

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

## Production Analytics of Certain Crops of India

G. Sunitha<sup>1</sup>, K. Sampath Kumar<sup>2</sup>, S. A. Jyothi Rani<sup>3</sup>, V.V. Haragopal<sup>4</sup>

Assistant Professor, Department of Statistics, Vivekananda Govt. Degree College, Vidyanagar, Hyderabad,  
Telangana, India<sup>1</sup>

Assistant Professor, Department of Applied Statistics, Telangana University, Dichpally, Nizamabad, Telangana, India<sup>2</sup>

Assistant Professor, Department of Statistics, Osmania University, Hyderabad, Telangana, India<sup>3</sup>

Professor, Department of Mathematics, BITS Pilani, Hyderabad Campus, Jawaharnagar, Shamirpet, Telangana, India<sup>4</sup>

**ABSTRACT:** Agriculture is the most important occupation for most of the Indian families and has a significant role in Indian economy. It is the primary sector towards GDP growth. In India, agriculture contributes about 18% of total GDP. Agricultural growth has direct impact on poverty eradication; it is an important factor in containing inflation; raising farm wages and employment generation. In this paper a study has been undertaken to evaluate the production of certain cash crops viz., Chillies, Maize, Millet, Rice and Wheat in India by using different statistical models like ARIMA and Regression and comparing them with Artificial Neural Network Models (ANNs) to highlight which procedure fits well with the secondary data. The data analytics has been carried out by using secondary data available from [www.fao.org/faostat/en/#data/QC](http://www.fao.org/faostat/en/#data/QC). In this study Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are evaluated for ARIMA, Regression and ANN models, then comparison of these methodologies are evaluated, for Chillies, Maize and Rice crops. ANN is performing better than the traditional statistical models viz., ARIMA and Regression while for Millets and Wheat crops, Regression and ARIMA respectively are performing better.

**KEYWORDS:** GDP, ARIMA, ANN, RMSE, MAPE

### I. INTRODUCTION

Agriculture is the primary sector towards GDP distribution. It plays a strategic role in the process of economic development of a country. Agriculture has already made a significant contribution to the economic prosperity of advanced countries and its role in the economic development of less developed countries is of vital importance. It is the basic source of food supply of all the countries of the world whether underdeveloped, developing or even developed. Agriculture is the most important occupation for most of the Indian families and has a significant role in India economy. It is the means of livelihood for two-thirds of our population and is an important source of raw materials for several industries. The Indian agriculture has witnessed profound changes during successive decades of its planning, more particularly in the wake of new agricultural strategies adopted in the mid sixties.

In India, agriculture contributes about 18% of total GDP. Agricultural growth has direct impact on poverty eradication; it is important factor in containing inflation; raising farm wages and employment generation. Productivity levels in Indian agriculture are still much lower than the global standards. The share of agriculture in GDP has been constantly declining over the years. This was highlighted yet again in Economic Survey 2013-14, In eight years from 2000 to 2008 it has declined to 6.4 percent.

Forecasting is one of the main objectives of time series analysis which predict the future values using various forecasting models. It also helps in providing useful input for proper planning, especially, in agriculture which is full of

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

uncertainties. Now a day's agriculture has become highly input and cost intensive. Under the changed scenario, today forecasting of various aspects relating to agriculture has become essential. Crop yield forecasts are extremely useful in formulation of policies regarding stock distribution and supply of agricultural produce to different areas in the country. It is also used for forewarning of incidence of crop pests and diseases, weather forecasts, price forecast etc. Forecasting is central in making food policy decisions in developing countries. In the computer era, forecasting can be done with the help of sophisticated statistical software more efficiently.

Keeping in view of the above context, the present study has been undertaken to evaluate the production of certain crops viz., Chillies, Maize, Millet, Rice and Wheat in India by using different advanced statistical models and comparing them with Artificial Neural Network Models (ANNs).

## II. LITERATURE REVIEW

A review of the past research helps the researchers for identifying the conceptual and methodological issues related to the study. It enables the researchers to collect relevant data and carry out their research in proper meaningful interpretations and enables them to draw conclusions.

Shah et al. (2014) analyzed growth and trend in area, production and yield of major crops of Khyber Pakhtunkhwa by using a time series data from 1980-81 to 2011-12 of major crops (wheat, maize, rice and sugarcane). The results showed that area, production and yield of maize was increased over the time due to more area was brought under hybrid and improved open pollinated maize varieties. The area under rice crop has decreased whereas their production increased due the corresponding increase in per hectare yield of rice crop.

Hossain and Hassan (2013) studied the growth pattern of milk, meat and egg production in Bangladesh during the time period from 1991-92 to 2011-12. By using appropriate deterministic type growth model using latest model selection criteria. They revealed that the cubic model was the best fitted model for milk, meat and linear model was the best for egg production. The selected model was used for succeeding four years forecast with a 95% confidence interval of milk, meat and egg production. The analysis found that if the present growth rates continue then the production of milk, meat and egg would be 4.55, 3.77, and 7544.67 million tons, respectively, in the year 2015-16.

Sudha et al. (2013) studied to measure the growth trends of area, production and productivity of maize between 1970-71 to 2008-09 and to estimate the future projections up to 2015 AD by using the growth functions like linear, logarithmic, inverse, quadratic, cubic, compound, power And exponential. Based on highest coefficient of determination ( $R^2$ ) and its adjusted  $R^2$  they concluded that among all the models cubic function was found to be best fitted model for future projections of maize area, production and productivity.

Khan et al. (2013) was made an attempt to develop nonlinear statistical models viz., monomolecular, logistic, Gompertz, Richards and MMF for studying area, production and productivity of rice for three agro-climatic zones as well as whole Chhattisgarh State level for the period of 1975-76 to 2000-01. Various goodness of fit statistics viz.,  $R^2$ , RMSE, MAE, MSE were considered. The results revealed that for modelling production of rice, logistic, Gompertz and Monomolecular models came to be the most appropriate for the Northern hills and whole Chhattisgarh State. Gompertz and logistic models for Bastar plateau. For modelling productivity of rice logistic and Gompertz models were found to be the most suitable for the Northern hill, logistic for whole Chhattisgarh. Monomolecular and logistic models were found to be the best model for describing the area data of Chhattisgarh State. They found that, for all state monomolecular and logistic model came out to be the best fitted models.

Arunachalam and Balakrishnan (2012) made an attempt to study the trends in area, production and productivity of wheat crop grown during the period from 1950-1951 to 2009-2010 in India by fitting of Different non-linear models. The results indicated that none of the non-linear model was found suitable to fit the trends in area data. The Sinusoidal model was found suitable to fit the trends in production as well as productivity of wheat crop.

Sarika et al. (2011) conducted modelling and forecasting time-series data of pigeon pea production [*Cajanus cajan* (L.) Millsp.] in India, by using Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) during the period from 1969-70 to 2007-08. RMSE, AIC and BIC were used to identify the best model. The performance of fitted model was examined using Mean Absolute Error, Mean Per cent Forecast Error, Root Mean Square Error and Theil's inequality

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

coefficients. They concluded that ARIMA (2, 1, 0) model performed better among other models of ARIMA for modelling as well as forecasting purpose. Awal and Siddique (2011) estimated the growth pattern of rice production of different seasons in Bangladesh and to develop appropriate ARIMA models for the time series of Aus, Aman, and Boro rice production in Bangladesh. Based on model selection criteria like R2, RMSE, AIC, BIC, MAE, and MAPE are used to select the best model. The study revealed that ARIMA (4, 1, 4), ARIMA (2, 1, 1), and ARIMA (2, 2, 3) were selected as best fitted models for forecasting of Aus, Aman, and Boro rice production, respectively.

In this study a detailed analysis is carried out for the forecasting performance of univariate time series models for agriculture production, focusing on certain crops for which very long time series is available. Our main aim is to establish whether simple autoregressive (AR) models can still be used, or whether they should be substituted for more sophisticated specifications. Of course, multivariate time series models could be also considered, but this would entail a substantial increase of complexity due to the presence of non-linearity and time-varying parameters. Moreover, univariate linear models are often more robust than their multivariate counterparts, see e.g. Banerjee and Marcellino (2005). Artificial neural networks provide a powerful tool to approximate any kind of non-linear behavior, e.g., Hornik, Stinchcombe and White (1989).

To realize the various objectives of the study, data analytics for 56 years from 1961 to 2016 is considered. The important crops selected for the study are Chillies, Maize, Millet, Rice & Paddy and Wheat of India. The data has been collected from <http://www.fao.org/faostat/en/#data/QC>

### III. OBJECTIVES

1. To apply the traditional statistical models and compare them with ANN.
2. To estimate the parameters on production of certain agricultural crops in India using these statistical models.
3. To forecast the production of these crops by using best fitted model.

### IV. METHODOLOGY

In the next section We briefly explain the concepts of the above said three methodologies.

#### 1. ARIMA Model:

Forecasting can be carried out by applying the various Time series models. For this study Auto Regressive Integrated Moving Average (ARIMA) models are evaluated to study the crop pattern of various crops across India.

An ARIMA model is a difference stationary model where, the non stationary time series is reduced to stationary by selecting suitable order of differencing. Therefore, it is an ARIMA model with suitable differencing (d) and is referred as ARIMA (p, d, q) model. The concept of ARIMA model was developed by Box and Jenkins. ARIMA model essentially require identification of three constants p, d, q i.e. the order of AR terms (p), order of differencing (d) and the order of MA terms (q). Box and Jenkins established that these parameters can be obtained through trial and error approach after examining the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF). The fitting of ARIMA model consists of four steps and the procedure is well established and can be obtained from Box and Jenkins method.

The ARIMA (p, d, q) model is then formulated as

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t$$

Where,  $Z_t = Y_t - \bar{Y}$  (deviation of  $Y_t$  from mean  $Y$ ).

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

## 2. Artificial Neural Networks (ANN)

Although the forecasting of time series has generally been made under the assumption of linearity, which has promoted the study and use of linear models such as the autoregressive (AR), Moving Averages (MA), Autoregressive Moving Averages (ARMA) and Autoregressive Integrated Moving Averages (ARIMA), it has been found that in reality the systems often have unknown nonlinear structure for which several nonlinear models have been proposed, such as the Bilinear models, Autoregressive Conditional Heteroskedasticity (ARCH) and its extensions, Smooth Transition Autoregressive (STAR), Nonlinear Autoregressive (NAR), Wavelet Networks and Artificial Neural Networks (ANN), etc., while with respect to the ANN, it has received a great deal of attention over the last years and that its theory is very wide. They are being used in the areas of prediction and classification, areas where regression and other related statistical techniques have traditionally been used for exact predictions.

In order to assess the conditions and predictions under which various techniques should be used there is a need for more comparative studies between ANN and traditional statistical approaches. In this direction the present study is taken up. For the study Multilayer Perceptron is used which is perhaps the most popular network architecture in use today. A learning mechanism called error "backpropagation" (abbreviated "backprop") for feedforward networks are referred to as "multilayer perceptrons" (MLPs). The phrase "backpropagation network" is sometimes used to describe feedforward neural networks trained using the back propagation learning method. It is based on gradient descent where weights are modified in a direction corresponds to the negative gradient of an error measure.

## 3. Simple Linear Regression

Simple linear regression, also known as Linear regression, is a statistical concept often applied to economic and psychological data. While regression analysis seeks to define the relationship between two or more variables, in linear regression there are only two variables, the explained variable, represented by  $y$ , and the explanatory variable, represented by  $x$ . The model used to describe the relationship between a single dependent variable  $y$  and a single independent variable  $x$  is

$$y = \beta_0 + \beta_1 x + \epsilon.$$

Here,  $\beta_0$  and  $\beta_1$  are referred to as the model parameters, and  $\epsilon$  is a probabilistic error term that accounts for the variability in  $y$  that cannot be explained by the linear relationship with  $x$ .

## V. RESULT

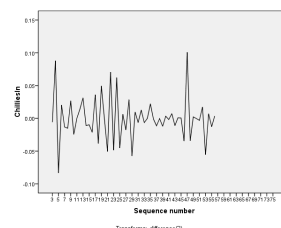
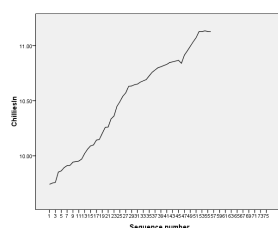
Keeping in view with the objectives formulated for the present study as discussed above, this section deals with the results of the performance concerning the forecasting methods. The predicted values are then used to estimate the errors and these are examined by computing RMSE and MAPE.

The data analytics obtained by the three methodologies for each of the five crops are compared and discussed here under:

### Chillies

#### 1) ARIMA :

The sequence plots for converting the non – stationary data into stationary are given in figure 1.



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

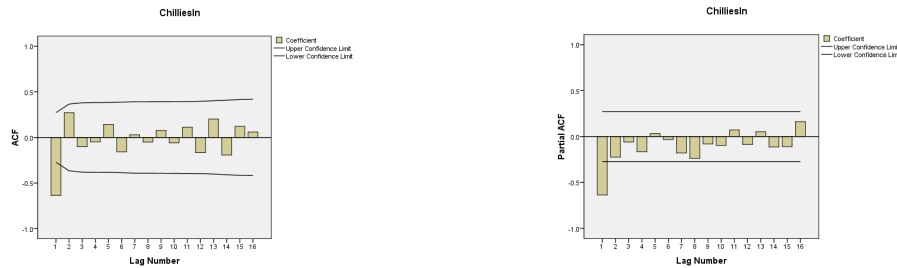


Figure (1)

Based on the above plots (Figure 1), for the production of Chillies ARIMA(2,1,2), ARIMA(1,2,2), ARIMA(2,2,1), ARIMA(1,2,1) and ARIMA(0,2,1) are considered as tentative models.

### Model Statistics

Model	Number of Predictors	Model Fit statistics					Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	RMSE	MAPE	Normalized BIC	Statistics	DF	Sig.	
(0,2,1)	0	.424	.996	.026	.192	-7.144	17.799	17	.402	0

Table 1

It is observed that none of these models are appropriate as these models are not showing any significance although the BIC value of ARIMA (0,2,1) model is minimum compared to the other tentative models.

### Model Parameters

				Estimate	SE	t	Sig.
ChilliesIn-Model_1	ChilliesIn	No Transformation	Constant	.000	.000	-1.046	.300
			Difference	2			
			MA Lag 1	.999	4.077	.245	.807

Table 2

After analysing the data on Chillies the model fitted is treated for forecasting the prediction of the crop and the same is given in Figure 2. In this figure the first half is the actual data representation while the second part is the prediction. Also, the ACF seem to be well within control.



Figure 2

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

From the results obtained by various ARIMA models it is observed that none of the models fits well for the chillies data. Hence, we tried estimating the forecasts with Regression method.

- 2) **Regression:** In this procedure we considered the Production as dependent variable and the Years of production as independent variable.

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson	
					R Square Change	F Change	df1	df2		Sig. F Change
1	.989 <sup>a</sup>	.978	.977	2446.49059	.978	2346.089	1	54	.000	.170

a. Predictors: (Constant), Years

b. Dependent Variable: Chillies and peppers green **Table 3**

### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.404E10	1	1.404E10	2.346E3	.000 <sup>a</sup>
	Residual	3.232E8	54	5985316.185		
	Total	1.437E10	55			

a. Predictors: (Constant), Years

b. Dependent Variable: Chillies and peppers green

**Table 4**

### Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	-1.909E6	40221.784		-47.455	.000			
	Years	979.701	20.227	.989	48.436	.000	.989	.989	.989

a. Dependent Variable: Chillies and peppers green

**Table 5**

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

From tables 3, 4 and 5 it is evident that the Chillies production over the years 1951 to 2016 is fitting perfectly and the production seems to be significantly different under the Indian context.

ANN:

Model Summary			
Training	Sum of Squares Error		15.964
	Average Overall Relative Error		.173
	Relative Error for Scale Dependents	Chilliesln	.007
Testing	Sum of Squares Error		4.411
	Average Overall Relative Error		.234
	Relative Error for Scale Dependents	Chilliesln	.020
Holdout	Average Overall Relative Error		.321
	Relative Error for Scale Dependents	Chilliesln	.030
a. Error computations are based on the testing sample.			

**Table 6**

Table 6 shows the summary of the results obtained by neural network analysis for the data. Now we compare the results obtained by the three methods using RMSE and MAPE.

**Comparison of the three methods:**

### CHILLIES

Methods	No. Of Observations	RMSE	MAPE
Regression	56	0.06325	0.49636
ANN	56	0.0313	1.1031

**Table 7**

From table 7, it is observed that the two techniques used viz., Regression and ANN fits well for the chillies data while the ARIMA model seem to underperform for the chillies data and it over fits the data. Moreover, since RMSE is minimum for ANN model, hence, it is concluded that the ANN model is the best model for forecasting the production of chillies crop.

Similar analysis is performed for Maize, Millet, Rice(paddy) and Wheat under Indian context, the results are presented here.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 2, February 2019

Crops	No. Of Observations	ARIMA		Regression		ANN	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Maize	56	0.158	0.746	0.145	0.730	0.0830	1.436
Millet	56	0.199	0.946	0.160	0.773	0.30	0.997
Ricepaddy	56	0.122	0.481	2.308	14.222	0.0891	1.2896
Wheat	56	0.131	0.571	0.1923	0.9636	0.0943	1.6685

## VI. CONCLUSIONS

From this data point of view, it is observed that the ANN methodologies seem to perform well for Chillies, Maize, Rice and Wheat production while Regression method gives a better performance for Chillies and Millets.

With these observations we can conclude that the production of various crops seems to be varying over the years 1961-2016 and the data out performs for ANN than Regression and ARIMA, which indicates that the production of these crops seem to be nonlinear and not linear thus, ANN is the best test for prediction.

## REFERENCES

- Anghelache, C., Sacală, C. Multiple linear regression used to analyse the correlation between GDP and some variables, Romanian Statistical Review, Supplement, no.9, pp. 94-99. (2016).
- Anghelache, C., Manole, A., Anghel, M.G. Unifactorial Econometric Model - Connection between the Final Consumption and the Private Consumption, Asian Academic Research Journal Of Social Science & Humanities, Volume 2, Issue 6, November 2015, pp.212-219. (2015).
- Anghelache, C., Anghel, M.G. GDP Analysis Methods through the Use of Statistical – Econometric Models, „ECONOMICA” Scientific and didactic journal, nr. 1 (91), Chişinău, Republica Moldova, pp. 124-130. (2015).
- Armstrong, J.S. (Ed), *Principles of forecasting, a handbook for researchers and practitioners*. Kluwer Acad.Press, Boston, Dordrecht, London, 2001.
- Atkinson, Tony. Atkinson Report: Final Report, Measurement of Government Output and Productivity for the National Accounts. New York: Palgrave Macmillan. 2005.
- Badmus, M.A., and Ariyo, O.S. 'Forecasting Cultivated Areas and Production of Maize in Nigeria Using ARIMA model', *Asian Journal of Agricultural Sciences* 3(3), 171- 176. (2011)
- Beck, T., Demirg-Kunt, A, Levine, R. Finance, Inequality and the Poor, *Journal of Economic Growth*, 12(1), pp. 27–49. (2007).
- Boken, V.K., Forecasting spring wheat yield using time series analysis: a case study for the Canadian Prairies, *Agronomy Journal*, 92(6),1047-1053. 2000.
- Bureau of Foreign and Domestic Commerce and National Bureau of Economic Research. National Income 1929–32, Report to Congress. Washington, DC: Government Printing Office. Available at <http://library.bea.gov>. 1934.
- Carbone, G. J., Kiechle, W., Locke, C., Mearns, L. O., McDaniel, L. and Downton, M. W. 'Response of soybean and sorghum to varying spatial scales of climate change scenarios in the southeastern United States', *Climatic Change* 60(1 - 2), 73–98. (2003).
- Dumitrescu, D., Anghel, M.G., Anghelache, C. Analysis Model of GDP Dependence on the Structural Variables, *Theoretical and Applied Economics*, Volume XXII, No.4 (605), Winter, pp. 151-158. (2015).
- Huang, B.N., Hwang, M.J., Yang, C.W. Causal relationship between energy consumption and GDP growth revisited: A dynamic panel data approach, *Ecological Economics*, Volume 67, Issue 1, 15 august 2008, pp. 41–54. (2008).
- Pal, S., Ramasubramanian, V. and Mehta, S.C. 'Statistical models for forecasting milk production in India', *Journal of Indian Society of Agricultural Statistics* 61(2), 80–83. (2007).
- Suleman, N. and Sarpong, S. 'Forecasting Milled Rice Production in Ghana Using Box-Jenkins Approach', *International Journal of Agricultural Management and Development*, 2(2), 79-84. (2012).
- Triplett, Jack E., and Barry P. Bosworth. *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Washington, DC: Brookings Institution Press. 2004.