Palestine Journal of Mathematics

Vol 14(Special Issue II)(2025) , 248–256

Design and Development of ARIMA Model for *Allium Cepa* Production in India

T. Jai Sankar and P. Pushpa

MSC 2010 Classifications: Primary 60Gxx; Secondary 62Fxx..

Keywords and phrases: ARIMA, Forecasting, Box-Ljung, BIC, A. *cepa*, Production.

The authors would like to thank the reviewers and editor for their constructive comments and valuable suggestions that improved the quality of our paper.

Abstract One of the world's most widely produced and consumed bulbous spices is *Allium Cepa (Onion)*, primarily grown in India. It can be used medicinally to treat a wide range of illnesses. India comes in second place globally in terms of A. *cepa* output, with China being the top producer. Based on historical data from 1978 to 2021, the current study focuses on the design and development of ARIMA model for A. *cepa* production in India. Using Autoregressive Integrated Moving Average (ARIMA), stochastic modelling and associated forecasting were carried out. Based on the lowest value of the Bayesian Information Criterion (BIC), ARIMA (1,1,0) was determined to be appropriate for A. *cepa* production data. Root Mean Square Error (RMSE), Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), Normalized BIC, Box-Ljung Q statistics, and Mean Absolute Percentage Error (MAPE) were used to examine the forecast accuracy of ARIMA models. According to the selected model, the production of A. *cepa* is predicted to increase from 31.27 million tonnes in 2021 to 46.74 million tonnes in 2030.

1 Introduction

Allium cepa (Onions) are one of the most extensively grown and eaten vegetables in the world. They are members of the Allium genus, which also contains chives, shallots, garlic, and leeks. A. *cepa* are a common element in a wide variety of meals due to their strong flavor and culinary adaptability. A. *cepa* have been cultivated for thousands of years; proof of its consumption may be found in the records of the ancient Egyptian, Greek, and Roman civilizations. A. *cepa* are now grown in a variety of temperatures and geographical areas, making them available to people all over the world. This adaptable vegetable is prized for its nutritional properties in addition to its flavor. Vitamins, minerals, and antioxidants abound in A. *cepa*, promoting general health and wellbeing. To further enhance their health advantages, they also have antibacterial and anti-inflammatory characteristics. A. *cepa* are not only important in food and nutrition, but they are also vital to agricultural economics. In many nations, the cultivation of A. *cepa* is a major industry that supports farmers' livelihoods and promotes both domestic and foreign trade. This introduction lays the groundwork for examining a range of A. *cepa* production-related topics, such as growing methods, worldwide production patterns, growers' obstacles, and industry advancements. A. *cepa* is richer in manganese, copper, phosphorus, potassium, magnesium, calcium, and zinc (Figure 1).

India is renowned for its rich agricultural heritage with A. *cepa* playing a pivotal role in sustaining its vast population and driving economic growth. The production of A. *cepa* in India holds immense significance, not only in ensuring food security but also in contributing significantly to the nation's GDP and employment generation. India has become one of the world's largest producers of various food grains. Historically A. *cep* have been the backbone of Indian agriculture, with staple crops like rice, wheat, maize, millets, and pulses forming the bedrock of the country's agricultural landscape. India's A. *cepa* production has witnessed significant growth

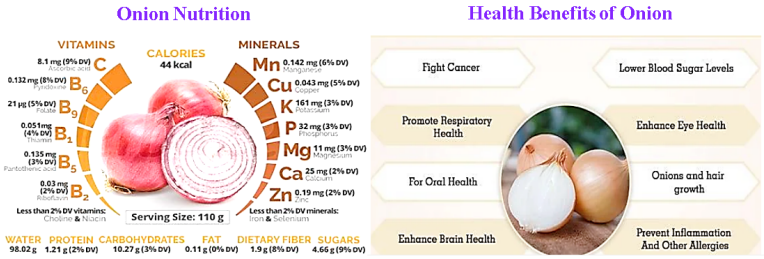

Figure 1. Nutrition and Health Benefits of A. *cepa.*

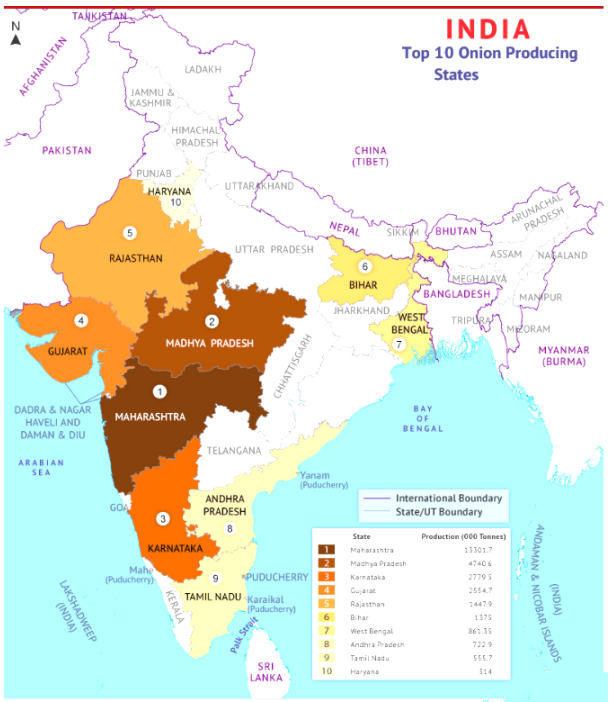

Figure 2. India's Growing States of A. *cepa.*

over the years, owing to various factors such as technological advancements, government policies, and agricultural reforms. The objective of the study is to develop appropriate ARIMA models for the stochastic modeling for forecasting of A. *cepa* production in India and to make seven year forecasts with appropriate prediction interval.

2 Material and Methods

The current behavior of the variable under investigation is described by the ARIMA (Auto Regressive Integrated Moving Average) model in terms of linear relationships with its historical values. All that is needed for this extrapolation method is historical time series data for the variable being studied. The main purpose of ARIMA models is to forecast the associated variable. ARIMA residual autocorrelations were measured by Box and Pierce (1970) [1]. Slutzky (1973) applied Moving Average (MA) models [2]. As described by Akaike (1983), the stationary time series is defined as being bounded by the same integer [3]. A model based on ARIMA (1,1,5), ARIMA (0,1,5) and ARIMA (1,1,4) was developed by Mishra et al. (2013) for forecasting area, production and productivity of A. cepa in India for the period from 1978 to 2008 [4]. According to Jai Sankar and Pushpa (2019), $Saccharumofficinarum$ production in India was validated into the period 1950-2017 and ARIMA (2,1,0) model was applied up to 2022 [5]. Jai Sankar and Pushpa (2020) considered ARIMA (0,1,1) model for stochastic forecasting analysis for peanut $(Arachishypogaea)$ production in India during the years 1950-2017 [6]. For tea production in

Assam from 1957 to 2016, Sakuntala Deka Umar et al. (2021) developed an ARIMA (0,2,1) model, which forecasts through the upcoming 10 years [7]. Balaga Divya and Abhiram Dash (2022) found ARIMA (1,1,2), ARIMA (1,1,0), ARIMA (1,1,0) model to forecast the area, yield and production of arhar in Odisha [8]. Rama Shankar Yadav et al. (2022) considered an application of time series ARIMA forecasting model for predicting nutri cereals area in India during the period of 1951 to 2020 [9]. According to Sameerabanu and Sekhar (2022) onion yield in India was validated during the periods 1978-2020 and an ARIMA (1,1,1) model was applied up to 2032 [10]. Sudhir Paswan et al. (2022) considered ARIMA (1,1,0) model for time series prediction for sugarcane production in Bihar using ARIMA and ANN model during the years from 1939-40 to 2019-20 [11]. The ARIMA (1,1,0) model was used by Tichaona W. Mapuwei et al. (2022) for an application of time series ARIMA forecasting model for predicting tobacco production in Zimbabwe for the period from 1980 to 2018 [12]. Yashpal Singh Raghav et al. (2022) applied ARIMA (0,1,1) and ARIMA (1,1,0) model to forecasting of pulses production in South Asian countries [13]. Bhusanar and Satyveer Singh Meena (2023) found with ARIMA (0,1,1) model to forecast groundnut area, production and productivity in Rajasthan from 1991 to 2019 [14]. Jai Sankar and Pushpa (2023) calculated ARIMA (0,1,2) model for implementation of stochastic time series forecasting ARIMA model for $Hordeumvulgare$ production in India during the years from 1960 to 2020 [15]. Statistically independent and normally distributed residuals were significant features of stochastic time-series ARIMA models (Alan Pankratz, 1983) which were widely used to analyze time series data [16].

In this study, a four-step ARIMA model was used, consisting of identification, estimation, diagnostic checking, and forecasting. Model parameters were considered to fit the ARIMA models.

$$\text{AR process of order (p) is,} Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t;$$

$$\text{MA process of order (q) is, } Y_t = \mu - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t;$$

and ARIMA process of order (p, d, q) is,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \mu - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t;$$

Where $Y_t - A.cepa$ production, ϵ_t's - independently and normally distributed with zero mean and constant variance σ^2 for $t = 1, 2, ..., n; d-$ the fraction differenced while interpreting AR and MA, and ϕ's and ϕ's - coefficients to be valued.

Trend Fitting: The Box-Ljung Q statistics was used to convert the non-stationary data into stationarity data and also to validate the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, a range of reliability statistics like R squared, Stationary R squared, RMSE, MAPE and BIC were applied. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_l)^2]^{\frac{1}{2}} \text{and} MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_l}{Y_i}\right|$$

$$BIC(p,q) = Inv^*(p+q) + (p+q)\left[\frac{In(n)}{n}\right]$$

where p and $q-$ order of AR and MA processes; n - number of observations; and v* - approximate of white noise variance σ^2.

$$Q = \frac{n(n+2)\sum_{i=1}^{k} rk^2}{(n-k)}$$

where n - number of residuals and rk - residuals autocorrelation at lag k.

In this analysis, the data on A. $cepa$ production in India were collected from the Annual Report Agricultural Statistics at a Glance 2022, Government of India for the period from 1978 to 2021 (Table 1) and were applied to fit the ARIMA model to predict the future production.

Table 1. Actual A. *cepa* production (million tonnes) in India

Year	Production	Year	Production	Year	Production
1978	2.20	1993	4.01	2008	13.48
1979	2.50	1994	4.04	2009	12.16
1980	2.50	1995	4.08	2010	15.12
1981	2.65	1996	4.18	2011	17.51
1982	2.43	1997	3.62	2012	16.81
1983	2.70	1998	5.33	2013	19.4
1984	3.10	1999	4.90	2014	18.93
1985	2.86	2000	4.55	2015	20.93
1986	2.53	2001	4.83	2016	22.43
1987	2.70	2002	4.21	2017	23.26
1988	3.35	2003	5.92	2018	22.82
1989	3.07	2004	6.43	2019	26.09
1990	3.23	2005	8.68	2020	26.64
1991	3.58	2006	8.89	2021	31.27
1992	3.49	2007	9.14		

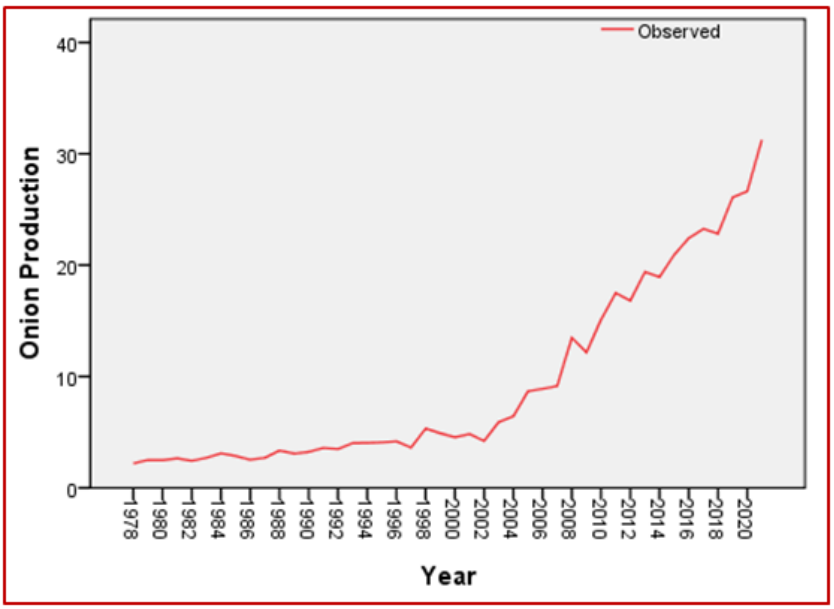

Figure 3. Time plot of A. *cepa.* Production

3 Results and Discussion

In this analysis, to fit an ARIMA model, the process for any variable involves four steps: identification, estimation, diagnostic and forecasting. ARIMA (p,d,q) is steady to make certain stationarity through reading the graph or time plot of the given data. Figure 3 suggests that the data is non-stationary. The autocorrelation and partial autocorrelation coefficients of various orders of Yt are calculated (Table 2). The graphs of ACF and PACF are produced (Figure 3). The models and corresponding BIC values are specified (Table 3). The value of normalized BIC is 0.333 and R squared value is 0.986 in the most appropriate model for A. cepa production is ARIMA(1,1,0) as this model has the lowly BIC value.

Table 2. ACF and PACF of A.*cepa* Production

Lag	AC	Std. Error (white noise)	Box-Ljung Statistic	PAC	Std. Error	Lag	AC	Std. Error (white noise)	Box-Ljung Statistic	PAC	Std. Error
	Value	Df	Sig. (Chi-Square Approx.)	Value	Df		Value	Df	Sig. (Chi-Square Approx.)	Value	Df
1	0.896	0.146	37.803	0.896	0.151	**17**	-0.128	0.115	189.278	-0.075	0.151
2	0.827	0.144	70.757	0.121	0.151	**18**	-0.154	0.113	191.130	0.017	0.151
3	0.748	0.142	98.360	-0.062	0.151	**19**	-0.180	0.111	193.752	-0.007	0.151
4	0.682	0.141	121.885	0.012	0.151	**20**	-0.193	0.109	196.907	-0.002	0.151
5	0.609	0.139	141.156	-0.054	0.151	**21**	-0.209	0.107	200.747	-0.031	0.151
6	0.529	0.137	156.071	-0.095	0.151	**22**	-0.224	0.104	205.380	-0.020	0.151
7	0.454	0.135	167.355	-0.034	0.151	**23**	-0.243	0.102	211.078	-0.091	0.151
8	0.388	0.133	175.805	0.000	0.151	**24**	-0.268	0.099	218.327	-0.065	0.151
9	0.307	0.132	181.268	-0.113	0.151	**25**	-0.281	0.097	226.744	0.000	0.151
10	0.243	0.130	184.786	0.006	0.151	**26**	-0.295	0.094	236.546	-0.056	0.151
11	0.167	0.128	186.491	-0.086	0.151	**27**	-0.308	0.092	247.855	-0.006	0.151
12	0.100	0.126	187.118	-0.045	0.151	**28**	-0.317	0.089	260.529	-0.031	0.151
13	0.051	0.124	187.287	0.056	0.151	**29**	-0.325	0.086	274.811	-0.034	0.151
14	-0.015	0.122	187.301	-0.112	0.151	**30**	-0.331	0.083	290.619	-0.014	0.151
15	-0.051	0.120	187.483	0.062	0.151	**31**	-0.326	0.080	307.212	0.013	0.151
16	-0.088	0.118	188.049	0.004	0.151	**32**	-0.324	0.077	324.895	-0.015	0.151

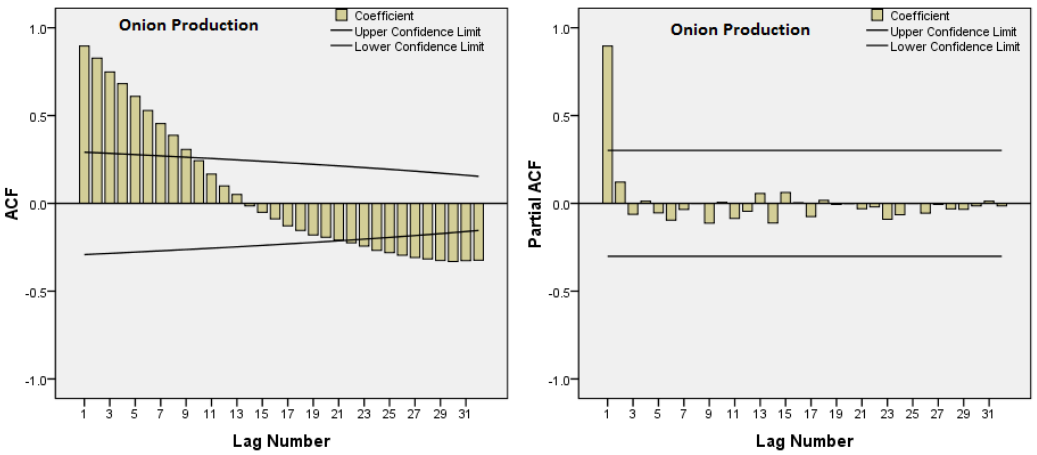

Figure 4. ACF and PACF of differenced data

Table 3. BIC values of ARIMA(p,d,q)

ARIMA (p,d,q)	BIC Values
0,1,0	0.563
0,1,1	0.417
0,1,2	0.426
1,1,0	**0.333**
1,1,1	0.441
1,1,2	0.521
2,1,0	0.437
2,1,1	0.513
2,1,2	0.567
3,1,0	0.487
3,1,1	0.587
3,1,2	0.681

Table 4. Estimated AR Model of A. *cepa* Production

	Estimate	SE	t	Sig.
Constant	-91.442	16.618	-5.503	0.000
AR 1	-0.573	0.144	-3.969	0.000

Table 5. Estimated AR Model Fit Statistics

ARIMA (p,d,q)	Stationary R^2	R^2	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
0,1,0	0.208	0.98	1.214	11.702	33.108	0.901	3.273	0.563
0,1,1	0.388	0.984	1.08	11.453	43.099	0.774	2.982	0.417
0,1,2	0.448	0.986	1.039	10.866	37.788	0.71	2.809	0.426
1,1,0	**0.437**	**0.986**	**1.036**	**11.033**	**37.84**	**0.719**	**2.915**	**0.333**
1,1,1	0.44	0.986	1.047	11.106	39.28	0.717	2.87	0.441
1,1,2	0.458	0.986	1.043	10.656	40.512	0.708	3.065	0.521
2,1,0	0.442	0.986	1.045	11.15	40.431	0.713	2.819	0.437
2,1,1	0.462	0.986	1.039	10.963	37.128	0.682	2.754	0.513
2,1,2	0.494	0.987	1.021	10.673	32.52	0.681	2.772	0.567
3,1,0	0.476	0.987	1.025	10.706	32.406	0.687	2.759	0.487
3,1,1	0.484	0.987	1.032	10.632	31.486	0.676	2.681	0.587
3,1,2	0.495	0.987	1.035	10.694	32.209	0.682	2.787	0.681

Model Estimation: Model parameters were found and accounted (Table 4 and Table 5). The model verification is concerned with examining the residuals of the model to progress on the chosen ARIMA (p,d,q). This is done through validating the autocorrelations and partial autocorrelations of the residuals of various orders, up to 32 lags were considered and the same along with their significance which is checked by Box-Ljung test are given (Table 6). This proves that the chosen ARIMA model is a suitable model.

Table 6. Residual of ACF and PACF of A. *cepa* Production

Lag	ACF		PACF		Lag	ACF		PACF	
	Mean	SE	Mean	SE		Mean	SE	Mean	SE
1	-0.053	0.152	-0.053	0.152	**17**	0.058	0.186	0.074	0.152
2	0.085	0.153	0.083	0.152	**18**	0.035	0.187	-0.062	0.152
3	0.124	0.154	0.134	0.152	**19**	-0.179	0.187	-0.171	0.152
4	-0.155	0.156	-0.152	0.152	**20**	-0.055	0.191	-0.075	0.152
5	0.222	0.160	0.194	0.152	**21**	-0.066	0.191	-0.088	0.152
6	-0.048	0.167	-0.026	0.152	**22**	-0.044	0.192	-0.061	0.152
7	-0.194	0.167	-0.213	0.152	**23**	0.068	0.192	0.009	0.152
8	-0.049	0.172	-0.133	0.152	**24**	-0.036	0.192	0.089	0.152
9	-0.262	0.173	-0.187	0.152	**25**	-0.009	0.193	0.013	0.152
10	0.026	0.182	0.015	0.152	**26**	0.001	0.193	-0.059	0.152
11	-0.024	0.182	-0.005	0.152	**27**	0.032	0.193	-0.018	0.152
12	-0.050	0.182	0.061	0.152	**28**	0.057	0.193	-0.098	0.152
13	0.080	0.182	0.070	0.152	**29**	0.021	0.193	-0.079	0.152
14	-0.154	0.183	-0.109	0.152	**30**	-0.024	0.193	-0.104	0.152
15	0.019	0.186	-0.075	0.152	**31**	0.016	0.193	0.046	0.152
16	0.044	0.186	-0.053	0.152	**32**	0.135	0.193	0.098	0.152

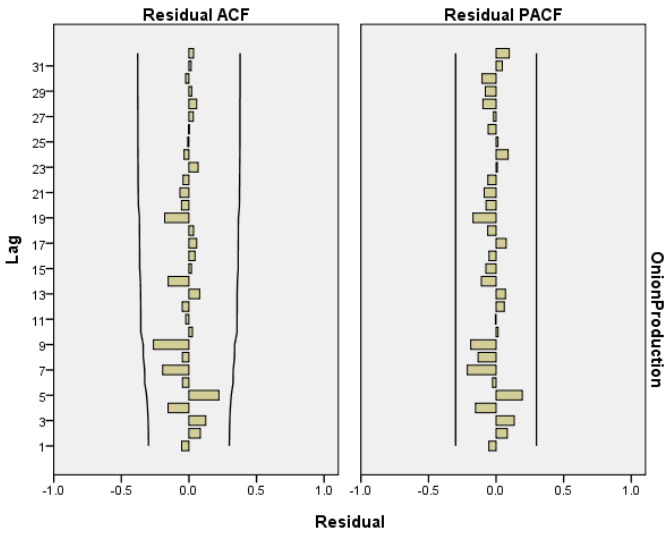

Figure 5. Residuals of ACF and PACF

Table 7. Forecast of A. *cepa* Production

Year	Predicted	LCL	UCL
2022	31.20	29.11	33.29
2023	33.89	31.62	36.17
2024	35.08	32.31	37.84
2025	37.20	34.19	40.21
2026	38.85	35.53	42.18
2027	40.85	37.29	44.41
2028	42.72	38.91	46.53
2029	44.74	40.71	48.77
2030	46.74	42.50	50.99

The ACF and PACF of the residuals are specified (Figure 5) and also indicated 'good fit' for the selected ARIMA model of the A. *cepa* production data is

$$Y_t = \mu + \phi_1 \epsilon_{t-1} + \epsilon_t$$

$$Y_t = -91.442 - 0.573\epsilon_{t-1} + \epsilon_t$$

The forecasted value of A. *cepa* production (quantity in million tonnes) for the years 2022 through 2030 respectively is given by 31.20, 33.89, 35.08, 37.20, 38.85, 40.85, 42.72, 44.74 and 46.74 in Table 7. We calculated significant measures of the forecasts' accuracy for the sample period in order to evaluate the fit of an ARIMA (p,d,q) model. This measure shows that the forecasting inaccuracy is low. Figure 6 indicates that the actual and forecasted value of A. *cepa* production data with 95% confidence limits.

4 Conclusion remarks

The most suitable ARIMA model for A. *cepa* production forecasting of data was establish to be ARIMA (1,1,0) and it can be found that forecasted production would raise from 32.27 million tonnes in 2021 to 46.74 million tonnes in 2030 in India for using time series data from 1978 to 2021 on A. *cepa* production, the results of this study gives an indication on future A. *cepa* production in India, which can be considered for future policy creation and preparing new strategies for increasing and supporting A. *cepa* production in India.

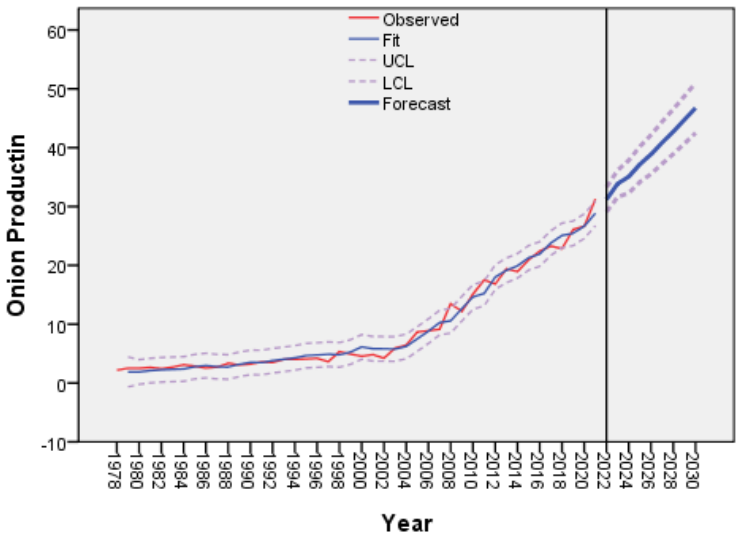

Figure 6. Actual and Estimate of A. *cepa* Production

References

[1] G.E.P. Box and D.A. Pierce, *Distribution of Residual Autocorrelations in ARIMA Models*, J. American Stat. Assoc., 65 , 1509-1526, (1970).

[2] E. Slutzky, *The summation of random causes as the source of cyclic processes,* Econometrica, 5 , 105-146, (1973).

[3] H. Akaike, *Statistical Predictor Identification,* Annals of Institute of Statistical Mathematics, 22 , 203-270, (1983).

[4] P. Mishra, C. Sarkar, K.P. Vishwajith, B.S. Dhekale and P.K. Sahu, *Instability and forecasting using ARIMA model in area, production and productivity of A. cepain India,* Journal of Crop and Weed, 9(2), 96-101, (2013).

[5] T. Jai Sankar and P. Pushpa, *Design and development of time series analysis for Saccharum officinarum production in India,* A Journal of Composition Theory, 12(9) , 203-211, (2019).

[6] T. Jai Sankar and P. Pushpa, *Stochastic forecasting analysis for peanut (Arachishy pogaea) production in India,* Journal of Critical Reviews, 7(12) , 2394-5125, (2020).

[7] Sakuntala Deka, P.J. Hazarika and A.N. Patowary, *Tea production in Assam: forecasting with ARIMA model. Advances in Space Research,* 33(1) , 48-56, (2021).

[8] Balaga Divya and Abhiram Dash, *Using ARIMA model to forecast the area, yield and production of arhar in Odisha,* Biological Forum – An International Journal, 14(3) , 1179-1185, (2022).

[9] Rama Shankar Yadav, Vishal Mehta and Ashish Tiwari, *An application of time series ARIMA forecasting model for predicting nutri cereals area in India,* The Pharma Innovation Journal. 11(3) , 1260-1267, (2022).

[10] P. Sameerabanu and C. Sekhar, *A stochastic process in modeling and forecasting of onion production in India,* International Education & Research Journal, 8(11) , 56-59, (2022).

[11] Sudhir Paswan, Anupriya Paul, Ajit Paul and Ashish S Noel, *Time series prediction for sugarcane production in Bihar using ARIMA & ANN model,* The Pharma Innovation Journal, 11(4) , 1947-1956, (2022).

[12] Tichaona W. Mapuwei, Jenias Ndava, Mellissa Kachaka and Brain Kusotera, *An application of time series ARIMA forecasting model for predicting tobacco production in Zimbabwe,* American Journal of Modeling and Optimization, 9(1) , 15-22, (2022).

[13] Yashpal Singh Raghav, Pradeep Mishra, Khder Mohammed Alakkari, Monika Singh, Abdullah Mohammad Ghazi AI Khatib and Ritisha Balloo, *Modelling and forecasting of pulses production in South Asian countries and its role in nutritional security,* Legume Research - An International Journal, 45(4) , 454-461 (2022).

[14] S.B. Bhusanar and Satyveer Singh Meena, *Forecasting groundnut area, production and productivity in Rajasthan, India using ARIMA model,* Asian Journal of Agricultural Extension, Economics & Sociology, 41(5) , 99-105, (2023).

[15] T. Jai Sankar and P. Pushpa, *Implementation of stochastic time series forecasting ARIMA model for Hordeum vulgare production in India*, International Journal of Agricultural and Statistical Sciences, 19(1) , 133-139, (2023).

[16] Alan Pankratz, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases,* John Wiley & Sons, New York, (1983).

Author information

T. Jai Sankar, Department of Statistics, Bharathidasan University, Tiruchirappalli, Tamilnadu, India.
E-mail: tjaisankar@gmail.com

P. Pushpa, St. Peter's Institute of Higher Education and Research, Avadi, Chennai, Tamil Nadu, India.
E-mail: pushpastats@gmail.com