Bulletin of Business and Economics, 13(2), 481-486
https://bbejournal.com
https://doi.org/10.61506/01.00358

Forecasting Mango Production and Area in Punjab, Pakistan for 2021-2029

Sana Arooj[1], Yasir Nawaz[2], Abdul Hafeez[3], Ghulam Mustafa[4]

Abstract
Production of Mango in Pakistan has increased due to use of improved farm inputs and better management practices. Despite an increased production and rising demand in the export market, the potential of mango export has, however, not been fully achieved. Pakistan has comparative advantage in the production of Mango and enormous potential exists for its export in the vast Middle East market. The Punjab province of Pakistan which account for 63% of the country's total mango output. The objective of this study is to predict the area and output of mangoes of Pakistan's Punjab province for the year 2029. For this purpose, the 63-year data from 1957-2019 has been collected from Crop Reporting Service Punjab. Box-Jenkins method has been used to build an appropriate Univaraite ARIMA model. The best model was chosen after a comparison of the models' mean absolute percentage error, root mean squared error, stationary R-square, and other properties. The area's acceptable model was ARIMA (1,1,0) and production's acceptable model was ARIMA (1,1,1). Additionally, this estimate predicts that by 2029, the area will be (246.46-279.63) hectors and output will be 1358.10-1538.47) million tons. The results of this study will be useful to government planners, marketers, businesses and purchasing nations.
Keywords: Area, ARIMA, Box Jenkins-approach, Correlograms, Forecast, Mango, Stationary

1. Introduction
Pakistan produces and exports a variety of fruity kinds; however, mangoes are the most popular owing to its flavor, color, size, quantity, and sale. Mango is the second major fruit crop in Pakistan (Khan *et al.*, 2008). For many generations mango has been regarded as the "King of fruits," and it is a prominent crop grown in nations such as with the Islands, Mexico, India, Brazil, Pakistan, China, and Thailand (Jam *et al.*, 2013; Kumar *et al.*, 2020). Pakistan is the fifth-largest mango producer in the world (Mohsin *et al.*, 2021). Punjab and Sindh, which together produce 99.7% of all the mangoes consumed in Pakistan (Khan *et al.*, 2008)Mangoes are farmed across Pakistan, although the provinces of Punjab and Sindh contribute the most. Mango production in Pakistan is mostly concentrated in two provinces: Punjab and Sindh, which account for 99.7% of the country's total mango output, with Punjab accounting for 63% and Sindh for 37% (Khan *et al.*, 2008). Punjab is the biggest producer in Pakistan. Chaunsa, Anwar Ratol, Fajrii Kalan, Langra, Duseri, Samer Bahisht, and Bidar are a few of the well-known mango varieties in Pakistan (Mohsin *et al.*, 2021).
According to (Qureshi *et al.*, 2014)Pakistan exports mango mostly to United Arab Emirates, United Kingdom, Afghanistan, Oman, Saudi Arabia, Azerbaijan, and Qatar. Mangoes have been transported, and their manufacturing costs are considerably lower than their retail expenses. The most well-known nations for the production of mangoes are Thailand, Maxico, the Netherlands, Peru, Brazil, India, Pakistan, Spain, and China (Kumar *et al.*, 2020). Due to its delicious flavor and beneficial nutritional qualities, Pakistani mangoes are the second most significant fruit after citrus in terms of consumption and industry worldwide (Jam *et al.*, 2013; Khan *et al.*, 2008; Mohsin *et al.*, 2021).
The Box-Jenkins (1970) ARIMA model is mostly used for forecasting variety of different agricultural outcomes (Kumari *et al.*, 2022) a literature is available in forecasting area and output of fruits including mango. For example, (Jam *et al.*, 2012) carried out study to develop an appropriate Box-Jenkins technique-based Univaraite ARIMA model for assessing the cropping intensity of mangoes in Pakistan. The ARIMA model is the most reliable for predicting sale dates in Pakistan (Naz *et al.*,2012). utilized the ARIMA model for forecasting Kinnow exports from Pakistan to other nations. They highlighted that Europe and the Arabian Peninsula are Pakistan's Kinnow markets. (Ullah *et al.*, (2018) predicted the area and output of peach in Pakistan. (Kumar *et al.*, 2020) used an ARIMA model to project the area and output of mangoes in Himachal Pradesh. (Garde *et al.*, 2023) constructed forecasting models for regional production as well as forecasting mango prices.
(Kumari *et al.*, 2022) used different models to forecast the area, production and productivity of mango in Gujrat, India. The researcher used different artificial neural network models for forecasting. (Varalakshmi *et al.*, 2023) analyze the trend and forecast in area, production and productivity of mango crop in Karnataka, India.to estimate the trend and forecast, the researcher used linear, quadratic, exponential, logistic and Gompartz models were fitted and the best-fitted model was selected based on lowest MAPE. (Kumari *et al.*, 2022) presented a report on different kinds of fruits like mango, citrus, olive, peaches and strawberry and banana. The researcher applied regression analysis and ARIMA techniques to check future trend of these mentioned crops.

2. Methodology
In this study, sixty-three (63) years' time series data starting from 1957-58 and ended in 2019-20 has been used to predict how much area and production will be available for mangoes in 2029. The time series data were examined using SPSS (version 26) and Eviews (version 10) software. Data was gathered from the Crop Reporting Service (CRS) Punjab Pakistan.

2.1. ARMA, ARIMA Model & Box Jenkins Methodology
To get the greatest match between a time series and previous values in order to generate predictions, the Box-Jenkins technique employs ARIMA models. This approach is sometimes employed when there are indications of non-stationary in the data and the non-stationary may be removed using a first integrated phase. (Box-Jenkins, 1976). Box and Jenkins is a well-known strategy that combines the moving average and autoregressive processes. Consequently, the model is known as the ARMA model. One of the key presumptions of the Box-Jenkins model is that the time series data must be stable. To achieve stationary, Box and Jenkins advocate differencing non-stationary series once or more. An ARMA model is defined as follows:

[1] Lecturer, Department of Statistics, Government Commerce College for Women, Rahim Yar Khan, Pakistan, aroojsana483@gmail.com
[2] Lecturer, Department of Statistics, University of Veterinary & Animal Sciences (UVAS), Lahore, Pakistan, yasirnawaz0300@gmail.com
[3] Assistant director, Crop Reporting Service Agriculture, Department, Punjab, Pakistan, hafeezstat05@gmail.com
[4] Deputy Director, Crop Reporting Service Agriculture, Department, Punjab, Pakistan, gmstatbzu@gmail.com

$$y_t = \theta + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \beta_o \mu_t + \beta_1 \mu_{t-1} + \cdots + \beta_q \mu_{t-q} \qquad (1)$$

where y_t has both characteristics of AR and MA model there for its known as ARMA model. In an ARMA(p,q) process, there will be p autoregressive and q moving average terms. θ represents a constant term. When we take difference of the ARMA model given in Eq. (1), it become an ARIMA (p, d, q). There is p autoregressive and q moving average, d is differencing term and I stand for "Integrated,". ARIMA model is given as follows:

$$\Delta y_t = \theta + \Delta a_1 y_{t-1} + \cdots + \Delta a_p y_{t-p} + \Delta\beta_o \mu_t + \Delta\beta_1 \mu_{t-1} + \cdots + \Delta\beta_q \mu_{t-q} \qquad (2)$$

where Δ represent the difference term. The Box Jenkins approach consists of four steps: model identification, parameter estimate, diagnostic verification, and forecasting.

2.2. Autocorrelation

"Autocorrelation refers to the way the observation in a time series are related to each other and is measured by a simple correlation between current observation y_t and the observation p periods from current one, i.e., $y_{(t-p)}$" (Gujarati & Porter 2008). The function used for autocorrelation is referred to as autocorrelation function (ACF).

2.3. Partial Autocorrelation

Partial Autocorrelation measures the correlation between time series observation that are k time periods apart after controlling for correlation at intermediate lags. 'I-e' lags less than k. In other words, it is used to measure the degree of association between y_t and $y_{(t-p)}$ when the effects at other time lags 1, 2, 3, … , $(p-1)$ are removed." (Gujarati & Porter 2008). Partial Autocorrelation function is denoted by PACF.

2.4. Unit root test

The most common test for determining whether data is stationary or non-stationary is the Augmented Dickey Fuller (ADF) unit root test. Apart from alternative hypotheses with strongly defined departures from equilibrium, the major advantage of utilizing panel unit root testing is that their effect is considerably stronger due to the low power of traditional time-series unit root tests in determinate samples. The ADF test numbers clearly fall into the acceptable range. As a result, the null hypothesis, that the data is non-stationary, will be confirmed, while the hypothesis 2, that the data is static, will be dismissed. As a result, our data set on mango area in Punjab is clearly non-stationary.

3. Results and Discussion

The first step in the time series analysis is to determine the stationarity of the data. The most popular methods are given below.

3.1. Sequence Plot

In econometric time series analysis, figuring out if the series is stationary is crucial. The presence of stationary can have a substantial influence on the behavior of a series. After differentiating the sequence plot, it is clear that the data meet the stationary assumption. During the span of time, the mean and variance of a stationary series stay constant. At ground level, the sequence of mango regions was found to be non-stationary, but at the first difference, it became stationary. Correlograms indicates the same results in first glance.

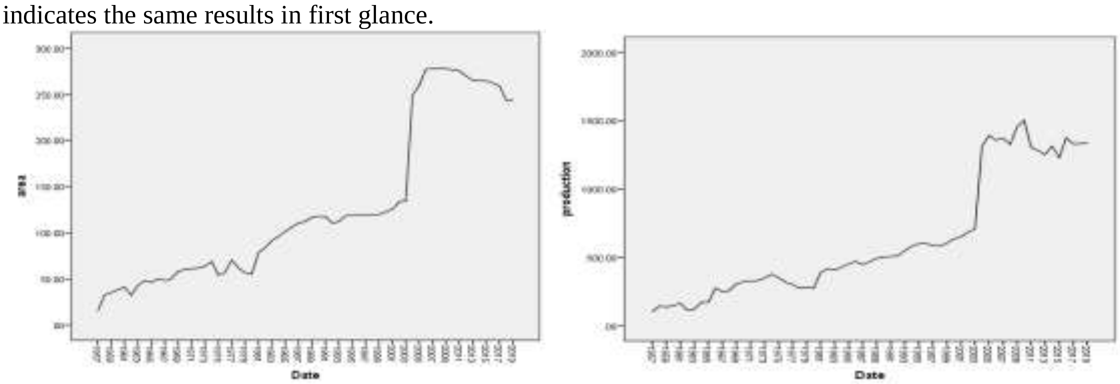

Figure 1: Sequence Plot of Mango Area (000) and production of Punjab (Original series)

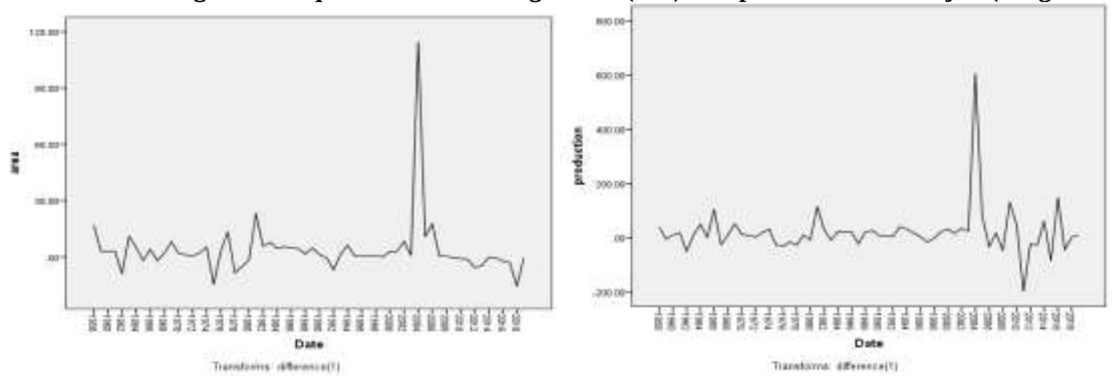

Figure 2: Sequence plot for mango area and production of Punjab after 1st difference

3.2. Augmented Dickey Fuller test (ADF)

The Augmented Dickey Fuller (ADF) unit root test is the most used method for detecting whether data is stationary or non-stationary. The main benefit of using panel unit root testing is that, in contrast to alternative hypotheses with clearly defined departures from equilibrium, their influence is much larger as a result of the limited power of conventional time-series unit root tests in determinate samples.

Table 1: ADF test of Mango Area (000) in Punjab after Difference 1st

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.235896	0.0000
Test critical values: 1%level		-3.542097	
	5%level	-2.910019	
	10%level	-2.592645	

Table 2: ADF test of Mango production in Punjab after Difference 1st

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.540439	0.0000
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

Unquestionably, the ADF test results are within acceptable limits. As a result, the second hypothesis—that the statistics were static—will be refuted, supporting the null hypothesis—that the data are not stationary. Therefore, it is clear that our data set on the Punjab mango region is non-stationary. Therefore, after gaining a major striker difference, our variable has become stationary at the initial difference.

It was determined that the data needed to be separated in order for the sequence chart to become stationary since the mean and variance in this chart do not demonstrate a constant value and show a trend. When the original difference is taken into account, the data become stationary. The Correlograms of ACF and PACF were examined in order to identify the order of Autocorrelation AR, the value of p, and for Moving average MA, the value of q, since the value of d difference is set at 1. Analysis revealed that p and q had identical values of zero.

The Correlograms of the autocorrelation function (ACF) of differenced series (Figure 2) demonstrates that the autocorrelation function (ACF) declines quickly after one lag, hence the value of "q" was selected to be "0." Additionally, the partial autocorrelation function (ACF) of the Correlograms for the differenced series (Figure 3) demonstrates that the autocorrelation function rapidly declines after one lag, The Correlograms of ACF and PACF has one lag significant so the ARIMA (1,1,0) may be chosen for the estimation of parameter and forecasting of area of mangoes of Punjab, Pakistan. The ARIMA method is used for parameter estimation, model validation, and predictions of the Punjabi mango area.

3.3. Correlograms

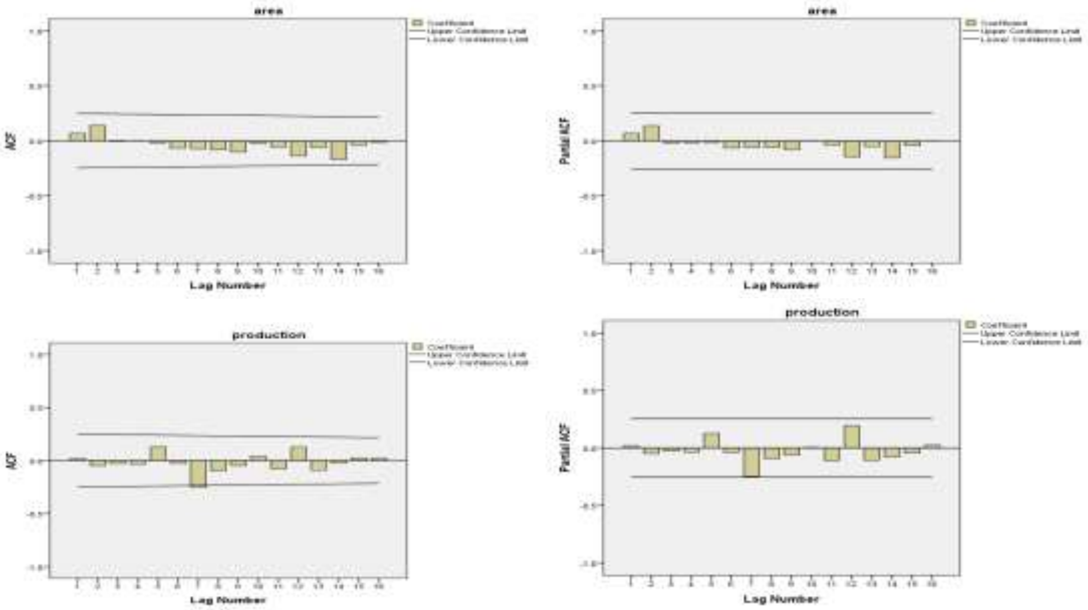

Figure 3: Correlograms of ACF and PACF of differenced series for mango area and yield in Punjab Pakistan

3.4. Comparison of Models

Table 3: ARIMA Models for Time Series Data of Mango Area and Corresponding Selection Criteria

Fit statistics	Mean (0,1,0)	Mean (0,1,1)	Mean (1,1,0)	Mean (1,1,1)
Stationary R^2	2.220E-16	.003	.004	.011
R-squared	966	.966	.967	.967
RMSE	15.730	15.835	15.827	15.906
MAPE	6.718	6.641	6.633	6.650
MaxAPE	44.544	44.605	44.617	44.597
MAE	6.692	6.518	6.475	6.421
MaxAE	110.732	110.884	110.913	110.863
Normalized BIC	5.578	5.658	5.657	5.733

Table 4: ARIMA models for Time Series data of mango production and Corresponding Selection Criteria

Fit Stationary	(0,1,0)	(0,1,1)	(1,1,0)	(1,1,1)
Stationary	1.1E- 16	.000	.000	0.002
R2	.959	.959	.959	.959
RMSE	90.089	90.819	90.819	90.519
MAPE	7.980	7.93	7.93	7.852
MaxAE	64.262	64.23	64.23	63.984
MAE	41.971	41.696	41.696	41.328
MaxAE	584.127	584.055	584.055	584.41
BIC	9.068	9.151	9.151	9.233

Four distinct models for area are evaluated in the table above to get the best fit model. Various models and parameters were calculated at this point. The estimated models are evaluated to a variety of criteria, including The value of R-squared and Stationary R-squared must be maximum, while the remaining values of errors, (RMSE), (MAPE), (MaxAPE), (MAE) and normalized BIC must be minimum. The parameters of ARIMA (1, 1, 0) were discovered when studying the preceding table, and this model was found to be an acceptable model since it had the highest R2- stationary and R-squared values. The flaws in this model, including MAPE and MAE, are minimal. We conclude that the model (1,1,0) is a suitable model for forecasting the area of mangoes in Pakistan's Punjab region based on these data.

Four distinct models are calculated and compared for mango production in the table above to identify the best fit model. The estimated models are compared to several criteria, such as the highest value of R-squared and Stationary R-squared, as well as the lowest values of (MAPE), (MAXAPE), (MAE), and Normalized BIC. The parameters of ARIMA (1, 1, 1) were discovered when evaluating the above table, and this model was found to be a suitable model since it has the least error, while the majority of errors, such as MAPE, MAXAPE, and MAXAE, are the least of this model. Based on the findings, we conclude that model (1,1,1) is an adequate model for forecasting mango output in Pakistan's Punjab province.

3.5. Estimates of Parameters

Table 5: Parameters of ARIMA model(area)

Area	Estimate	SE	t	Sig.
Constant	3.688	2.147	1.718	.047
AR lag 1	.065	.129	.502	.058
Difference	1			

At this stage we estimate parameters. It was concluded that parameters are significant. Because p-value is less than 10% level of significant.

Table 6: Parameters of ARIMA model (production)

production	Estimate	SE	t	Sig
Constant	-565.801	1326.962	-426	.0671
AR lag 1	-583	2.292	-254	.0800
Difference	-651	2.229	-278	.0783
MA lag1	.295	.667	.441	.0661

At this stage we estimate parameters. It was concluded that parameters are significant. Because p-value is less than 10% level of significant.

3.6. Residual Plots

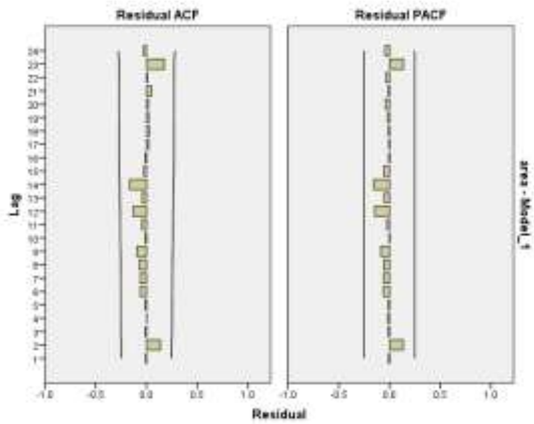

Figure 4: Residual Plots (Correlograms) of ACF and PACF for identified ARIMA model (1,10)

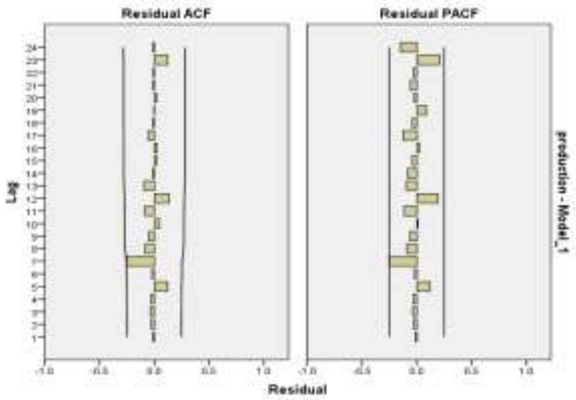

Figure 5: Residual Plots (Correlograms) of ACF and PACF for identified ARIMA model (1,1,1).

As shown by the autocorrelation function (ACF) and partial autocorrelation function (PACF), almost all coefficient values are first found to be within the upper and lower confidence limits. The diagnostic check is also used to determine if the residuals of the autocorrelation and partial autocorrelation functions must be contained within the confidence intervals. We can now say that mistake is being decreased.
Forecasting of area and production

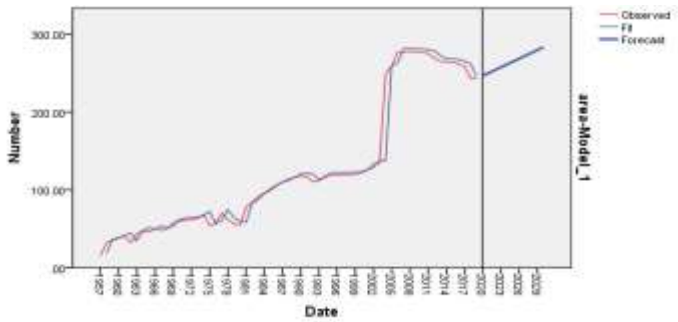

Figure 6: Forecasted Value of Mango Area Punjab, Pakistan

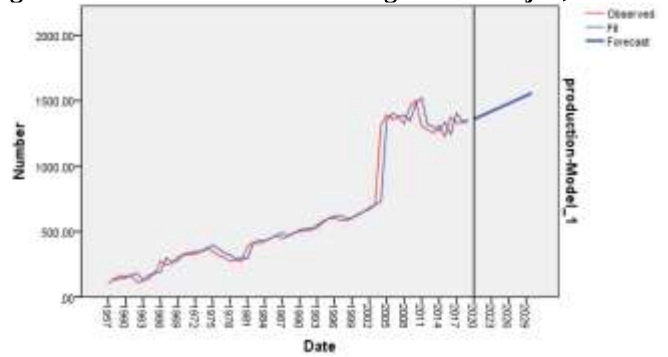

Figure 7: Forecasting graph for mango production of Punjab

Table 7: Forecasting of ARIMA model

Years	Forecast	95% Lower Confidence limit	95% Upper Confidence limit
2020	246.46	214.80	278.12
2021	250.13	203.88	296.37
2022	253.82	196.51	311.12
2023	257.50	190.95	324.06
2024	261.19	186.52	335.86
2025	264.88	182.90	346.88
2026	268.57	179.87	357.27
2027	272.26	177.32	367.19
2028	275.95	175.15	376.74
2029	279.63	173.31	385.96

The current status of mango orchards and the anticipated future state of mangoes are shown in the table above. The ARIMA (1, 1, 0) and (1,1,1) model was used to determine mango area and production forecasts (with 95 percent confidence intervals) for the years 2020 to 2029. Predictions are displayed in this table (along with upper and lower limits at 95 percent confidence intervals). The figures in the table demonstrate an increase in mango area and production between 2020 and 2029, with predicted mango area (forecast values) ranging from 246.46 to 279.63 hector tones and 946.97 to 2129.97 million tons. This suggests that there will be more mangoes accessible for both home and international consumption in the future. In this case, our predicted values are more in line with the observed data, and the model provides the best match.

Table 8: Forecasting of mango production of Punjab

Years	Forecast	95% Lower Confidence limit	95% Upper Confidence limit
2020	1358.01	1174.88	1541.14
2021	1379.23	1115.63	1642.84
2022	1398.50	1075.85	1721.15
2023	1418.84	1045.42	1792.27
2024	1438.59	1020.99	1856.19
2025	1458.67	1000.90	1916.44
2026	1478.56	984.00	1973.13
2027	1498.56	969.69	2027.43
2028	1518.50	957.45	2079.55
2029	1538.47	946.97	2129.97

4. Conclusion

The ARIMA model was fitted to analyze the trend in area and production of mangoes in Punjab, Pakistan. Result by the current study revealed that ARIMA is a useful model for predicting the production and area of mangoes in Punjab, Pakistan. The best accurate model for forecasting mangoes area and production in Punjab, Pakistan is ARIMA (1,1,0) and (1,1,1) respectively. According to projections, Punjab's mango farming would cover 279,63 hectors in 2029. The forecast said that in 2029, Punjab will produce 1538.47 tonnes of mangoes. Mango output, area, and productivity have all been seen to increase during the course of the study.

4.1. Recommendations

Mango plantations that are well-managed and well-managed may assure long-term fruit production and can help Pakistan increase its mango exports. The following recommendations are made for enhancing mango output in the country.

- According to the prediction, there will be a rise in mango production in the future. Given the current trend, governments should invest in infrastructure to handle excess fruit output.
- Stakekeholders should be given incentives to encourage the expansion of the packaging and processing businesses. Increased marketable surplus will be ensured by improved infrastructure (better highways, refrigerated shipping, and cold storage). A well-thought-out manufacturing and marketing strategy may ensure a bright future for mango.
- A campaign to increase mango exports should be launched. New markets should be explored, and a value-addiction culture should be encouraged. Governments should devise a strategy to deal with the new problems posed by the WTO regime's norms and international standards.

References

Garde, Y., Chavda, R. R., Thorat, V. S., Pisal, R. R., Shrivastava, A., & Varshney, N. (2023). Forecasting area, productivity and prices of mango in Valsad District of Gujarat: Time series analysis. *Environment Conservation Journal*, *24*(2), 218–227.

Khan, M., Mustafa, K., Shah, M., Khan, N., & Khan, J. Z. (2008). Forecasting Mango Production in Pakistan an Econometric Model Approach. *Sarnahad J. Agric*, *24*(2), 363–370.

Kumari, P., Parmar, D. J., M, S. K., Lad, Y. A., & Mahera, A. B. (2022). *Forecasting area , production and productivity of mango in Gujarat by using an artificial neural network model*. *11*(4), 822–826.

Qureshi, M. N., Bilal, M., Ayyub, R. M., Ayyub, S., Degree, G., & Sharaqpur, C. (2014). Modeling of Mango Production in Pakistan. *Sci.Int,(Lahore)*, *26*(3), 1227–1231.