Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 11, 2025

FORECASTING WHEAT YIELD IN PAKISTAN (2023-2030), WITH STATISTICAL AND MACHINE LEARNING MODELS

Muhammad Waseem[*1], Darshan Jee[2], Muhammad Zikriya[3], Muhammad Hasnain Qasim[4], Asif Nawaz[5]

*[*1,3,5]Department of Statistics Abdul wali khan University Mardan, Kpk, Pakistan*
[2]Department of Statistics the Islamia University of Bahawalpur.
[4]Department of Statistics and Data Science, University of Mianwali

[*1]waseemstat111@gmail.com, [2]djee864@gmail.com, [3]zikriyadurrani666@gmail.com, [4]hasnainqasim1234@gmail.com, [5]asifnawaz10010@gmail.com

DOI: https://doi.org/10.5281/zenodo.17731033

Keywords
Wheat production, Forecasting, Pakistan, Time series, ARIMA, Exponential Smoothing, TBATS, ANN

Article History
Received: 07 October 2025
Accepted: 15 November 2025
Published: 27 November 2025

Corresponding Author: * Muhammad Waseem

Abstract
Wheat is one of the most important crops in Pakistan, it plays a major role in feeding the population and supporting the economy. This study focusses on predicting wheat production in Pakistan from 2023 to 2030 using four models: Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing(ES), Trigonometric Box-Cox ARIMA Trend Seasonal(TBATS) and Artificial Neural Network(ANN). We used wheat production data from 1960 to 2023 and analyzed it using the R programming language. The performance of each model was tested using common error measures like RMSE, MAE and MAPE. Among all models TBATS give most accurate results by handling seasonal and trends patterns more effectively. The results show wheat production in Pakistan is likely to increase over next seven years. This forecast can help government and farmers plan better for future food needs and agriculture development.

INTRODUCTION

Wheat is one of the earliest domesticated crops and has remained a fundamental component of human diets for over 8000 years, particularly in Europe, West Asia, and North Africa [1]. Today, it stands among the most cultivated cereal crops globally and serves as a major source of carbohydrates and protein for billions of people. In 2023, global wheat production reached approximately 799 million tonnes, with China leading at 136.6 million tonnes (17.1%), followed by India with 110.6 million tonnes (13.8%), while Pakistan produced 28.2 million tonnes, positioning it among the top ten producers worldwide[2]. Despite this, wheat productivity in Pakistan remains below potential levels due to water scarcity, outdated farming practices, and climatic challenges. In Pakistan, wheat is the most important food crop, accounting for nearly 60% of daily caloric intake and contributing 10.1% to agricultural value and 2.2% to GDP. The average per capita wheat consumption is approximately 125 kg per year, making it a staple food source for most of the population. The country has undergone three major agricultural phases: the traditional pre-1965 period with low yields, the "Green Revolution" (1966–1976) characterized by improved seed and fertilizer use, and the post-1977 era marked by high-yield and disease-resistant varieties. Despite these advancements, yield disparities persist, with some farmers achieving up to 7 tonnes per hectare, while the national average remains around 2.8 tonnes per hectare[2].Pakistan's wheat sector continues to face critical challenges, including water scarcity, climate change, land fragmentation, and inefficient irrigation systems. Although the country possesses one of the world's largest canal networks, mismanagement and outdated infrastructure have significantly reduced water-use efficiency. Moreover, dependence on imported fertilizers, limited access to modern farming technologies, and inadequate agricultural training have constrained productivity. These challenges are compounded by social issues such as land ownership inequality and gender disparities in rural farming communities. To address these concerns, it is essential to adopt data-driven forecasting techniques that can help policymakers plan for sustainable wheat production and ensure food security. Accurate forecasting of wheat production is vital for planning agricultural policies, ensuring food security, and stabilizing markets. Time series forecasting models such as ARIMA (Auto-Regressive Integrated Moving Average), ETS (Error, Trend, and Seasonality), and TBATS (Trigonometric, Box-Cox, ARIMA, Trend, and Seasonal) have been extensively applied in agricultural forecasting due to their ability to capture temporal patterns. In recent years, Artificial Neural Networks (ANN) and hybrid models combining statistical and machine learning techniques have gained prominence for modeling complex nonlinear relationships in crop yield data. The main objective of this study are to select the best model for forecasting among the ARIMA, ETS, TBATS and ANN models to forecast the wheat production of Pakistan for the next 7 years by best model.

Several studies have contributed to the advancement of wheat forecasting models. Singh et al. (2016) applied ARIMA models to forecast wheat area, production, and productivity in Gujarat, India, identifying ARIMA(0,1,1) and ARIMA(1,1,0) as optimal models based on AIC and residual diagnostics, emphasizing the importance of regular model updates for policy decisions[3]. Similarly, Saeed et al. (2000) used the Box–Jenkins ARIMA approach to predict Pakistan's wheat production trajectory from 1947 to 2013, selecting ARIMA(2,2,1) as the best-fitting model and highlighting the model's effectiveness in supporting food security policies[4].Recent research has integrated machine learning and hybrid modeling approaches for greater accuracy. Devi et al. (2021) proposed a hybrid ARIMA–ANN model to analyze wheat production trends in Haryana, India, finding that the hybrid approach outperformed individual models by capturing both linear and nonlinear trends[5].Likewise, Zulfiqar et al. (2024) developed a hybrid ARIMA–IIS model that integrated Impulse Indicator Saturation to account for structural changes in Pakistan's wheat data, significantly improving forecasting accuracy[6]. Furthermore,reference 7 Shrivastvi et al. (2022) compared ARIMA and ETS models for Indian wheat production, finding that ETS performed better for certain states, demonstrating the context-dependent performance of forecasting methods[7]. Fajar and Nonalisa (2021) applied the TBATS model to predict agricultural price trends in Indonesia, achieving high accuracy for short-term forecasts, illustrating the suitability of TBATS for seasonal agricultural data[8].Additionally,Gawdiya et al. (2024) applied ensemble machine learning models, including Random Forest and ANN, to predict wheat yield under climate change scenarios, showing that hybrid and AI-based models significantly enhance predictive reliability.Building upon these findings[9].This study aims to apply and compare four forecasting models ARIMA, ETS, TBATS, and ANN to predict Pakistan's wheat production for the next seven years.

Methodology

This section provides a clear and detailed account of the methods, models, algorithms, and computational procedures used in the study This study employs a quantitative time-series forecasting approach to predict Pakistan's wheat production using both linear and nonlinear models. The core objective is to identify the most accurate model among ARIMA, Exponential Smoothing (ES), TBATS, and Artificial Neural Networks (ANN). The methodology begins by assessing data stationarity and model identification through the Box–Jenkins procedure, which includes model specification, parameter estimation, diagnostic checking, and model validation. The selection of the optimal ARIMA model follows a detailed evaluation of Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions. Nonlinear modeling techniques, such as ANN and TBATS, are also explored to capture complex and nonlinear dependencies in the data. The study further defines the framework for variable selection, data preprocessing, and performance evaluation based on Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The methodology concludes by presenting a comprehensive procedure for model comparison and selection, enabling the identification of the most robust forecasting approach for future wheat production trends in Pakistan.

ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is a world widely use statistical technique for future forecasting. It was introduced by George Box and Gwilym Jenkins in their seminal work, Time Series Analysis: Forecasting and Control (1970) [10]. The ARIMA model is particularly useful for datasets where future values are dependent on past observations, making it suitable for economic, financial, and scientific forecasting.

Statistically The ARIMA (p, d, q) model can be written as follow

$$Y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots\dots + \alpha_p y_{t-p} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \dots\dots + \beta_q e_{t-q}$$

Where, p is AR term order q denote MA term order d show number of differencing y_t denotes dependent variable(time series), y_{t-p} denotes P_{th} the "AR term" α show AR term coefficient. e_{t-q} shows the q_{th} "MA term" while coefficient of "MA" is denoted by β. e_t is known as white noise error term, which assumes, IID(0 , σ^2).

Artificial Neural Network

It is usually assumed that the work of [11], which offered a mathematical description of the nervous system's structure as a simple logically element network (today known as a neural network), was the start of the neural network journey. The primary benefit of neural networks is their capacity to simulate complicated, nonlinear interactions without making any assumptions about their nature. The ANN model allows for nonlinear operational map from past data ($X_{t-1}, X_{t-2}, \dots\ _{t-p}$) to the future value X_t.

$$X_t = f(X_{t-1}, X_{t-2}, \dots\dots X_{t-p,\omega}) + e_t \qquad (2)$$

where ω denotes the vector of the parameters, The coefficient of the function f is defined by the network's structure and connect weight. Thus a neural network is similar to a nonlinear autoregressive approach.

Network train is a critical component of developing a successful neural network. Back-propagation is the most commonly and widely implemented learning technique. As a result, this paper uses the back-propagation algorithm in the following experiment. For analyzing time series, the method known as ANN is primarily responsible for determining the appropriate amount for input vectors, q and p, which represent the nodes that are hidden and the length of lag data, respectively. However, the work is practically complicated, with no overarching guidelines. Thus, effort is required to determine the optimal p and q values.

Input parameter and output parameter

The end result of the finalize model is deeply depend heavily on important targetDvariables practiced for model's building. The collection of a top fitted and very most proper regressor variables is a very and most very crucial mechanism to be able to correctly build a model the network in matter and accelerating the productivity of a machines [12].

Multilayer Perception (MLP)

The Multilayer-Perceptron(MLP) is a forecast control neural network containing a staring layers, one or more hidden(unseen) layers, and a output(decision) layer. It is prepared using a supervised approach with the backpropagation algorithm, which iteratively adjusts weights to minimize output errors. MLP operate knowledge in a one direction without reaction, manufacturing them a well suited for a many forecasting modeling. Each neuron receives weighted inputs from the preceding layer, applies an activation function (typically sigmoid or hyperbolic tangent in hidden layers and linear in the output layer), and produces an output. In time series forecasting, MLPs require past values as input since they lack inherent time dependence, unlike recurrent models like LSTM. The input data is structured into a matrix format with past observations, allowing the model to generate accurate predictions [13].

Exponential Smoothing (ES) Model

Exponential Smoothing (ES) is a method of time series prediction approach that provides weights that decrease exponentially to previous observations, preferring most recent data. It is easy to use and useful for short-term prediction since it smoothes off stochastic changes in data. The model appears mathematically as follows:

$$S_\circ = X_\circ, \quad S_t = ax_t + (1-\alpha)S_{t-1}, \quad 0 < \alpha$$

Here s_t is a smoothed data point at a time t, x_t is a true data point, and at the end α is the smoothing factor and that smoothing factor controls how much weight are given to a ongoing data. If this formula is expanded repeatedly, it shows that older values contribute less over time, leading to an exponential decay in their influence. While simple exponential smoothing is useful for smoothing data, it cannot predict future trends. Double and triple exponential smoothing methods extend this approach by incorporating trend and seasonality components for better forecasting [14] 2.4.Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) Model Time-series forecasting is used to predict future values based on past observations, especially when data exhibits seasonality (e.g., daily, weekly, monthly, or yearly patterns). Many traditional models, such as ARIMA and exponential smoothing, struggle with multiple or complex seasonalities. The TBATS model is specifically designed to handle such cases. It is built using exponential smoothing and can model complex seasonal patterns, including non-integer, non-nested, and long-period seasonalities[15]. Unlike other models, TBATS is more flexible in dealing with seasonality constraints, allowing for better long-term forecasting. TBATS model is expressed as

$$Y_t = l_t + b_t + \sum_{j=1}^{k} S_{t,j} + d_t$$

y_t = forecast value at time t , l_t =level (Smoothed value) , b_t =Trend Component , $s_{t,}$ =Season Component for Period p , d_t =ARIMA Term Error.

2.4.1 TBATS Components:

T: Trigonometric seasonality (models repeating seasonalpatterns)

B:Box-Cox transformation (stabilizes variance)

A: ARIMA errors (captures short-term dependencies)

T: Trend component (models overall direction of data)

S: Seasonal components (captures periodic patterns) 2.5. Forecast Evaluation Metrics This study used three types of error-based prediction evaluation measures: mean absolute error in percentages (MAPE), mean absolute error (MAE), and a mean square error (RMSE).

i. Mean Absolute Error (MAE): Determines the average absolute forecasting error.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \widehat{Y}_t\right|$$

ii. RMSE is an important precision metric to evaluate different models. In the terminology of mathematics, it appears as:

$$RMSE = \frac{1}{n}\sum_{t=1}^{n} e^2{}_t$$

iii. Mean Absolute Percentage Error (MAPE): Determines how much accuracy of the predictions.

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_{t-}\widehat{Y}_t}{Y_t}\right|$$

2.6 Data

This research used the pakistan year-wise wheat production dataset from (1960 -2023) for analysis to find the best model for future wheat production prediction for Pakistan. The models under consideration are : ARIMA, ES, TBATA and ANN.

2.7 Data Preprocessing

Proper data preprocessing is essential for ensuring reliable model estimation and facilitating accurate forecasting. The subsequent procedures were applied to Pakistan year-wise wheat data set before forecasting through the TBATS model. Table 1 presents data preprocessing procedures

Table 1. Data preprocessing procedures

Step	Description	Purpose
Data Cleaning	The actual index data were examined for value absences, disparities, and date inconsistencies.	confirm the dataset's completion and accuracy.

Stationarity Tests	The augmented Dickey-Fuller tests were used for stationary behavior.	Ensures stationary behavior, which is required for ARIMA modeling.
Data Partitioning	The set of data was split into two subsets: training (80%) and testing (20%).	This provides a solid basis for modeling and assessment across the original sample.

Theoretical Results

This section describes the core mathematical structure and statistical formulation used in the investigation. The TBATS (Trigonometric, Box-Cox, ARMA errors, Trend, and Seasonal components) model was found as the most accurate predicting approach for Pakistan's yearly crop of wheat from 1960 to 2023. De Livera, Hyndman, and Snyder (2011) established the TBATS method, which extends the exponential smoothing space of states approach to accommodate complicated, variety, or non-integer seasonally. It also includes Box-Cox transformations, trend and damping parameters, trigonometry seasonality terms, and short-run ARMA errors. [16]

Theoretical Framework of the TBATS Model

The TBATS approach integrates multiple components to create a single time-series architecture containing trend, seasonality, and short-term autocorrelation all at once. The fundamental TBATS formula is given as

$$Yt^{(\lambda)} = l_{t-1} + \emptyset b_{t-1} + \sum_{j=1}^{k} s_{t-1,j} + d_t \ , \quad (8)$$

$$l_t = l_{t-1} + \emptyset b_{t-1} + \alpha d_t \ , \quad (9)$$

$$b_t = \emptyset b_{t-1} + \beta d_t \ , \quad (10)$$

$$S_{t,j} = S_{t-1,j} \cos(\omega_j) + S^*{}_{t-1,j} \sin(\omega_j) + \gamma_j d_t \ , \quad (11)$$

$$S^*{}_{t,j} = -S_{t-1,j} \sin(\omega_j) + S^*{}_{t-1,j} \cos(\omega_j) + \gamma^*{}_j d_t \ ,$$

Where $Yt^{(\lambda)}$ is the Box-Cox transformed observation with parameter λ

l_t is the level component

b_t is the trend component with damping parameter $\emptyset$ $(0 < \emptyset < 1)$

$S_{t,j}$ *and* $S^*{}_{t,j}$ represent the pair of trigonometric seasonal components

$\omega_j = {}^{2\pi}/_{m_j}$ corresponds to the seasonal frequency for period m_j

d_t is an ARMA (p,q) error term $d_t = \sum_{i=1}^{p} \emptyset_i \, d_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t$

$\varepsilon_t \sim i.i.d.(0, \sigma^2)$ represent random error term.

For implementing the TBATS model, divide the data into two sets: training and testing. The training set is used to fit the model, while the testing set evaluates its performance. Then, apply the TBATS model to the training data to capture trends and seasonal components. Finally, run the trained model to obtain projections for the selected future period. Finally, when dealing with time-series data with various and dynamic seasonalities, TBATS outperforms established approaches.

Application

This part provides the major outcomes of the study

ADF TEST FOR ARIMA MODEL STATIONARITY

It is important to consider to the stationary of data before utalizing the model ARIMA to conform if the given time series data is stationary or not. The table 1 resulted values tells that the given data is not stationary, because P-vale =0.1126991 which is exceed than 0.05

Table 1 ADF test for ARIMA model

Test	Dickey_Fuller_Statistic	Lag_Order	P_Value	Alternative_Hypothesis
Augmented Dickey-Fuller Test	-3.140859	3	0.1126991	Non-stationary

After taking the 1st difference of the data the table 2 represent that the data become stationary p-value is 0.01 obtained by the ADF test, which is less than 0.05, and that's outcomes tells that the data become perfect stationary, after taking the first difference.

Table 2 1st Difference outcomes of ADF test

Test	Dickey_Fuller_Statistic	Lag_Order	P_Value	Alternative_Hypothesis
Augmented Dickey-Fuller Test	-5.774517	3	0.01	Stationary

Significance of Parameters

The ADF test table shows the estimated coefficients, standard errors, t-values, and pvalues for the ARIMA model applied to Pakistan's wheat production data. The ma1 and drift parameters are statistically significant, as their p-values is very so small, highlighting so strong evidence against the null hypothesis. So the negative coefficient of ma1 suggests a significant moving average component in the model, while the positive drift parameter highlights an increasing trend in wheat production. The low standard errors indicate precise parameter estimates. These findings validate the ARIMA model's suitability for forecasting wheat production trends.

Table 3: Significance test of ARIMA model parameters

	Parameter	Coefficient	Std_Error	T_value	P_Value
ma1	ma1	-0.7938644	0.1002697	-7.917288	$2.442491e^{-15}$
drift	Drift	384.4710262	27.7882914	13.835720	$0.000000e^{+00}$

4.3 Actual Vs. Predicted Values of the Models

In table 4, Four model have been used. The comparison of ARIMA, ES, TBATS and ANN models. The predicted values of Pakistan dataset by ES model is the most close to the actual values as compared to other three mentioned models. ARIMA is the second-best model, while ANN and TBATS have larger deviation making them less reliable.

Table 4: Comparison of ARIMA, ES, TBATS and ANN models

Year	Actual Value	ES Prediction	ARIMA Prediction	TBATS Prediction	ANN Prediction
2004	19500	19999.09	20029.14	20532.27	19825.05
2005	21612	20267.41	20304.54	20997.16	20131.79
2006	21277	20944.94	20958.52	21383.17	22114.95
2007	23295	21397.85	21408.64	21798.52	21807.84
2008	20959	22198.21	22181.96	22139.97	23611.28
2009	24033	22302.55	22314.33	22599.07	21513.62
2010	23311	23065.97	23053.08	22936.1	24241.71
2011	25214	23499.83	23490.72	23316.27	23625.12
2012	23473	24259.84	24230.42	23626.89	25216.26
2013	24211	24464.84	24458.76	24024.2	23764.82
2014	25979	24788.07	24792.16	24401.94	24391.33
2015	25086	25432.0	25421.28	24714.86	25824.37
2016	25633	25734.87	25736.64	25076.95	25112.71
2017	26674	26091.9	26099.75	25426.71	25551.61
2018	25076	26600.75	26602.59	25742.22	26360.73

2019	24349	26641.97	26672.38	26139.43	25104.6
2020	25248	26512.51	26577.92	26584.57	24506.66
2021	27464	26611.12	26688.25	27002.14	25243.67
2022	26400	27179.63	27232.63	27331.78	26951.6
2023	26810	27385.83	27445.46	27720.28	26151.12

Model Evaluations

The performance of the models: ARIMA, ES,TBATS and ANN Model was evaluated using three error-based metrics: Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The results are summarised in Table 5

Table 5 Model Evaluation

Country	Models	RMSE	MAE	MAPE
Pakistan	ARIMA Model	1158.14	979.55	4.06
Pakistan	ES Model	1155.71	977.65	4.06
Pakistan	TBATS Model	1067.30	916.35	3.78
Pakistan	ANN Model	1380.86	1207.80	5.02

Table 5 present the forecast performance based on the MAE, RMSE, and MAPE metrics, indicates that the TBATS model outperforms all other models in terms of forecasting accuracy due to it minimum values as compare to other model. We conclude that the TBATS model provides the best results for the yearly Wheat Production and is the optimal choice for forecasting in this analysis.

Forecast values (24-2030) using TBATS model

Table 6 : Forecast values using TBATS model

Year	TBATS Forecast (1000 MT)
2024	28102.3
2025	28437.24
2026	28767.36
2027	29092.64
2028	29413.07
2029	29728.65
2030	30039.36

Table 6 that shows the seven year forecast values of wheat production for Pakistan using proposed TBATS model. The forecast values are increasing from one to another, which means that the wheat production of Pakistan will be increased by 2030 upto 6.89% according to the TBATS model which help the country to plan their future strategies and agricultural development.

4.6 Forecasting Plots

Figures [2–3] show the forecasting outcomes of ARIMA, ES, TBATS and ANN for Pakistan and the proposed TBATS models, respectively These graphics illustrate each model's capability to detect patterns and variability in the dataset. Figure 2 illustrate that the most accurate among the four models ANN shows significant fluctuations and deviations, while TBATS provides a smoother trend but fails to capture sudden variations in actual values. The actual data exhibits noticeable fluctuations, and while ES and ARIMA effectively track these changes, ANN and TBATS show larger deviations.

Figure 3 illustrate TBATS model, the values are having an upward trend which means that the wheat production of Pakistan for the next seven years will be increasing according to the forecasting evaluation of proposed model (TBATS) model. While figure 1 shows the multilayer perception graphical representation for ANN.

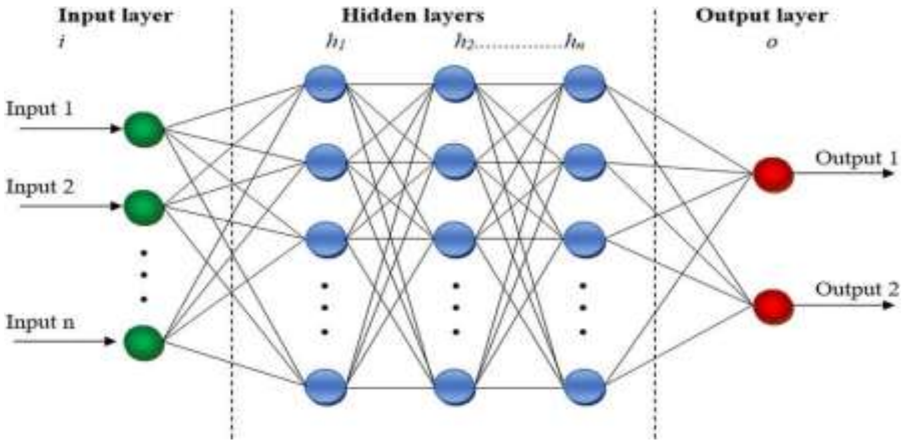

Figure 1 : Architecture of multilayer perception artificial neural network (MLP-ANN)

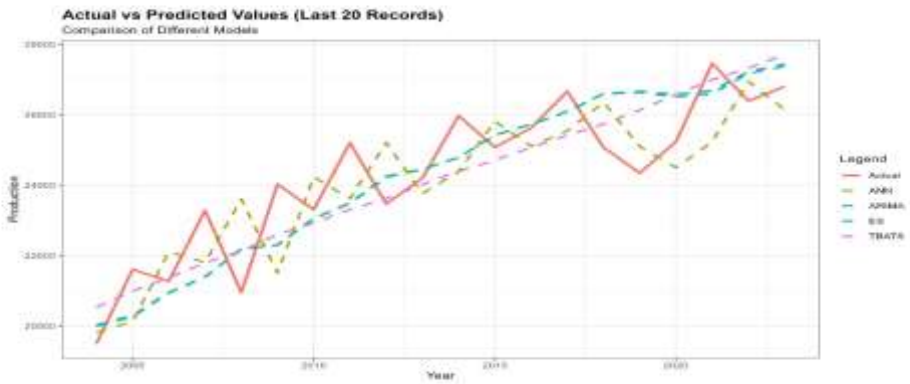

Figure 2: Testing Set Actual Values vs ARIMA, ES, TBATS and ANN for Pakistan

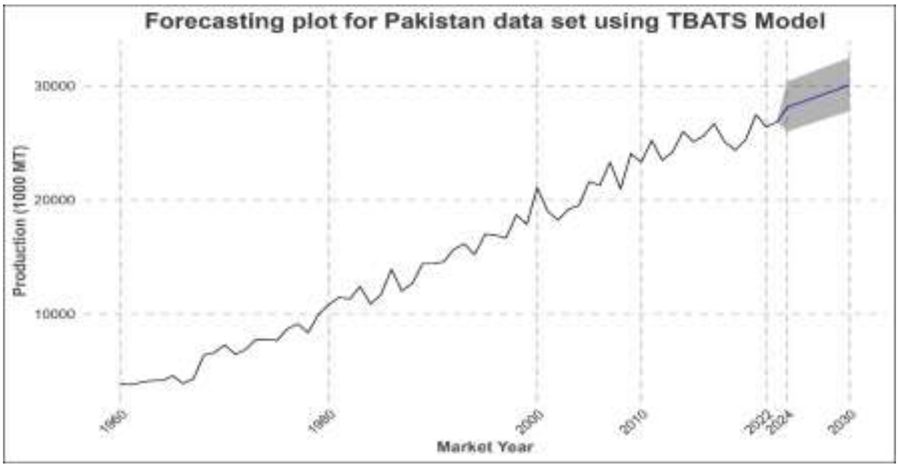

Figure 3 : Forecasting plot for Pakistan data set using TBATS Model

Result and Discussion

The ARIMA, ES, TBATS and ANN Models is fit on wheat Pakistan production data. The Box-Jenkin methodology was used in ARIMA Model and its result stationary after taking 1st difference as shown in table 2 Several plots and tables including results of the models and visual depiction of models outcomes. The finding from evaluation metrics table 4 examine that the TBATS model having minimum evaluaton metrics outperform well across all , Based on TBATS model the table 5 represent the forecast prediction of wheat will be increase gradually year wise. The models is contrasted on their performance indicators derived from the result. The data analysis is conducted with the help of R-studio programming software. The analysis proceeded after dividing the original data into training and testing components. this research enhances forecasting accuracy and supports data-driven policy decisions for sustainable wheat production in Pakistan.

Conclusion

The study investigated Pakistan's annual wheat crop from 1960 to 2023 using four prediction models: ARIMA, Exponential Smoothing (ES), TBATS, and Artificial Neural Network (ANN) to discover the most accurate technique for future production forecasting. The TBATS (Trigonometric, Box-Cox, ARMA errors, Trend, and Seasonal) model outperformed all others, achieving the lowest RMSE (1067.30), MAE (916.35), and MAPE (3.78%). The TBATS approach outperformed linear statistical and computational methods in identifying complex seasonal fluctuations and nonlinear patterns. The finding shows a steady 6.9% growth in wheat output through 2030, indicating significant crop growth and the potential benefits of technological and policy developments. These findings have significant implications for economic and agricultural planning, as exact output projections enable policymakers to develop effective policies for food safety, resource usage, and balanced agricultural expansion.

Future Research work

The forecasting of these data based on under normal condition and may not fully capture the impact of unexpected shocks such as floods, earthquake any other kind of natural disasters. For future, it is suggested to investigate some advanced and hybrid models to cover this drawback and make better and more effective forecasts. Advanced models include machine learning and deep learning techniques i.e. Recurrent Neural Networks (RNNs), Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Long Short Term- Memory (LSTM) networks, Support Vector Machines (SVM), Support Vector Regression (SVR), Naive Bayes and Random Forests, which are competent to capture complex and non- linear patterns in the data. Additionally, hybrid approaches that combine the techniques of traditional time series with artificial intelligence like ARIMA-RNN, Prophet-LSTM and ETS-SVM may further improve the performance of predictions.

References

M.A. Khan, “Report on wheat (for Expo Pakistan 2017): Product: Wheat, meslin, & flour (HS Code: 1001 & 1101),” Trade Development Authority of Pakistan (TDAP), 2017.

L.N. Singh, V.B. Darji, and Y.S. Singh, “Area, production and productivity of wheat (Triticum aestivum) in Gujarat State: Forecasting by using ARIMA models,” International Journal of Bio-resource and Stress Management, vol. 7, no. 5, pp. 1093–1098, 2016.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. Journal of Asian Development Studies, 14(2), 1518-1527.

N. Saeed, A. Saeed, M. Zakria, and T.M. Bajwa, “Forecasting of Wheat Production in Pakistan using ARIMA Models,” International Journal of Agriculture & Biology, vol. 2, no. 4, pp. 352–353, 2000.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). Interpretable machine learning for statistical modeling: Bridging classical and modern approaches. International Journal of Social Sciences Bulletin, 3(8), 43-50.

M. Devi, J. Kumar, D.P. Malik, and P. Mishra, “Forecasting of Wheat Production in Haryana Using Hybrid Time Series Model,” Journal of Agriculture and Food Research, 2021.

H. Zulfiqar, R. Ahmad, and U. Shahzad, “Hybrid ARIMA-IIS Approach for Wheat Yield Forecasting: An Integrated Approach,” iRASD Journal of Economics, vol. 6, no. 1, pp. 109–127, 2024.

KHAN, R., SHAH, A. M., & KHAN, H. U. (2025). Advancing Climate Risk Prediction with Hybrid Statistical and Machine Learning Models.

S. Shrivastri, K.M. Alakkari, P. Lal, A. Yonar, and S. Yadav, “A Comparative Study between (ARIMA-ETS) Models to Forecast Wheat Production and its Importance in Nutritional Security,” Journal of Agriculture, Biology and Applied Statistics, vol. 1, no. 1, pp. 25–37, 2022.

Ahmad, M., Khan, I. A., Khan, R., Saleem, M., & Ullah, I. (2025). Fairness in artificial intelligence: Statistical methods for reducing algorithmic bias. Journal of Media Horizons, 6(3), 2206-2214.

M. Fajar and S. Nonalisa, “Forecasting Chili Prices Using TBATS,” International Journal of Scientific Research in Multidisciplinary Studies, vol. 7, no. 2, pp. 1–5, 2021.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. Policy Research Journal, 3(8), 55-62.

S. Gawdiya, D. Kumar, B. Ahmed, R.K. Sharma, P. Das, M. Choudhary, and M.A. Mattar, “Field scale wheat yield prediction using ensemble machine learning techniques,” Smart Agricultural Technology, vol. 9, p. 100543, 2024.

Ahmad, M., Rehman, A. A., Khan, R., & Bibi, H. (2025). Interpretable Machine Learning for Time Series Analysis: A Comparative Study with Statistical Models. ACADEMIA International Journal for Social Sciences, 4(3), 4001-4009.

W.S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” Bulletin of Mathematical Biophysics, 2025.

Ahmad, M., & Ahmad, R. W. (2025). Statistical Process Control for Real-Time Industrial Data Streams. Annual Methodological Archive Research Review, 3(8), 1039-1049.

A. Rezrazi, S. Hanini, and M. Laidi, “An optimization methodology of artificial neural network models for predicting solar radiation: a case study,” Theoretical and Applied Climatology, vol. 123, no. 3–4, pp. 769–783, 2016.

Hanif, M. A., Wadood, A., Ahmad, R. W., Shah, S. A., & Khan, R. (2025). Real-Time Anomaly Detection in IoT Sensor Data Using Statistical and Machine Learning Methods. ACADEMIA International Journal for Social Sciences, 4(3), 5203-5227.

Lazcano, M.A. Jaramillo-Morán, and J.E. Sandubete, “Back to Basics: The Power of the Multilayer Perceptron in Financial Time Series Forecasting,” Mathematics, vol. 12, no. 12, p. 1920, 2024.

Ahmad, M., Khan, R., Ahmad, R. W., Wahab, F., & Nizamani, S. (2025). Quantifying the Impact of Dot Balls on Winning Probability in T20 Cricket. ACADEMIA International Journal for Social Sciences, 4(3), 4865-4885.

Ullah, A. (2025). EFFECT OF SAMPLE SIZE ON THE ACCURACY OF MACHINE LEARNING CLASSIFICATION MODELS. Spectrum of Engineering Sciences, 826-834.

Ahmad, M., Saleem, M., & Memon, B. A. (2025). EFFECT OF OUTLIERS ON CLASSICAL VS. ROBUST REGRESSION TECHNIQUES. International Journal of Social Sciences Bulletin, 3(8), 686-692.

A.M. De Livera, R.J. Hyndman, and R.D. Snyder, “Forecasting time series with complex seasonal patterns using exponential smoothing,” Journal of the American Statistical Association, vol. 106, no. 496, pp. 1513–1527, 2011.