Environment and Ecology 43 (3A) : 888—904, July—September 2025
Article DOI: https://doi.org/10.60151/envec/WEVG2444
ISSN: 0970-0420 (Print), ISSN: 3049-3625 (Online)

Advanced Time Series Forecasting of Turmeric Production using ARIMA Modelling: An Empirical Analysis in Selected Southern States of India

Deepa V., Vishal Mehta, Abhishek Pal

Received 12 June 2025, Accepted 23 August 2025, Published on 23 September 2025

ABSTRACT

Turmeric, an important part of India's agricultural heritage, significantly contributes to its export economy. Southern states are the main hubs of turmeric cultivation. This study analyzes 43 years (1980-2022) of data from the Indiastat database to examine trends in area, production and yield across Southern India. ARIMA models were utilized to forecast turmeric production trends across the Southern states of India, with model selection guided by minimizing the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the number of residual spikes exceeding the confidence interval. For Andhra Pradesh, the ARIMA (2,1,3) model was identified as the best fit, reflecting fluctuations in production. In Karnataka, ARIMA (1,1,3) provided the most suitable model, while Tamil Nadu and Kerala showed increasing trends best captured by ARIMA (2,1,2) and ARIMA (1,1,1), respectively. This approach ensured that the selected models offered both statistical robustness and reliable forecasting for each state's production dynamics.

Keywords Turmeric, Andhra Pradesh, Tamil Nadu, Karnataka, Kerala, Autoregressive integrated moving average.

INTRODUCTION

India, known as the "Spice Bowl of the World," has a rich tradition of cultivating and exporting key spices like pepper, cardamom, chilies, turmeric and ginger. Turmeric (*Curcuma longa*), or "Indian Saffron," native to Southeast Asia, is mainly grown in Kerala, Tamil Nadu, Maharashtra, and Andhra Pradesh and used for its golden color in foods and beverages (Srinivas 2021).

India is the leading turmeric exporter, earning US$226.6 million in 2023-24 and US$237 million from April to November 2024-25. It accounts for over 62% of global turmeric trade and produced

Deepa V.[1], Vishal Mehta[2]*, Abhishek Pal[3]

[2]Assistant Professor
[1,2,3]Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture & Technology, Kumarganj 224229, Ayodhya, Uttar Pradesh, India

Email: mehta.vishal@nduat.org
*Corresponding author

11.61 lakh tonnes in 2022-23, representing 75% of world output across 18 states, including Maharashtra, Telangana, Tamil Nadu, Karnataka, Madhya Pradesh, Andhra Pradesh and Odisha. Over 30 varieties, both traditional and research-developed, are grown in more than 20 states (Indus Food 2024).

After bifurcation, turmeric cultivation in Andhra Pradesh declined but has recently rebounded. In 2021-22, Telangana led with 14.33% of the area, 24.75% of production, and 6,600 kg/ha productivity, while Andhra Pradesh reported 26,000 hectares and 75,000 tonnes, with low yields due to cyclone damage. By December 2022, turmeric covered 0.18 lakh hectares, slightly below the previous year (ANGRAU, crop outlook, 2022, India. Ministry of Agriculture and Farmers Welfare 2021). In Telangana, turmeric is mainly grown in Nizamabad report, Nirmal and Jagtial. Nationally, turmeric area rose from 1.02 lakh ha in 1980-81 to 3.3 lakh ha in 2021–22, and production from 2.17 to 12.21 lakh tonnes. Telangana's area expanded from 22,000 to 34,000 ha and production from 51,000 to 2.16 lakh tonnes. Key varieties include Nizamabad bulb, Amruthapani, Armoor, Duggirala and Tekurpeta (Anusha *et al.* 2023). In 2020, Tamil Nadu contributed 17% to national production, cultivating 25,000 ha for 1.5 lakh tonnes annually (Sangeetha & Brindha 2021). Leading varieties like Erode Local, Salem, BRS-1 and Prathibha are known for color and curcumin. Erode district dominates with the "manjal" variety on 12,570 acres (Rubasri & Ramanathan 2024). Karnataka ranked fourth in 2022-23 with 19.44 thousand ha and 1.1 lakh tonnes. Key districts include Belagavi, Chamarajanagar, Bagalkote and Mysuru, growing varieties like Sugandham, Salem, Duggirala and Erode Local (Naik 2022). Kerala, though a smaller producer, cultivates turmeric on 6,000 ha with 40,000 tonnes annually. Major areas are Wayanad, Idukki and Palakkad, known for high-curcumin varieties like Alleppey Finger, Kasturi and Prathibha (CEIC Data 2023).

Bhattacharyya (2024) found that ARIMAGA (2,1,1) outperformed standard ARIMA in forecasting chilli and turmeric production by achieving lower MAPE and RMSE. Krishnan *et al.* (2022) concluded that the ANN model best predicted rainfall in the Northern and central zones, while the SARIMA (1,0,1), (2,0,0) model was most accurate for the Southern zone. Kumar and Baishya (2020) identified ARIMA models tailored to each state and India as best-fitted for forecasting potato prices, based on lowest AIC, BIC, MAPE and RMSE values. Padhan (2012) found that ARIMA models effectively forecasted annual productivity for 34 agricultural products, with tea showing the lowest MAPE and cardamom the lowest AIC among the selected crops. Yadav *et al.* (2022) concluded that the ARIMA (0,1,1) model predicts the area under nutri-cereals in India will decline by about 1% annually over the next decade. Kumar *et al.* (2024) analyzed ARIMA modelling, which showed continued declines in ragi area and production in both Karnataka and Uttar Pradesh, with ARIMA (2,2,1) and (2,1,2) providing the best fit for forecasting. Supriya *et al.* (2023) studied and concluded that ARIMA models best forecast lentil production trends in major Indian states, with increasing instability in most regions except Uttar Pradesh, where yield improvements drove production changes. Kumar (2016) studied ARIMA modelling for potato area and production in India and concluded that both area and production under cultivation would increase during the forthcoming years. Prabhu *et al.* (2022) analyzed the forecasting behavior of millet in India and showed that the area and production showed a decreasing trend. Sathish Kumar *et al.* (2022) studied trend analysis on area, production, and productivity of minor millets in India and showed that the area, production showed a decreasing trend and productivity showed an increasing trend.

This research is designed to deliver an in-depth exploration of Autoregressive Integrated Moving Average (ARIMA) modelling as applied to turmeric cultivation data from key Southern states. Utilizing secondary datasets and robust analytical methods, the study intends to produce practical findings that will assist policymakers, agricultural extension bodies, and farmer groups in promoting greater stability and growth within the turmeric industry.

MATERIALS AND METHODS

All data used in this analysis is sourced from secondary data on turmeric area, production and yield from 1980 to 2022 were obtained from www.

indiastat.com. It is important to note that, following the creation of Telangana as a separate state on June 2, 2014, statistics for Telangana's area, production, and productivity are included under Andhra Pradesh.

Autoregressive integrated moving average (ARIMA)

To accomplish the objective, a range of predictive models have been formulated to estimate turmeric production and efficiency in the South regions of India utilizing ARIMA modelling. The ARIMA (p, d, q) method is described by

$$y_t=\theta_0 +\phi_1 y_{t-1}+\phi_2 y_{t-2}+...+\phi_p y_{t-p}+\varepsilon_t -\phi_1\theta\varepsilon_{t-1} -\theta_2 \varepsilon_{t-2} -.. \phi_q \varepsilon_{t-q}$$

Where y_t and ε_t are the actual value and random error at Time period t, p and q are model parameters, respectively ϕ_i (i = 1,2,.....*p*) and ϕ_j (*j*=1,2....*q*) are integers that are frequently referred to as model orders. Random mistakes ε_t are considered to be distributed randomly and identically, with a mean of zero and a constant variance σ^2.

Model approach

The ARIMA (Box-Jenkins model) is created in sequential stages as follows:

Model identification and model selection

It is essential to confirm that the variables are stationary by detecting seasonality in the dependent series, applying seasonal differencing if needed. Additionally, plots of the lagged correlation and sequential correlation functions of the dependent time series should be utilized to identify which components are autoregressive or moving average in the model.

Detecting stationary

The initial phase of developing the ARIMA (Box-Jenkins) model during the model identification stage involves assessing the stationarity of the time series data related to turmeric production and productivity in Southern states of India. Stationarity was confirmed through a sequential plot of runs, which indicated a consistent location and scale, signifying data stationarity. For instances of non-stationarity, a differencing approach was employed to achieve stationarity. Additionally, Box-Jenkins models utilize curve fitting by subtracting the fitted values from the original data.

Order of autoregressive process (p) and of moving-average process (q)

After ensuring the stability and accounting for periodic variation in the time series, the next step is to determine the lag orders for the self-regressive (p) and smoothing (q) components. Since the conditional serial correlation for an autoregressive process of order p approaches zero after lag (p+1), we reviewed the sample partial autocorrelation function for any significant values beyond this point. Likewise, because the serial correlation function for a moving average process of order q drops to zero after lag (q+1), we assessed the sample autocorrelation function to confirm its convergence to zero. These evaluations are usually carried out by visualizing the sample partial autocorrelation together with a 95% confidence interval—a feature commonly available in statistical software that produces these graphs. In cases where the software does not display a confidence band, it can be estimated as ±2/N divided by the square root of the sample size (N). The observed autocorrelation and partial autocorrelation values are then compared with their plotted confidence intervals to determine their statistical significance. Theoretical ACFs and PACFs for various models can be sourced from literature (Pankratz 2012) for different values of autoregressive and moving average components (p and q). Consequently, a comparison is made between the correlograms derived from the time series data and the conceptual ACF/PACFs to identify a suitable match, allowing for a tentative selection of one or more ARIMA models.

Estimation stage

The process of parameter assessment relies on computational algorithms to identify the coefficients that optimize the fit of the ARIMA model. Among the most prevalent approaches for this purpose are maximum likelihood estimation and non-linear least squares estimation.

Model estimation

Once a provisional model structure was determined, the model coefficients including autoregressive (AR) and moving average (MA) terms as well as other relevant parameters were calculated using maximum likelihood estimation or standard least squares methods. To further refine model selection, statistical penalty criteria such as the Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC), and adjusted Akaike Information Criterion (AICc) were utilized. These criteria impose penalties on model complexity, promoting simpler and more efficient models in line with the principle of parsimony. These statistics serve as one of several evaluations to ensure the adequacy of the selected models. Models with the lowest AIC and BIC values are considered to have residuals that approximate a series of uncorrelated random errors. The Akaike Information Criterion (AIC) for a model is determined by multiplying the count of estimated coefficients by two and subtracting twice the natural logarithm of the likelihood. In contrast, the Bayesian Information Criterion (BIC) is calculated as negative two times the log-likelihood plus the product of the natural logarithm of the number of observations and the number of independent variables. Here, the likelihood is represented by L, the sample size by n, and the quantity of free parameters by k. Following this, a statistical significance assessment (t-test) is performed using the estimated coefficients and their standard deviations. For a parameter to be included in the final model, the absolute value of its t-ratio should be greater than 1.96 or 2, indicating statistical significance; otherwise, non-significant parameters are trimmed from the model. Moreover, the estimated AR and MA parameters must adhere to specific boundary conditions, remaining between-1 and 1. If these parameters fall outside of this range, they are re-estimated, or an alternative candidate model is considered for estimation.

Diagnostic checking

Model checking confirms if the fitted ARIMA model's residuals are white noise-meaning independent, with constant mean and variance. This is tested by plotting residuals, examining ACF and PACF plots, and using the Ljung-Box test. ACF values should mostly fall within ±1.96/n bounds; otherwise, the model may be misspecified. The Ljung-Box test formally checks for residual autocorrelation, ensuring the model truly fits the data. The test statistic is formulated as: $Q = n\,(n+2) \sum_{k=1}^{h} \frac{r_k^2}{(n-k)}$ where n stands for the count of data entries within the series, r_k^2 is the square of the autocorrelation at lag k and h is the maximum lag considered. The hypotheses tested are:

H_0 : The set of autocorrelations for residuals is white noise (the model fits well).
H_1 : The set of autocorrelations for residuals differs from white noise.

The test statistic Q is compared against a chi-square distribution denoted as, $\chi^2_{\alpha,\,(h-p-q)}$, where α is set at 5% (0.05), *h* is the maximum lag considered and p and q are the orders of the AR and MA processes, respectively. If $Q < \chi^2_{\alpha,\,(h-p-q)}$, we accept the null hypothesis (H_0) and rject the alternative hypothesis, conversely, if $Q > \chi^2_{\alpha,\,(h-p-q)}$, it indicates that the residuals are not white noise if the test statistic falls within the extreme 5% (0.05) tail of the chi-square distribution. If these estimations prove inadequate, it necessitates revisiting earlier steps to construct a more suitable model.

Model selection criteria

Optimal ARIMA models were picked based on the following evaluation measures:

Akaike information criterion (AIC)

Akaike (1973) developed the Akaike Information Criterion (AIC) is used to compare models by balancing fit and complexity. For example, when analyzing factors like price, apple type, or region affecting apple juice ratings, researchers can fit different regression models and use AIC to select the most efficient one. The model with the lowest AIC is preferred. Once chosen, standard hypothesis tests can further examine specific variable impacts. The AIC value is calculated as follows:

$$AIC = -2 * \ln L + 2 * k$$

Where L represents the likelihood value, N denotes the number of recorded measurements, and k indicates the number of estimated parameters (Burnham and Anderson 2002).

Hurvich and Tsai (1989) subsequently improved this estimate to account for small sample sizes:

$$AICc = AIC + \frac{2k\,(k+1)}{n-k-1}$$

Where n is the sample size and k is the parameter count, AIC and AICc are used for model comparison. When n is much larger than k, AIC is sufficient, but AICc is more flexible and generally preferred. The model with the lowest AIC or AICc is best. These scores only rank models and have no absolute value. While AIC is common in environmental and biological sciences, it is less used elsewhere, though applications do exist in pharmacology and other fields.

Bayesian information criterion (BIC)

The Bayesian Information Criterion (BIC), introduced by Gideon E. Schwarz and also called the Schwarz criterion, is a model evaluation tool grounded in Bayesian reasoning. Similar to AIC, BIC uses the likelihood function but imposes a stricter penalty for extra parameters to prevent overfitting. This heavier penalty makes BIC especially useful when seeking parsimonious models in statistical analyses.

$$BIC = -2 - lNL + 2lnN * kAIC = -2 * \ln L + 2 * K$$

Where, L is the value of the likelihood, N is the number of recorded measurements, and k is the number of estimated parameters.

BIC is widely used to select time series and linear regression models, yet it is versatile enough for any model estimated by maximum likelihood. In practice, when competing models have identical parameter counts, BIC's selection process essentially reduces to choosing the model with the highest likelihood-making it a robust tool for comparing model fit across various contexts.

Forecasting

To ensure sustainable agricultural productivity, accurate time series forecasts for agricultural commodities are crucial. Such forecasting plays a vital role in stabilizing production and supporting long-term sustainability. Equipping decision-makers with reliable price forecasts enables better management of supply chains and can guide consumer purchasing patterns. Agricultural prices are influenced by consumer choices, supplier strategies and seasonal fluctuations-particularly those driven by climate change. Although human-driven factors may not always follow predictable seasonal patterns, they can be analyzed as relationships between historical trends and present conditions.

RESULTS AND DISCUSSION

Applying ARIMA time series modelling to forecast upcoming turmeric production levels in the Southern states of India

The aspect of the study has been illustrated with time series data on production of turmeric from 1980-2022.

Analysis of turmeric production in Andhra Pradesh by ARIMA modelling

Model-identification of area data

The initial phase of ARIMA modelling involves verifying if the chosen forecasting variable forms a stationary time series-meaning its mean and variance remain consistent over time. As illustrated in Fig.1,

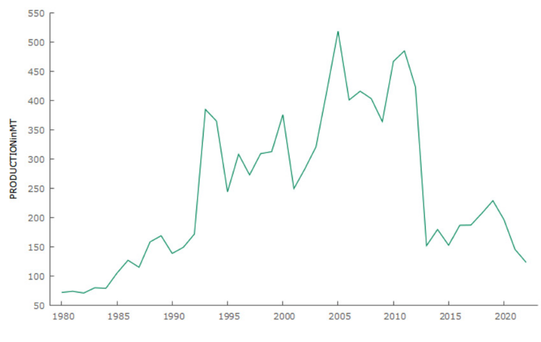

Fig. 1. Observed production of turmeric in AP.

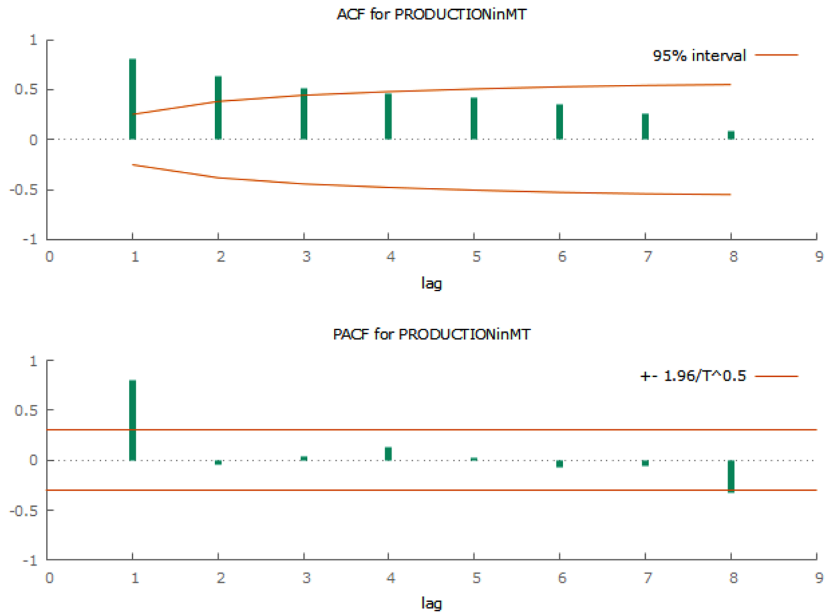

Fig. 2. First differenced series by lag area data ACF and PACF.

turmeric production data in Southern India exhibit non-stationarity, evident from a clear upward trend. Applying a unit root test with first-order differencing (d=1) can significantly enhance the efficiency of this assessment.

Augmented dickey-fuller test

The baseline hypothesis proposes that the dataset lacks stationarity. After properly differencing the chronological data, this assumption is examined using the augmented Dickey-Fuller (ADF) procedure with first-order differencing (d=1).

Dickey-Fuller = –1.882, Lag order = 7, p-value = 4e-09

Correlogram and partial correlogram of production

The first-order differenced data for turmeric production in Andhra Pradesh is represented in the correlogram ACF shown in Fig. 2, covering lags 1 to 18. Notably, the autocorrelation values for these lags do not exceed the significance level and gradually decline toward zero. Across lags 0 to 8, the autocorrelation coefficients remain within acceptable limits, indicating a lack of significant autocorrelation in this range.

Forecasting model selection

After transforming the non-stationary series to stationary through first-order differencing (d=1), the optimal ARIMA model parameters were identified by analyzing significant spikes in the autocorrelation and partial autocorrelation functions (ACF and PACF). Although various combinations of autoregressive (p) and moving average (q) terms were considered, the ARIMA (2,1,3) specification was uniquely selected based on the number of spikes exceeding confidence thresholds in the correlogram, indicating its suitability for the analysis.

An ARIMA (2, 1, 3) model applied to turmeric production data for this region means the model includes two autoregressive (AR) coefficients and three moving average (MA) coefficients. Only first-order differencing (d=1) is required to address non-stationarity in the series. Coefficient of the model and AIC

Table 1. Coefficient of turmeric production model.

ARIMA	Cons-tant	AR1	AR2	MA1	MA2	MA3
(2, 1, 3)	0.762	-0.4955	-0.922	-0.953	1.204	-0.324
SE	8.165	0.1307	0.071	0.175	0.159	0.181

Table 2. BIC and AIC values of ARIMA model applied to production data.

ARIMA model	σ^2 (Esti-mated)	Log-likeli-hood	AIC	BIC	AICc
(2, 1, 3)	5378	-242.913	499.825	482.07	495.51

and BIC values of the model for turmeric production has been illustrated in Tables 1— 2.

The ARIMA (2, 1, 3) model, characterized by parameters p=2, d=1, and q=3, is considered the most effective for forecasting. Its superiority over other models is demonstrated by its lower AIC and BIC values.

Forecasting used for selecting ARIMA model

To estimate the for next decade, ARIMA (2,1,3) model chosen is identified as the best-performing model was employed as outlined in Table 3.

The aim was to conduct time series forecasting utilizing the ARIMA (2, 1, 3) model. As depicted in Fig. 3, an 8-year prediction based on this model indicates a consistent upward and downward trend in Turmeric Production in AP.

Evaluating whether the forecast errors generated by the ARIMA (2, 1, 3) model are normally distributed with a mean of zero and consistent variance is a critical step. Additionally, conducting a residual autocorrelation check is necessary to confirm that the model adequately captures the underlying patterns in the data.

Metrics such as ME, RMSE, MAE, MPE, MASE, ACF and MAPE have been calculated to assess the correlation of the fitted ARIMA model. The insights are recapped in Table. 4.

Sequential prediction errors are indeed ACF and PACF. Fig. 4 provided visualization of the predicted ACF and PACF, respectively.

Although the ARIMA (2,1,3) model's residuals

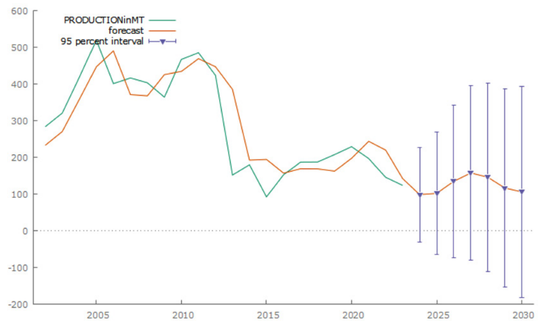

Fig. 3. Forecasts from ARIMA (2, 1, 3).

Table 3. Forecasted table for turmeric production in AP.

Year	2023	2024	2025	2026	2027	2028	2029	2030
Forecasted production	142.14	98.393	101.829	134.595	157.531	145.84	116.558	105.678

Table 4. Accuracy of production ARIMA (2, 1, 3).

ARIMA model	ME	RMSE	Accuracy of production model MAE	MPE	MAPE	MASE	ACF1
(2, 1, 3)	0.61324	65.839	44.972	4.257	20.774	40.976	-0.054

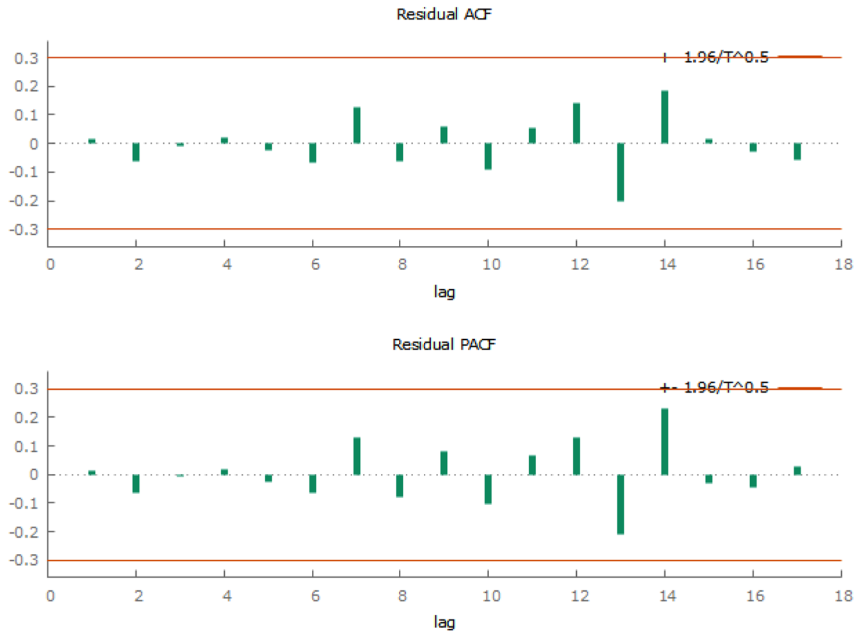

Fig. 4. Estimated ACF and PACF of residuals for the fitted ARIMA (2, 1, 3).

showed some significant coefficients in PACF and ACF plots up to lag 18, no significant autocorrelation was actually detected in the residuals.

Analysis of turmeric production in Tamil Nadu by ARIMA modelling

Fig. 5 illustrates the observed turmeric production in Tamil Nadu from 1980 to 2022.

Model-identification of production data

To begin ARIMA modeling, it's crucial to determine if the variable being forecasted is stationary, meaning its values vary around a stable mean and variance over time. Fig. 5, which presents turmeric production data, clearly demonstrates non-stationarity due to a noticeable upward and downward trend. Using a unit root test with first-order differencing (d=1) can significantly improve the efficiency of addressing this non-stationarity.

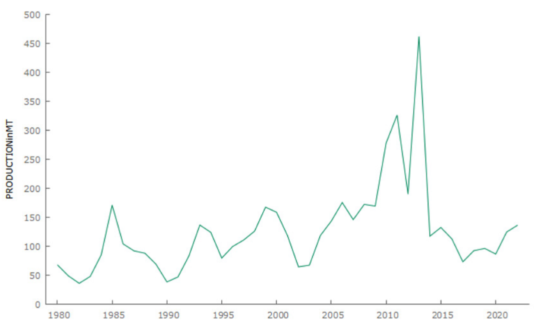

Fig. 5. Observed production of turmeric in Tamil Nadu.

Augmented dickey-fuller test

The baseline hypothesis proposes that the dataset lacks stationarity. After properly differencing the chronological data, this assumption is examined using the augmented Dickey-Fuller (ADF) procedure with first-order differencing (d=1).

Dickey-Fuller = -2.371, Lag order = 7, p-value = 3.875e-24

Correlogram and partial correlogram of turmeric

The correlogram ACF plot in Fig. 6 illustrates the autocorrelation for lags 1 through 8 of the time series transformed by subtracting each value from its

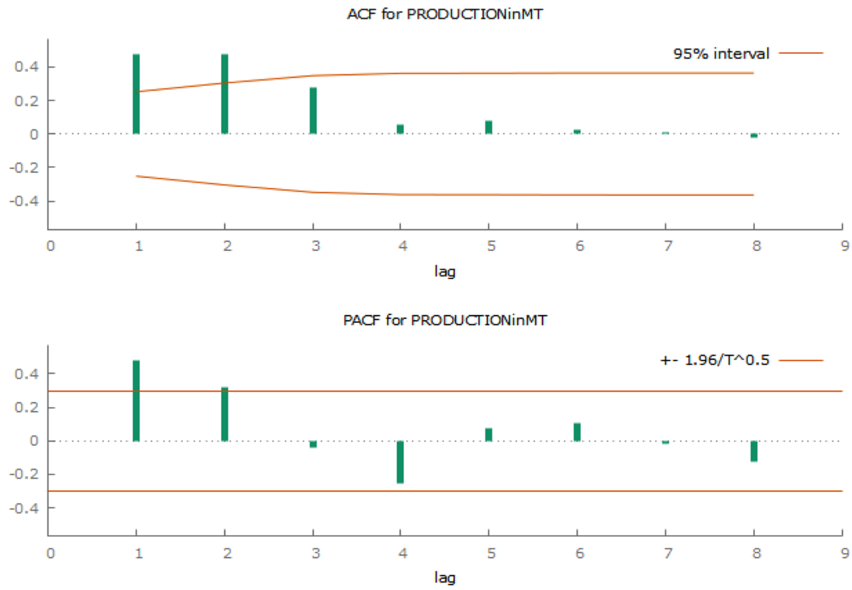

Fig. 6. ACF and PCAF first differenced series by lag in production.

previous value data for turmeric production in Tamil Nadu. As noted earlier, the autocorrelation values for these lags do not exceed the significance threshold and gradually decline towards zero, remaining consistent from lag 0 to 8.

Fig. 6 displays the Partial Autocorrelation Function (PACF) for lags 1 through 8, revealing a pattern of divergence over time.

Forecasting model selection

After converting the non-stationary series to a stationary one using first-order differencing (d=1), the optimal ARIMA model parameters were determined by examining the significant spikes in the autocorrelation (ACF) and partial autocorrelation (PACF) plots.

While multiple combinations of autoregressive (p) and moving average (q) terms were explored, The ARIMA (2,1,2) model was selected because the autocorrelation plot revealed several peaks exceeding the confidence intervals, indicating its suitability for the analysis. Coefficient of the model and AIC and BIC values of the model for turmeric production has been illustrated in Tables 5–6.

Using an ARIMA (2, 1, 2) model would mean there are 2 coefficients in MA term and 2 coefficients in AR term related to turmeric production for this particular state. The model would only involve differencing of order 1 (d=1).

Having evaluated several models, the ARIMA (2, 1, 2) model (p=2, d=1, q=2) was identified as the best for generating time series forecasts. This selection was based on its achieving the lowest BIC

Table 5. Coefficient of turmeric production model.

ARIMA	Constant	AR1	AR2	MA1	MA2
(2, 1, 2)	1.651	-0.957	-0.509	0.457	0.324
SE	7.491	0.528	0.276	0.518	0.341

Table 6. AIC and BIC values of the ARIMA model applied to production data in TN.

ARIMA model	σ (Estimated)	Log-likelihood	AIC	BIC	AICc
(2, 1, 2)	4790	-236.331	484.661	478.69	481.86

Table 7. Forecasted table for turmeric production in TN.

Year	2023	2024	2025	2026	2027	2028	2029	2030
Forecasted production	129.48	142.85	137.62	139.89	144.45	143.00	146.14	147.94

Table 8. Accuracy of ARIMA (2,1,2) model.

ARIMA model	ME	RMSE	Accuracy of production model MAE	MPE	MAPE	MASE	ACF1
(2, 1, 2)	-0.048	66.893	47.314	-11.682	39.798	44.029	0.456

and AIC values.

Selecting ARIMA model for forecasting

Utilizing the ARIMA (2, 1, 2) model, we can forecast the forthcoming of turmeric product ion in TN. Table 7 depicts a 8-year forecast of these predictions.

Fig. 7 shows forecast projection intervals. It's important to assess if forecast errors from the ARIMA (2,1,2) model follow a zero mean for normal distribution and constant variance.

It is crucial to evaluate the distribution of forecast errors from the ARIMA (2,1,2) model to determine if they conform to a normal distribution characterized by a zero mean and consistent variance. Additionally, conducting a diagnostic analysis of autocorrelation in the residuals (forecast errors) is essential.

Metrics such as ME, RMSE, MAE, MPE, MASE, ACF, and MAPE have been calculated to assess the

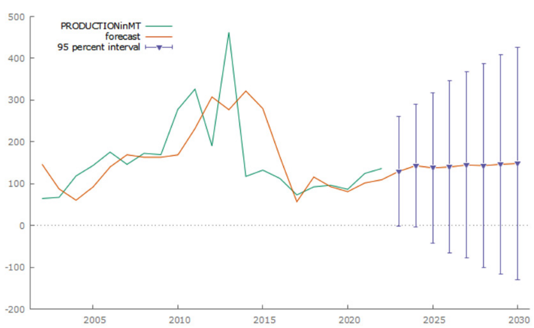

Fig. 7. Forecasts from ARIMA (2, 1, 2).

correlation of the fitted ARIMA model. The insights are outlined in Table 8.

Fig. 8 illustrates the ACF and PACF plots, respectively, aiming to discern the presence or absence of correlations between error values.

As observed in the ACF Figure above, the autocorrelation coefficient between lags 1 to 8 are fit inside the important bounds.

For the ARIMA (2,1,2) model, all PACF and ACF coefficients of the residuals are significant for lags 1 to 8. Nevertheless, neither ACF nor PACF was able to detect any autocorrelation in the residual predictions for these lags.

Analysis of turmeric production in Karnataka by ARIMA modelling

Fig. 9 illustrates the observed turmeric production in Karnataka from 1980 to 2022.

Model-identification of production data

The initial step in ARIMA modeling involves determining whether the target variable for forecasting is a stationary time series. A stationary series indicates that the values of the variable oscillate around a consistent mean and variance over time. In Fig. 9, the turmeric production data exhibits non-stationarity, as evidenced by the upward trend in the time series. Conducting a unit root test with the lowest-order differencing (d=1) can significantly enhance the efficiency of addressing this concern.

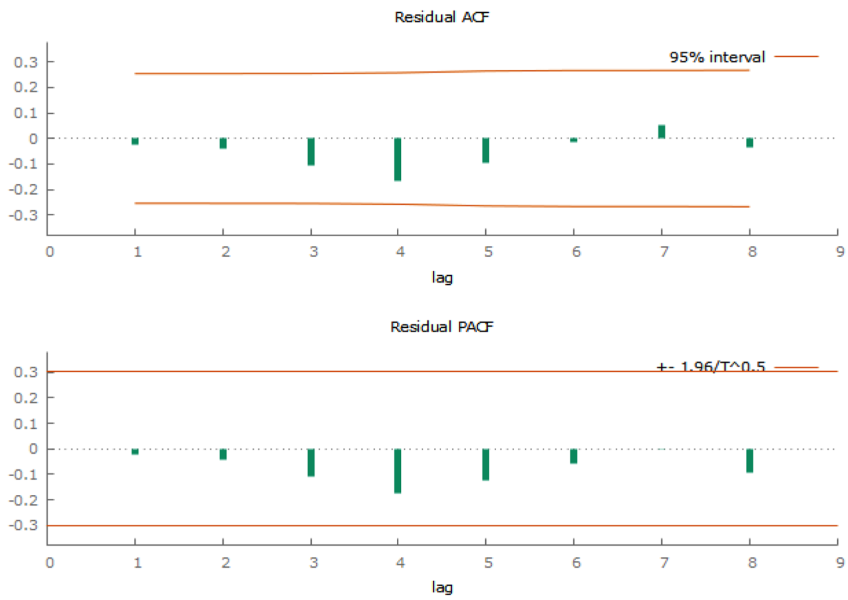

Fig. 8. Estimated ACF and PACF of residuals for ARIMA (2, 1, 2).

Augmented dickey-fuller test

The starting hypothesis is that the time series lacks a stable mean and variance. After transforming the series by first-order differencing to correct for instability, the Augmented Dickey-Fuller test is applied to check if a unit root is still present in the data.

Dickey-Fuller = -7.579, Lag order = 7, p-value = 8.157e-12

Correlogram and partial correlogram of turmeric

The correlogram ACF plot in Fig. 10. illustrates the autocorrelation for lags 1 through 8 of the first-order differenced time series data for turmeric production in Karnataka. As noted earlier, the autocorrelation values for these lags do not exceed the significance threshold and gradually decline towards zero, remaining consistent from lag 0 to 8. Fig. 10 displays the Partial Autocorrelation Function (PACF) for lags 1 through 8, revealing a pattern of divergence over time.

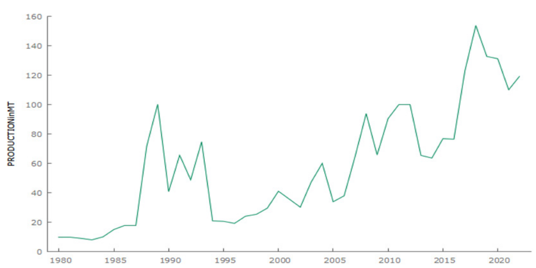

Fig. 9. Observed production of turmeric in Karnataka.

Forecasting model selection

Following the transformation of the non-stationary series into a stationary one via first-order differencing (d=1), the selection of optimal ARIMA model parameters was guided by a systematic analysis of the autocorrelation (ACF) and partial autocorrelation (PACF) functions. Various configurations of autoregressive (p) and moving average (q) terms were evaluated; however, the ARIMA (1,1,3) specification was ultimately chosen. This decision was based on the observation that the number and pattern of significant spikes exceeding the confidence intervals in the correlogram most strongly supported the suitability of this model for the data under consideration.

Using an ARIMA (1, 1, 3) model would mean there is 1 coefficient in the MA term and 3 coefficients

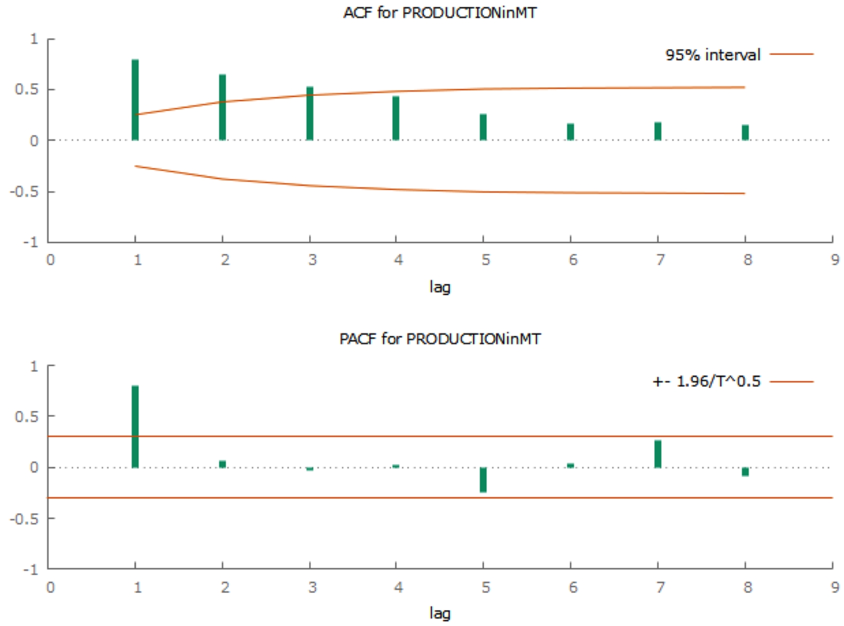

Fig. 10. ACF and PACF first differenced series by lag in production data.

Table 9. Coefficient of turmeric production model.

ARIMA	Constant	AR1	MA1	MA2	MA3
(1, 1, 3)	2.153	-0.3311	0.0727	-0.2875	-0.4621
SE	0.647	0.2671	0.2321	0.1387	0.2133

in the AR term related to turmeric production for this particular state. The model would only involve differencing of order 1 (d=1). Coefficient of the model and AIC and BIC values of the model for turmeric production has been illustrated in Tables 9–10.

The ARIMA (1,1,3) model has the smallest AIC and BIC values (p =1, d =1, q =3) making it the finest model for projecting future time series values.

Table 10. AIC and BIC values of the ARIMA model applied to production data.

ARIMA model	σ^2 (Estimated)	Log likelihood	AIC	BIC	AICc
(1, 1, 3)	514.6	-245.288	502.576	499.254	497.554

Selecting ARIMA model for forecasting

Utilizing the ARIMA (1, 1, 3) model, we can forecast the forthcoming turmeric production in KA. Table 11 depicts an 8-year forecast of these predictions.

Fig. 11 shows forecast projection intervals. It's important to assess if forecast errors from the ARIMA (1,1,3) model follow a zero mean for normal distri-

Table 11. Forecasted for turmeric production in Karnataka.

Year	2023	2024	2025	2026	2027	2028	2029	2030
Forecasted production	112.91	122.10	121.10	121.05	124.27	128.34	130.46	132.62

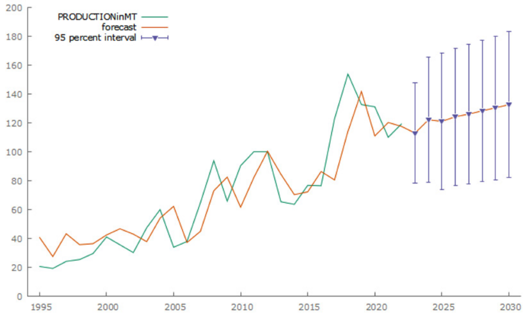

Fig. 11. Forecasts from ARIMA (1, 1, 3).

bution and constant variance.

Metrics such as ME, RMSE, MAE, MPE, MASE, ACF, and MAPE have been calculated to assess the correlation of the fitted ARIMA model. The findings are summarized in Table 12.

To determine the presence or absence of correlations between error values ACF and PACF plots are constructed for the same which is depicted in Fig. 12.

For the ARIMA (1, 1, 3) model, all PACF and ACF coefficients of the residuals are significant for lags 1 to 8. Nevertheless, neither ACF nor PACF was able to detect any autocorrelation in the residual predictions for these lags.

Analysis of turmeric production in Kerala by using ARIMA model

Model-identification of production data

The first phase in ARIMA analysis is to check if the variable to be forecasted has a stable mean and variance. In Fig. 13, the wheat production data displays a changing mean or variance, as seen in the increasing pattern over time. Using a statistical test for non-stationarity with first-order differencing can greatly improve the process of correcting this issue.

Augmented Dickey-Fuller test

The baseline hypothesis proposes that the dataset lacks stationarity. After properly differencing the chronological data, this assumption is examined using the Augmented Dickey-Fuller (ADF) procedure with first-order differencing (d=1).

Table 12. Validation of ARIMA (0, 1, 0) model.

ARIMA Model	ME	RMSE	Accuracy of production model MAE	MPE	MAPE	MASE	ACF1
(1, 1, 3)	-0.4698	17.737	12..623	-44.908	58.767	13.266	0.0478

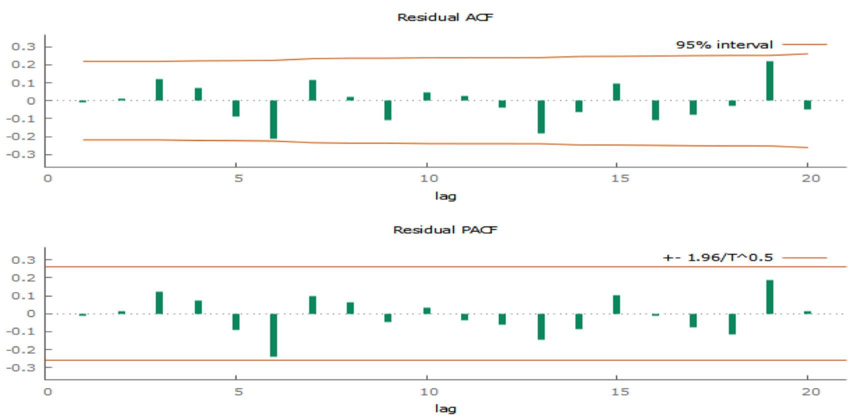

Fig. 12. Estimated ACF and PACF of residuals - ARIMA (1, 1, 3).

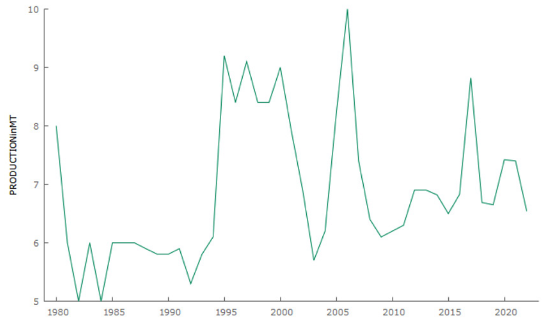

Fig. 13. Observed production of turmeric in Kerala.

Dickey-Fuller = -5.047, Lag order = 7, p-value = 1.627e-05

Correlogram and partial correlogram of Kerala

PACF and ACF of first differenced series by lag in production data has been illustrated in Fig. 14.

Forecasting model selection

After rendering the non-stationary series stationary through first-order differencing (d=1), the practice of selecting the optimal ARIMA model parameters involved a thorough examination of the autocorrelation (ACF) and partial autocorrelation (PACF) plots. Multiple combinations of autoregressive (p) and moving average (q) components were assessed. Ultimately, the ARIMA (1,1,1) model was elected, as the distribution and magnitude of significant spikes surpassing the confidence limits in the correlogram provided the strongest evidence for its appropriateness for the dataset in question. Coefficient of the model and AIC and BIC values of the model for turmeric production has been illustrated in Tables 13–14.

R software chose the forecasting model with the lowest BIC and AIC values out of the nine provided. Table 14 presents the findings of the ARIMA model.

The ARIMA (1,1,1), the model with the least AIC and BIC values (p =1, d =1, q =1) makes it the best model for estimating future time series values.

Selecting ARIMA model for forecasting

The forecasted values for turmeric production in

Table 13. Coefficient of turmeric production model.

ARIMA	Constant	AR1	MA1
(1, 1, 1)	0.0157	0.5964	-1.000
SE	0.0281	0.1200	0.0695

Table 14. AIC and BIC values of the ARIMA model applied to production data.

ARIMA model	σ^2 (Estimated)	Log-likelihood	AIC	BIC	AICc
(1, 1, 1)	0.9834	-59.779	127.558	130.98	126.32

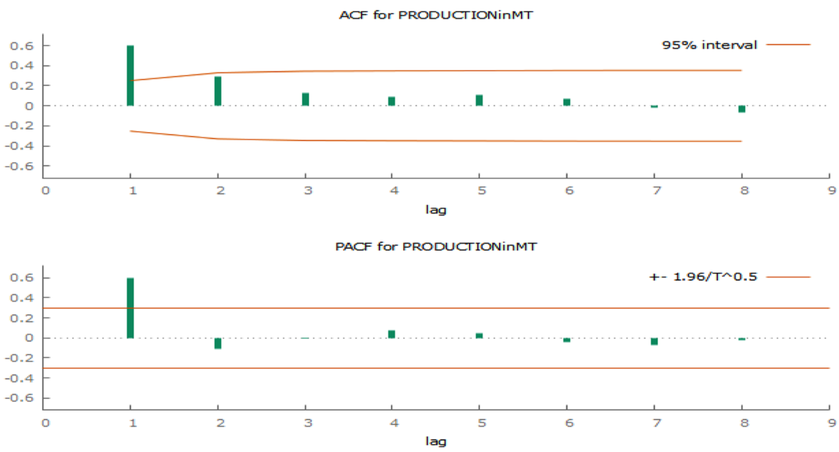

Fig. 14. PACF and ACF of first differenced series by lag in production data.

Table 15. Forecasted for turmeric production in Kerala.

Year	2023	2024	2025	2026	2027	2028	2029	2030
Forecasted production	6.84	7.02	7.14	7.22	7.28	7.32	7.35	7.37

Table 16. Validation of ARIMA (1,1,1) model.

ARIMA model	Accuracy of production model						
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
(1, 1, 1)	-0.031	0.975	0.699	-2.086	9.921	0.924	0.075

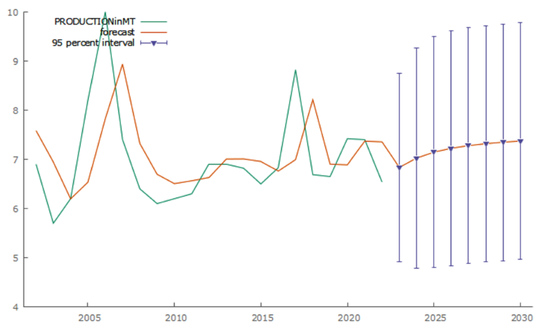

Fig. 15. Forecasts from ARIMA (1, 1, 1).

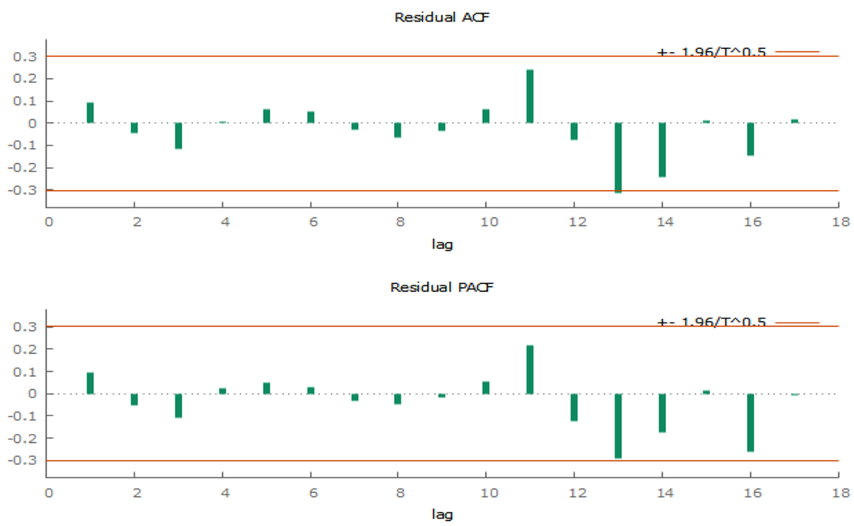

Fig. 16. Estimated ACF and PACF of residuals – ARIMA (1, 1, 1).

Kerala has been shown in Table 15.

The ARIMA (1, 1, 1) model is used to forecast the Turmeric Production during the next years, shown in Fig. 15.

For checking the accuracy of the fitted ARIMA model, MSE, RMSE, MAE, MPE, MASE, ACF, and MAPE are calculated which is obtainable in Table 16.

To determine the presence or absence of correlations between error values ACF and PACF plots are constructed for the same which is depicted in Fig. 16.

All ACFs and PACFs coefficients of residuals are significant for lags 1 to 17. For the ARIMA (1,1,1) model, ACF and PACF were unable to detect any autocorrelation in the prediction of residuals for lags 1 to 17.

CONCLUSION

The study provides a comprehensive examination of the principal determinants that have shaped the trajectory of turmeric cultivation in Southern States of India between 1980 and 2022. ARIMA models were applied to forecast turmeric production in Southern India using 43 years of data (1980-81 to 2022-23), with state-specific variations. For Andhra Pradesh ARIMA (2,1,3) demonstrated optimal performance, capturing complex autocorrelation patterns identified through correlogram analysis. For Tamil Nadu ARIMA (2,1,2) effectively modelled production trends, balancing autoregressive and moving average components. Karnataka ARIMA (1,1,3) addressed non-stationarity while maintaining predictive accuracy. Kerala. The simpler ARIMA (1,1,1) structure sufficed for its production dynamics.

The models were chosen following residual diagnostic criteria surpassing established confidence bounds, confirming their effectiveness in addressing temporal dependencies. Differences in parameter settings capture regional variations in production volatility and the complexity of underlying trends. State-specific ARIMA forecasts thus offer reliable guidance for optimizing supply chain management and informing export strategy development.

ACKNOWLEDGMENT

The authors are highly grateful to the editor and referees for their constructive comments/suggestions that help in the improvement of the revised version of the paper.

REFERENCES

Acharya, N. G., Ranga Agricultural University. (2022). Crop Outlook Reports of Andhra Pradesh. Guntur, ANGRAU, pp 7.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov B.N., Csaki F., (eds). Akademiai Kiado, Budapest, pp, 267—281. (Second international symposium on information theory).

Anusha, J., Lavanya, T., Vijaya Kumari, R., Radhika, P., & Supriya, K. (2023). Growth and instability of turmeric in Telangana. *International Journal of Statistics and Applied Mathematics*, 8 (5), 120—124.

Bhattacharyya, B. (2024). Autoregressive Integrated Moving Average (ARIMA) Model with genetic algorithm to forecast the chilli and turmeric productions in India. *Journal of Scientific Research and Reports*, 30 (6), 127—135.

Burnham, K. P., & Anderson, D. R. (2002). Advanced Issues and Deeper Insights. In: Burnham, K.P., Anderson, D. R. (eds). Model Selection and Multimodel Inference. Springer, New York, NY. https://doi.org/10.1007/978-0-387-22456-5_6

CEIC Data. (2023). https://www.ceicdata.com/en/india/production-of-horticulture-crops. Production: Horticulture Crops: Spices: Turmeric: Kerala – Economic Indicators. Downloaded on 23/10/2024.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76 (2), 297—307.

India. Ministry of Agriculture & Farmers Welfare, Agricultural Census Division. (2021). All India Report on Number and Area of Operational Holdings - T1 Report. New Delhi, Ministry of Agriculture and Farmers Welfare, pp, 120.

Indus Food. (2024). https://indusfood.co.in/article/golden-horizons-future-of-turmeric-in-global-trade Golden Horizons: Future of Turmeric in Global Trade. Downloaded on 25/10/2024.

Krishnan, K. B., Mehta, V., & Yadav, R. S. (2022). Assessment of future pattern of rainfall in different zones of Kerala using incorporation of SARIMA, ANN and Hybrid SARIMA-ANN models. *Economic Affairs*, 67 (5), 823—832.

Kumar, G. (2016). Forecasting production and area of potato cultivation in India using ARIMA model. *International Journal Social Science Development Policy*, 2 (2), 15—28.

Kumar, R. R., & Baishya, M. (2020). Forecasting of potato prices in India: An application of ARIMA model. *Economic Affairs*, 65 (4), 473—479.

Kumar, S., Mehta, V., Mourya, K. K., & Kumar, A. (2024). Comparative Study of Forecasting for Finger Millet (Ragi) According to their Area and Production in Different States of India through ARIMA Model. *Environment and Ecology*, 42 (4C), 1994—2003.

Naik, V. (2022). Value chain analysis of turmeric in Southern Karnataka. *Mysore Journal Agricultural Sciences*, 56 (1), 407—416.

Padhan, P. C. (2012). Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture and Social Sciences*, 8 (2), 50—56.

Pankratz, A. (2012). Forecasting with Dynamic Regression Models. New York, John Wiley & Sons. Available at: https://books.google.co.in/books/about/Forecasting_with_Dynamic_Regression_Mode.html?id=ZCmFI5U1j6cC&redir_esc=y.

Prabhu, R., Gowri, M. U., Gayathri, R., Govindaraj, M., & Manikandan, G. (2022). Growth dynamics and forecasting of minor millets in India: A time series analysis. *Journal of Applied and Natural Science*, 14, 145—150.

Rubasri, S. A., Ramanathan, G. (2024). Economics of production of turmeric in Erode district of Tamil Nadu. *Indian Journal Applied & Pure Biology*, 39 (1), 60—98.

Sangeetha, M., & Brindha, K. (2021). Present status and future prospects of turmeric production in Tamil Nadu. *Interna-*

tional Journal of Commerce and Management Research, 7 (6), 45—46.

Sathish Kumar, M., Lad, Y. A., & Mahera, A. B. (2022). Trend analysis of area, production and productivity of minor millets in India. *In Biological Forum–An International Journal*, 14 (2), 14—18.

Srinivas, B. A. (2021). Turmeric production in India. Thesis, MSc. Dr. Panjabrao Deshmukh Krishi Vidhyapeeth, Akola, pp, 123, In press.

Supriya, A. B., Raghav, Y. S., Devi, M., Kumari, P., Yadav, S., Mishra, P., Gautam, R., Gupta, B.K., Verma, S.K. & Bohra, D. (2023). Modeling and forecasting of lentil production in India and its instability. *Journal of Animal and Plant Sciences-JAPS*, 33 (4), 817—828.

Yadav, R. S., Mehta, V.,& Tiwari, A. (2022). An application of time series ARIMA forecasting model for predicting nutri-cereals area in India.*The Pharma Innovation Journal*, 11(3), 1260—1267.