e-ISSN: 2394-5532
p-ISSN: 2394-823X

International Journal of Applied And Pure Science and Agriculture
www.ijapsa.com

FORECASTING OF AREA, PRODUCTION AND PRODUCITIVITY OF RAGI IN ANDHRA PRADESH USING ARIMA MODELS

V Nireesha[1], V Srinivasa Rao[2], DVS Rao[3] and G Raghunadha Reddy[4]

[1] P.G Student, Department of Agricultural Statistics, Agricultural College, Bapatla.
[2] Professor & Univ.Head, Department of Agricultural Statistics, Agricultural College, ANGRAU.
[3] Professor & Univ.Head, Department of Agricultural Economics, Agricultural College, ANGRAU.
[4]Associate Professor, Department of Agricultural Economics, Agricultural College,ANGRAU.

ABSTRACT

This paper attempted to identify the trend of area, production and productivity of Ragi (Figer millet) in Andhra Pradesh through fitting different Time series models like ARIMA, Exponential Smoothing techniques. Influence of weather parameters on area, production and productivity of Ragi crop by using Karl Pearson's correlation and Multiple Linear Regression Analysis was also studied. From the best fitted model forecasting of area, production and productivity was also done. ARIMA (1, 1, 1) model was identified as the best model for area, production and Linear trend was identified as the best model for productivity of the observed data and forecasting was done for the Ragi area, production and productivity up to 2020 AD. It was observed that there was an increasing trend in the productivity, but both area and production was in the decreasing trend during the study period.

Key words: ARIMA (Auto Regressive Integrated Moving Average), Exponential Smoothing, Karl Pearson's correlation, Multiple Linear regression, Theil's U-Statistic, R^2, Adj R^2 and MAPE.

I. Introduction

The millets are a group of small seeded species of cereal crops or grains widely grown around the world for food and fodder. They have been in cultivation in East Asia for the last 10,000 years. There are a variety of millets and widely cultivated species are the Pearl millet (*Pennisetum glaucum*) known as Bajra in India, Foxtail millet (*Setaria italica*), Proso millet or common millet or broom corn millet or hog millet or white millet (*Panicum miliaceum*) and the Finger millet (*Eleusine coracana*) known as Ragi or Mandwa in India.

Ragi (Finger millet) is a staple food crop in many hilly regions of the country and predominantly grown as a dry land crop in Karnataka, Andhra Pradesh and Tamil Nadu. It is extensively cultivated in India, Africa, Ceylon, Malaysia, China and Japan. In India it is cultivated over an area of 2.65 Mha with total production of about 2.9 Mt.

In India the area under ragi has decreased from 1175.8 thousand hectares in 2011 to 1128.0 thousand hectares in 2012 and production from 1929.2 thousand tonnes in 2011 to 1574.4 thousand tonnes. Similarly the yield has marginally decreased from 1641 kg ha^{-1} to 1396 kg ha^{-1}. (Ministry of Agriculture, Govt of India).

Andhra Pradesh is one of the important ragi growing states. The area under sorghum was 10 thousand hectares, production was 12 thousand tonnes and productivity was 1200 kg ha^{-1} during 2012 in Andhra Pradesh.

In the present study ragi area, production and productivity of Andhra Pradesh have been forecasted by using Auto Regressive Integrated Moving Average (ARIMA) models.

Lot of work was done on forecasting by using ARIMA models on various crops. Iqbal *et al*. (2005) studied forecasting of wheat area and production in Pakistan by using ARIMA model. They developed models such as ARIMA (1, 1, 1) for area and ARIMA (2, 1, 2) for production by using data of 1950-2000 to forecast up to 2022. The forecasted production of wheat showed an increasing trend. Nehru and Rajaram (2009) studied on prediction of wheat production in India by ARIMA technique. ARIMA (1, 1, 0) was selected as best fitted model based on AIC and SBC. The predicted values of wheat production showed that there will be steady increase from 2008-09 to 2014-15. Ahmed *et al.* (2011) determined to forecasts milk production in Pakistan using ARIMA model covering time series data from 1990-91 to 2010-11. They concluded that ARIMA (1, 1, 1) was the best fitted model and forecasting of milk production for the year 2015 would be 47494.2 thousand tonnes with lower and upper limits 38826.3 thousand tonnes and 56162.1 thousand tonnes respectively. Forecast of milk production showed an increasing trend.

Singh *et al*. (2008) analysed the effect of rainfall and temperature effect on wheat yield in south western region of Punjab. Maximum, minimum temperature and rainfall from December to March for each period of five years 1977-81 to 1997-2001 were analyzed. The results revealed that temperatures during February and March showed significant effect on wheat yield. The grain yield revealed positive correlation with minimum temperature but no trends observed for other parameters. Rankja *et al*. (2010) made an attempt to study the quantitative relationship between weather parameters and district level yield of cotton and to develop pre harvest forecast models for cotton yield. For this purpose 32 years data of weather parameters and crop yield was collected. The results showed that the 26 week crop period model was recommended for pre harvest forecast due to highest coefficient of determination (R^2) and lower forecast error. The analysis also revealed that time trend, maximum temperature, morning and evening relative humidity have significantly affected on crop yield.

II. MATERIALS AND METHODS

The data on Ragi crop area, production and productivity was collected for a period of 47 years (1966–2012) from Ministry of Agriculture, Govt of India and www.Indiastat.com (Statistical database), official website. By using SAS 9.3 software fitted the time series models i.e., ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0) and ARIMA(1,1,1) was selected based on trial and error approach method and Exponential smoothing models also fitted. Best model was selected based on the diagnostics like highest Theil's U-Statistic, highest R^2, adjusted R^2 having least MAPE values. The projections were also estimated up to 2020 AD. The association between weather parameters on area, production and productivity also analysed by using correlation and regression analysis. The time series models, correlation and regression analysis taken under consideration are as follows.

Description of the model:

Auto Regressive Integrated Moving Average (ARIMA) model was introduced by Box and Jenkins. In general, an ARIMA model is characterized by the notation ARIMA (p, d, q) where p, d, q denote orders of auto-regression, integration (differencing) and moving average respectively. In ARIMA, time series is a liner function of past actual values and random shocks. A stationary ARIMA (p,d, q) process is defined by the equation

$$\mathbf{Yt = \varphi\ 1\ Yt\text{-}1+\varphi 2\ Y_{t\text{-}2}+\ldots+\varphi p\ Y_{t\text{-}p} -\ \omega 1\varepsilon t\text{-}1 -\ \omega\ 2\varepsilon\ t\text{-}2\text{-}\ldots\text{-}\ \omega\ q\ \varepsilon\ t\text{-}q + \varepsilon\ t}$$

Where Yt is the response (dependent) variable at time t. $Y_{t-1}, Y_{t-2}, Y_{t-p}$ are the response (dependent) variable at lags t-1, t-2….t-p, these Y's are independent variables. An assumption about error term is same as standard regression model. $\omega_1, \omega_2 \ldots \omega_q$ are the coefficients to be estimated.

III. ARIMA MODEL BUILDING

Model identification:

ARIMA model was estimated only after transforming the variable into a stationary series. The stationary series was the one whose values vary over time around a constant mean and constant variance. There are several ways to ascertain this.

The most common method is to check stationary through examining the graph or time plot of the data. Non-stationary in mean is corrected through appropriate differencing of the data. The newly constructed differenced variable can now be examined for stationary. The graph of differenced variable was stationary.

The next step in the identification process is to find the initial values for the orders of non-seasonal parameters, p and q. They are obtained by looking for significant autocorrelation and partial autocorrelation coefficients from ACF and PACF charts.

Estimation:

At the identification stage, one or more models are chosen that seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model are obtained by least-squares. Standard computer packages like SAS, SPSS etc…are available for finding the estimates of relevant parameters using iterative procedures.

Exponential Smoothing:

A simple exponential smoothing model is a very popular model used to produce a smoothed time series. Where in simple Moving Average models the past observations are weighted equally, Exponential Smoothing assigns exponentially decreasing weights as the observations get older. However, a "smoothing parameter" or "smoothing constant" – is used to determine the weights assigned to the observations (Gardner, 1985).

The time series data is $Y_1, Y_2 \ldots\ldots Y_t$ 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 weighing the most recent observations Y_t with a weight value α and weighting the most recent forecast F_t with a weight of (1- α). 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 α is a smoothing constant.

Karl Pearson's Correlation:
Correlation coefficient is a measure of degree of linear association between two variables.

Correlation coefficient between two variables X and Y

$$r(X,Y) = \frac{Cov(X\,Y)}{\sqrt{Var(x).Var(y)}}$$

For test the significance of correlation coefficient (r) t-test is used.
Null Hypothesis (H_0): $\rho = 0$
Alternate Hypothesis (H_1): $\rho \neq 0$

Test the Significance of correlation coefficient:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} \text{ degrees of freedom}$$

Multiple Linear Regression:

Prediction of an unknown dependent variable with the help of independent variables is called linear regression. Multiple Linear Regression Analysis is if more than one independent variable is to be used in the model, linear regression can be extended to multiple linear regression to accommodate several independent variables (Pearson 2011). It is of the form

$$Y_i = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

Where

a = intercept in the model,
x_1, $x_2 + \ldots\ldots + x_n$ are independent variables,
b_1, $b_2 + \ldots\ldots + b_n$ are regression coefficients.

Model selection:

The choice of time series models amongst the available alternatives is very crucial. Many researchers uses Theil's U-Statistic, coefficient of multiple determination (R^2) or adjusted R^2 ($\bar{R}^2$) and least MAPE as the criterion of model selection.

Theil's U – Statistic:

This statistics allows a relative comparison of normal forecasting methods with naive approaches and also squares the errors involved so that large errors are given much more weight than small errors. The positive characteristic that is given up in moving to Theil's U-statistic as a measure of accuracy is that of intuitive interpretation. The difficulty will become apparent as the computation of this statistic and its application are examined. Mathematically, it is defined as,

$$U = \sqrt{\sum_{t=1}^{n-1} \frac{\left(\frac{F_{t+1} - Y_{t+1}}{Y_t}\right)^2}{\left(\frac{Y_{t+1} - Y_t}{Y_t}\right)^2}} * 100$$

If U = 100 both models forecast with equal accuracy

U < 100 model forecast is best

Coefficient of Determination (R^2):

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well, the regression line approximates the real data points. An R^2 of 1.0 indicates that the regression line perfectly fits the data. It provides a measure of how well future outcomes are likely to be predicted by the model.

$$R^2 = \frac{\text{Regression sum of squares}}{\text{Total sum of squares}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Adjusted R^2 ($\bar{R}^2$)

Adjusted R^2 is a modification of R^2 that adjusts for the number of explanatory terms in a model. The adjusted R^2 can be negative, and will always be less than or equal to R^2. The adjusted R^2 is defined as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Where,

p is the number of parameters in the equation

n is the total number of observations

Mean Absolute Percent Error (MAPE): It is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It is usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE = \left(\sum_{t=1}^{n} \left(\left| \frac{A_t - F_t}{A_t} \right| \right) \times 100 \right) / n$$

Where,

A_t is the actual value and

F_t is the forecast value

IV. RESULTS AND DISCUSSION

It was observed that the time series data on Ragi crop area, production and productivity was not stationary. It was made stationary by using the first order differencing technique. The appropriate model was chosen based on highest Theil's U-Statistic (model accuracy), highest R^2, highest Adjusted R^2 and least MAPE values.

From the study it is revealed that ARIMA (1,1,1) model was selected as best in area, production of ragi but in productivity linear trend exponential smoothing was selected as best fitted model because of its highest Theil's U-Statistic, highest R^2, adjusted R^2 and least MAPE values. Hence forecasting of area, production and productivity upto 2020 AD were fitted based on these models and results are presented in table 1.

Projections:

The future projections of area, production and productivity of Ragi in Andhra Pradesh by 2020 AD were calculated and the results were presented in Table 2. The projected area, production and productivity under Ragi by 2020 AD would be 0.91thousand hectares, 1.62 thousand tonnes and 1303 kg ha^{-1} respectively. The projections of Ragi area, production would be decreasing and productivity would be slightly increasing by 2020 A.D.

Correlation Analysis:

From correlation analysis it was observed that Maximum Temperature and Minimum Temperature were positively correlated with area, production but on productivity these parameters

were negatively correlated. Rainfall, morning relative humidity (RH1) and evening relative humidity (RH2) were negatively correlated with area, production but with yield positively correlated and test of significance analysis was done where Minimum temperature, RH1and RH2 was found to be significant at 5 % level in area, production but in productivity morning relative humidity (RH1) was only significant and all the other parameters were non-significant at 5% level. Even though the yield was increased because of less area the production was decreased. Correlation coefficient values and test of significance values are presented in table 3.

Multiple Linear Regression Analysis:

From multiple linear regression analysis it was observed that among all the weather parameters Morning relative humidity (RH1) was negatively significant at 1% level on area and production. But in yield morning relative humidity (RH1) was positively significant at 5% level. Because of these models lowest coefficient of determination (59.9%, 58.0%, and 24.6%) it can be concluded that other than weather parameters many factors are influencing on area, production and yield of Ragi. The predicted models are presented in table 4.

Table 1. Time Series models in Area, Production and Productivity of Ragi in Andhra Pradesh

Time Series Models	Theil's U-Statistic (%)	R^2	Adj R^2	MAPE
AREA				
ARIMA (0, 1, 0)	95.50	0.966	0.965	12.21
ARIMA (0, 1, 1)	95.49	0.966	0.965	12.20
ARIMA (1, 1, 0)	95.49	0.966	0.965	12.20
ARIMA (1, 1, 1)	**95.69**	**0.968**	**0.967**	**11.38**
Linear (Holt) Exponential Smoothing	95.60	0.968	0.967	11.94
PRODUCTION				
ARIMA (0, 1, 0)	92.39	0.889	0.887	17.49
ARIMA (0, 1, 1)	92.81	0.901	0.899	15.80
ARIMA (1, 1, 0)	92.80	0.901	0.899	16.38
ARIMA (1, 1, 1)	**92.97**	**0.907**	**0.905**	**15.12**
Damped Trend Exponential Smoothing	92.75	0.899	0.897	16.25
PRODUCTIVITY				
ARIMA (0, 1, 0)	94.44	0.169	0.151	9.09

ARIMA (0, 1, 1)	95.49	0.457	0.445	7.55
ARIMA (1, 1, 0)	95.24	0.393	0.380	7.92
ARIMA (1, 1, 1)	95.61	0.486	0.475	7.46
Linear Trend Exponential Smoothing	**95.55**	**0.520**	**0.509**	**7.77**

Table2. Future Projections of Area, Production and Productivity of Ragi in Andhra Pradesh

YEAR	AREA ('000 ha)	PRODUCTION ('000 tonnes)	PRODUCTIVITY (kg ha^{-1})
2013	9.44	12.96	1252
2014	8.03	11.40	1259
2015	6.27	8.93	1267
2016	4.49	6.37	1274
2017	2.93	4.16	1281
2018	1.74	2.55	1289
2019	1.04	1.67	1296
2020	0.91	1.62	1303

Table 3. Correlation coefficient values of Area, Production and Productivity (Yield) of Ragi With Weather Parameters

Weather Parameters	Area	Production	Yield	Test of significance (t) for Area	Test of significance (t) for production	Test of significance (t) for yield
Max temp (^{0}C)	0.19	0.22	-0.03	1.071	1.242	0.203
Min temp (^{0}C)	0.43	0.46	-0.05	2.551*	2.816*	0.279
Rainfall (mm)	-0.20	-0.14	0.23	1.133	0.774	1.254
RH1 (%)	-0.72	-0.68	0.47	5.621*	4.988*	2.895*
RH2 (%)	-0.36	-0.36	0.24	2.084*	2.049*	1.316

(* significant at 5% level) (t tab value = 2.048)

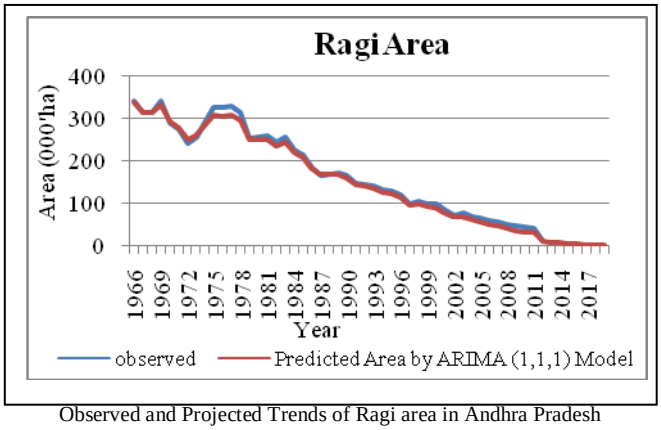
Table 4. Predicted models for Area, Production and Yield of Ragi on Weather Parameters

Dependent Variable	Model	R^2
Area	$\hat{Y}$ = 2087.9 + 0.75 (Max Temp) + 2.45 (Min Temp) + 0.42 (Rainfall) - 26.37** (RH1) + 0.36 (RH2)	0.599*

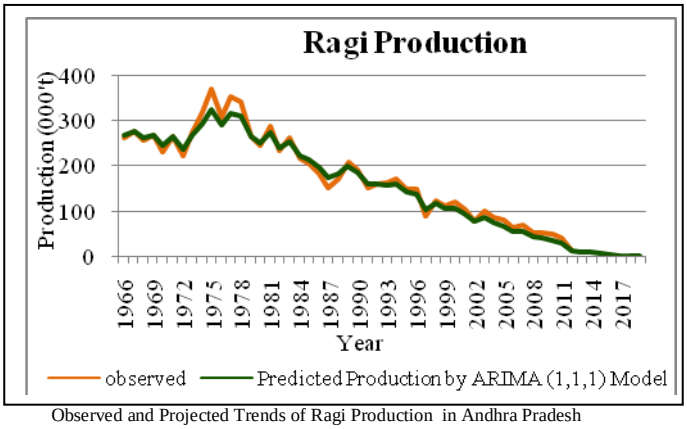
Production	$\hat{Y}$ = 2030.5 + 0.80 (Max temp) + 3.10 (Min temp) + 0.57 (Rainfall) - 25.04** (RH1) - 0.66 (RH2)	0.580*
Yield	$\hat{Y}$ = -1699.6 - 0.95 (Max temp) + 4.19 (Min temp) -0.21 (Rainfall) + 35.28* (RH1) - 0.80 (RH2)	0.246*

(**,*significant at 1% and 5% level respectively)

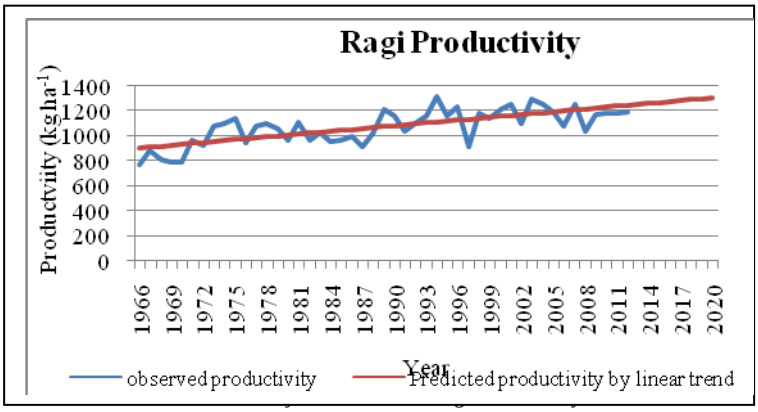
Charts for Area, Production and Productivity



Observed and Projected Trends of Ragi area in Andhra Pradesh



Observed and Projected Trends of Ragi Production in Andhra Pradesh



Bibliography

Ahmed, F., Shah, H., Raza, I and Saboor, A. 2011. Forecasting milk production in Pakistan. *Pakistan Journal of Agriculture Research.* 24 (1-4): 82-85.

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.

Nehru, S and Rajaram, S. 2009. ARIMA model to predict agricultural production. *Southern Economist.* 48 (5): 11-13.

Rankja N J, Upadhyaya S, Pandya H R, Parmar H R and Varmora S L 2010 Estimation of cotton based on weather of Banaskantha district in Gujarat state. *Journal of Agrometeorology.* 12 (1): 47-52.

Singh S S H, Hndl, Kaur P 2008 Effect of temperature and rainfall on wheat yield in south western of Punjab. *Journal of Agrometeorology.* 10 (1): 70-74.

