International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Ten Year Out of Sample Point Forecast of Area, Production and Productivity of winter rice in Assam, India

Rabijita Buragohain[1], Raju Prasad Paswan[2], Hemanta Saikia[3], Jugabrat Sarma[4]

[1]Young Professional II, Assam Agricultural University
[2]Professor, Assam Agricultural University
[3]Assistant Professor, Assam Agricultural University
[4]SRF, ICAR-ATARI, Zone-VI, Guwahati

Abstract

The present study was carried out in the state Assam to forecast area, production and productivity of winter rice of the state. The study area was purposively selected as being the major contributor towards agricultural Gross Domestic Product (GDP), the state also provides a major contribution to the economy of India. The prime objective is to estimate ten years ahead forecast of area, production and productivity of winter rice of Assam. To estimate the point forecast a time series analysis was carried on. An ARIMA model was endeavoured using the r-studio auto-Arima statistical package. Study revealed the best suitable p, d, q of area, production and productivity as ARIMA (0,0,4), ARIMA (1,1,2), ARIMA (0,1,1) respectively with minimum AIC, BIC, RMSE, MASE parameter criterion, on the basis on which ten years out of sample point forecast for area, production and productivity of winter rice in Assam has been done.

Keywords: Time series, ARIMA, winter rice, Assam, forecast

1. Introduction:

Rice *(Oryza sativa)* is the main staple food of Assam, a North-Eastern state of India. It occupies about two-third of total cropped area in the state. Rice production is fundamental to Assam's agrarian culture and the state's economy, making up over two thirds of the state's total cultivated land. Traditional significance of this crop is that, it has been grown throughout the year *viz.* Winter (Sali rice), Autumn (Ahu) and Summer (Boro rice). The crop accounts for nearly 41 percent of the total area under production (Barah et.al, 2009). This paper mainly emphasizes on the Winter rice, which is one of the most important varieties of rice that is farmed during the month of June/July and harvested in November - December. Planting in the monsoon and harvesting in the winter, this type not only rules the state's agricultural terrain but also makes a significant contribution to the total yield of crops. Rice production in Assam plays an important role because it has a significant contribution to countries Gross Domestic Product (GDP) (Bhowmick, B. C et al, 2005). There are a number of reasons why winter rice is so important to Assam's agriculture. First off, rice farming is greatly aided by the region's climate, which is marked by copious amounts of rainfall and fertile alluvial soil. Furthermore, winter rice is a good fit for the state's traditional cropping patterns and agricultural techniques, which guarantee food security and a means of subsistence for a sizable portion of the rural population. In spite of its vital function, Assam's winter rice productivity suffers many obstacles, including as unpredictable weather patterns, pest infestations, and socioeconomic limitations. For efficient agricultural planning and resource allocation, agricultural characteristics including acreage, production and productivity must be forecasted. Time series plays a pivotal role in forecasting agricultural events. It is a collection of observations of well-defined data obtained through repeated measurements over time. When it comes to time series forecasting, the Autoregressive Integrated Moving Average (ARIMA) model is one of the most well-known statistical techniques because of its dependability and robustness. The popularity of the ARIMA model is due to its statistical properties as well as the known Box-Jenkins methodology in the model building process. Although ARIMA model are quite flexible and it represents several different types of time series, *i.e* autoregressive (AR), moving average (MA) and combined AR and MA (ARMA) series, but the major limitation is pre-assumed linear form of the model. Thus, the approximation of linear models to complex real-world problem is not always satisfactory (Khashei, M.,2009). The study aims to produce precise forecasts that can aid in strategic planning and the execution of agricultural policy by examining historical data and spotting trends. It is anticipated that the knowledge acquired from this study will improve the socioeconomic circumstances of farmers of the state and also optimise resource utilisation and agricultural output. The study considered the secondary data for the sixty years *i.e* from 1962-63 to 2021-22 and a ten year ahead forecast from 2022-23 to 2032-33.

2. Review of Literature:

A dataset of 57 years of area, production and productivity of sugarcane was prepared for forecasting sugarcane yield in Tamil Nadu. The study was conducted using fitting of univariate Auto Regressive Integrated Moving Average (ARIMA) model, which was validated by comparing original data values with predicted values. Study found that ARIMA (1,1,1) model was best fitted for forecasting sugarcane area and productivity, while ARIMA (2,1,2) was found suitable for production of sugarcane (Suresh.K.,2011). The ARIMA model used for forecasting maize production in India for the year 2018 to 2022. By evaluating ACF & PACF, AIC, SBC of differenced series revealed that ARIMA (2,1,0) was best fitted model for forecasting maize production. The study found that there was 13.76 percentage of increment in the near future *w.e.f* 2017 to 2022 in India (Sharma P.K, 2018). There was an attempt to forecast cultivable area and production of cotton in India by using Box-Jenkins Stochastic autoregressive integrated moving average model. Dataset contains from 1950-51 to 2010-11 was collected was done by the method of maximum likelihood estimation. The best fitted model is selected using multiple R^2, root mean square error (RMSE), Akaike information criterion (AIC), schwartz bayesian criterion (SBC), normalized BIC, mean absolute error (MAE) and mean absolute proportion percentage (MAPPE). It revealed that ARIMA (0,1,0), ARIMA (1,1,4) and ARIMA (0,1,1) were best suitable for forecasting area, production and yield of cotton. They also estimated that if present growth rate continued then it could be 10.92 million hectare, 39.19 million bales of 170 kg of each and 527 kg/ha respectively (Debnath K.M.2013). There was an attempt to minimize the error gap in smart farming by forecasting production demand to estimate grain price of rice. Researchers formulated autoregressive integrated moving average method. The results were tested by mean square error and mean absolute percentage error (MAPE) standards. As MSE and MAPE values were lower, the efficacy of formulated model for price forecasting of rice was effectively demonstrated and projected for the year 2020 (Wawale G.S., 2022). Kavasseri G.R, 2009 et.al formulated fractional ARIMA (f-ARIMA) model to forecast wind speeds on the day ahead and two day ahead horizons. The f-ARIMA models were applied to wind speed records collected from four potential wind generation sites. The forecast errors in wind speed are examined and compared with the existing model. Results showed that significant improvements in forecasting accuracy were obtained by f-ARIMA. Almeida et al. (2020) utilized the ARIMA model to predict soybean yields in Brazil, evaluating historical data for variances and trends. Their results demonstrated that the ARIMA (2,1,2) model provided the most accurate forecasts, significantly aiding farmers in making informed planting decisions. Similarly, Kumar and Singh (2019) applied the ARIMA model for wheat yield prediction in India, finding that the model's efficiency was enhanced when seasonal effects were included in the analysis. Several studies have focused on regional comparisons. For instance, Oyedokun and Obafemi (2021) examined the predictive accuracy of ARIMA in different Nigerian states regarding maize production. Their research highlighted that ARIMA (1,0,1) was optimal for certain regions, indicating localized adaptability of the model. Conversely, Ramteke et al. (2022) used ARIMA to predict rice production across various states in India, concluding that local climatic factors should be integrated into model selection to improve accuracy. When compared to other forecasting techniques, the ARIMA model often shows superior performance. For example, Anis et al. (2021) conducted a comparative analysis of ARIMA and artificial neural networks (ANNs) for forecasting canola yields in Canada. They found that while both models performed adequately, ARIMA outperformed ANNs in terms of mean absolute percentage error (MAPE). This demonstrates ARIMA's robustness, particularly in environments with less complex data patterns. The effectiveness of ARIMA can be further enhanced by incorporating exogenous variables or using seasonal decomposition. In their study, Gupta et al. (2020) implemented ARIMAX (ARIMA with exogenous variables) to forecast sugar cane production in India. Their findings indicated that including variables such as rainfall and fertilizer application improved the accuracy of the forecasts significantly. Seasonality plays a critical role in agricultural forecasting. Rada and Anda (2022) accessed the impact of seasonal components on corn yield predictions using seasonal ARIMA. They found that including seasonal terms led to more accurate forecasts, which help farmers optimize planting schedules and resource allocation. Similarly, Chaudhary et al. (2023) emphasized the importance of considering seasonal variations in the ARIMA model for forecasting the yields of pulse crops, noting increased precision when adjustments were made for harvest cycles. The implications of accurate crop production forecasts extend to sustainability practices and economic planning. A study conducted by Zubair and Hashmi (2021) explored the economic viability of adopting ARIMA forecasting techniques in sustainable agricultural programs in Pakistan. Their research showed that timely and accurate yield predictions could lead to better resource management and food security initiatives. Despite its advantages, the ARIMA model has limitations, particularly when dealing with non-linear data patterns. Ali et al. (2021) reported challenges in implementing ARIMA models on complex agricultural datasets characterized by volatility and irregularities. They suggested adopting a hybrid approach, combining ARIMA with machine learning techniques to improve forecasting accuracy and reliability.

3. Materials and Methods:

3.1. Description of study:

The yearly data of area(ha), production (tonnes) and productivity (kg/ha) of winter rice from 1962-62 to 2021-22 were collected from Statistical Handbook of Assam, published by Directorate of Economics and Statistics, Government of Assam. In time series model data from 1962-63 to 2021-22 were used for model building and 2022-23 to 2032-33 were used for ten years ahead forecasting of the model. The statistical software R-studio, packages ‘timeseries’ and ‘forecasting’ was used for modelling and forecasting using ARIMA.

3.2 Methodology :

3.2.1 Autoregressive Integrated Moving Average (ARIMA): Box-Jenkins methodology refers to a systematic method of identifying, fitting, checking and using integrated autoregressive moving average (ARIMA) time series model. The general Box-Jenkins ARIMA (p,d,q) model for w is written as (George et al., 2008)

$$\omega_t = \phi_1\omega_{t-1} + \phi_2\omega_{t-2} + \cdots + \phi_p\omega_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_q a_{t-q} \text{ ……….(1)}$$

where, ϕ and θ are unknown parameters and the 'a' are independent and identically distributed normal errors with zero mean, p is the number of lagged value of ω_t, it represents the order of autoregressive (AR) dimensions, d is the number of times ω is differed and q is the number of lagged values of the error terms representing the order of moving average (MA) dimension of the model. The term integrated means that to obtain a forecast for ω from this model it is necessary to integrate the forecast ω_t. The ARIMA methodology involves three different phases- identification, estimation and testing.

3.2.2 Model identification: Identification gives us nothing except tentative consideration of a class of ARIMA models that will later be efficiently fitted and checked (George et. al.,2008). The Box-Jenkins assumes that the time series is stationary. A series is said to be strictly stationary if it has a fixed mean, fixed constant variance, and a constant auto-covariance structure. If the last condition is not satisfied, then the series is said to be stationary in the weak sense, or second order (Yafee and McGee, 1999). When data are tested for non-stationarity, if the autocorrelation starts high and decline slowly, then the series is non-stationary, and the Box-Jenkins methodology recommended differencing one or more times to get stationarity. The variance of the error underlying model must be invariant (*i.e* constant). This means that the variance for each subgroup of data is the same and does not depend on the level or the point in time. If it is violated it should be stabilized through Box-Cox test (Box & Cox, 1982). The identification phase for choosing and building the appropriate (p, q) values of the ARMA model for the stationary series ω_t is carried out on the grounds of its characteristics, i.e., the mean, the autocorrelation function (ACF), and the partial autocorrelation function (PACF).

Estimation and Testing: Not only does the Box-Jenkins model have to be stationary, it also has to be invertible *i.e* recent observations are more heavily weighted than more remote observations. To select the best model model from several adequate models we should use suitable criteria that deal with measures of accuracy and also with measures of goodness of fit of a model. In this study, the following measures were selected to check the model accuracy.

Root Mean Square Error (RMSE): It is a good measure of accuracy, but only to compare forecasting errors of different models or model configurations for a particular variable and not between variables, as it is scale dependent.

$$RMSE = \sqrt{\frac{\sum(Predicted_i - Actual_i)^2}{N}}$$

Akaike Information Criterion (AIC): AIC operates by assessing the models fit to the training set of data and adding a penalty term for model complexity. Finding the lowest AIC, which shows the model's best fit, is the desired outcome.

AIC = 2K - 2ln(L),

where, k is the number of model parameters and L is the Log-Likelihood function of the model.

Bayesian Information criteria (BIC): It is a criteria for model selection among a finite set of models. Models are compared with the Bayesian information statistics, lowest the BIC value, better is the time series model.

$$BIC = -2 * \log (L) + k * \log (n),$$

where log (L) is the log-likelihood of the model, k is the number of parameters, and n is the sample size. A lower BIC score indicates a better model.

Mean Absolute Scaled Error (MASE): It is the mean absolute error of the forecast values, divided by the mean absolute error of the in-sample one-step naïve forecast. This parameter is highly recommended for determining comparative accuracy of forecasts. It is independent of the scale of the data, so it can be used to compare forecasts across data sets with different scales[1].

$$MASE = \frac{MAE}{MAE_{naive}}$$

4. Results and Discussion:

4.1 Summery for Area (ha), Production (tonnes) and Productivity:

In case of area under winter rice cultivation, Autoregressive Moving Average (ARIMA) analysis has found that ARIMA (0,0,4) was the best fitted model for forecasting area. The model has been selected on the basis of lowest AIC and BIC criteria along with ACF, PACF. The model accuracy showed in the table 1. The model is good with RMSE and MASE of 327923 and 1.41 respectively. The model ARIMA (1,1,2) was determined as the best for winter rice production on the basis of lowest AIC & BIC of 1786.36 and 1787.08 respectively along with ACF and PACF plots, which results no significance lag existence. The anticipated outcomes for the production of winter rice revealed a trend towards declining growth. The model is best fitted with RMSE and MASE of 642025 and 0.89 respectively. The ARIMA model (0,1,1) was the best fitted model for forecasting productivity of winter rice with RMSE and MASE value of 142.96 and 0.83 respectively.

Table 1: ARIMA model summary for Area(ha), Production(tonnes) and Productivity(kg/ha) of winter rice

	Area(ha)	Production(tonnes)	Productivity(kg/ha)
Model	ARIMA (0,0,4)	ARIMA (1,1,2)	ARIMA (0,1,1)
AIC	1735.84	1786.36	774.03
BIC	1748.51	1794.73	780.31
RMSE	327923	642025	142.9507
MASE	1.41	0.89	0.83

4.2 Forecasted values for Area(ha), Production(tonnes) and Productivity(kg/ha) of winter rice:

The ARIMA (0,0,4) projected the area under winter rice production from the year 2022-23 to 2031-32 is presented in Table 2. Initially, the projected values have shown an upward tendency, reaching a maximum of 1682839 hectares in 2025–2026, but from 2026–2027 to 2031–2032 it showed a stable value depicting a high chance of declining growth in the near future. The fact that, currently significant decrease in cultivable land is a huge problem. People change the livelihood preferences. There are numerous

[1] MASE mathematical formula for - Search (bing.com)

International Journal for Multidisciplinary Research (IJFMR)

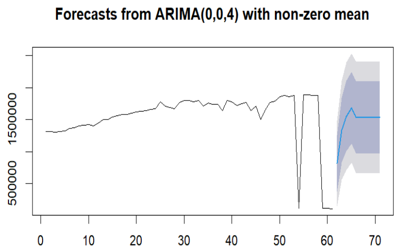
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

government initiatives, programs, and NGOs that educate people worldwide about the importance of soil and the absence of life without it.

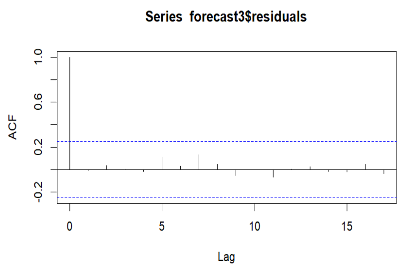
Table 2: Forecasted values for Area(ha), Production(tonnes) and Productivity(kg/ha) of winter rice from 2022-23 to 2031-32

Forecasted year	Area (ha)	Production (tonnes)	Productivity (kg/ha)
2022-23	815773.3	3269234	1989.697
2023-24	1333957.1	3203786	2006.446
2024-25	1553490.3	3164625	2023.196
2025-26	1682839.0	3141193	2039.946
2026-27	1529928.3	3127172	2056.695
2027-28	1529928.3	3118782	2073.445
2028-29	1529928.3	3113762	2090.195
2029-30	1529928.3	3110758	2106.944
2030-31	1529928.3	3108961	2123.694
2031-32	1529928.3	3107886	2140.444

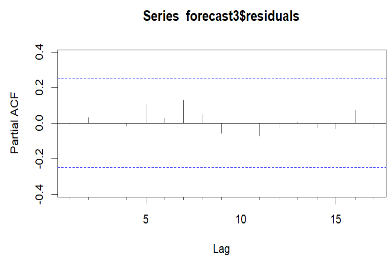
According to the decadal forecasts, the area used to produce winter rice would expand for a short period of time before possibly levelling off or even declining after that. It presented a serious threat to human life. [fig 1]. The 10 year advance winter rice production projection was created using ARIMA (1,1,2) and is shown in table 2. The minimum production projection with 95% confidence showed that winter rice production may decrease in the coming years and will reach 3107886 tons during 2031-32 with is at par with current production level. This may occur due to adverse effect of temperature rise, government attention to the crop and decrease in cultivable land [fig 2]. Point forecast for productivity (kg/ha) of winter rice was forecasted with the best fitted model of ARIMA (0,1,1) with lowest AIC and BIC of 774.03 & 780.31 and non-significant ACF and PACF plot [fig 3].



(a)

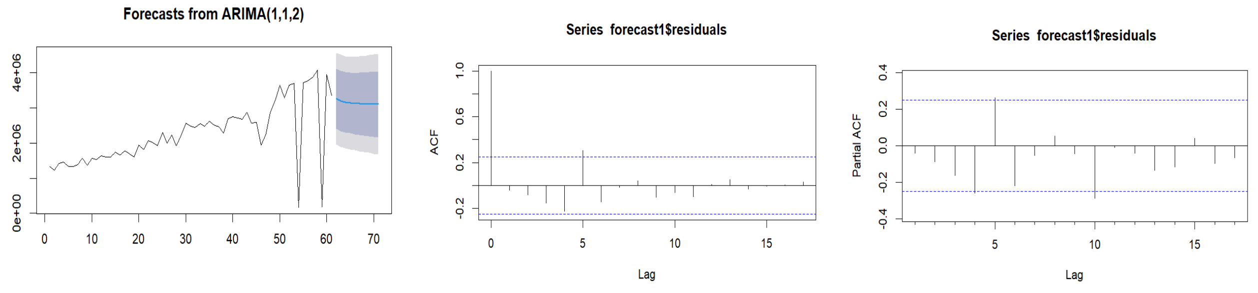


(b)



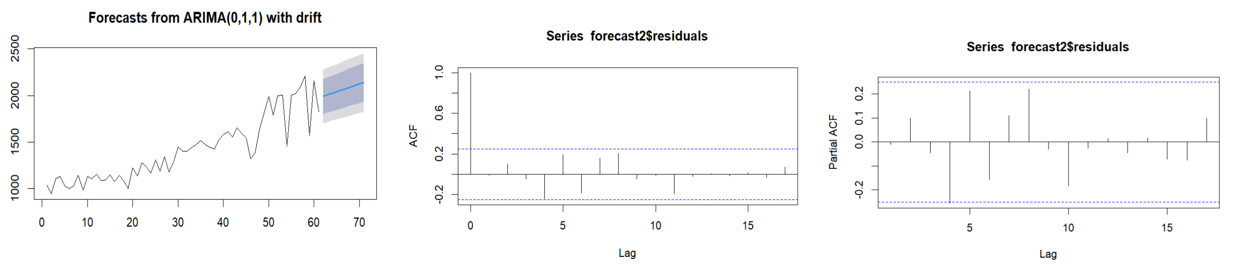
(c)

Fig1: Area: (a) Forecasting graph of ARIMA (0,0,4), (b) ACF plot and (c) PACF plot



(a) (b) (c)

Fig2: Production: (a) Forecasting graph of ARIMA (1,1,2), (b) ACF plot and (c) PACF plot



(a) (b) (c)

Fig3: Productivity: (a) Forecasting graph of ARIMA (0,1,1), (b) ACF plot and (c) PACF plot

According to the estimated forecast, the average winter rice productivity per hectare will rise from 1989.68 kg/ha to 2140.44 kg/ha between 2022–2023 and 2031–2022. Even with climate change, the policy interventions to promote cultivation can result in a significant improvement in winter rice yield. Though, at 5% probability, the increase in minimum and maximum productivity is due to better technology availability, suitable government policies and irrigation facilities.

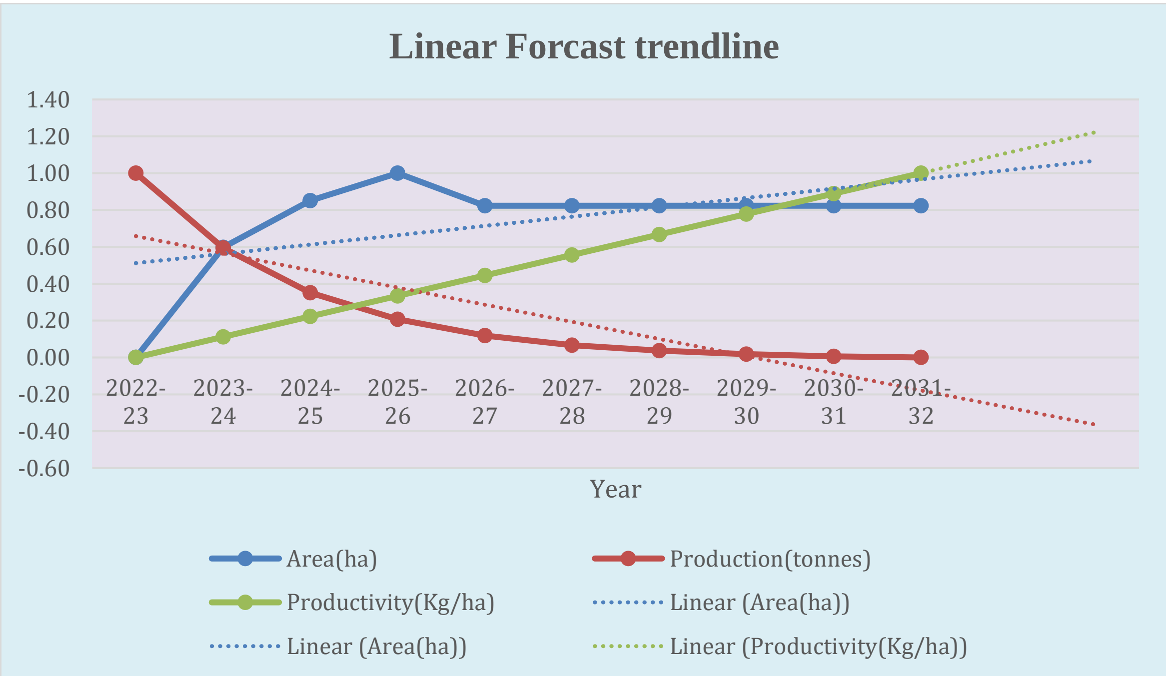


Fig:4 Linear forecasted trendline for area, production and productivity of winter rice

A linear forecast trendline is fitted for all the three variables which is presented in Fig 4. From the linear forecast graph the changing pattern of ten years ahead forecasts visibly stated that area and productivity following an increasing trend, while area follows a relatively downward trend.

5. CONCLUSION:

Forecast helps farmers manage risk, it is also used to identify and form policy responses to emerging issues. ARIMA model offers a good technique for predicting the magnitude of any variable. Its strength lays the fact that the method is suitable for any time series with any pattern of changes. Its limitation includes the requirement of a long time series. Above analysis showed a clear direction changes of area, production and productivity of winter rice in the following decades. Expansion of area under winter rice would directly increases its production in near future. Total cultivable area can only be increased through suitable government policies like increase in minimum support price (MSP), crop weather insurance, available irrigation facility and adequate supply of farm inputs etc. For better planning of agricultural production these prediction model will be extremely helpful for the farming community. These real insights will be helpful for the policy makers to bring the significant changes in major areas of production and productivity in Assam. It has been argued that analysis of time series help us analysing the changing pattern of the series. In agriculture, forecasting based on time series is an important task and facilitates efficient planning of cropping systems, thereby increasing the output. However, the reviewed literature demonstrates the widespread application and effectiveness of the ARIMA model in predicting agricultural crop production. It serves as a valuable tool for farmers, economists, and policymakers to make informed decisions regarding agricultural practices and resource allocation. While the ARIMA model has proven effective, researchers should continue exploring avenues for improving its application, including integrating external variables and hybrid forecasting methods, to enhance its predictive accuracy in the dynamic field of agriculture.

References:

Ali, S., Khan, M., & Nadeem, M. S. (2021). Evaluation of the ARIMA model for forecasting the yield of various crops in Pakistan. *Journal of Agricultural Economics*, 67(2), 234-250.

Almeida, J., Ribeiro, J., & Martins, M. (2020). Soybean yield prediction using ARIMA models in Brazil. *International Journal of Agricultural Research*, 15(4), 467-482.

Anis, B., Rahman, M., & Waseem, A. (2021). A comparison of ARIMA and artificial neural network models for canola yield forecasting. *Computers and Electronics in Agriculture*, 178, 105663.

Barah, B. C., Betne, R., & Bhowmick, B. C. (2001). Status of rice production system in Assam: a research perspective. *Prioritization of strategies for agriculture development in north-eastern India: October 2001; New Delhi*, 50-68.

Bhowmick, B. C., Barah, B. C., Pandey, S., & Barthakur, N. (2005). Changing Pattern of Rice Production Systems and Technology in Assam: A spatio-temporal analysis of performance and prospects.

Box G.E.P. and Cox D.R. (1982). An analysis of transformations, revisited, rebutted. Journal of American Statistical Association, 77, 209 210.

Chaudhary, M. M., Sabu, J., & Yadav, R. (2023). Seasonal ARIMA forecasting model for pulse crop yields. *Agricultural Economics Research*, 45(3), 611-620.

Debnath, M. K., Bera, K., & Mishra, P. (2013). Forecasting area, production and yield of cotton in India using ARIMA model. *Research & Reviews: Journal of Space Science & Technology*, *2*(1), 16-20.
George E., Jenkins G. M. and Reinsel G. C. (2008). Time series analysis: Forecasting and control, (4th ed.). John Wiley & Sons, INC. Yafee and McGee, 1999
Gupta, R., Sharma, P., & Verma, A. (2020). Incorporating exogenous variables in ARIMA models for predicting sugar cane production. *Journal of Supply Chain Management Science*, 9(1), 18-29.
Kartika, N. D., Astika, I. W., &Santosa, E. (2016). Oil palm yield forecasting based on weather variables using artificial neural network. *Indones. J. Electr. Eng. Comput. Sci*, *3*, 626-633.
Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy*, *34*(5), 1388-1393.
Khashei, M., Bijari, M., & Ardali, G. A. R. (2009). Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing*, *72*(4-6), 956-967.
Kumar, S., Kumar, V., & Sharma, R. K. (2015). Sugarcane yield forecasting using artificial neural network models. *International Journal of Artificial Intelligence & Applications (IJAIA)*, *6*(5), 51-68.
Kumar, V., & Singh, A. (2019). Wheat yield forecasting in India: An ARIMA approach. *Journal of Agricultural Science*, 12(1), 12-27.
Oyedokun, R. A., & Obafemi, F. (2021). ARIMA modeling for maize production forecast in Nigeria. *Nigerian Journal of Economic and Financial Studies*, 14(2), 55-70.
Rada, M. A., & Anda, S. (2022). Seasonal ARIMA model for forecasting corn yields in favorable climatic zones. *Crop Science*, 62(5), 1232-1240.
Ramteke, L. D., Rao, M. S., & Kesavan, S. (2022). Rice production forecasting in India using ARIMA model. *Asian Journal of Agricultural Research*, 16(1), 94-103.
Shabri, A., Samsudin, R., & Ismail, Z. (2009). Forecasting of the rice yields time series forecasting using artificial neural network and statistical model. *Journal of Applied Sciences*, *9*(23), 4168-4173.
Sharma, P. K., Dwivedi, S., Ali, L., & Arora, R. K. (2018). Forecasting maize production in India using ARIMA model. *Agro-Economist*, *5*(1), 1-6.
Suresh, K. K., & Krishna Priya, S. R. (2011). Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Tech*, *13*, 23-26.
Wawale, S. G., Jawarneh, M., Kumar, P. N., Felix, T., Bhola, J., Raj, R., ... &Boddu, R. (2022). Minimizing the error gap in smart framing by forecasting production and demand using ARIMA model. *Journal of Food Quality*, *2022*.
Zubair, M., & Hashmi, M. (2021). Economic impacts of ARIMA model forecasting in sustainable agriculture development. *Sustainable Agriculture Reviews*, 35(1), 219-233.