DR-2103
[1-6]

RESEARCH ARTICLE Asian Journal of Dairy and Food Research

Modeling and Forecasting of Milk Production in the Western Zone of Tamil Nadu

S. Vishnu Shankar[1], R. Ajaykumar[2], S. Ananthakrishnan[3],
A. Aravinthkumar[4], K. Harishankar[5], T. Sakthiselvi[6], C. Navinkumar[7]

10.18805/ajdfr.DR-2103

ABSTRACT

Background: In India, the dairy business is expanding dramatically. Tamil Nadu milk cooperatives significantly contribute to the growth of the dairy sector in the state. In terms of delivering economic income for dairy smallholders and satisfying customer demand, the identification of milk production is one of the primary financial operations made in India. Considering this, it is crucial to understand future production to enhance and sustain the sector's growth and development.

Methods: The present investigation attempts to predict and forecast milk production in Tamil Nadu using time series models. Yearly milk data from 1976 to 2020 was taken. The study considered Auto-Regressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) to select the appropriate stochastic model for forecasting milk production in Tamil Nadu. Further statistical modeling procedures employed for milk production reveal that the selection of a suitable time series model will always depend on the nature of the data.

Result: Results revealed that the ARIMA model is selected as the best model despite ANN, even if it is considered the most powerful model. The CAGR for forecasted milk production from 2020-2025 was 0.02%. Model adequacy criteria like RMSE, MAPE and MAE are used. Based on observation ARIMA model (1, 1, 2) is chosen as the best model over the ANN model.

Key words: ANN, ARIMA, Milk Production, Time series models.

INTRODUCTION

Milk is closely linked to our civilization, as the domestication of dairy animals started thousands of years ago. It is now well connected to several sectors like food, health, pharmaceuticals, cosmetics, *etc*. In most countries, milk ranks among the top five agricultural commodities in terms of monetary returns. Besides their dietary purpose, it plays a vital role in the growth of the Indian economy employment generation, nutritional security and protection of native breeds. Among agricultural products, milk and other dairy products account for 14% of global trade (Mottet *et al*., 2018). The estimated global milk production was about 906 million tonnes (MT) in 2020 (an increase of 2.0% from 2019). At the beginning of India's independence in 1950, milk production was around 17 MT annually. However, the launch of Operation Flood in 1970 resulted in a rise in annual production of 23.2 MT in 1973-74 to 209.96 MT in 2020-21, an eight-fold growth in less than five decades (Khera *et al*., 2022). Now, India is the world's leading milk producer, accounting for 23% globally and 52% of Asia's total output. As the single-largest agricultural commodity in India, the dairy industry contributes around 26 % to the entire agriculture GDP and more than 5% to the national economy (Chatellier, 2021). The per capita availability of milk in India has also increased to about 427 grams per day (2021) against the ICMR recommendation of 300 grams per day. This shows the sustained growth in milk availability for the burgeoning subcontinent's population.

The United Nations reported that the global population is expected to reach 9 billion and India will have a population of 1.66 billion by 2050. India should produce around 401.4 million metric tonnes of milk to meet the demand in 2050 (Aatralarasi *et al*., 2021). The emerging population growth

[1]Department of Basic Sciences, Dr. Y.S. Parmar University of Horticulture and Forestry, Solan-173 230, Himachal Pradesh, India.
[2]Department of Agronomy, Vanavarayar Institute of Agriculture, Pollachi-642 103, Tamil Nadu, India.
[3]Department of Soil Science and Water Management, Dr. Y.S. Parmar University of Horticulture and Forestry, Solan-173 230, Himachal Pradesh, India.
[4]Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi-110 012, India.
[5]Department of Agricultural Economics, S. Thangapazham Agricultural College, Tenkasi-627 758, Tamil Nadu, India.
[6]Department of Soil Science, Kerala Agricultural University, Vellayani-695 522, Kerala, India.
[7]Department of Agricultural Meteorology, Vanavarayar Institute of Agriculture, Pollachi- 642 103, Tamil Nadu, India.

Corresponding Author: S. Vishnu Shankar, Department of Basic Sciences, Dr. Y.S. Parmar University of Horticulture and Forestry, Solan-173 230, Himachal Pradesh, India.
Email: s.vishnushankar55@gmail.com

How to cite this article: Shankar, S.V., Ajaykumar, R., Ananthakrishnan, S., Aravinthkumar, A., Harishankar, K., Sakthiselvi, T. and Navinkumar, C. (2023). Modeling and Forecasting of Milk Production in the Western Zone of Tamil Nadu. Asian Journal of Dairy and Food Research. doi:10.18805/ajdfr.DR-2103.

Submitted: 13-04-2023 **Accepted:** 07-06-2023 **Online:** 11-07-2023

is expected to elevate the need for all agricultural products, especially animal protein. The increase in the production of resources should majorly come from intensification rather than expansion. The increase in milk production has to come from productivity improvements. Still, the number of interactive factors related to breeding, feeding, health and management continue to yield less milk. At least in the short run, there is a scope to raise milk yield with the existing milch stock and its quality (Patel *et al*., 2022). However, all these studies were confined to a particular zone or region, state or production system, which widely lacks policy implications. To control market demands, stakeholders and authorities had to formulate some effective short- and long-term plans. These plans aid in managing milk supply using precise estimations for milk production derived from well-accepted prediction methods. Hence there is a need to forecast the milk production to foresight if any demand would create in the future.

Forecasting is an important aspect in a developing economy like India. It aids in decision-making and planning for sustainable growth, poverty alleviation and overall development (Zhang, 2003). Different forecasting techniques are performed using statistical analysis. These statistical predicting models are used to develop suitable time series models by identifying the trends and patterns of past data. This research objects to model and forecast the milk production of the western zone of Tamil Nadu using ARIMA and ANN techniques.

MATERIALS AND METHODS

Data collection

The study aims to forecast the milk production of the western zone of Tamil Nadu, India. Yearly data on milk production from 1976 to 2020 (National Dairy Development Board, Anand, Gujarat) was used for the forecasting purpose. To fit the model, the data was split into training and testing data *i.e*., data from 1976 -2015 was taken for model building and data from 2016-2020 was accepted for model validation purposes. R software was used to fit and forecast the study data.

Compound growth rate (CGR)

Exponential function was to analyse the compound growth rate, which was initially transformed into a linear form by taking the log function (Kalidas *et al*., 2020). The final equation for estimating the growth rate was given by,

$$\text{CGR } (r) = [\text{Antilog } (\log b) - 1] * 100$$

Where,
r = Compound growth rate.
b = Regression coefficient.

Time series models

Forecasting is a technique where future observations are predicted using past observations. It is made possible using time series models, extending its application to all fields (Hamid *et al*., 2016). Generally, no time series models are assumed to be the best model for data, as the data don't follow similar patterns in all cases. So, the suitable method for data is determined based on its nature. Among several time series models, some were tried fitting for the data. The best-fitted model will be used for forecasting future milk production (Demir and Kirisci, 2022). Different time series models employed in the study are detailed below.

Auto-regressive integrated moving average model

Auto-regressive integrated moving average (ARIMA) is a statistical time series model used to estimate and forecast the time series data (Punyapornwithaya *et al*., 2022). It is a linear model used for handling univariate data. The model comprises three major parts. The first part of the model is auto-regressive (p), the second part is integration (d) and the third part is moving average (q). The ARIMA model can be expressed as follows,

$$\varphi (B)(1-B)^d y_t = \theta(B)[\varepsilon_t \;\; \varphi$$

Autoregressive operator: $\varphi (B) = 1 - \varphi_1 (B) - \varphi_2 (B)^2 -...- \varphi_p (B)^p$
Moving average operator: $\theta(B) = 1 - \theta_1(B) - \theta_2(B)^2 -... - \theta_p(B)^q$

Where,
ε_t = White noise or error term.
d = Differencing term.
B = Backshift operator, *i.e*.
$B^aY_t = Y_{t-a}$

Being a linear model, it can be applied only to the data unless it is stationary. If the data is non-stationary, it should be converted into stationarity by differencing it to the respective order (Deshmukh and Paramasivam, 2016). The major steps involved in building the ARIMA model.

Identification

The values of *q* and *p* can be found using the plots of ACF (Auto Correlation Function) plot and PACF (Partial Auto Correlation Function) plot. The p and q denote the model information from past values and errors. The *d* in Integration indicates the number of times the data should integrate to convert into stationary.

Parameter estimation

The parameters of ARIMA models are estimated using Maximum Likelihood Estimation (MLE) method by which the values of AR and MA values are found.

Diagnostics

The model which gives the low AIC and BIC values is chosen to be the best fitted model for the data (Naveena and Subedar, 2017). Finally, the data is forecasted if the residual of fitted model is uncorrelated, *i.e.*, white noise.

Artificial neural network model

An artificial Neural Network (ANN) is a computational model inspired by the structure and functionality of biological neural networks. They are the network of interconnected neurons mimicking the function of the human brain. A feed-forward neural network is one of the basic neural networks and serves as a non-linear time series model for forecasting purposes (Rathod *et al*., 2017). They are made of input, hidden and output nodes. Every unit in a layer is related to every unit in the previous layer. The data is given through the input node and the result is obtained from the output node. The in-between hidden layer is the place where processing is done. Each layer consists of weights and biases. The numbers of input and hidden nodes are determined by experimentation, as there is no theoretical base for finding these parameters (Mishra *et al*., 2021). The mathematical representation of the ANN model is given as:

$$y_t = Ø_0 + \Sigma^{q}_{j=1} Øjq + [\theta_{oj} + \Sigma^{p}_{t=1} \theta_{ij}\ y_{t-1}] + \varepsilon_t$$

Where,

j (j = 0,1,2,...,q) , ij(i = 0,1,2,..., p and j = 1,2,...,q) = Weights.

0_j = Bias terms.

ε_t = White noise.

Estimation of model performance

Model validation helps estimate the performance of different models by comparing the actual and fitted values. The various error measures used for model selection are Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The model with a low error rate is the optimum model for the study data.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum^{n}_{t=1}\left|Yt - \hat{Y}t\right|^2}$$

$$\text{MAE} = \frac{1}{n}\sum^{n}_{t=1}\left|Yt - \hat{Y}t\right|$$

$$\text{MAPE} = \frac{1}{n}\sum^{n}_{t=1}\frac{\left|Yt - \hat{Y}t\right|}{Yt}$$

Y_t = Denotes the actual values.

v_t = Represents the predicted values.

n = Indicates the number of observations.

RESULTS AND DISCUSSION

Descriptive statistics is calculated for the study, which is given in Table 1. The maximum milk production was recorded in 2020 (833.12 MT), whereas the minimum milk production was recorded in 1982 (165.20 MT). The mean milk production for the last 45 years is 423.49 MT, with a standard deviation of 175.30. The kurtosis is -0.04 with positively skewed. The coefficient of variation is 41.39%.

The compound growth rate was calculated to inspect the growth rate percentage in every prefixed period (Fig 1). This study estimated the growth rate for every five years of milk production using an exponential function (Anjum, 2018). The highest growth rate was recorded during 1991-95 (6.31%), followed by 2011-15 (6.20%), whereas the lowest growth rate was recorded during 2001-05. During 1976-80 a negative growth rate (-1.83%) was observed and an overall growth rate of 2.71% was registered over the last 45 years.

This study used two different time series models for model building and model validation. The model that fits better for the data with a low error rate was used to predict future milk production. Both the models were fitted to the data initially split into training and testing datasets. The Autoregressive Integrated Moving Average (ARIMA) is a linear model, fitted only when the datasets are stationary. If the datasets are not stationary, they are converted into stationary using differencing (Kour *et al*., 2017). Augmented Dickey-Fuller (ADF) test and Philips-perron test were two autocorrelation tests used for testing stationarity in the data. Table 2 revealed that the p-value of the two tests was greater than 0.05, *i.e*., they are not stationary. The graph (Fig 2.) shows that the spikes came out of significant lines in both ACF and PACF plots. So, the data was differenced one time to convert them into stationarity.

Table 1: Descriptive statistics of milk production in western zone of Tamil Nadu.

Statistics	Milk production
Observations	45.00
Mean	423.49
Median	371.99
Standard deviation	175.30
Kurtosis	-0.04
Skewness	1.01
Range	667.92
Minimum	165.20
Maximum	833.12
Coefficient of variation (%)	41.39

Table 2: Autocorrelation test for stationarity.

Milk	Augmented dickey-fuller test		Phillips-perron test	
production	t-Statistic	Probability	t-Statistic	Probability
	-0.472	0.978	-2.517	0.951
Remarks	Non-stationary		Non-stationary	
No. of differencing required		1		

Table 3: Parameter estimation of ARIMA model using MLE method for milk production in western zone of Tamil Nadu.

Model	AR1	MA1	MA2	Log-likelihood	AIC	BIC	Box pierce test	Shapiro wilk test
ARIMA (1,1,2)	-0.16 (1.10)	0.12 (0.07)	0.08 (0.21)	-204.09	416.18	422.84	0.09	0.96

Table 3 revealed that the ARIMA model (1, 1, 2) was fitted to the training datasets with the lowest AIC and BIC values, *i.e.*, 416.18 and 422.84, respectively. The test statistics of AR1 was -0.16 with a standard error 1.10, MA1 was 0.12 with a standard error of 0.07 and MA2 was 0.08 with a standard error of 0.21. In addition, the box-pierce test and Shapiro-wilk test were conducted to test the autocorrelation and normality of the fitted model (Pardhi *et al*., 2018). The p-value of both tests was greater than 0.05, meaning that the residuals are white noise and normal.

Feed-forward neural network (FFNN) is a type of ANN used to fit the data in this study (Table 4). The initial step is to find the network of the ANN model. Different combination input nodes (1:10) and hidden nodes (1:10) were tried for the training dataset and the appropriate node was found after several iterations (Tealab *et al*., 2017). The output node is always one. Finally, the network 6-2-1 was found to be the best fit for the data with weights 17. Linear sigmoidal activation function was used in model building. The Box-pierce test and Shapiro-wilk test show that the fitted ANN model's residuals are not autocorrelation and are normal.

Since both the fitted models are checked for white noise and normality tests at the model-building phase, they can be taken for model validation. Table 5 indicates the predictive performance of ARIMA and ANN models. Error measures like RMSE, MAPE and MAE were used for comparing the model (Sankar and Prabakaran, 2012). The ARIMA model

Table 4: Parameter estimation of artificial neural network (ANN) model.

Particulars	Parameters
Network	6-2-1
Optimum lag	6
Optimum hidden node	2
Cross validation	25
Network type	Feed Forward network
Activation function	Linear Sigmoidal
Weights	17
Box pierce test	0.19 (0.67)
Shapiro wilk test	0.98 (0.91)

Table 5: Predictive performance of different models (ARIMA and ANN Model).

Measures	ARIMA		ANN	
	Training	Testing	Training	Testing
RMSE	44.75	49.06	45.01	51.47
MAPE	5.11	5.36	5.39	5.60
MAE	34.02	41.57	34.90	43.88

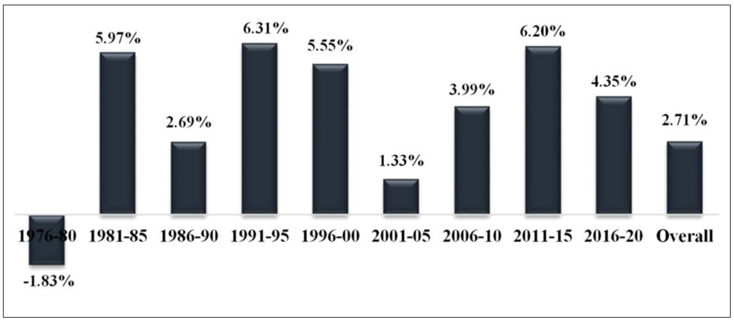

Fig 1: Compound growth rate of milk production in western zone of Tamil Nadu.

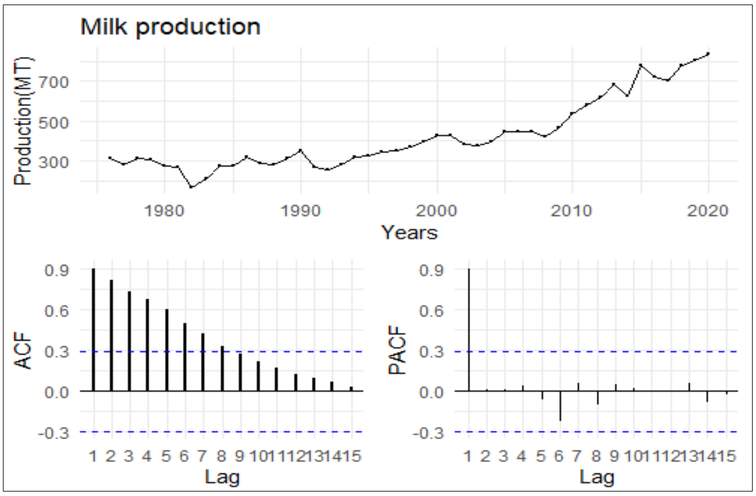

Fig 2: Time series plot of milk production along with ACF and PACF plots.

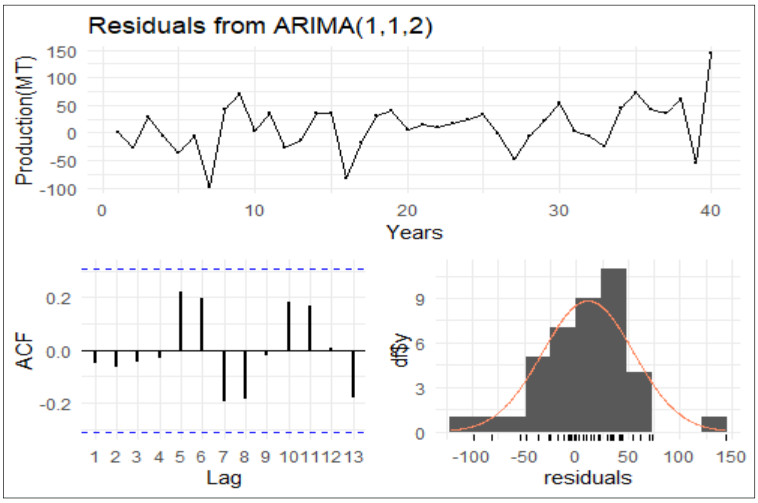

Fig 3: Residual plot, corresponding ACF plot and histogram from ARIMA model.

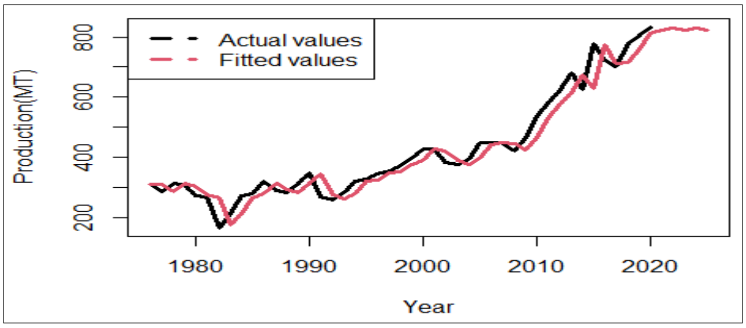

Fig 4: Plot showing actual values and fitted values.

Table 6: Forecasting of milk production using the ARIMA model.

Year	Milk production (MT)
2021	821.12
2022	829.92
2023	821.70
2024	829.38
2025	822.20
CAGR	0.02%

was found to have a low error in training and testing, followed by ANN with minor differences. Even though the error rate is calculated for both training and testing datasets, the testing datasets, *i.e.*, model evaluation, are deciding factor for the model section (Aslanargun *et al*., 2007). Based on that criterion, the ARIMA model's error rate is comparatively lower than ANN in all aspects. Therefore, the ARIMA model is selected as the best fitted model for study and used for forecasting purposes.

Since ARIMA models fit well for the study data with a low error rate, satisfying all necessary conditions of a time series model (Fig 3), it can be used for forecasting the future milk production (MT) (Yonar *et al*., 2022). Table 6 indicates that the forecasted values of milk production from 2021 to 2025. The CGR for the forecasted milk production is 0.02%. Fig 4 gives the graphs showing actual and fitted values, along with predicted values obtained using the ARIMA model.

CONCLUSION

The Indian dairy sector covers a sizeable portion of the world's dairy resources. The livestock industry supports the national economy and the nation's socio-economic development. The overall growth rate for milk production obtained by CAGR for 45 years is 2.10%. Further statistical modelling procedures employed for milk production reveal that the selection of a suitable time series model will always depend on the nature of the data. Likely ARIMA model is selected as the best model in spite of ANN even if it is considered the powerful model. The CAGR for forecasted milk production from 2020-2025 was 0.02%. When considering the approximate projected population growth from various studies, the forecasted growth rate could be insufficient to meet our demand. Thus, we need more and more research in the dairy sector to increase milk production.

Conflict of interest: None.

REFERENCES

Aatralarasi, S., Dhaliwal1 L.K., Kingra P.K. and Jain G. (2021). Prediction of future milk production trend in India and Central Punjab. Journal of Animal Research. 11(6): 1051-1058. DOI: 10.30954/2277-940X.06.2021.15.

Anjum, S. (2018). Growth and instability analysis in Indian agriculture. International Journal of Multidisciplinary Research and Development. 5(11): 119-125.

Aslanargun, A., Mammadov, M., Yazici, B. and Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: Tourist arrival forecasting. Journal of Statistical Computation and Simulation. 77(1): 29-53.

Chatellier, V. (2021). International trade in animal products and the place of the European Union: Main trends over the last 20 years. Animal. 15: 100289. https://doi.org/10.1016/j.animal.2021.100289.

Demir, I. and Kirisci, M. (2022). Forecasting COVID-19 disease cases using the SARIMA-NNAR hybrid model. Universal Journal of Mathematics and Applications. 5(1): 15-23.

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. DOI: 10.18805/ajdfr.v35i1.9246.

Hamid, M.A., Ahmed, S., Rahman, M.A. and Hossain, K.M. (2016). Status of buffalo production in Bangladesh compared to SAARC countries. Asian Journal of Animal Sciences. 10(6): 313-329. DOI: 10.3923/ajas.2016.313.329.

Kalidas, K., Mahendran, K. and Akila, K. (2020). Growth, instability and decomposition analysis of coconut in India and Tamil Nadu, Western Tamil Nadu, India: A time series comparative approach. J. Econ. Manag. Trade. 26: 59-66.

Khera, P.K., Hussain, J., Bordoloi, J.P., Saharia, J.T., Gohain, A.K. and Borpuzari, L. (2022). Effect of dietary supplementation of Shatavari (*Asparagus racemosus*) on the production performance of crossbred cows. The Pharma Innovation Journal. 11(11): 491-495.

Kour, S., Pradhan, U.K., Paul, R.K. and Vaishnav, P.R. (2017). Forecasting of pearl millet productivity in Gujarat under time series framework. Economic Affairs. 62(1): 121-127.

Mishra, P., Matuka, A., Abotaleb, M.S.A., Weerasinghe, W.P.M.C.N., Karakaya, K. and Das, S. S. (2021). Modeling and forecasting of milk production in the SAARC countries and China. Modeling Earth Systems and Environment. 8: 947-959.

Mottet, A., Teillard, F., Boettcher, P., De'Besi, G. and Besbes, B. (2018). Domestic herbivores and food security: Current contribution, trends and challenges for a sustainable development. Animal. 12(2): 188-198.

Naveena, K. and Subedar, S. (2017). Hybrid time series modelling for forecasting the price of washed coffee (Arabica Plantation Coffee) in India. International Journal of Agriculture Sciences. 9(10): 4004-4007.

Pardhi, R., Singh, R. and Paul, R.K. (2018). Price forecasting of mango in Lucknow market of Uttar Pradesh. International Journal of Agriculture, Environment and Biotechnology. 11(2): 357-363.

Patel, H.A., Odedra, M.D., Ahlawat, A.R., Gamit, V.V., Prajapati, V.S. and P.H.A. (2022). A review on milking management practices of dairy animal in India. The Pharma Innovation Journal. 11(10): 109-112

Punyapornwithaya, V., Mishra, P., Sansamur, C., Pfeiffer, D., Arjkumpa, O., Prakotcheo, R. and Jampachaisri, K. (2022). Time-series analysis for the number of foot and mouth disease outbreak episodes in cattle farms in Thailand using data from 2010-2020. Viruses. 14(7): 1367. https://doi.org/10.3390/v14071367.

Rathod, S., Mishra, G.C. and Singh, K.N. (2017). Hybrid time series models for forecasting banana production in Karnataka State, India. Journal of the Indian Society of Agricultural Statistics. 71(3): 193-200.

Sankar, T.J. and Prabakaran, R. (2012). Forecasting milk production in Tamil Nadu. International Multidisciplinary Research Journal. 2(1): 10-15.

Tealab, A., Hefny, H. and Badr, A. (2017). Forecasting of nonlinear time series using ANN. Future Computing and Informatics Journal. 2(1): 39-47.

Yonar, H., Yonar, A.Y.N.U.R., Mishra, P., Abotaleb, M., Al Khatib, A.M.G., Makarovskikh, T.A.T.I.A.N.A. and Cam, M.U.S.T.A.F.A. (2022). Modeling and forecasting of milk production in different breeds in Turkey. Indian J. Anim. Sci. 92(1): 105-111.

Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 50(1): 159-175.