iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XII **Month of publication:** December 2020

DOI: https://doi.org/10.22214/ijraset.2020.32613

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

Modeling and Forecasting of Rice Production in Some Major States of India using ARIMA

Tinni Chaudhuri[1], Prashant Verma[2], Mukti Khetan[3]
[1]Department of Statistics, Amity University Kolkata, India
[2]Department of Statistics, Banaras Hindu University, India
[3]Department of Mathematics, Indian Institute of Technology Bombay, India

Abstract: This paper focuses on the production of rice for last 61 years in India primarily dealing with the production variable of the ten highest rice harvesting states of the country. Using the Box Jenkins approach of model identification, the best possible ARIMA model is fitted among various competing models. Goodness of fit of each of these models is tested and the best model is used to forecast the production of rice for the coming years. In this regard, checking for series stationary is an important aspect which has been taken in account before proceeding with any analysis.
Keywords: ARIMA, forecasting, model identification, stationary.

I. INTRODUCTION

India ranks second in rice production in the world. It is also among the top rice consuming countries as rice is the staple crop in many Asian countries. It is believed to be the first cultivated crop in the world. Asia is leading in rice production accounting for over 90% of the total world production of milled rice. It is the second largest cereal produced in the world after corn. According to the Economic Times, for 2018-19, the agriculture ministry has fixed the production target for rice at 113 million tons, against 112.9 million tons last year. Currently, most of India's top rice suppliers and rice exporters are mainly based in regions such as West Bengal, Uttar Pradesh, Andhra Pradesh, Punjab, Tamil Nadu, Orissa, Bihar, and Chhattisgarh. These large rice producing states hold about 72% of the total rice-growing area in India. India exported around 8 million tons of rice in its last financial year. Saudi Arabia, UAE, Iran, South Africa and Senegal are some of the regular rice importing customers of India. Rice plantation is seen as a serious business module in the country. Though agriculture contributes to 18% of India's GDP, increase is rice production is mandatory for the agricultural GDP to grow. Muhammad et. al (1992) employed ARIMA model to forecast exports and internal consumption requirements.

Contreras et. al (2003) used ARIMA to forecast electricity prices in central Spain and other markets, though Stergiou (1989) used ARIMA modeling in a time series data for forecasting purpose. Hasan et. al (2008) measured change and instability in area and production of wheat and maize in Bangladesh. Mishra P et. al (2014) used the same for maize production in India. Nath B. et. al (2019) suggested in their paper of forecasting wheat production to check for the best ARIMA model after stationary has BEEN ACHIEVED FOR PREDICTION. THERE ARE ABOUT 20 STATES IN INDIA THAT PRODUCES RICE ACTIVELY COVERING AN ESTIMATED AREA OF 4000 Lakh Hectares. The following Table 1 gives top 10 largest rice producing states in India 2017-18, which contribute to 80% of total rice production in India.

Table I
Top 10 Rice Producing States of India (2017-18)

State	Production (Lakh tons)
Andhra Pradesh	128.95
Bihar	72.68
Assam	48.18
Haryana	41.68
Odisha	60.48
Madhya Pradesh	64.28
Punjab	105.42
Tamil Nadu	75.85
Uttar Pradesh	140.34
West Bengal	146.05

II. RESEARCH METHOD

A. Stationary

A time series Xt is said to be stationary if for any (t_1, t_2,………, t_k) and h>0,

$$F_{Xt1,Xt2,\ldots\ldots,Xtk}(x_{t1}, x_{t2},\ldots\ldots, x_{tk}) = F_{Xt1+h,Xt2+h,\ldots\ldots,Xtk+h}(x_{t1}, x_{t2},\ldots\ldots, x_{tk}).$$

More generally, if,

1) E(Xt) is independent of t.
2) The covariance matrix of (X_{t1}, X_{t2},………, X_{tk}) is independent of t.

To test the stationarity of each series the Augmented Dickey Fuller (ADF) test is performed. Also note that plotting each series state-wise helps in visual inspection as whether they are stationary. If the raw series fails, the unit root test of stationarity we use the different methods of changing it to a stationary series before proceeding to model fit.

B. Augmented Dickey Fuller Test/ Unit root test

It tests the presence of unit root in a time series, hence checking its stationarity. Given the data in hand, it is necessary to test for every state whether the data is stationary after having plotted them to confirm our visual interpretation. If the level of significance is 5%, reject Ho if p-value is less than 0.05, otherwise accept Ho which says that the series is non-stationary. If the series turns out to be non-stationary, we proceed to fit the ARIMA (p,q,d) model directly whereas if the data is stationary , we fit an ARMA (p,q) model.

C. Box-Jenkins method

Proceeding to Box-Jenkins method, the best model is explored by executing few major steps like model identification (roughly by looking at the ACF and PACF) followed by model fitting and estimation. It is of utmost importance to carry out diagnostic checking and finally forecasting or prediction concludes it.

D. Auto-Regressive Integrated Moving Average (ARIMA) Model

The most common and widely used stochastic model for time series forecasting is the ARIMA model. For a given, non-seasonal, non-stationary time series Xt having trend, the ARIMA model can be written as:

$Ø (B) (1-B)^d X_t = \Psi (B) \varepsilon_t$

where

d: No of times we need to differentiate Xt to detrend it, i.e. make it trend stationary

B: Backward shift operator, so,

$Ø (B): 1-\alpha_1 B-\alpha_2 B^2-\alpha_3 B^3-\ldots\ldots-\alpha_p B^P$,

$\Psi (B): 1+\beta_1 B+\beta_2 B^2+\beta_3 B^3-\ldots\ldots+\beta p B^P$, are characteristic polynomials in B

B: is the backward shift operator.

ε_t ~ WN (0,1); WN: White noise.

It is essential to plot the auto-correlation function (ACF) and partial autocorrelation function (PACF) of a time series. Since these are calculated from the given sample of data so it is often referred to as estimated ACF and PACF. These graphs for the differenced series help to understand the correlations between observations at different time lags, hence indicating the values of parameters p and q that we may consider while fitting AR(p) and MA(q) models. It is done by noting the cut-offs in ACF and PACF graphs of the stationary series. But this is only a tentative model selection procedure and we need to take various combinations of (p,q,d) to obtain the best possible model based on goodness of fit criteria.

E. Diagnostic check

In order to select the best model among the various combination of models , there can be various measures like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) , Mean Square Error (MSE) , Akaike's Information Criteria (AIC) , Bayesian Information Criteria(BIC) , estimated standard deviation etc. The Portmanteau test and Ljung-Box test are performed on the residuals to test the goodness of fit of the models.

1) *Portmanteau test:* The null hypothesis tested in this is the fit is good against an alternative of not a good model fit._The test statistic is given by:

Qp= $n\sum_{h=1}^{h*}\widehat{\varphi_r}^{2}(h) \sim \chi^2_{h*-(p+q)}$; asymptotically.

We reject the null hypothesis if Qp > $\chi^2_{\alpha,h*-(p+q)}$ where α= level of significance.

Here $\hat{\varphi}$r(h) = sample auto correlation function of lag h of the residual series.

p and q are the orders of AR and MA processes, h* is chosen such that h≤h* ∀ h.

2) *Ljung Box test:* Another similar test used to check goodness of fit in time series analysis is Ljumg Box test where the test statistic is given by:

Q*p=(n+2)$n\sum_{h=1}^{h*}\widehat{\varphi_r}^{2}(h)$ /(n-h) $\sim \chi^2_{h*-(p+q)}$; asymptotically.

The fit is not good if Q*p>$\chi^2_{\alpha,h*-(p+q)}$.

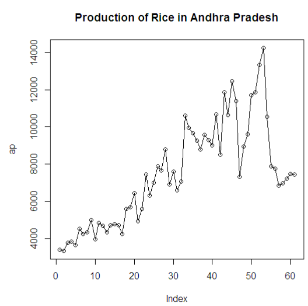
F. Model selection and Forecasting

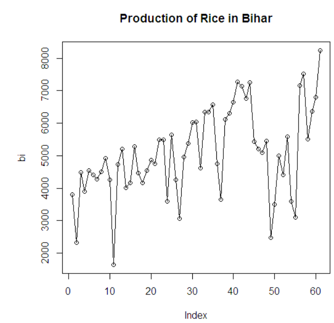
Based on the above measures we choose the best ARIMA model for each state and India. Our inference of choosing the best model is done based on the AIC value among the competing ones.

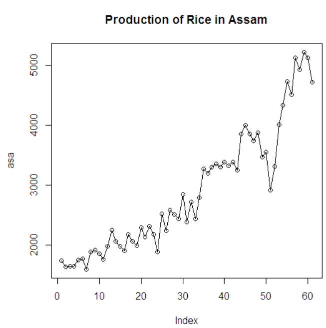
Forecasting is done based on the best selected model for the upcoming 10 years.

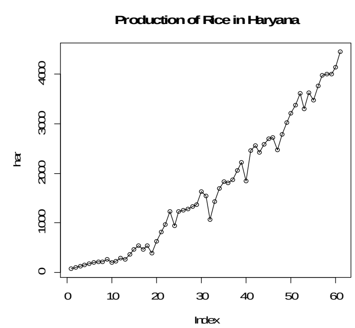
III. RESULTS AND DISCUSSION

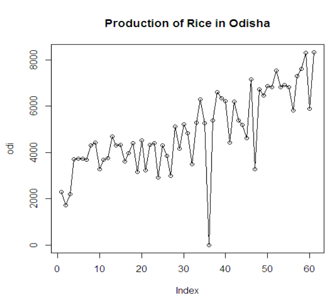
The series of rice production of the top 10 states in India and for entire "India" from 1957 to 2017 is shown below:

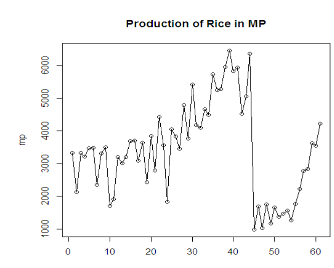












International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

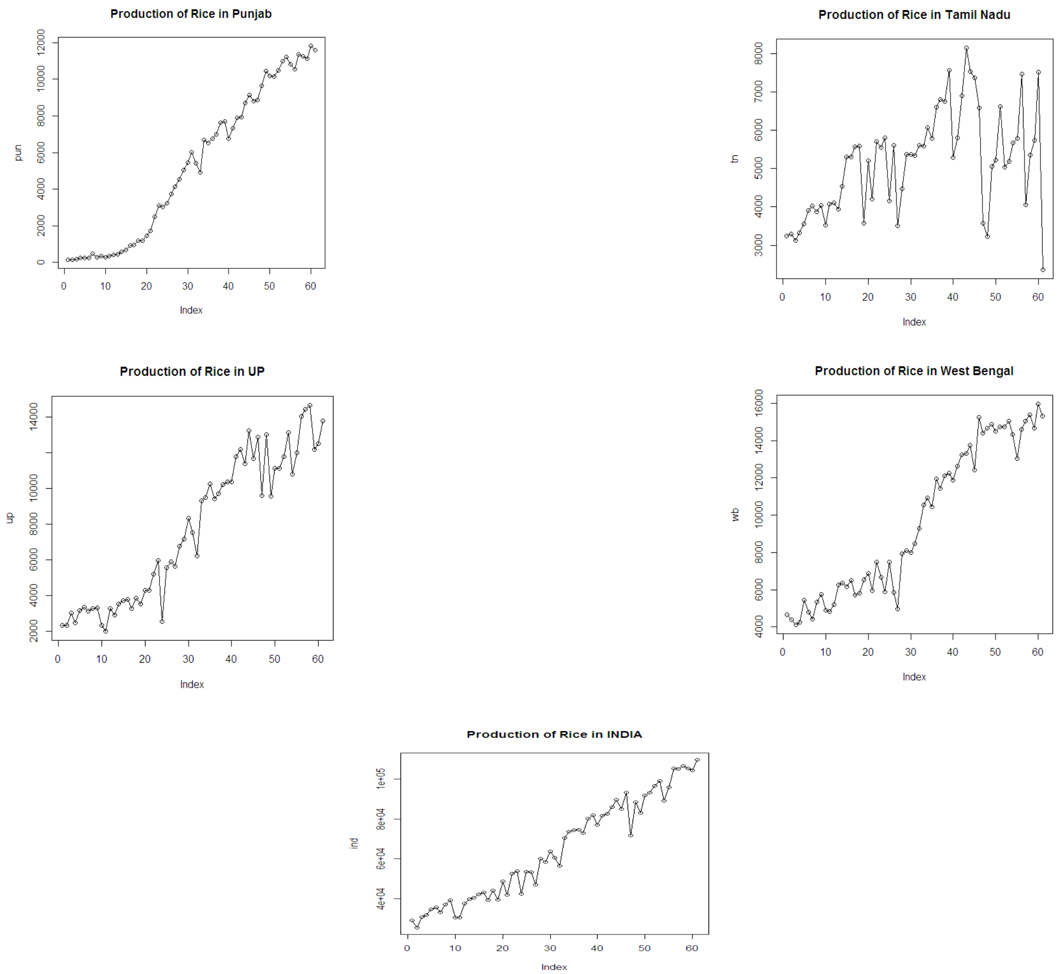


Fig. 1 Graphs of state wise rice production (1960-2018)

It is clear from the above graphs that all the series are *non-stationary*. We carry out the Augmented Dickey Fuller / Unit root test to be certain about this.

A. Unit root test/ADF Test

We want to test whether each of the series is stationary under this test. The null hypothesis is presence of unit root i.e. non-stationary.

TABLE 2. Unit Root Test

States	Test statistic value	P-value
Andhra Pradesh	-2.3717	0.4285
Bihar	-2.5748	0.3423
Assam	-3.071	0.1417
Haryana	-1.6253	0.7264
Odisha	-3.2616	0.08613
Madhya Pradesh	-1.8415	0.6389
Punjab	-2.5028	0.3715
Tamil Nadu	-2.2234	0.4845
Uttar Pradesh	-2.1434	0.5168
West Bengal	-1.5927	0.7396
India	-2.8789	0.2193

As it can be seen from Table 2, the p-value corresponding to each series is greater than 0.05 implying that all the series are non-stationary which establishes the correctness of our conclusion from previous graphs.

B. ACF and PACF Analysis

Here the main objective is to plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series to get a rough idea about the combination of models we need to consider while fitting ARIMA.

C. Model Fitting

The cut offs (spikes) from the ACF and PACF function of the stationary series we can take various combinations of p and q to fit ARIMA (p,d,q). Here d is 1, as we difference each of the series once to make it stationary (for all the states and India). We choose the best model looking at the value of Akaike's Information Criteria (AIC). Lesser the value of AIC, better is the fit.The following Table 3 gives the best fitted ARIMA model for each state along with the estimated parameters and some diagnostic measures.

Table 3

Best fitted model with parameter estimation and diagnostic values

State	Best fitted ARIMA model	Estimated parameters along with its standard error (s.e)	AIC	Normalized BIC	MAPE	RMSE	R-sqaure
Andhra Pradesh	(0,1,1)	ma1 -0.5981 s.e. 0.1247	1024.85	14.537	11.536	1339.823	0.762
Bihar	(1,1,2)	ar1 ma1 ma2 -0.6864 0.3175 -0.6825 s.e. 0.1091 0.1421 0.1343	1021.19	14.375	20.796	1154.299	0.310
Assam	(1,1,0)	ar1 -0.2642 s.e. 0.1255	848.77	11.350	7.389	272.263	0.932
Haryana	(1,1,2)	ar1 ma1 ma2 0.9949 -1.5645 0.6109 s.e. 0.0081 0.1329 0.1293	796.54	10.620	17.517	176.516	0.983
Odisha	(0,1,1)	ma1 -0.7675 s.e. 0.0689	1026.89	14.214	20.119	1139.878	0.540
Madhya Pradesh	(0,1,1)	ma1 -0.4318 s.e. 0.0998	1003.64	13.993	30.700	1020.774	0.514
Punjab	(0,1,3)	ma1 ma2 ma3 0.0465 -0.2121 0.3398 s.e. 0.1301 0.1110 0.1122	904.59	12.202	21.980	389.606	0.991
Tamil Nadu	(0,1,1)	ma1 -0.6989 s.e. 0.1166	1016.94	14.195	17.423	1129.030	0.289
Uttar Pradesh	(1,1,0)	ar1 -0.5109 s.e. 0.1099	1022.42	14.24	13.715	1155.676	0.920
West Bengal	(0,1,3)	ma1 ma2 ma3 -0.4485 -0.0390 0.3116 s.e. 0.1242 0.1316 0.1091	982.93	13.648	8.045	802.490	0.962
India	(1,1,2)	ar1 ma1 ma2 0.9985 -1.693 0.7073 s.e. 0.0067 0.090 0.0836	1206.12	17.716	4.299	5472.601	0.793

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

D. Diagnostic Check

1) *Ljung Box test for residuals:* Here given at 5% level of significance, if p-value is less than 0.05 we reject Ho implying that the SSE is large (or fit is not good) , otherwise we accept Ho i.e. SSE is small or fit is good. In the following Table 4 we have given the value of Ljung-Box test statistic along with its corresponding p-value for each series.

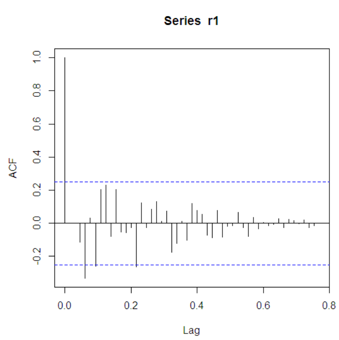
Table 4. Ljung-Box Test

States	Ljung Box test statistic	P-value
Andhra Pradesh	48.0257	0.553
Bihar	65.1785	0.0732
Assam	60.4439	0.148
Haryana	40.7768	0.8208
Odisha	36.2435	0.9277
Madhya Pradesh	37.5955	0.902
Punjab	36.2983	0.9267
Tamil Nadu	29.4834	0.9908
Uttar Pradesh	46.0762	0.6316
West Bengal	37.2823	0.9084
India	37.1049,	0.9119

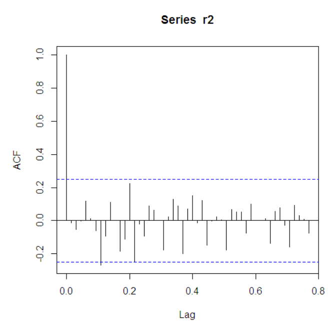
Clearly as we can see from Table 4, Ho is accepted in all the cases which mean that the models we have chosen for each state is appropriate.

2) *Residual Acf:* As we know a primary measure of goodness of fit is that errors of the fitted models should be uncorrelated. This is a desired property that ensures that the model fitted is apt. If the spikes of residual ACF are on both sides of the horizontal axis in a random pattern and gradually decreases to zero, we may call it an ideal situation. Hence, eventually as the lag increases the spikes get within the control band. We can see from the residual ACFs plotted in Fig.2, that all the ARIMA models fitted are proper.

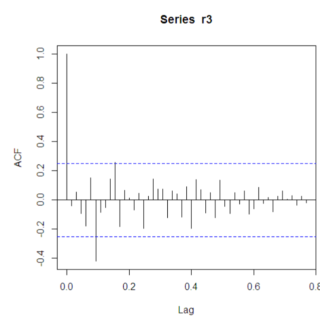
1. Andhra Pradesh



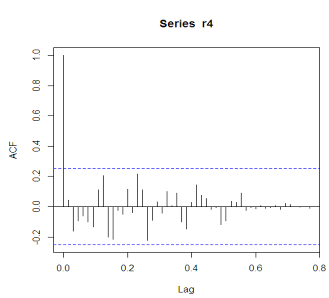
2. Bihar



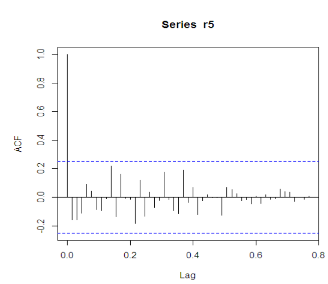
3. Assam



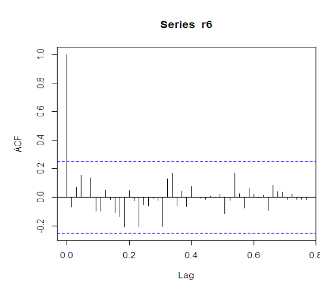
4. Haryana



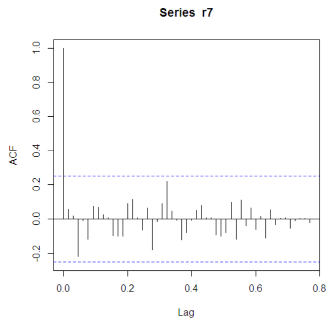
5. Odisha



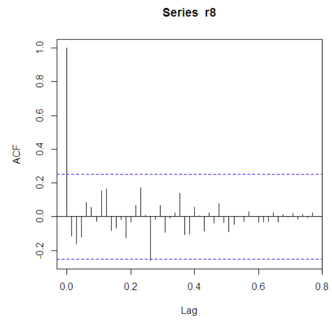
6. Madhya Pradesh



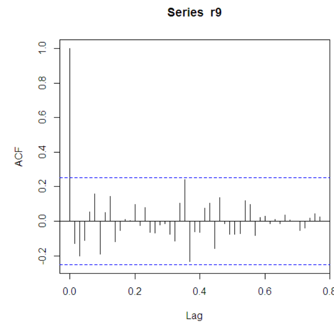
7.Punjab



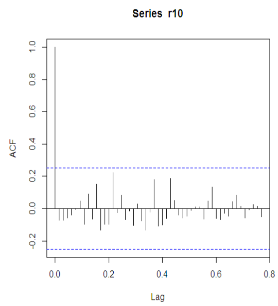
8. Tamil Nadu



9. Uttar Pradesh



10. West Bengal



11. India

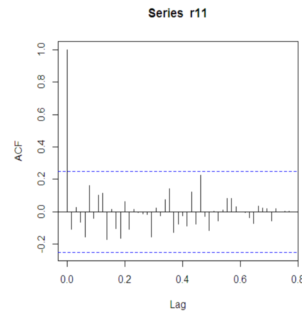


Fig. 2 Residual ACFs.

E. Forecasting

The forecasted values of rice production (in lakh tons) till 2026 is given in Table 5 for the top 10 states of India along with their upper and lower control limits.

Table 5
Forecasts with control limits

States		*2017-18*	*2018-19*	*2019-20*	2020-21	2021-22	2022-23	2023-24	2024-25	2025-26
Andhra Pradesh	U	10214.43	11001.54	11661.11	12243.66	12773.18	13263.30	13722.67	14157.16	14570.99
	F	7532.49	7600.31	7668.13	7735.95	7803.77	7871.60	7939.42	8007.24	8075.06
	L	4850.56	4199.09	3675.16	3228.25	2834.37	2479.89	2156.16	1857.31	1579.12
Bihar	U	9423.21	9499.50	9807.13	9893.42	10109.70	10220.45	10395.80	10519.88	10674.47
	F	7150.97	6866.26	7149.44	7046.69	7206.22	7187.50	7289.91	7310.01	7386.05
	L	4878.73	4233.03	4491.76	4199.97	4302.73	4154.55	4184.02	4100.13	4097.62
Assam	U	5466.10	5589.52	5782.00	5930.48	6076.70	6212.42	6342.77	6467.88	6588.94
	F	4921.24	4928.97	4995.26	5043.12	5096.78	5148.62	5201.03	5253.25	5305.54
	L	4376.38	4268.42	4208.51	4155.76	4116.86	4084.81	4059.28	4038.63	4022.14
Haryana	U	4760.07	4865.01	4979.50	5083.51	5188.84	5290.36	5391.14	5490.08	5588.02
	F	4406.55	4458.38	4539.31	4606.44	4680.12	4750.69	4822.73	4894.08	4965.76
	L	4053.02	4051.76	4099.11	4129.37	4171.39	4211.02	4254.32	4298.08	4343.49
Odisha	U	9367.77	9439.63	9511.47	9583.31	9655.14	9726.97	9798.79	9870.60	9942.41
	F	7118.54	7190.40	7262.26	7334.11	7405.97	7477.83	7549.68	7621.54	7693.40
	L	4869.32	4941.17	5013.04	5084.92	5156.80	5228.69	5300.58	5372.48	5444.39
Madhya Pradesh	U	5935.22	6256.23	6541.56	6801.36	7041.72	7266.64	7478.92	7680.58	7873.17
	F	3896.08	3910.51	3924.93	3939.36	3953.78	3968.21	3982.63	3997.06	4011.49

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XII Dec 2020- Available at www.ijraset.com

	L	1856.94	1564.78	1308.30	1077.36	865.84	669.77	486.35	313.54	149.80
Punjab	U	12475.03	13106.93	13262.65	13603.25	13927.45	14239.70	14542.73	14838.36	15127.89
	F	11695.67	12089.76	12186.09	12376.95	12567.81	12758.67	12949.53	13140.39	13331.25
	L	10916.30	11072.60	11109.53	11150.65	11208.16	11277.63	11356.33	11442.42	11534.60
Tamil Nadu	U	7438.53	7551.01	7660.69	7767.82	7872.65	7975.37	8076.14	8175.12	8272.43
	F	5178.75	5208.79	5238.82	5268.85	5298.88	5328.91	5358.94	5388.97	5419.00
	L	2918.98	2866.56	2816.95	2769.87	2725.10	2682.45	2641.74	2602.82	2565.57
Uttar Pradesh	U	15678.87	16408.02	16963.95	17532.54	18017.83	18507.69	18959.35	19405.07	19831.07
	F	13365.65	13859.16	13878.87	14153.14	14290.64	14501.62	14673.12	14865.83	15047.15
	L	11052.44	11310.30	10793.78	10773.73	10563.45	10495.55	10386.89	10326.59	10263.23
West Bengal	U	17285.55	17898.66	18010.98	18471.11	18900.14	19306.73	19696.16	20071.89	20436.39
	F	15680.92	16128.30	16132.70	16319.00	16505.30	16691.60	16877.90	17064.20	17250.50
	L	14076.29	14357.94	14254.42	14166.89	14110.46	14076.47	14059.64	14056.50	14064.61
India	U	119741.9	121542.2	123254.6	124967.7	126664.2	128352.6	130031.2	131701.7	133364.2
	F	109415.9	111030.9	112305.3	113681.5	115027.3	116382.2	117734.4	119087.4	120440.2
	L	99089.93	100519.5	101356.0	102395.3	103390.5	104411.9	105437.6	106473.2	107516.2

U: UCL F: Forecast L: LCL

We plot the forecasted values state-wise from 2018-2026.

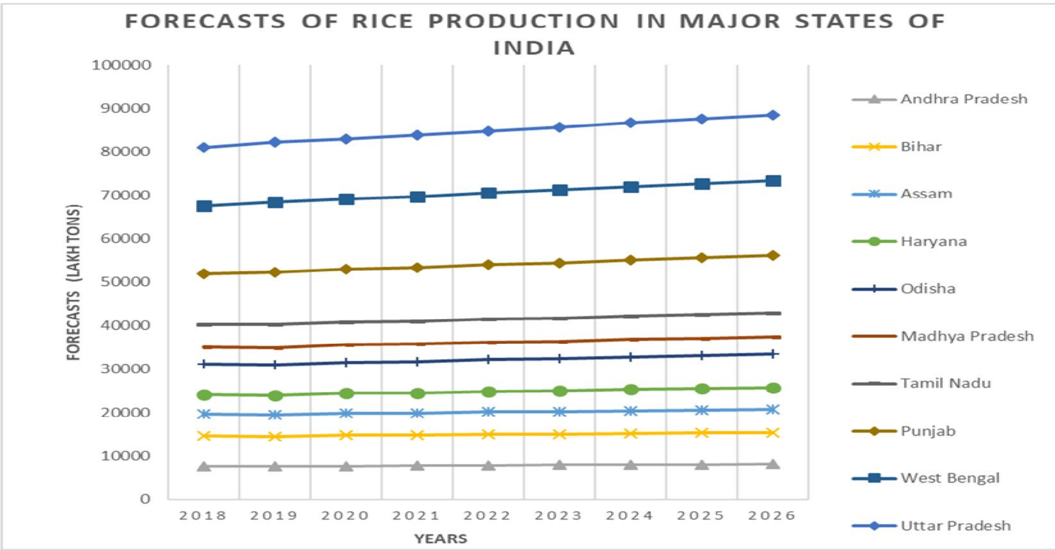


Fig. 3 Forecasts of Rice Production.

IV. CONCLUSIONS AND RECOMMENDATIONS

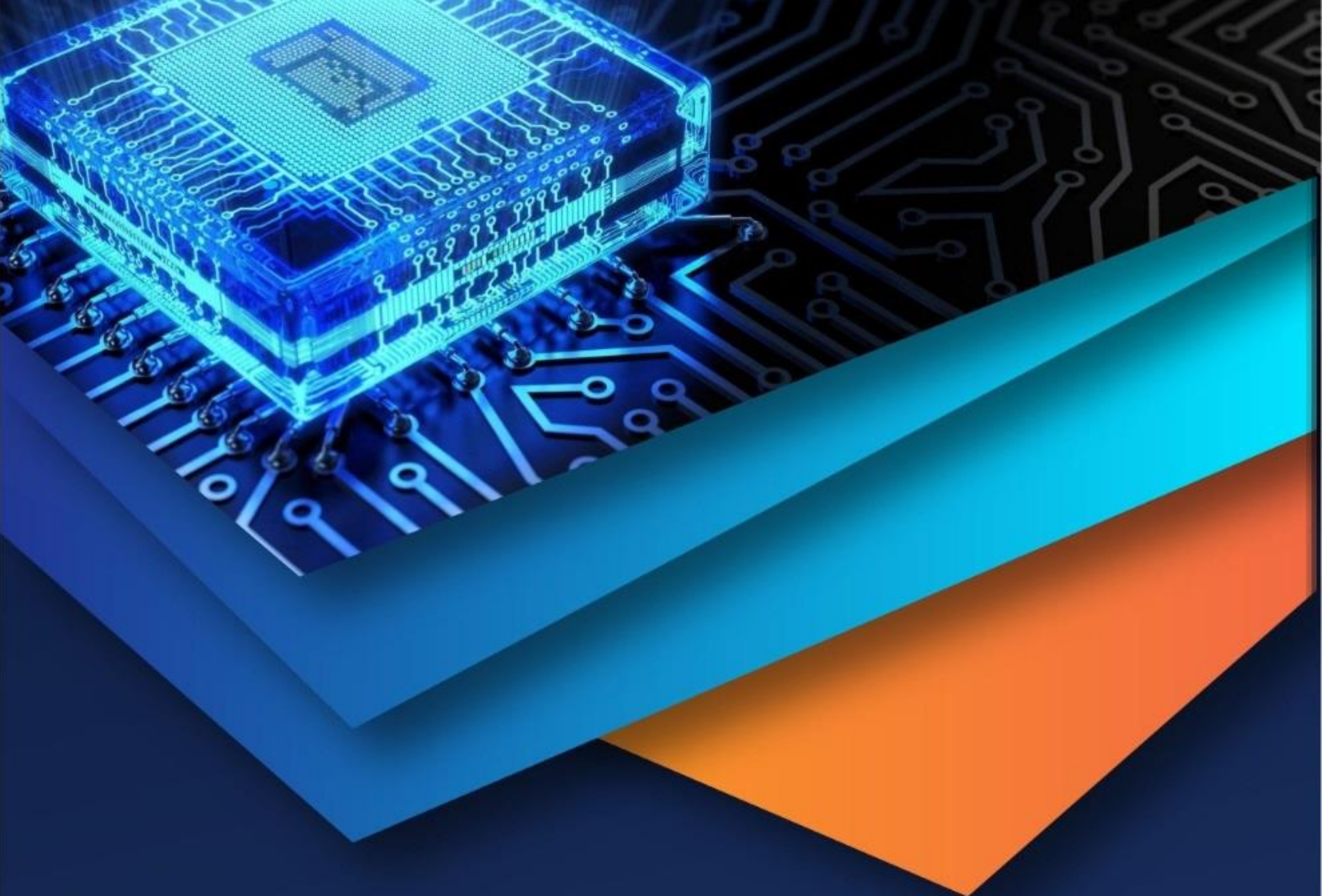
The condition of rice production in India over the past years has been quite consistent. Due to increased demand of rice, there has been steady and continuous growth in the production over the years. States like Andhra Pradesh, Assam, Haryana, Punjab, Uttar Pradesh, West Bengal have seen major and proportionate rise in the yield of this major crop as population has increased gradually over the years. The future prediction trend lines of every state also provide evidence in support of the same. We have used the best ARIMA fitted model for prediction which is different for every state. This asserts the fact that production will increase in a congruous manner over the future years till 2026. However, there has been significant contribution in production of rice from other states as well. Agriculture being an essential part of our national economy, government has implemented various measures and improved technologies to increase the productivity of rice in order to meet the need. However, this demand will go further up in future and to suffice for that cultivation of hybrid rice needs to be promoted.

V. ACKNOWLEDGEMENT

We wish to extend our heartfelt gratitude to Sweta Dixit of School of Management, SRM Institute of Science and Technology and Vijay Sharma of Department of Mathematics, Indian Institute of Technology Bombay for helping us to get the dataset and proceed with the work. We also thank all our colleagues and friends for their help and support.

REFERENCES

[1] Box, G E P and Jenkins G M. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco. 1976.
[2] Cuddy J D A and Della Valle P A. Measuring the instability of time series data. Oxford Bulletin of Economics and Statistics 40 (1): 79-85. 1978.
[3] Fischer R A, Derek Byerlee and Edmeades G O. Can technology deliver on the yield challenge to 2050? Expert meeting on how to feed the world in 2050, Food and Agriculture Organisation of the United Nations Economic and Social Development Department. 2009.
[4] Gujarati D N. Basic Econometrics, McGraw Hill, New York. 2003
[5] Nasurudeen P, Anil Kuruvila, Sendhil R and Chandrasekar V. The dynamics and inequality of nutrient consumption in India. Indian Journal of Agricultural Economics 61 (3): 362-370. 2006
[6] Hasan, M N, Mohayem Miah M A, Islam M S, Islamand Q M, Hossain M I. Change and Instability in area and production of Wheat and Maize in Bangladesh. Bangladesh Journal of Agricultural Research, 33(3), 409-417. 2008
[7] Paroda, R S and Kumar P. Food production and demand in south Asia. Agriculture Economics Research Review, 13, 1-24. 2000
[8] Mishra P., Sahu P.K. Forecasting Area, Production and Yield of Maize in India using ARIMA Model. Int. J.Agricult.Stat, Vol 10,No.2, pp.425-435.2014
[9] Nath B, Dhakre D.S and Bhattacharya D. Forecasting wheat production in India: An ARIMA modelling approach. Journal of Pharmacognosy and Phytochemistry 8(1):2158-2165. 2019
[10] Kumar M, Anand M. An application of time series ARIMA forecasting model for predicting sugarcane production in India, Studies in Business and Economics. ; 9(1):81-94. 2014.
[11] Ljung GM, Box GEP. On a measure of lack of fit in time series models, Biometrika; 67:297-303. 1978
[12] Tripathi A, Mishra AK. The Wheat Sector in India: Production, Policies and Food Security, the Eurasian Wheat Belt and Food Security. 275-296. 2017
[13] Ramdas S., Singh R and Sharma I. Exploring the performance of wheat production in India. J. Wheat Res. 4(2): 37-44.2014.
[14] Contreras J, Espinola R, Nogales FJ, Conejo AJ. ARIMA Models to Predict Next- day Electricity Prices, IEEE Transactions on Power Systems, 18(3):1014-1020; 2003.
[15] Stergiou KI. Modeling and forecasting the fishery for pilchard (Sardina pilchardus) in Greek waters using ARIMA time-series models, ICES Journal of Marine Science. 46(1):16-23;1989.
[16] Muhammad F, M Siddique, M Bashir, S Ahmad. Forecasting rice production in Pakistan-using ARIMA Models, J of Animal Plant Sci; 2:27-31. 1992.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY