Journal of AgriSearch, 6(4):205-210
An Open Access International Peer Reviewed Quarterly
ISSN : 2348-8808 (Print), 2348-8867 (Online)
https://doi.org/10.21921/jas.v6i04.16905

Predicting Rice Production using Autoregressive Integrated Moving Average Model

RC BHARATI* AND ANIL KUMAR SINGH

ARTICLE INFO

Received on	:	19-06-2019
Accepted on	:	25-10-2019
Published online	:	03-12-2019

ABSTRACT

A study was conducted on time-series data of rice production in India. Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) time-series process was considered for predicting country's rice production using the time series data from 1950–51 to 2017–18. Data from 1950–51 to 2014–15 were used for model development and three years data from 2015–16 and 2017–18 were kept for validation The augmented Dicky Fuller test was applied to test stationarity in data set. Based on ACF and PACF, the model was defined and tested for its suitability. Akaike information criterion and Bayesian information criterion were used to judge the suitability of the model to be fitted. The performance of the fitted model was examined using mean absolute error, mean percent forecast error, root mean square error and Theil's inequality coefficients. IMA (0, 1, 1) model performed well for forecasting purpose. The percent prediction error for the last three years i.e. from 2015–16 and 2017–18, was below 3%. The predicted values along with their standard errors up to the year 2099, were also obtained using the model.

KEYWORDS

ARIMA model, Box-Jenkins, Prediction, Rice, Time series data

INTRODUCTION

Rice (Oriza sativa) is first mentioned in the Yajur Veda (c. 1500-800 BC) and then is frequently referred to in Sanskrit texts. In India, there is a saying that grains of rice should be like two brothers, close but not stuck together. Rice is often directly associated with prosperity and fertility. Hence there is the custom of throwing rice at newlyweds. In India, rice is always the first food offered to the babies when they start eating solids or to husband by his new bride, to ensure they will have children. Rice production in India has increased during the last 68 years by nearly 5.48 times from 20.58 million tonnes in 1951 to almost 104.32 million tonnes during 2017-18 (Singh *et al.*, 2017). Rice is the most important crop of India and it occupies 23.3 percent of the gross cropped area of the country. Rice contributes 43 percent of total food grain production and 46 percent of total cereal production. It continues to play a vital role in the national food grain supply. It is the staple food of nearly half of the world population. It ranks third after wheat and maize in terms of worldwide production (Singh *et al.*, 2010).

There are many varieties of rice which are cultivated with differential response to climatic factors, such as temperature, rainfall and day length (Singh *et al.*, 2008). The soil types and different physiographic factors are also quite relevant in the cultivation of rice crop. The indica varieties of rice (Oryza sativa) are grown mainly in tropical countries . These varieties are photosensitive and the maturity period is affected with the date of planting. Rainfall is the most important weather element for the successful cultivation of rice. The distribution of rainfall in different regions of the country is greatly influenced by the physical features of the terrain, the situation of the mountains and plateau (Kumar *et al.*, 2013). Rainfall is the most critical weather factor to determine the paddy production because rain during the active phase of the initiation of panicle primordia is significantly beneficial. Thus, rain always gives beneficial effects even when this factor is taken jointly with other climatic elements, such as the mean temperature and sunshine. Therefore, rainfall is one of the most important climatic elements to determine the growth and yield of rice crop (Singh *et al.*, 2012). Temperature is another climatic elements which have favorable and in some cases, unfavorable influence on the development, growth and yield of rice (Singh *et al.*, 2012). Rice is a tropical and sub-tropical plant that requires a reasonably high temperature, ranging from 20° to 40°C. The optimum temperature of 30°C during day time and 20°C during night time seems to be more favorable for the development and growth of rice crops (Kumar *et al.*, 2013). The low temperature affects the tillering rate. The period of tillering is prolonged due to low temperature but low temperature gives more tillers and more panicles than higher temperatures. Low temperature depresses the internodal elongation and thereby induces the partial emergence of panicles (Singh and Singh, 2007). This phenomenon further affects the rate of photosynthesis and also causes partial sterility (Singh *et al.*, 2012).

However, low temperature during the period of ripening prolongs the ripening period and enables the plant to maintain green leaves. Such a condition contributes to the accumulation of carbohydrates in the grains. Sunlight is essential for the development and growth of the plants (Singh and Singh, 2007). In fact, sunlight is the source of energy for plant life. The response to solar radiation is a varietal character. The yield of rice is influenced by solar radiation particularly during the last 35 to 45 days of its ripening period. The effect of solar radiation is more profound where water, temperature and nitrogenous nutrients are not limiting

ICAR Research Complex for Eastern Region, Patna, Bihar, India
*Corresponding author email: drrcbharati@yahoo.com

factors. Bright sunshine with low temperatures during the ripening period of the crop helps in the development of carbohydrates in the grains. Solar radiation is a limiting factor for upland rice because upland rice is grown during the rainy season. Therefore, low productivity in the case of upland rice is a problem in the tropics. As rice production is greatly affected by climate, which is changing and is bound to change in the future, the prediction process is not an easy task. Accurate forecasting is important to both government and industry that needs to predict future production of foodgrains (Kumar *et al.,* 2013).

Such kind of exercise would enable the policy-makers to foresee the future requirements of rice, its import/export, thereby supporting them to take appropriate measures in this regard (Gujarati, 2003). The forecast would thus help save much of the precious resources of our country, which otherwise might be wasted. In many scientific or technical applications, data is generated in the form of time-series, thus making time-series analysis one of the significant tools in research and development. Since its inception, the univariate Box-Jenkins ARIMA approach is widely used throughout the world for different types of agricultural and industrial time-series analysis (Box *et al.,* 1994). The most significant point of this approach is that the explanatory variables in these models are the past values of the same variable. The models are constructed as a linear function of past values of the series and/or previous random shocks (or errors). It can be used when the series is stationary and there is no missing data within the time-series. Forecasts are generated under the assumption that the past. History can be translated into predictions for the future. This paper aims to develop a model from the observed rice data applying ARIMA methodology for uses in future forecasts.

MATERIALS AND METHODS

For the present study, India's rice time series production data from 1950–51 to 2017–18 were obtained from http://dacnet.nic.in (Agricultural Statistics at a Glance, 2008). Data from 1950–51 to 2014–15 were used for model development and three years data from 2015–16 and 2017–18 were kept for validation.

Autoregressive integrated moving average (ARIMA) methodology

In agricultural research, data are usually collected over time. Each observation of the observed data series, y_t was considered as a realization of a stochastic process $\{Y_t\}$, which is a family of random variables $\{Y_t, t\ T\}$, where T = { 0, ± 1, ± 2 ±,…}. Standard time-series approach was applied to develop an ideal model, which adequately represented the set of realizations and also their statistical relationships in a satisfactory manner. There are number of approaches available for forecasting time-series. In our study, we applied Box-Jenkins ARIMA modelling (Kumar 1990, Hossain *et al.,* 2006, Koutroumanidis *et al.,* 2009), which is one of the most widely used time-series prediction methods. This method uses a systematic procedure to select an appropriate model from a rich family of ARIMA models. Such models amalgamate three types of processes, viz autoregressive (AR) of order p, differencing of degree d to make the series stationary and moving average (MA) of order q, and is written as ARIMA (p, d, q). In general, its mathematical form is represented as follows:

$$\phi_p(B)(1-B)^d Y_t = c + \theta_q(B)\varepsilon_t$$

where, ϕ_p (B) and θ_q (B) are polynomials in B of degrees p and q respectively, c = constant; B = a backshift operator; d = order of difference operator; p = order of non-seasonal AR operator; and q = order of non-seasonal MA operator.

The conditions of stationarity and invertibility of the data under study were met only if all the roots of the characteristic equations ϕ_p (B)= 0, θ_q (B) =0 lie outside the unit circle. Choice of the most appropriate values for p, d and q is major problem in ARIMA modeling technique. In our study, this problem is partially resolved by performing prediction through the following stages: Model Identification Parameter estimation Diagnostic checking If model is satisfactory, forecasting is done else go to identification stage for further evaluation.

Model identification: Testing for stationarity and estimation of parameters as well as Orders of AR and MA components were determined. This is the preliminary, but very important step considered at this stage. This is to check whether or not the time-series under study met the condition of stationarity, since univariate ARIMA models are only applicable to stationary series (time-series with no systematic change in mean and variance). In order to test the stationarity (Pankratz, 1983), autocorrelation function (ACF) of difference series are computed. The series, in general is considered to be stationary if ACF for first and higher differences drop abruptly to zero, which is a heuristic approach. In our study, more statistically sound technique, viz augmented Dickey Fuller (ADF) test (Dickey and Fuller 1979) was applied to the data as such for testing the stationarity. SAS JMP V9.2 software package was used to calculate ADF test statistic.

Model estimation: Linear model coefficients were estimated. After identifying appropriate models in the identification stage, precise estimates of the parameters for chosen models were derived. For estimation of parameters Principles of Least Squares technique was used. Briefly, this estimation process produces new coefficients from some given initial values of coefficients in order to minimize residual sum of square. SAS JMP V9.2 software package was used to calculate ADF test statistic.

Model validation: Certain diagnostic methods were used to test the suitability of the estimated model.

Forecasting: The best model chosen was used for forecasting.

RESULTS AND DISCUSSION

Box-Jenkins ARIMA methodology resolved the problem of deciding appropriate values for p, d and q partially by following the steps described earlier. The preliminary step for fitting ARIMA model started with the stationarity test. A time series plot of rice production in India indicated that the series is non stationary (Fig. 1).

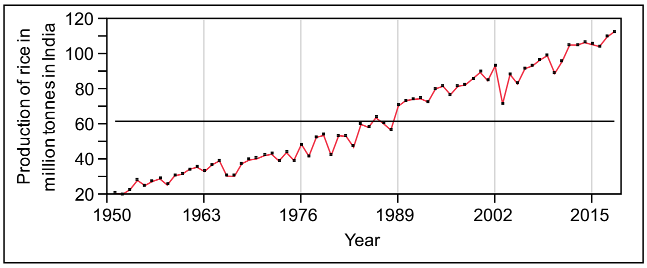
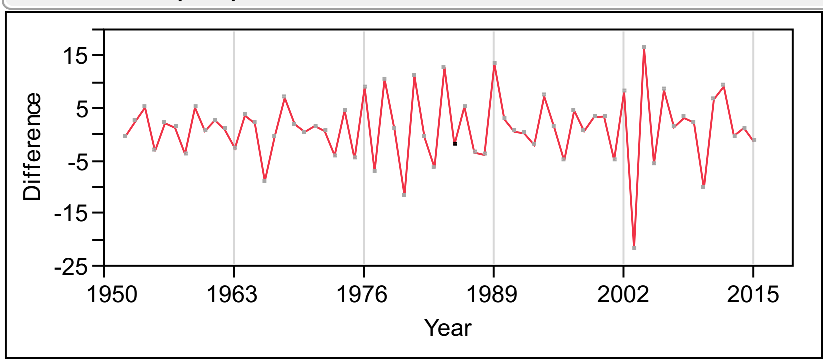


Fig.1: Time Series Production of rice in million tonnes in India

Since univariate ARIMA models are only applicable to stationary time-series data with no systematic change in mean and variance, a first order difference of the above data resulted in a stationary time-series data (Table 1). The Box-Ljung statistic reported insignificant values that were consistent with the hypothesis that residuals are random (Table 1).

Table 1: Out put Data of ACF and PACF of first order differenced time series data for rice production in India

Difference: (1-B)^1



Mean	1.3265625
Std	6.257563
N	64
Zero Mean ADF	-13.24997
Single Mean ADF	-14.42444
Trend ADF	-14.33281

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.5459	19.9819	<.0001*	1	-0.5459
2	0.1010	20.6774	<.0001*	2	-0.2806
3	0.0104	20.6848	0.0001*	3	-0.1116
4	-0.1816	23.0062	0.0001*	4	-0.3112
5	0.2270	26.6941	<.0001*	5	-0.0741
6	-0.1676	28.7392	<.0001*	6	-0.1361
7	0.0886	29.3214	0.0001*	7	-0.0886
8	0.1392	30.7830	0.0002*	8	0.1988
9	-0.2276	34.7613	<.0001*	9	0.0412
10	0.1332	36.1492	<.0001*	10	0.0178
11	-0.0270	36.2073	0.0002*	11	0.1245
12	-0.1096	37.1839	0.0002*	12	-0.0629
13	0.1058	38.1107	0.0003*	13	-0.1226
14	-0.0648	38.4652	0.0004*	14	-0.0691
15	0.0395	38.5995	0.0007*	15	-0.1323
16	0.0058	38.6024	0.0012*	16	-0.1428

The graph of the first difference time series data indicates the stationarity of the data. However, this was also confirmed by suitable statistical test values in different form of augmented Dickey Fuller (ADF) test (Dickey and Fuller, 1979) which was applied to the data as such for testing the stationarity ADFs. After taking the first difference, ADF test statistic 'tau' were found to be smaller than critical value, which made the series stationary. Hence, the value of d was assumed to be 1. The number of lags for this study was taken as n/4, is 16, where, n is the number of years for the time series. The dies out ACF after lag 1 and tails off PACF after lag 1 suggested ARIMA of the order 1,1,1. Since this approach is based on empirical values, the pattern of which may differ in other data set, various ARIMA models up to order 1 were considered for the selection of better model. In this case the different ARIMA models are ARIMA(0,0,0), I(1), AR(1), MA(1), IMA(1,1),ARI(1,1), ARMA(1,1) and ARIMA(1,1,1). The SAS JMP 9.2 ARIMA model comparison statistics along with residuals ACF and PACF of best model are depicted in Table 2.

Based on variances, AIC, SBC and R square of the group of models, it was found that IMA(1,1) was found to perform better. The pattern of ACF and PACF of the IMA(1,1) also suggested the statistical validity of the selected model. The statistical summary of the selected model indicated that invertibility bounds of the model had not been violated (Table 3) and the series is stationary (Table 1), it model was found to be stable and its parameters, i.e . MA1 and intercept significant at p<0.0001 (Table 3).

Table 2 : JMP software output for comparing ARIMA model

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare
✓	✓	IMA(1, 1)	62	24.349988	388.69596	393.01372	0.964
		ARIMA(1, 1, 1)	61	24.533336	390.14943	396.62607	0.964
		ARI(1, 1)	62	28.33315	397.95432	402.27209	0.958
		I(1)	63	39.778636	418.34936	420.50824	0.940
		ARMA(1, 1)	62	33.974989	420.63020	427.15336	0.911
		AR(1)	63	42.020649	432.79717	437.14594	0.902
		MA(1)	63	266.50707	550.55556	554.90433	0.597
		ARIMA(0, 0, 0)	64	682.29631	609.60940	611.78379	0.000

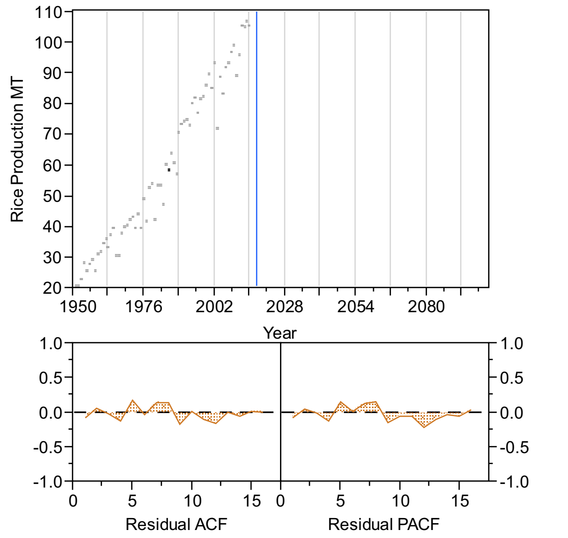


Table 3: Model summary and its parameter estimate of IMA(1,1) for predicting rice production in India

Model Summary

DF	62	Stable	Yes
Sum of Squared Errors	1509.69929	Invertible	Yes
Variance Estimate	24.3499885		
Standard Deviation	4.93457075		
Akaike's 'A' Information Criterion	388.695956		
Schwarz's Bayesian Criterion	393.013722		
RSquare	0.96409142		
RSquare Adj	0.96351225		
MAPE	6.53156129		
MAE	3.52037339		
-2LogLikelihood	384.695956		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
MA1	1	0.7364747	0.0802305	9.18	<.0001*	1.33303418
Intercept	0	1.3330342	0.1675676	7.96	<.0001*	

As a diagnostic check, the percent forecast error was computed with the actual values which were not included in the model. The model was developed with 65 years data and the last three years data i.e. data for the year 2015-16, 2016-17 and 2017-18 were used for the computation of forecast error (Table 4).

The model forecasted future production with a very low forecasting error. Hence, the model was used to predict rice production in India from the year 2015-16 to 2098-99 (Table 5).

Table 4 : Percent forecast error of rice production in India during 2015-16 to 2017-018

Year	Actual production of rice in India (million tonnes)	Forecast for rice production in India (million tonnes)	Forecast error (%)
2015 -16	104.41	106.97	2.45
2016 -17	109.70	108.31	1.27
2017 -18	112.91	109.64	2.90

Table 5: Table showing the predicted production of rice in India in million tonnes up to the year 2099.

Year	Prediction	Std. Err.	Year	Prediction	Std. Err.	Year	Prediction	Std. Err.
2016	106.97	4.93	2044	144.30	8.47	2072	181.62	10.91
2017	108.31	5.10	2045	145.63	8.57	2073	182.96	10.99
2018	109.64	5.27	2046	146.97	8.66	2074	184.29	11.06
2019	110.97	5.42	2047	148.30	8.76	2075	185.62	11.14

Year	Prediction	Std. Err.	Year	Prediction	Std. Err.	Year	Prediction	Std. Err.
2020	112.31	5.58	2048	149.63	8.86	2076	186.96	11.22
2021	113.64	5.73	2049	150.96	8.95	2077	188.29	11.29
2022	114.97	5.87	2050	152.30	9.05	2078	189.62	11.37
2023	116.31	6.02	2051	153.63	9.14	2079	190.96	11.44
2024	117.64	6.15	2052	154.96	9.23	2080	192.29	11.51
2025	118.97	6.29	2053	156.30	9.32	2081	193.62	11.59
2026	120.31	6.42	2054	157.63	9.41	2082	194.95	11.66
2027	121.64	6.55	2055	158.96	9.50	2083	196.29	11.73
2028	122.97	6.68	2056	160.30	9.59	2084	197.62	11.80
2029	124.30	6.81	2057	161.63	9.68	2085	198.95	11.88
2030	125.64	6.93	2058	162.96	9.77	2086	200.29	11.95
2031	126.97	7.05	2059	164.30	9.85	2087	201.62	12.02
2032	128.30	7.17	2060	165.63	9.94	2088	202.95	12.09
2033	129.64	7.29	2061	166.96	10.02	2089	204.29	12.16
2034	130.97	7.40	2062	168.29	10.11	2090	205.62	12.23
2035	132.30	7.52	2063	169.63	10.19	2091	206.95	12.30
2036	133.64	7.63	2064	170.96	10.27	2092	208.29	12.36
2037	134.97	7.74	2065	172.29	10.35	2093	209.62	12.43
2038	136.30	7.85	2066	173.63	10.44	2094	210.95	12.50
2039	137.63	7.95	2067	174.96	10.52	2095	212.28	12.57
2040	138.97	8.06	2068	176.29	10.60	2096	213.62	12.63
2041	140.30	8.16	2069	177.63	10.68	2097	214.95	12.70
2042	141.63	8.27	2070	178.96	10.75	2098	216.28	12.77
2043	142.97	8.37	2071	180.29	10.83	2099	217.62	12.83

Since uncertainty increases as prediction is made further from the data we have, the standard errors associated with predictions increases. Thus, it is advisable to use ARIMA methodology for the short-term forecast. The graph of prediction over the years is depicted in Fig. 2.

CONCLUSION

The results obtained from the above study is based on sound statistical foundation of Bo-Jenkins ARIMA modelling. The diagnostic checks has been performed at each level of model development. The result indicated that the increase in rice production in India will continue with same rate as it was in past. However, the confidence interval of the forecasted values will widen year after year. It was also seen that the production of rice in India will double by the end of century, which may be more that requirement.

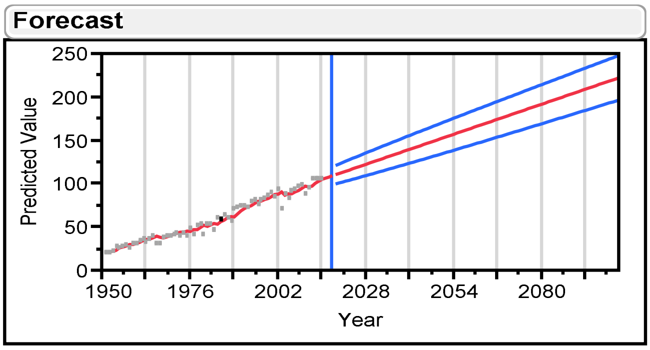


Fig. 2: Graph showing time series and prediction up to the year 2099

REFERENCES

Agricultural Statistics at a Glance. 2008. http://dacnet.nic.in

Box GEP, Jenkins GM and Reinsel GC. 1994. Time Aeries Analysis: Forecasting and Control, 3rd edn. pp 21–83. Prentice Hall, USA.

Dickey D A and Fuller W A. 1979. Distribution of the estimators for autoregressive time series with a unit root. Jounal of American Statistical Association 74: 427–31.

Gujarati D N. 2003. Basic Econometrics, 4th edn. pp 465–7. McGraw-Hill Companies Inc., New York.

Hossain M Z, Samad Q A and Ali M Z. 2006. ARIMA model and forecasting with three types of pulse prices in Bangladesh: a case study. International Journal of Social Economics 33 (4): 344–53.

Koutroumanidis T, Ioannou K and Arabatzis G. 2009. Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and hybrid ARIMA-ANN model. Energy Policy 37: 3627–34.

Kumar K. 1990. Some recent developments in time series analysis. Singapore Journal of Statistics 1: 45–73.

Kumar R , Pandey AK, Singh AK and Verma AK. 2013. Performance of rice genotypes under low land ecosystems of Jharkhand. Envi. & Ecol.31 (4): 1801-1805.

Pankratz A.1983. Forecasting with Univariate Box-Jenkins Models–Concepts and Cases, pp 119–54. John Willey, New

Singh AK and Singh Lal. 2007. Role of thermal time in rice phenology. Envi. & Ecol.25 (1): 46-49.

Singh AK, Chandra N and Bharti RC. 2012. Effects of Genotype and Planting Time on Phenology and Performance of Rice (Oryza sativa L.). Vegetos. 25 (1): 151-156.

Singh AK, Manibhushan, Chandra Naresh and Bharati RC. 2008. Suitable crop varieties for limited irrigated conditions in different agro climatic zones of India. Int. J. Trop. Agri. 26 (3-4): 491-496.

Singh AK, Singh AK, Kumar R, Prakash V Sundaram PK and Yadav SK. 2017. Indian Cereals Saga: Standpoint and Way Forward. Journal of AgriSearch 4 (1): 1-10.

Singh AK, Verma Nidhi, Tyagi Vandana and Dimree S. 2010. Indian needs of crop genetic resources – setting priorities. Prog. Agri. 10 (2): 1-16.

Citation:
Bharati RC and Singh AK. 2019. Predicting rice production using autoregressive integrated moving average model. *Journal of AgriSearch* 6(4): 204-210