Environment and Ecology 42 (2) : 383—390, April—June 2024
Article DOI: https://doi.org/10.60151/envec/UVWC8442
ISSN 0970-0420

Using ARIMA Model to Forecast Production of *kharif* Sweet Potato in Odisha

Jayashree Pradhan, Abhiram Dash

Received 12 August 2023, Accepted 19 December 2023, Published on 5 April 2024

ABSTRACT

Sweet potato, which is frequently referred to as a small farmers crop, is a major vegetable crop of Odisha. In India, Odisha is producing the most sweet potatoes. During the *kharif* season (June to August), sweet potatoes are grown as rain fed crop. A timely and accurate forecast of the area and production of such major vegetable crop is helpful for making agricultural policy decisions and giving nutrients to the population as the sweet potato occupies a key place among vegetable crops in Odisha. The objective of the present research is to predict the area, yield, and production of *kharif* sweet potato in Odisha by applying the most widely used forecasting model, ARIMA model. ACF and PACF plots and secondary data on the area, yield, and production of *kharif* sweet potato were collected from 1970–1971 to 2019–20 to fit the models that were determined to be appropriate. The best fit model was chosen based on the importance of the estimated coefficients, model diagnostic tests, and model fit statistics. By refitting the model with data from the most recent 4 years, 3 years, 2 years and 1 year, as well as by making one step ahead forecasts for the years 2016–17 to 2019–20, and the best fit model was cross-validated. The models with the best fits for the area, yield, and production of *kharif* sweet potatoes were found to be ARIMA (1,0,0) with constant model, ARIMA (0,1,1) without constant model, and ARIMA (0,1,1) without constant without constant. The forecast values show that the area, yield, and production of *kharif* sweet potatoes in Odisha will remain constant in future years, regardless of variation in the lower and upper class interval of the forecast values.

Keywords ARIMA, Cross validation, Forecast, MAPE, RMSE, Stationary.

INTRODUCTION

Sweet potato is the primary vegetable crop of Odisha and commonly referred to as a small farmers crop. After wheat, rice, maize, potato, barely, and cassava, sweet potatoes are the seventh-largest food crop in the world, according to ASHS (2007). Although it is grown in practically all states of India, the majority of sweet potatoes are produced in the eastern states of India such as Odisha, Kerala, West Bengal, and Uttar

Jayashree Pradhan[1], Abhiram Dash[2]*

[2]Assistant Professor and Head

[1,2]Department of Agricultural Statistics, College of Agriculture, Bhubaneswar, (OUAT), Odisha University of Agriculture and Technology 751003, India

Email : abhiramdash@ouat.ac.in
*Corresponding author

Pradesh. In India, there are 130 thousand hectares of sweet potatoes being grown, with a production of 1470 thousand tonnes. When it comes to sweet potato production in India, Odisha is the top producing state. Odisha has the most land planted in sweet potato, producing 420.275 thousand tonnes from 404.10 thousand hectares. Both the *kharif* (the rainy southwest monsoon) and the *rabi* (the dry post-monsoon) seasons are used in Odisha for growing sweet potatoes. Sweet potato is grown as a rain fed crop during *kharif* season (June- August). *Kharif* sweet potato is cultivated on about 80.18% of total area under sweet potato and 81.20% of total production of sweet potato. Sweet potato holds a significant position among the vegetable crops in Odisha, thus a timely and accurate forecast of the area and production of such important vegetable crops is useful for making agricultural policy decisions and supplying nutrients to the people. Many researchers have used different methods to forecast area, yield and production of different crops. Abah (2022) studied the growth of sweet potato output and yield in Nigeria and its implications on food production using ARIMA model. Singh and Verma (2022) used ARIMA model and Ordinary least square method for forecasting of vegetable production in Haryana. The present research is focused on forecasting the area, yield, and production of the *kharif* sweet potato in Odisha using ARIMA model. Divya and Dash (2022) conducted research on forecasting area, yield and production of arhar in Odisha through fitting of Auto Regressive Integrated Moving Average (ARIMA) models.

MATERIALS AND METHODS

The secondary data about the area, yield, and production of *kharif* sweet potato in Odisha is collected from the Five Decades of Odisha Agriculture Statistics, published by Directorate of Agriculture and Food Production, Odisha which consists of the years 1970–71 to 2019–20.

The Auto Regressive Integrated Moving Average (ARIMA) is a mathematical technique for predicting future trends. The ARIMA models are ARMA models that contain the order of differencing (which stationarise the data). “ARIMA (p, d, q)” refers to a non-seasonal ARIMA model, where p, d, and q stand for the number of auto regressive terms, the number of non-seasonal differences needed to stationarise the data, and the number of moving average components, respectively.

The general forecasting equation used to express the ARIMA (p, d, q) model:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$ (Dash *et al*. 2022)

Where μ is the constant term, $\phi_1, \phi_2, \ldots \phi_i$ and $\theta_1, \theta_2, \ldots \theta_j$ are the parametric coefficients of the autoregressive term and moving average terms, respectively of the model.

The ARMA model can only be estimated with stationary data. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test can be used to determine whether the data are stationary. The null hypothesis of KPSS test is H_0: the data are stationary. The alternative hypothesis, H_1: The data are non-stationary.

If the data are not stationary, then it is necessary to differentiate the data at an appropriate lag in order to make it stationary. After stationarizing the data, the most likely Moving Average (MA) and Auto Regression (AR) orders are determined using the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots. A number of ARIMA models are fitted and their parameters are determined on the basis of the observed AR and MA orders.

The Shapiro-wilk test and the Box-pierce test, respectively, are employed as model diagnostics tests to establish the normality and independence of the residuals of the fitted models.

The fitted models are then compared using model fit statistics such as the mean absolute percentage error (MAPE), root mean square error (RMSE), and Akaike’s Information Criteria corrected (AICc), which are as follows:

Root mean square error (RMSE): $\sqrt{\dfrac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}}$

Mean absolute percentage error (MAPE):

$$\frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Where $\hat{y}_t$ = forecasted value, y_t = actual value and n = number of times the summation iteration happens.

After a successful cross validation, the model with the lowest RMSE and MAPE values among the fitted ARIMA models is chosen as the most suitable ARIMA model and used for forecasting. The chosen ARIMA model is cross-validated using the one-step ahead forecast. Five years forecast from 2016-17 to 2019-20 are used for the purpose.

The MAPE of the forecasted values for four years is calculated using the formula below:

$$\text{MAPE} = \sum_{i=1}^{4} \frac{APE_i}{4}$$

Where, APE_i is the absolute percentage error for the i^{th} period

$$APE_i = \left(\frac{Y - \hat{Y}}{Y_i}\right)$$

Where, Y_i is the observed value of i^{th} year in the left out period

$\hat{Y}$ is the forecast values of i^{th} year in the left out period

The best fit ARIMA model is used to predict the *kharif* sweet potato area, yield, and production of Odisha for the next years from 2020–21 to 2024–25 after a successful cross validation.

Table 1. Test of stationary of data on area, yield and production of *kharif* sweet potato in Odisha.

Variable	Original series		First order differenced series	
	KPSS test statistics	p value	KPSS test statistics	p value
Area	0.19	0.1	----	----
Yield	1.05	0.01	0.05	0.1
Production	0.77	0.01	0.09	0.1

RESULTS AND DISCUSSION

The KPSS test statistics of the data on area, yield, and production of *kharif* sweet potato in Odisha are shown in Table 1. The data of area made stationary after first order difference while the data for yield and production are stationary so no need to be differenced.

The next step involved using the ACF and PACF charts to determine the order of MA and AR variables like p and q.

The ordering of AR and MA terms were used to identify various fitted ARIMA models. The ACF and PACF plots showing raw data of area of the *kharif* sweet potato in Odisha are shown in Fig. 1. The ACF and PACF plots of First order difference data of yield and production of *kharif* sweet potato are shown in Figs. 2 - 3.

The fitted models to the data on area under *kharif* sweet potato and their estimated coefficients are shown in Table 2. The analysis of the table shows that

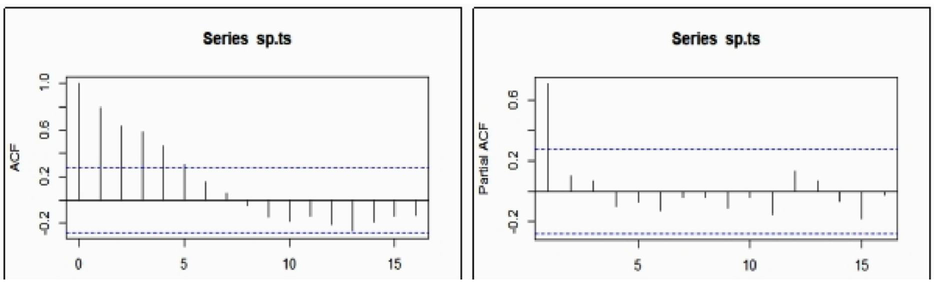

Fig. 1. ACF and PACF plot of area under *kharif* sweet potato in Odisha.

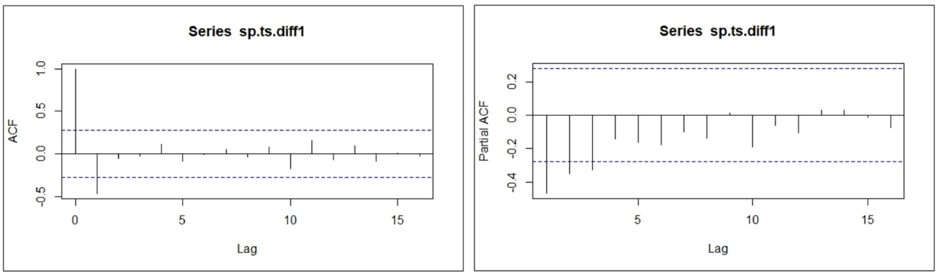

Fig. 2. ACF and PACF plot of first order difference of yield of *kharif* sweet potato in Odisha.

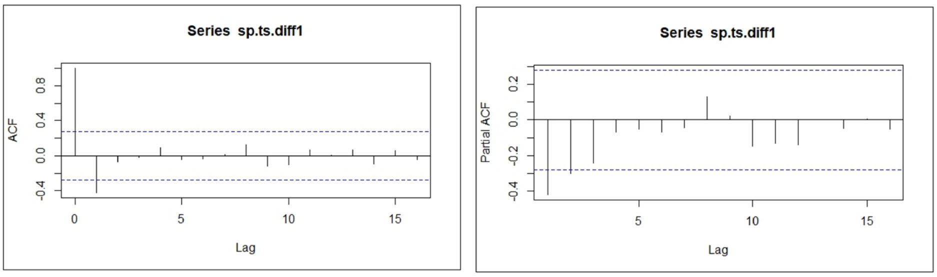

Fig. 3. ACF and PACF plot of first order difference of production of *kharif* sweet potato in Odisha.

the ARIMA(1,0,0) with constant and ARIMA(1,0,1) with constant models have significant estimates for the constants and first auto regressive parameter, but the first order moving average value of ARIMA(1,0,1) with constant model is found to be not significant.

The model diagnostics test results and model fit statistics for the fitted ARIMA models are shown in Table 3. The ARIMA (1,0,0) with constant model satisfies the tests for normality and independence of residuals and the model has low RMSE, MAPE, AIC values as compared to other models. Therefore The ARIMA (1,0,0) with constant model is chosen as the one that fits the *kharif* sweet potato area the best. Figure 4 further confirms the normality and independence of residuals.

Table 4 displays the fitted ARIMA models for

Table 2. Parameter estimates of the ARIMA (p,d,q) model fitted to area under *kharif* sweet potato.

ARIMA (p,d,q)	Constant	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2
ARIMA (1,0,0)	32.049**	0.795**	----	----	----	----
with constant	(2.334)	(0.093)				
ARIMA (1,0,1)	31.178**	0.896**	----	----	-0.249	----
with constant	(3.341)	(0.084)			(0.163)	
ARIMA (2,0,0)	31.413**	0.656**	0.1917	----	----	----
with constant	(3.048)	(0.142)	(0.1499)			
ARIMA (3,0,0)	30.951**	0.633**	0.1219	0.118	----	----
with constant	(3.570)	(0.144)	(0.1726)	(0.149)		
ARIMA(1,0,2)	31.203**	0.892**	----	----	-0.262	0.031
with constant	(3.316)	(0.089)			(0.181)	(0.161)

Figures inside the parenthesis indicate the standard error.
** Significant at 1% level *Significant at 5% level.

Table 3. Model diagnostic test with model fit statistics of the ARIMA (p,d,q) model fitted to area under *kharif* sweet potato.

ARIMA (p,d,q)	Model diagnostic test				Model fit statistics		
	Shapiro-wilk test		Box- pierce test				
	W	p-value	χ^2	p-value	RMSE	MAPE	AIC
ARIMA (1,0,0) with constant	0.964	0.131	8.533	0.969	3.504	8.380	274.816
ARIMA (1,0,1) with constant	0.964	0.133	6.910	0.997	3.435	8.290	275.380
ARIMA (2,0,0) with constant	0.965	0.154	7.512	0.995	3.444	8.308	275.595
ARIMA (3,0,0) with constant	0.961	0.095	6.409	0.998	3.419	8.303	277.447
ARIMA (1,0,2) with constant	0.966	0.153	7.103	0.996	3.434	8.304	277.818

Table 4. Parameter estimates of the ARIMA (p,d,q) model fitted to yield of *kharif* sweet potato.

ARIMA (p,d,q)	Constant	ϕ_1	ϕ_2	ϕ_3	Φ4	θ_1	θ_2
ARIMA (0,1,1) without constant	----	----	----	----	----	-0.788** (0.075)	
ARIMA (0,1,2) without constant	----	----	----	----	----	-0.818** (0.157)	0.032 (0.149)
ARIMA (1,1,1) without constant	----	-0.032 (0.165)	----	----	----	-0.779** (0.090)	----
ARIMA (4,1,0) with constant	64.481 (76.742)	-0.775** (0.141)	-0.609** (0.168)	-0.406* (0.165)	-0.131 (0.136)	----	----
ARIMA (3,1,0) without constant	----	-0.723** (0.135)	0.524** (0.149)	-0.299* (0.131)	----	----	----
ARIMA (3,1,0) with constant	65.186 (87.692)	-0.730** (0.134)	-0.533** (0.149)	-0.306* (0.131)	----	----	----

Figures inside the parenthesis indicate the standard error.
** Significant at 1% level, *Significant at 5% level.

yield and their estimated coefficients.The analysis of the table shows that all of the computed coefficients for the ARIMA (0,1,1) without constant and ARIMA (3,1,0) without constant models are significant.

The result of model diagnostics test and model fit statistics for the fitted ARIMA models for the yield of *kharif* sweet potato are shown in Table 5. In the case of ARIMA (0,1,1) without constant model has lower

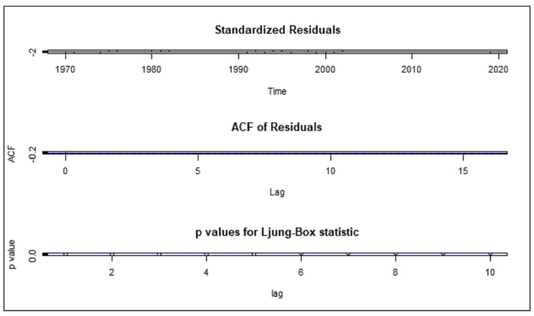

Fig 4. Residual plot for fitting of ARIMA (1,0,0) with constant to *kharif* area of sweet potato.

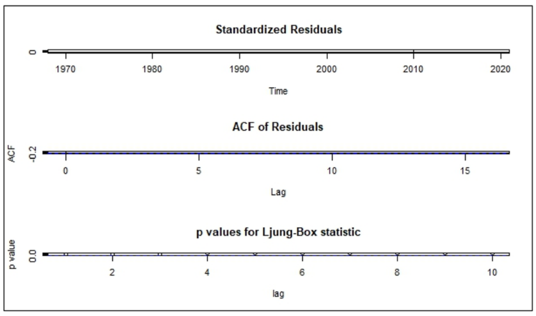

Fig. 5. Residual plot for fitting of ARIMA(0,1,1) without constant to yield of kharif sweet potato.

Table 5. Model diagnostic test with model fit statistics of the ARIMA (p,d,q) model fitted to yield of *kharif* Sweet potato.

ARIMA (p,d,q)	Model diagnostic test				Model fit statistics		
	Shapiro-wilk test		Box- pierce test				
	W	p-value	χ^2	p-value	RMSE	MAPE	AIC
ARIMA (0,1,1) without constant	0.533	0.955	4.719	0.999	1503.788	8.173	862.221
ARIMA (0,1,2) without constant	0.532	0.943	4.664	0.999	1502.764	8.153	864.446
ARIMA (1,1,1) without constant	0.532	0.935	4.673	0.999	1503.005	8.157	864.457
ARIMA (4,1,0) with constant	0.576	0.963	5.831	0.999	1512.151	8.543	872.402
ARIMA (3,1,0) without constant	0.579	0.957	6.114	0.998	1536.407	8.412	868.767
ARIMA(3,1,0) with constant	0.581	0.932	6.168	0.998	1527.548	8.537	870.709

RMSE, MAPE, and AICc values than ARIMA (3,1,0) without constant model. So ARIMA (0,1,1) without constant model is chosen as the one that best fits the yield of the sweet potato in the *kharif* season. Figure

Table 6. Parameter estimates of the ARIMA(p,d,q) model fitted to production of *kharif* sweet potato.

ARIMA (p,d,q)	Constant	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2
ARIMA(0,1,1) without constant	----	----	----	----	-0.676** (0.116)	----
ARIMA(0,1,2) without constant	----	----	----	----	-0.645** (0.148)	-0.049 (0.157)
ARIMA(0,1,1) with constant	3.464 (2.409)	----	----	----	-0.759** (0.154)	----
ARIMA(1,1,1) without constant	----	0.063 (0.220)	----	----	-0.713** (0.167)	----
ARIMA(3,1,0) without constant	----	-0.617** (0.139)	-0.426** (0.151)	-0.235 (0.136)		
ARIMA(3,1,0) with constant	3.256 (4.171)	-0.628** (0.139)	-0.439* (0.151)	-0.245 (0.135)	----	----

Figures inside the parenthesis indicate the standard error.
** Significant at 1% level, *Significant at 5% level.

Table 7. Model diagnostic test with model fit statistics of the ARIMA (p,d,q) model fitted to production of *kharif* sweet potato.

ARIMA (p,d,q)	Model diagnostic test				Model fit statistics		
	Shapiro-wilk test		Box- pierce test				
	W	p-value	χ^2	p-value	RMSE	MAPE	AIC
ARIMA(0,1,1) without constant	0.673	0.987	6.931	0.991	65.859	13.249	555.294
ARIMA (0,1,2) without constant	0.674	0.965	6.870	0.997	65.785	13.144	557.468
ARIMA(0,1,1) with constant	0.674	0.973	6.786	0.997	64.744	13.132	556.142
ARIMA(1,1,1) without constant	0.674	0.964	6.887	0.997	65.794	13.154	557.482
ARIMA(3,1,0) without constant	0.664	0.972	6.687	0.997	65.911	13.251	559.961
ARIMA (3,1,0) with constant	0.665	0.968	6.632	0.998	65.494	13.111	561.851

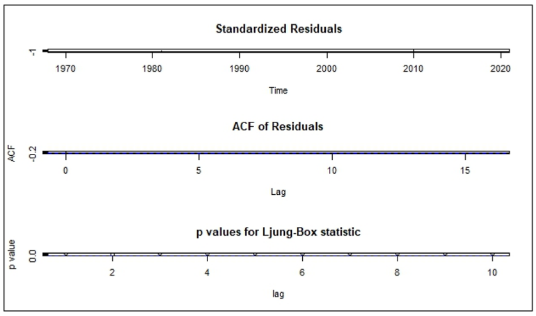

Fig. 6. Residual plot for fitting of ARIMA (0,1,1) without constant to production of *kharif* sweet potato.

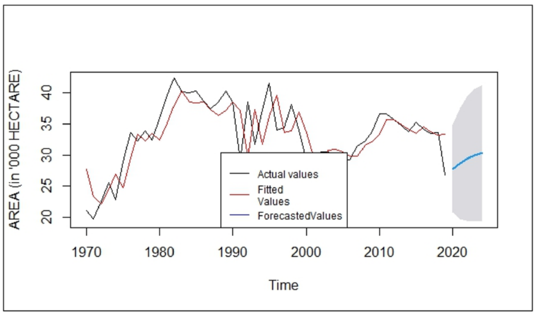

Fig. 7. Actual with fitted and forecasted values of area under *kharif* sweet potato from ARIMA (1,0,0) with constant model.

5 further validates the normality and independence of residuals.

Table 6 displays the fitted models and their estimated coefficients. The analysis of the table shows that the computed higher coefficients for the ARIMA (0,1,1) without constant model and ARIMA (0,1,1) with constant models are significant, but the constant of ARIMA (0,1,1) with constant model is found to be not significant.

Table 7 displays the results of model diagnostic test and model fit statistics for the fitted ARIMA models for *kharif* sweet potato production. The ARIMA (0,1,1) without constant model satisfies the tests for residual independence and normality and has lower RMSE, MAPE and AICc value than ARIMA (3,1,0) without constant model. So The ARIMA (0,1,1) without constant model is chosen as the one that best fits the production of the *kharif* sweet potato.Figure. 6 Further confirms the normality and independence of residuals.

The results of cross validation for each variable related to the *kharif* sweet potato in Odisha shown in Table 8. The MAPE (mean APE) for the area under the

Table 8. Cross validation of area, yield and production of *kharif* sweet potato.

Year	Area			Yield			Production		
	Actual	Predicted	APE	Actual	Predicted	APE	Actual	Predicted	APE
2016-17	34.23	34.71	1.417	9527	9623.847	1.016	326.12	336.484	3.178
2017-18	33.5	33.90	1.197	9427	9602.246	1.859	315.79	333.190	5.509
2018-19	33.67	33.31	1.063	9451	9564.18	1.197	318.23	327.651	2.960
2019-20	26.75	33.45	25.062	9559	9539.854	0.200	255.7	327.651	28.138
MAPE		7.185			1.068			9.947	

Table 9. Forecast values of area, yield and production of *kharif* sweet potato for the year 2020-21 to 2024-25 using ARIMA model.

Year	Area ('000 ha)			Yield (kg/ha)			Production ('000 tonnes)		
		95% confidence interval			95% confidence interval			95% confidence interval	
	Forecasted	Lower CI	Upper CI	Forecasted	Lower CI	Upper CI	Forecasted	Lower CI	Upper CI
2020-21	27.836	20.826	34.846	9543.808	6535.661	12551.96	302.196	170.452	433.939
2021-22	28.700	19.745	37.655	9543.808	6468.981	12618.64	302.196	163.705	440.687
2022-23	29.387	19.395	39.378	9543.808	6404.716	12683.90	302.196	157.272	447.120
2023-24	29.933	19.339	40.526	9543.808	6339.781	12747.84	302.196	151,112	453.279
2024-25	30.366	19.409	41,324	9543.808	6277.097	12810.52	302.196	145.194	459.198

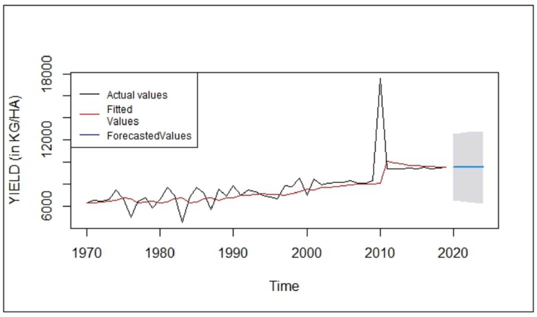

Fig. 8. Actual with fitted and forecasted values of yield of *kharif* sweet potato from ARIMA (0,1,1) with constant model.

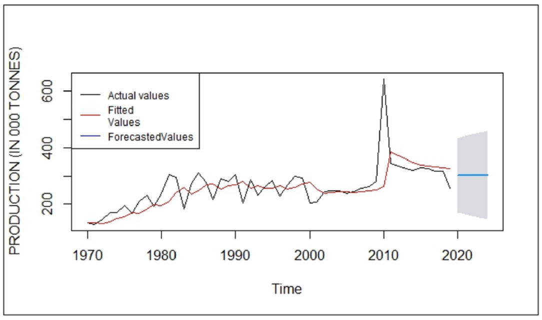

Fig. 9. Actual with fitted and forecasted values of production of *kharif* sweet potato from ARIMA (0,1,1) without constant model.

kharif sweet potato is found to be 7.185 and the APE (absolute percentage error) is found to be in the range of 1 to 25. Similar to yield, where the APE ranges from 0 to 1 and the MAPE is 1.068, and production, where the APE ranges from 2 to 28 and the MAPE is 9.947. These findings demonstrate that cross validation of the chosen ARIMA models was successful.

The area, yield, and production of the *kharif* sweet potato in Odisha were predicted using respective selected best fit ARIMA models, which are shown in the Table 9, for the years 2020–2021, 2021–2022, 2022–23, 2023–24, and 2024–25.

Figures. 7–9 shows the actual, fitted and forecast values of area, yield and production of *kharif* sweet potato in Odisha. The fitted values of area under *kharif* sweet potato are closer to the actual values and the forecasted values are likely increasing from 2020-21 to 2024-25. The forecasted values are found to be constant in the case of yield and production of *kharif* sweet potatoes, with fitted values being closer to actual values.

CONCLUSION

The models ARIMA (1,0,0) with constant model, ARIMA (0,1,1) without constant model, and ARIMA (0,1,1) without constant model are found to be the best fits for the area, yield, and production of *kharif* sweet potato in Odisha, respectively. Some models have been selected to predict the *kharif* sweet potato area, yield, and output in Odisha. According to the forecast values, the area, yield, and subsequently production of *kharif* sweet potatoes in Odisha will stay stable in the following years, irrespective of variations in the lower and upper class intervals of the forecast values.

REFERENCES

Abah D (2022) Assessment of the growth of sweet potato output and yield in Nigeria (1970-2020) and its implications on food production. *Nigerian J Agricult Technol* 2 (1): 9-22. https://doi.org/10.59331/njaat.v2i1.13.

Dash A, Bhattacharya D, Dhakre DS, Kishan M (2022) On selection of best model from spline regression and ARIMA models for forecasting *rabi* food grain production of Odisha. *Environ Ecol* 40 (3B) : 1428—1437.

Divya B, Dash A (2022) Using ARIMA Model to forecast the area, yield and production of Arhar in Odisha. *Biol Forum – An Int J* 14 (3): 1179-1185.

Five decades of agriculture statistics (2020) Directorate of agriculture and food production, Odisha.(1970-71 to 2019-20).

Singh S, Verma RK (2022) Forecasting of vegetable production in Haryana by Ordinary Least Square Method and ARIMA Model. *Ind J Exten Educ* 58 (4): 71-75.