Full Length Article

MODELLING AND FORECASTING IN DAIRY MILK PRODUCTION OF INDIA USING TIME SERIES APPROACHES

H.G. Koladiya[1], D.M. Mane[2]*, A.R. Ahlawat[3] and M.V. Darji[4]

Department of Animal Genetics and Breeding
College of Veterinary Science and Animal Husbandry
Kamdhenu University,
Anand, Gujarat - 388 001.

ABSTRACT

This study investigates the application of time series models for forecasting milk production in India. Using secondary data from 2001–02 to 2022–23, we evaluated four forecasting techniques, Exponential Smoothing (ES), Moving Average (MA), Vector Auto Regression (VAR), and Auto Regressive Integrated Moving Average (ARIMA). The ARIMA (1,1,0) model emerged as the best fit for short-term forecasting, while MA demonstrated better performance for long-term projections. The VAR model achieved an adjusted R^2 of 1.00, indicating a near-perfect fit, although its predictive capacity needs careful interpretation. Forecasting using ARIMA projected India's milk production to reach approximately 394 million tonnes by 2032–33, with a Compound Annual Growth Rate (CAGR) of 5.29%. These findings hold significant relevance for policymakers aiming to devise data-driven strategies to support sustainable dairy sector growth.

Keywords: *ARIMA, MA, VAR, Forecasting, Milk production*

Received :17.04.2025 Revised :10.09.2025 Accepted : 12.09.2025

INTRODUCTION

Milk production plays a crucial role in the Indian economy, particularly in enhancing rural livelihoods and ensuring nutritional security. As the world's largest milk producer, India contributed over 239.3 million tonnes in 2023–24 (NDDB, 2025), with a per capita milk availability of 471 grams per day in 2024 against the global average of 329 grams (Food Outlook, 2024). In this context, accurate forecasting of milk production is vital for informed planning, policy formulation, and sectoral investment (Chand, 2017).

As the dairy sector is a major contributor to the country's agricultural GDP, precise forecasting of milk output is essential to assess supply-demand dynamics and enable timely policy interventions (Mishra *et al.*, 2019). Against this backdrop, the present study aims to identify the most appropriate forecasting method for projecting milk production in India

[1]PhD Scholar,
[2]Post-Graduate Scholar, * Corresponding Author
E mail: digvijaym462@gmail.com,
[3]Professor & Head,
[4]Research Associate, Department of Veterinary Biotechnology

to support sustainable development and effective policy design.

MATERIALS AND METHODS

Data Source

Secondary data on annual milk production (2001–02 to 2022–23) was obtained from official government records (Table 1). The data was organized and analyzed to identify trends and fluctuations over time. This dataset has been previously utilized in similar modeling efforts in Indian agriculture (Kumar *et al*., 2012).

Methodology

Time series analysis has been extensively applied in agricultural economics to identify and project temporal trends. Among these, the Autoregressive Integrated Moving Average (ARIMA) model has been widely used due to its ability to model autocorrelated data (Box and Jenkins, 1976). Moving average and exponential smoothing techniques are effective for identifying underlying trends, while Vector Autoregression (VAR) models are suitable for examining interdependencies among multiple time series variables. Previous studies have employed ARIMA models to forecast milk production in India (Paul *et al.*, 2014; Deshmukh and Paramasivam, 2016; Mishra *et al.*, 2019), Pakistan (Ahmed *et al*., 2011), Ethiopia (Taye *et al*., 2020) and Brazil (Saude *et al.*, 2020). However, these studies primarily relied on linear modeling approaches, limiting their applicability to non-linear or complex time series data.

Four time series forecasting models were applied using python and spreadsheet-based statistical computations. The following models were evaluated:

ARIMA

ARIMA is commonly referred to as the Box–Jenkins (BJ) methodology (Deshmukh and Paramasivam, 2016). It was evaluated based on AIC and forecast accuracy (Gujarati and Porter, 2009). Time series when differentiated follows both AR and MA models and thus is known as autoregressive integrated moving average. In ARIMA (p, d, q) time series, p denotes the number of autoregressive terms (AR), d the number of times the series has to be differenced before it becomes stationary (I), and q the number of moving average terms (MA) (Chaudhari and Tingre, 2013).

The general form of ARIMA model of order (p, d, q) is,

$$Yt = Ø1Yt\text{-}1 + Ø2Yt\text{-}2 + \ldots + ØpYt\text{-}p + \mu - \Theta 1t\text{-}1 - \Theta 2t\text{-}2 - \ldots - \Theta qt\text{-}q + t$$

Where, Yt is milk production, t are independently and normally distributed with zero mean and constant variance for t = 1, 2,….., n; Ø p and Θq and are also estimated.

Moving Average (MA)

A technique for long-term trend analysis. This model computes future values based on average past observations, effectively smoothing short-term volatility. This method still considered as the best method by many people due to its easiness, objectiveness, reliability, and usefulness

(Hansun, 2013). The forecasting of milk production was calculated using following Moving average model,

Yt= μ - Ɵ1t-1 - Ɵ2t-2 -……. - Ɵnt-n + t

Exponential Smoothing

A method that assigns exponentially decreasing weights to past observations, offering a quick and responsive forecast for near-term projections. This technique is appropriate for a series that moves randomly above and below a constant mean (stationary series) (Ostertagova and Ostertag, 2012).

Vector Auto Regression (VAR)

A multivariate approach that models milk production simultaneously to account for potential interdependencies. This method is capable of producing forecasts of interrelated variables, examining the effects of interrelated time series variable's shocks, and analyzing the dynamic impact of random disturbances (Shahin *et al.*, 2014). The following model of the VAR method was used for determining the efficiency of predicting value,

Yt= A + B1Yt-1 + B2Yt-2 ……. BpYt-p + t

Where, Yt = (y_{t1}, y_{t2}, …, y_{tn})' is an (nx1) vector of time series variables, A is an (nx1) vector of intercepts, Bi (i=1, 2, …, p) is (nxn) coefficient matrices, Σ is an (nx1) zero mean error term (white noise).

All models were assessed using diagnostic checks such as residual analysis, AIC scores, and adjusted R^2 values. The final models were selected based on predictive robustness and ease of interpretability (Hyndman and Athanasopoulos, 2018).

Results and Discussion

The stationarity of the data assessed using the Autocorrelation function (ACF) and Partial autocorrelation function (PACF) plots. As shown in Fig. 1, the ACF and PACF values fall within the range of -0.5 to 0.5, staying within the confidence limits, which indicates that the data is stationary. Table 2 summarizes the comparative performance of the tested models, including their AIC, BIC, RMSE, and MAPE values. Among the evaluated models, ARIMA (1,1,0) emerged as the most suitable, exhibiting the lowest AIC (1.66), BIC (5.84), and RMSE (0.86), along with the second-lowest MAPE (0.39). These values indicate that ARIMA (1,1,0) provides a robust fit and can be considered an appropriate model for forecasting in the current context.

These findings are consistent with the earlier studies by Sankar and Prabakaran (2012) and Chaudhari and Tingre (2013), both of which identified ARIMA (1,1,0) as the most appropriate model for time series forecasting in their respective analyses. However, the present results diverge from those reported by Pal *et al.* (2007) and Deshmukh and Paramasivam, (2016), who found ARIMA (1,1,1) to be the most suitable model. Such variations across studies may be attributed to differences in dataset characteristics, time series structures, or model selection criteria employed.

The forecasting performance of three distinct time series models, Moving Average (MA), Exponential Smoothing (ES), and ARIMA, is illustrated in Figures 2, 3, and 4, respectively. Each model offers unique insights into the dynamics of milk production in India, with varying levels of forecasting accuracy and applicability.

The Moving Average model effectively captured the smoothed long-term trend in milk production, reflecting a steady upward trajectory across the observed years. This method is beneficial for identifying overall growth patterns by minimizing short-term fluctuations. However, the model's simplicity renders it less responsive to recent structural changes or shocks in the production system, limiting its sensitivity to real-time deviations or non-linear shifts. In contrast, the Exponential Smoothing model, although responsive to recent trends, demonstrated limited long-term predictive power. As shown in Figure 2, the forecast diverged from actual values in later years, particularly underestimating production growth. This suggests that ES may not be suitable for forecasting complex, non-linear agricultural outputs like milk production over extended horizons, where historical inertia and recent volatility interact.

Vector Autoregression method was used with the help of MS Excel 2019. The VAR model, achieved an adjusted R^2 of 1.00 with 0 Standard error, indicating an excellent in-sample fit. Similarly, Deshmukh and Paramasivam, (2016) reported an R^2 value of 0.99, suggesting a well-fitted model suitable for further analysis of milk production data. However, it may be prone to overfitting without external validation.

Forecast results indicate that India's milk production is projected to increase from 230.6 million tonnes in 2022–23 to approximately 394 million tonnes by 2032–33, representing a Compound Annual Growth Rate (CAGR) of 5.29%.

While the projections are promising, several constraints threaten the sustainability of this growth. These include climatic variability, feed and fodder shortages, animal health challenges, and market volatility. Future forecasting efforts may benefit from the integration of exogenous variables such as rainfall patterns, feed availability, input costs, and global price trends. Moreover, combining machine learning algorithms with classical statistical approaches could further enhance model adaptability and predictive accuracy under dynamic conditions. Consistent with earlier findings by Padhan (2012), time series models like ARIMA and VAR continue to offer practical and interpretable tools for agricultural forecasting. However, model performance must be contextualized within evolving production environments, and supplemented with ground-level data and interdisciplinary approaches.

CONCLUSION

The study assessed various time series models, ARIMA, Moving Average (MA), Exponential Smoothing (ES), and Vector Autoregression (VAR), to forecast milk production in India. Among the ARIMA models tested, ARIMA (1,1,0) outperformed others based on the lowest AIC, BIC, and RMSE values, indicating its suitability for short to medium-term forecasting. The MA model effectively highlighted long-term trends but lacked sensitivity to recent fluctuations, while ES was more responsive but less accurate for long-term projections. The VAR model exhibited an excellent in-sample fit (adjusted $R^2 = 1.00$). Forecasts suggest that India's milk production will rise from 230.6 million tonnes in 2022–23 to approximately 394 million tonnes by 2032–33, with a CAGR of 5.29%. To enhance future predictions, integrating exogenous factors and hybridizing classical models with machine learning approaches is recommended, especially under dynamic and complex production environments.

Table.1. Annual milk production (in Tonnes)

Year	All India
2001	84406
2005	97066
2010	116425
2015	146313
2020	209960
2022	230582

Table.2. Model comparison of ARIMA configurations based on AIC, BIC, RMSE, MAPE and adjusted R2

ARIMA (p,q,d)	AIC	BIC	RMSE	MAPE
ARIMA (1,0,0)	23.67	25.76	1.68	0.00
ARIMA (0,1,0)	152.46	155.73	27.90	3.81
ARIMA (0,1,1)	34.38	37.52	1.97	0.28
ARIMA (1,1,0)	1.66	5.84	0.86	0.39
ARIMA (1,1,1)	18.66	22.64	67.58	8.89
ARIMA (1,0,1)	16.24	18.23	1.43	0.00

In ARIMA (p,q,d): p=autoregressive terms; d=differencing order q=moving average terms.

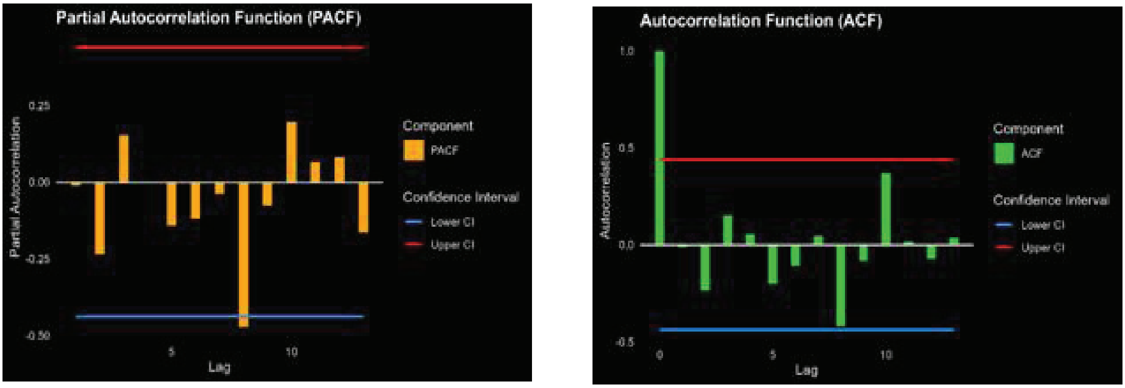

Fig.1. ACF and PACF value

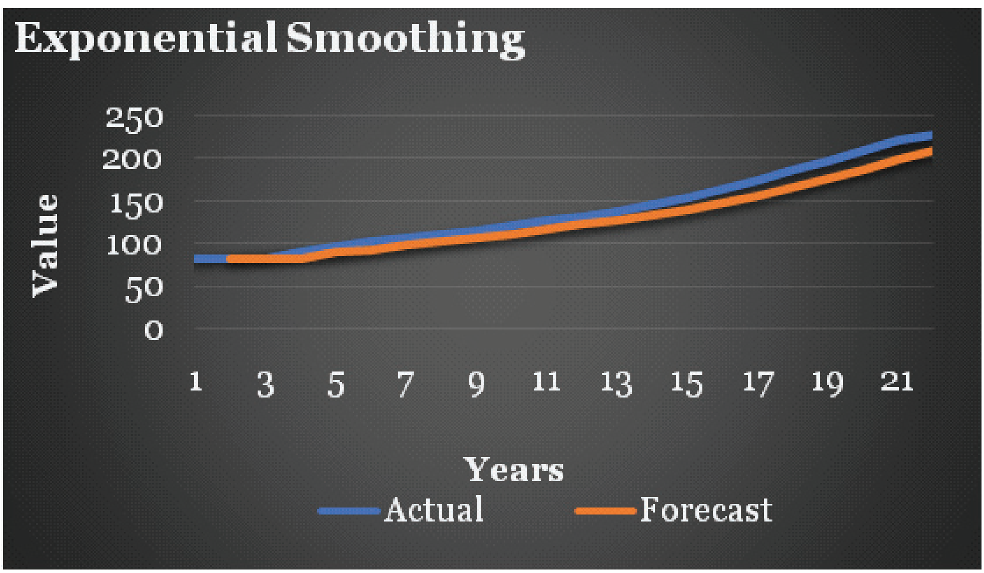

Fig.2. Milk production forecast by MA

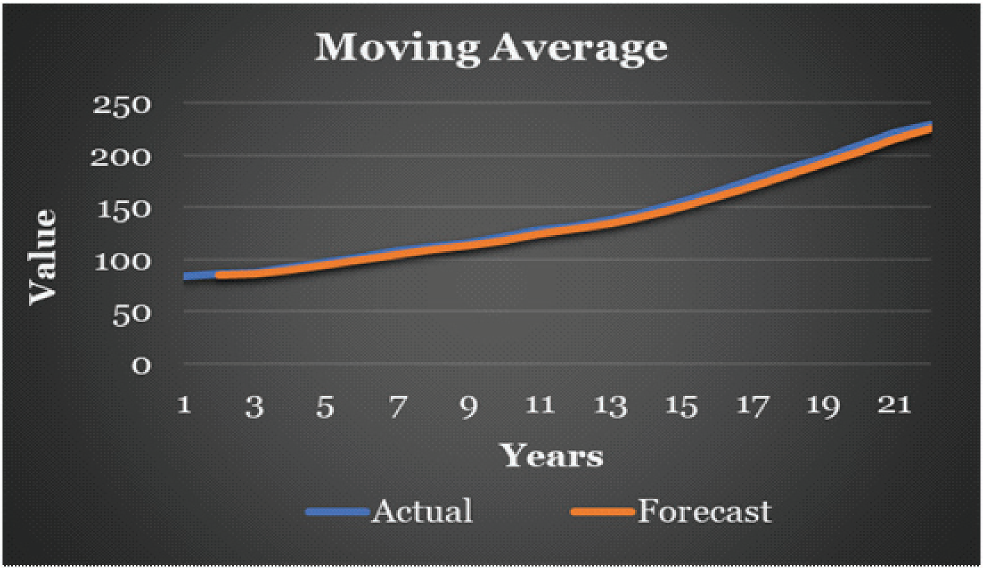

Fig.3. Milk production forecast by ES

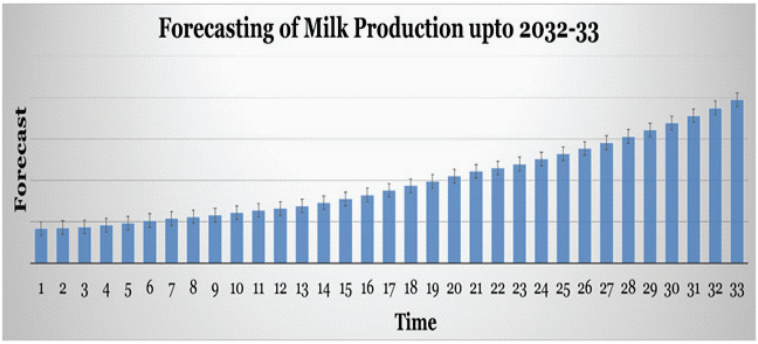

Fig.4. Milk production forecast by ARIMA

REFERENCES

Ahmed, F., Shah, H., Raza, I. and Saboor, A. (2011). Forecasting milk production in Pakistan. *Pakistan Journal of Agricultural Research*, **24**(1-4): 82-85.

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

Chand, R. (2017). Doubling farmers' income: Rationale, strategy, prospects and action plan. NITI Policy Paper.

Chaudhari D.J. and Tingre A.S. (2013). Forecasting of milk production in India: An application of ARIMA model. *Indian Journal of Dairy Science*, **66**: 72-78.

Deshmukh, S.S. and Paramasivam, R. (2016). Forecasting of milk production in India with ARIMA and VAR time series models. *Asian Journal of Dairy and Food Research*, **35**(1), 17-22.

Food Outlook. (2014). Food outlook - global market analysis food and agriculture organization of the United Nations **8**: 51–54. Press Release:Press Information Bureau

Gujarati, D.N. and Porter, D.C. (2009). Basic Econometrics (5th ed.). McGraw-Hill.

Hansun, S. (2013). A new approach of moving average method in time series analysis. In 2013 *conference on new media studies* (CoNMedia) (pp. 1-4). IEEE.

Hyndman, R.J. and Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

Kumar, A., Staal, S. and Singh, D.K. (2012). Small holder dairy farmers' access to modern milk marketing chains in India. *Agricultural Economics Research Review*, **25**: 243–253.

Mishra, P., Fatih, C., Niranjan, H.K., Tiwari, S., Devi, M. and Dubey, A. (2019). Modelling and forecasting of milk production in Chhattisgarh and India. *Indian Journal of Animal Research*, **54**: 912-917. DOI: 10.18805/ijar.B-3918.

NDDB, National Dairy Development Board, 2025, Milk Production in India | nddb.coop

Ostertagova, E. and Ostertag, O. (2012). Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica,* **12**(3): 62.

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

Pal S., Ramasubramanian V. and Mehata S.C. (2007). Statistical models for forecasting milk production in India. *Journal of the Indian Society of Agricultural Statistics*, **61**: 80–83.

Paul, R.K., Alam, W. and Paul, A.K. (2014). Prospects of livestock and dairy production in India under time series framework. *Indian Journal of Animal Sciences*, **84**(4): 462-466.

Sankar T.J. and Prabakaran R. (2012). Forecasting milk production in Tamil Nadu. *International Multidisciplinary Research Journal*, **2**: 10–15.

Saude, L.M.S., Gabriel, G.T. and Balestrassi, P.P. (2020). Forecasting of buffalo milk in a Brazilian diary using the ARIMA model. *Buffalo Bulletin,* **39**(2): 201-213.

Shahin, M.A., Ali, M.A. and Ali, A.S., (2014). Vector autoregression (VAR) modeling and forecasting of temperature, humidity, and cloud coverage. *Computational Intelligence Techniques in Earth and Environmental Sciences*, pp.29-51.

Taye, B.A., Alene, A.A., Nega, A.K. and Yirsaw, B.G. (2020). Time series analysis of cow milk production at Andassa dairy farm, West Gojam zone, Amhara region, Ethiopia. *Modeling Earth Systems and Environment*, **7**: 181-189. DOI: 10.1007/s40808-020 00946-z.