | $R^2$ | 0.52 | 0.61 | 0.64 | 0.61 | 0.60 |
| | MAPE | 27.67 | 25.31 | 24.31 | 25.39 | 25.67 |

Table 2. Forecasted values of area, production and productivity of maize in North coastal zone of Andhra Pradesh

| Year | Predicted Area('000 hectares | Predicted Production ('000 tonnes) | Predicted Productivity (kg $ha^{-1}$) |
|---|---|---|---|
| 2012 | 34.32 | 140.76 | 4031 |
| 2013 | 34.56 | 140.76 | 4067 |
| 2014 | 35.29 | 140.76 | 4039 |
| 2015 | 34.75 | 140.76 | 4061 |
| 2016 | 34.75 | 140.76 | 4044 |
| 2017 | 34.75 | 140.76 | 4057 |
| 2018 | 34.75 | 140.76 | 4047 |
| 2019 | 34.75 | 140.76 | 4055 |
| 2020 | 34.75 | 140.76 | 4048 |

**MODEL SELECTION**:

The choice of the trend equation amongst the available alternatives is very crucial. Many researchers use coefficient of multiple determination ($R^2$) and MAPE as the criterion of model selection are given below

**$R^2$-Criteria**

$$R^2 \equiv 1 - \frac{SS_{\text{err}}}{SS_{\text{tot}}}. \quad \text{or} \quad R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}}.$$

**Mean Absolute Percentage Error (MAPE)**

$$\mathrm{M} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|.$$

Where, $A_t$ is the actual value and $F_t$ is the forecast value. The difference between $A_t$ and $F_t$ is divided by the actual value $A_t$ again.

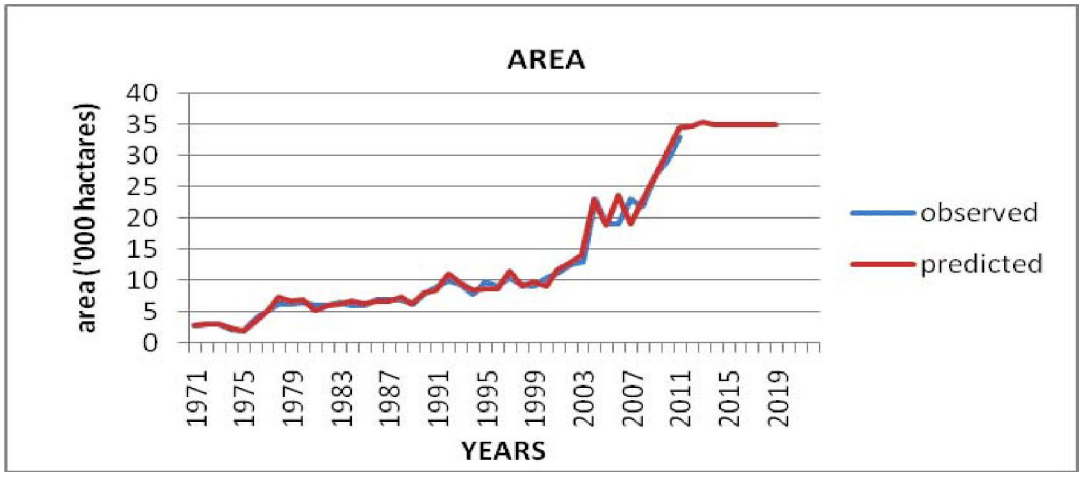
**RESULTS AND DISCUSSION**

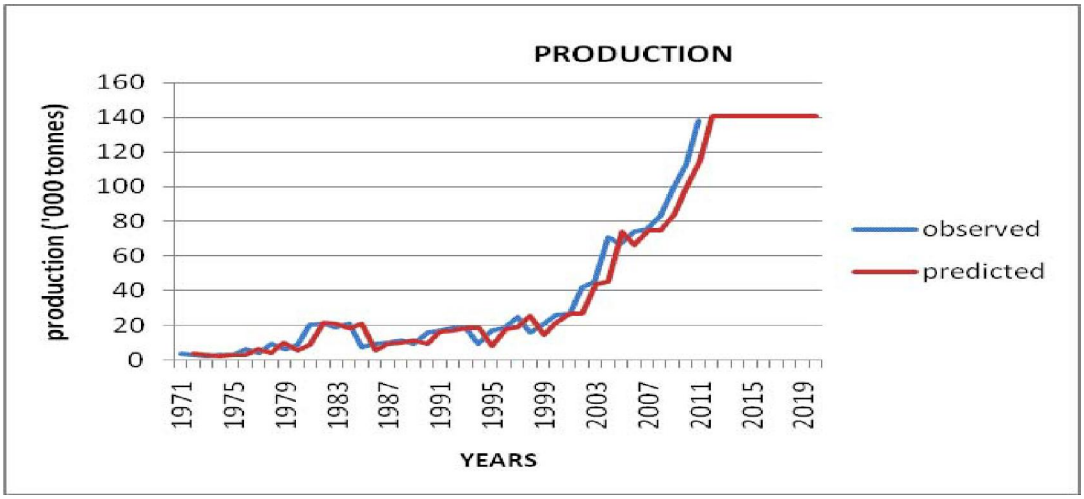
To understand performance of the time series models of area, production and productivity of maize in North coastal zone during the period 1971 to 2011, time series data was analyzed by using linear, non linear and time series models *viz.,* linear, exponential smoothing, ARIMA, ARCH, GARCH models were fitted and the results are presented in table 1.

**Maize area**

The average area of the Maize crop for the entire period of study was 10.52 thousand hectares. It was also observed that the area of the maize crop was tremendously increased from 1.9 thousand hectares in 1975 to 33 thousand hectares in the year 2011. The area of the maize crop ARIMA (0,1,4) model with least MAPE value of

Projections of maize area, production and productivity in north coastal zone of Andhra Pradesh




12.05 and highest $R^2$ value of 0.92 was considered as the best fit among all the models considered.

**Maize production**

The average production of the Maize crop for the entire period was 29.57 thousand tonnes. It was also observed that the production was increased enormously from 2.3 thousand tonnes in the year 1973 to 138 thousand tonnes in the year 2011. From the production model ARIMA (2,1,1) with least MAPE value of 28.98 and highest $R^2$ value of 0.96 was considered as the best fit among all the models considered.

**Maize Productivity**

The average productivity of maize crop for the entire period was 2257 kg $ha^{-1}$. It was also observed that the productivity is also increasing from 793 kg $ha^{-1}$ in the year 1933 to 2257 kg $ha^{-1}$ in the year 2011. From productivity the ARIMA (1,1,2) model with least MAPE value of 24.31 and highest $R^2$ value of 0.64 was considered as the best fit among all the models considered.

**Projections**:

The future projections of area, production and productivity of maize by 2020 AD were calculated based on the selected models. The projected area of maize by 2020 AD with ARIMA (0,1,4) would be 34.75 thousand hectares and it was presented in Table 2.

For production ARIMA (2,1,1) model was identified as the best model for future projections as it has highest $R^2$ value with the least MAPE value for long term projections. The projected production of maize by 2020 A.D would be 140.75 thousand tones.

ARIMA (1,1,2) model was found to be best fitted model for the projection of maize productivity as it has highest $R^2$ and least MAPE values. So, the productivity of maize projected by 2020 AD would be 4048 kg $ha^{-1}$.

## CONCLUSIONS

The study was under taken to obtain a suitable time series models for forecasting the area, production and productivity of the maize crop in the North coastal zone of Andhra Pradesh and are projected up to 2020 AD. The average area under maize was 10.52 thousand hectares with an average production of 29.57 thousand tones and productivity of 6967.43 kg $ha^{-1}$. ARIMA model indicated that there would be substantial increasing in area and production and productivity in the future.

Selection of an appropriate model is made by comparing the highest $R^2$ and lowest MAPE values. When there is volatility in the data linear, exponential smoothing and ARIMA models may not capture its trend. Hence ARCH and GARCH models were used to capture the volatility in the data. For the maize crop there was not much variation in the area but there was light change in the production and productivity.

## LITERATURE CITED

**Badmus M A and Ariyo, O S 2011** Forecasting cultivated areas and production of maize in Nigeria using ARIMA models. *Asian Journal of Agricultural Sciences*, 3(3): 171-176.

**Bollerslev T 1986** Generalized Auto Regressive Conditional Heteroscedasticity. *Journal of Econometrics,*. 31: 307-327.

**Engle R F 1982** Auto Regressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica,* 50 (4): 987-1007.

**Iqbal N, Bakhsh K, Maqbool A and Ahamad A S 2005** Use of the ARIMA model for forecasting wheat area and production in Pakistan. *Journal of Agriculture and Social Sciences*, 1(2): 120-122.

**Irfan M, Irfan M and Tahir M 2011** Modelling of province wise yield of rice crop in Pakistan using GARCH model. *International Journal of Science and Technology,* 6(1): 224-228.

**Muhammad F, Bashir M and Ahmad S 1992** Forecasting rice production in Pakistan using ARIMA models. *Journal of Animal Plant Science,* 2: 27-31.

**Pankratz A 1983** *Forecasting with Univariate Box – Jenkins Models. Concepts and Cases.* John Wiley and sons, New York.

**Reddy V M 1978** Growth rates. *Economic and Political Weekly,* 132: 806-812.

(Received on 29.06.2013 and revised on 10.01.2014)