STATISTICAL MODELING FOR AREA OF PEARL MILLET IN GUJARAT

## Abstract:

The study focuses on modeling the area under pearl millet cultivation in Gujarat, India, using various statistical methods. The analysis employs both parametric (ARIMA, linear, non-linear and Exponential Smoothing State Space models) and non-parametric (LOESS smoothing) approaches. Data from 1980-81 to 2021-22 were used, with models trained on data up to 2013-14 and tested on subsequent data. The study found that the cubic and logistic models provided the best fit among linear and non-linear models, respectively, while ARIMA (0,1,1) with drift and ETS (A,A,N) were optimal among time series models. Ultimately, the cubic model was slightly superior, though ARIMA (011) with drift showed the best forecasting ability. Despite efforts like the National Millet Mission and International Year of Millets, the area under pearl millet cultivation continues to decline, indicating a need for enhanced strategies to support pearl millet farming in Gujarat.

***Key words***: Non-linear and linear models, Auto-correlation, Band width, Kernel, Pearl Millet, Statistical Modeling, ARIMA, Exponential Smoothing, LOESS Smoothing,

# INTRODUCTION

Agriculture is the backbone of the economy, providing food security, raw materials, and livelihoods to a significant portion of the global population (Dodiya and Barad, 2022). Millets are a collective group of small-seeded annual grasses grown primarily as grain crops on marginal lands in dry areas of temperate, sub-tropical and tropical regions (Anon., 1996). Bajra (pearl millet) is one of the most significant millets cultivated widely in India due to its nutritional value and adaptability to harsh climatic conditions (Satyavathi *et al*., 2021; Dodiya *et al.,* 2024). India holds a prominent position in the cultivation and production of millets. Gujarat, in particular, plays a vital role in India's millet scenario. Gujarat occupies 6.65% of the total area under pearl millet cultivation in India, respectively. Moreover, Gujarat contributes 10.97% of the total Bajra (Chaudhari *et al*., 2023) production in India, amounting to 1089.83 metric tonnes (Anon., 2022). Pearl millet occupies almost 80% of the total area under millets in Gujarat. The importance of timely and reliable forecasts for the area, production and productivity of millet like pearl millet cannot be overstated for a country like India (Dodiya *et al.*, 2024). India proposed to the United Nations to declare 2023 as the International Year of Millets (IYOM), highlighting the significance of these crops (Anon., 2023). This information is crucial for planning and allocating resources within the agricultural sector. Accurate data on crop area, production and productivity helps planners and policy-makers make informed decisions regarding procurement, storage, public distribution, export, import and other related issues (Nath, 2008). Regional data analysis is vital as it forms the basis for economic and policy planning by both State and Central governments. The revival of millet cultivation in India and the global recognition of their benefits underscore the importance of continued research and development (Dodiya *et al.*, 2023). By leveraging reliable crop area, production and productivity data and employing statistical methods to analyze trends, stakeholders can enhance the agricultural sector's resilience and sustainability (Panwar and Dimri, 2018).

Keeping all these points in view, the objective of the present paper is to fit appropriate statistical models to study the trends in area of pearl millet grown in Gujarat State and to compare the best-fitted models using a test dataset. This study employs both parametric and non-parametric methods (PROC LOSSES) to analyze the data. Specifically, the study includes fitting ARIMA, linear, non-linear and Exponential Smoothing State Space (ETS) models as parametric methods, alongside a non-parametric loess smoothing approach.

# MATERIALS AND METHODS

## Data Collection

The secondary data on the area under pearl millet in Gujarat was collected from the Directorate of Agriculture, Government of Gujarat [(www.dag.gujarat.gov.in),](http://www.dag.gujarat.gov.in/) covering the period from 1980-81 to 2021-22. The data was divided into two sets: 1980-81 to 2013-14 for training the models and 2013-14 to 2021-22 for testing the models. After fitting models on the training dataset, the best model from each category was selected based on satisfactory goodness of fit criteria. Predictions were then made on the testing dataset using the best model from each category to compare their predictive abilities.

## Linear and Non-linear Models

Different linear and non-linear models given in Table 1 were employed. The evaluation criteria included adjusted R2, significance of F-value, RMSE and MAE values. Among all the fitted linear and non-linear models, the best-fitted model was selected based on the highest adjusted R2 with significant F-value and the lowest RMSE and MAE values, indicating its superior goodness of fit (Montgomery *et al*., 2003). Additionally, the Shapiro-Wilk test for normality of residuals and the run test procedure for randomness of error developed in the literature (Sidney and Castellan, 1988) were leveraged to diagnose the residuals of the fitted model. Whenever there was a comparable goodness of fit (adjusted R2, RMSE and MAE), a model with less and statistically significant predictors were chosen. The linear, non-linear and time series models (ARIMA) were fitted to the time series data using R/R studio.

## Table 1: List of linear and nonlinear models

|  |  |  |
| --- | --- | --- |
| **Model No.** | **Model** | **Name of the Model** |
| I. | Y=a + bt + e | Linear equation |
| II. | Y=a + bt + ct2 + e | Quadratic equation |
| III. | Y=a + bt + ct2 + dt3 + e | Cubic equation |
| IV. | Y= aebt + e | Exponential equation |
| V. | Y=a/[1+be(ct)] + e | Logistic equation |

Where, Y was the area and t were the time points, a, b, c and d are the parameters and e is the error term. In case of logistic model, the parameter ‘a’ represents carrying capacity, ‘c’ is the intrinsic growth rate and b is parameter.

## ARIMA (Autoregressive Integrated Moving Average) Model

ARIMA (Autoregressive Integrated Moving Average) model introduced by Box and Jenkins (1976), is a powerful time-series forecasting method that incorporates autoregression (AR), differencing (I), and moving average (MA) components. The model is denoted as ARIMA (p,d,q), where p is the number of lag observations, d is the degree of differencing needed to make the series stationary, and q is the size of the moving average window.

AR model is the linear function of lagged value of the variable, which can be denoted as; AR (p) = 𝑌𝑡 = 𝑐 + 𝜙1𝑌𝑡−1 + 𝜙2𝑌𝑡−2 + ⋯ + 𝜙𝑝𝑌𝑡−𝑝 + 𝑒𝑡

MA model is the function of random error and lagged value of the variable, which can be written as;

MA (q) = Yt = 𝜇0 + 𝑒𝑡+𝜃1𝑒𝑡−1 + 𝜃2𝑒𝑡−2 + ⋯ + 𝜃𝑞𝑒𝑡−𝑞

A complete ARIMA model equation without differencing becomes ARMA model and it can be written as (Hyndman and Athnasopoulos, 2018),

𝑌𝑡 = 𝑐 + 𝜙1𝑌𝑡−1 + … + 𝜙𝑝𝑌𝑡−𝑝 + 𝑒𝑡 +𝜃1𝑒𝑡−1 + … + 𝜃𝑞𝑒𝑡−𝑞

A complete ARIMA model equation with differencing and without drift can be written as (when d = 1),

𝑌𝑡= 𝑐 + 𝑌𝑡−1 + 𝜙1𝑌𝑡−1 + … + 𝜙𝑝𝑌𝑡−𝑝 + 𝑒𝑡+𝜃1𝑒𝑡−1 + … + 𝜃𝑞𝑒𝑡−𝑞

The same ARIMA model equation with differencing and drift term can be written as,

𝑌𝑡= c + δ𝑡 + 𝑌𝑡−1 + 𝜙1𝑌𝑡−1 + … + 𝜙𝑝𝑌𝑡−𝑝 + 𝑒𝑡+𝜃1𝑒𝑡−1 + … + 𝜃𝑞𝑒𝑡−𝑞

Where, c = constant

δt = the drift term where δ is the drift coefficient

𝜙1, 𝜙2, … , 𝜙𝑝 = regression coefficients of AR part

𝑌𝑡 = dependent variable at time t

𝑒𝑡 = random error term

𝜃1, 𝜃2, … , 𝜃q = regression coefficients of MA part

The methodology for developing both types of ARIMA models involves several key steps. Identification uses the autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine the order of the model. Parameters are then estimated, often using maximum likelihood estimation (MLE). Diagnostic checking ensures the model's adequacy through residual analysis and tests such as the Ljung-Box test (Cromwell *et al.,* 1994) for independence of residuals, the Shapiro-Wilk test (Shapiro and Wilk, 1965) for normality. In the case of time series (ARIMA) models, the first step involved checking the stationarity conditions using the Augmented Dickey- Fuller (ADF) test (Dickey and Fuller, 1981). Autocorrelations up to 15 lags were examined. If the autocorrelation plot did not show a sharp tail-off towards zero and visual inspection of time series reveals that the mean and variance were not constant over time, the time series was deemed non- stationary. Consequently, a new variable Xt was constructed by taking the difference of original time series (i.e. d = 1, 2, …) once, twice and so on to make the series stationary.

The Arima () function in R allows for the inclusion of constant or drift term by setting the argument include.drift = TRUE for d = 1 and automatically considers the constant term for d = 0. When d > 1, no constant is allowed, as a quadratic or higher-order trend is hazardous when forecasting. When d = 1, different ARIMA models with and without a drift parameter were fitted to check whether including the drift parameter improved the fit of the model (Hyndman and Athnasopoulos, 2018). After fitting series of ARIMA models (with and without drift) to the data, the one with the lowest Akaike’s Information Criterion (AIC), Bayesian Information Criterion (BIC), RMSE and significant 't' values for the estimated parameters was considered the best fit.

The fitted models were also assessed for normality and independence of error terms.

## Exponential Smoothing State Space (ETS) Model

The Exponential Smoothing State Space model constructs time series forecasts using three main components: Error, Trend, and Seasonal. The ETS framework allows for both additive and multiplicative models, accommodating various types of time series data (Gardner, 2006). Since the data used in this study is annual, the seasonal component (S) is ignored. The model can be either additive (Yt = T + E) or multiplicative (Yt = T \* E). The individual components of model are classified as:

* + - E (Error): [A - Additive, M - Multiplicative]
		- T (Trend): [N - None, A - Additive, M - Multiplicative, AD - Additive Dampened, MD - Multiplicative Dampened]
		- S (Seasonal): [N - None, A - Additive, M - Multiplicative] The parameters involved in the ETS model include:
		- α (alpha): Smoothing factor for level, β (beta): Smoothing factor for trend, ϕ (phi): Damping coefficient

The initial states required for the model are:

* + - l: Initial level components, b: Initial growth components

The Table 2 describes the models that we are working on (Yonar *et al*., 2022):

The ‘ets()’ function in R/R Studio was used to fit the ETS models (Ashfaq *et al*., 2014). After fitting different combinations, the best four models were retained based on lower AICc values. These retained models were evaluated using RMSE, AIC, and BIC. The residuals of the fitted models were assessed for normality and independence using the Shapiro-Wilk test and the Box-Ljung test, respectively. The best model was then selected based on the values of RMSE, AIC and BIC.

## Table 2: Probabilities of the model shape in state space

|  |  |  |  |
| --- | --- | --- | --- |
| **Trend** | **Additive Error Models** | **Trend** | **Multiplicative Error Models** |
| **N** | 𝑌𝑡 = 𝓁𝑡−1 + 𝑒𝑡𝓁𝑡 = 𝓁𝑡−1 + 𝛼𝑒𝑡 | **N** | 𝑌𝑡 = 𝓁𝑡−1(1 + 𝑒𝑡)𝓁𝑡 = 𝓁𝑡−1(1 + 𝛼𝑒𝑡) |
| **A** | 𝑌𝑡 = 𝓁𝑡−1 + 𝑏𝑡−1 + 𝑒𝑡𝓁𝑡 = 𝓁𝑡−1 + 𝑏𝑡−1 + 𝛼𝑒𝑡𝑏𝑡 = 𝑏𝑡−1 + 𝛽𝑒𝑡 | **M** | 𝑌𝑡 = (𝓁𝑡−1 + 𝑏𝑡−1)(1 + 𝑒𝑡)𝓁𝑡 = (𝓁𝑡−1 + 𝑏𝑡−1)(1 + 𝛼𝑒𝑡)𝑏𝑡 = 𝑏𝑡−1 + 𝛽(𝓁𝑡−1 + 𝑏𝑡−1)𝑒𝑡 |
| **AD** | 𝑌𝑡 = 𝓁𝑡−1 + ∅𝑏𝑡−𝑞 + 𝛽𝑒𝑡𝓁𝑡 = 𝓁𝑡−1 + ∅𝑏𝑡−𝑞 + 𝛼𝑒𝑡𝑏𝑡 = ∅𝑏𝑡−𝑞 + 𝛽𝑒𝑡 | **MD** | 𝑌𝑡 = (𝓁𝑡−1 + ∅𝑏𝑡−𝑞)(1 + 𝑒𝑡)𝓁𝑡 = (𝓁𝑡−1 + ∅𝑏𝑡−1)(1 + 𝛼𝑒𝑡)𝑏𝑡 = ∅𝑏𝑡−1 + 𝛽(𝓁𝑡−1 + ∅𝑏𝑡−1)𝑒𝑡 |

* 1. **Nonparametric Regression (LOESS Smoothing)**

In general, a nonparametric regression model can be expressed as follows:

Y = m (X) + Ɛ

Where Y is the response variable, the function of unknown form m(X) is the mean response or regression function assumed to be smooth and Ɛ represents the independently and identically distributed random error with a mean of zero.

Non-parametric regression using smoothing techniques, such as LOESS, estimates the regression function without assuming a specific model form (Hardle, 1990). A regression smoother depicts trends in predictors (X) as a function of the response (Y) by averaging Y-values for observations near the target X-value, with the neighborhood size determined by a smoothing parameter or bandwidth (Cleveland & Devlin, 1988).

The nonparametric regression model was fitted to the data of area of pearl millet using the PROC LOESS procedure in SAS. In the LOESS method, weighted least squares is used to fit linear or quadratic function of the predictors at the centers of neighborhoods. The radius of each neighborhood is chosen so that neighborhood contains specified percentage of the data points. The Estimation of optimum bandwidth was facilitated by the default method implemented in PROC LOESS, which selects the smoothing parameter that minimizes the AIC criterion aiding in the determination of an optimal bandwidth for the model (Hurvich *et al*., 1998). Validation for predictive accuracy was not performed using nonparametric LOESS because LOESS is designed to fit local segments of data without assuming a specific functional form, rendering it unreliable for extrapolation and forecasting beyond the training data range. The technique is implemented in the SAS procedure LOESS (SAS Institute Inc., 2021).

PROC LOESS DATA=SAS-data-set;

MODEL dependents = regressors / options;

OUTPUT OUT=SAS-data-set keyword=name … keyword=name / options; ID variables;

BY variables;

WEIGHT variable;

SCORE DATA=SAS-data-set ID=(variable-list) / options; RUN;

# RESULTS AND DISCUSSION

Different parametric and nonparametric models were utilized to fit time series data concerning the area of pearl millet. The statistics of fitted linear, non-linear, ARIMA, exponential smoothing state space (ETS) model and nonparametric models were collected and the results are discussed in the following sequence.

## Fitting of Linear and Non-Linear Models for Area

The analysis for the area of pearl millet evaluated five linear and non-linear models, as detailed in Table 3. The statistics presented indicate that all the models, both linear and non-linear, exhibited notably high and statistically significant coefficients of determination (R²). Furthermore, the adjusted R² values were substantial, suggesting that these models effectively account for a considerable portion of the variation in the dependent variable (area). The error terms of all the fitted linear and non-linear models were not normally distributed, but the randomness of the residuals was observed in the case of fitted quadratic, cubic and logistic models. The results states that cubic model (linear) and logistic model (non-linear) were best fits based on lower RMSE, MAE and higher adjusted R², with the cubic model being slightly superior than logistic model.

* The fitted cubic model;

𝑌̂𝑡 = 14653.84 + 187.23𝑡 − 21.25𝑡2 + 0.25𝑡3 (Adj.R2 = 93.44%, RMSE=771.67)

* The fitted logistic model;

𝑌̂𝑡

15935.63

= (1 + 0.03 × 𝑒𝑥𝑝(0.11𝑡))

(Adj.R2 = 92.41%, RMSE = 780.82)

## Fitting of ETS Models for Area

The analysis of the Exponential Smoothing State Space (ETS) models for the area under pearl millet yielded significant insights as detailed in Table 4. Four models were evaluated: (A,A,N), (A,N,N), (M,M,N) and (M,A,N). The evaluation criteria included AIC, BIC, RMSE and tests for normality and randomness of residuals. The (A,A,N) model performed the best, with the lowest AIC (595.88) and BIC (603.51) values and the lowest RMSE (946.42), indicating it is highly accurate. The error terms for this model were normally distributed (Shapiro-Wilk test value of 0.96) and the residuals were random (Box-Ljung test value of 14.57). In comparison, the other models - (A,N,N), (M,M,N) and (M,A,N) - had higher AIC, BIC and RMSE values, indicating they were less accurate. Additionally, these models had non-normally distributed residuals, further reducing their reliability compared to the (A,A,N) model.

## Table 3: Statistics of fitted linear and non-linear models for area under pearl millet

|  |  |  |  |
| --- | --- | --- | --- |
| **Aspects** | **Model** | **Parameters** | **Goodness of fit** |
| **a** | **b** | **c** | **d** | **R2 (%)** | **S-W****Test** | **Run Test (|Z|)** | **RMSE** | **MAE** |
| **Adj.R2(%)** |
| **Linear** | Linear | 16902.95\*\* [377.45] | -278.58\*\* [18.81] | - | - | 87.26\*\* | 0.93\* | 3.80\*\* | 1044.12 | 853.58 |
| 86.87 |
| Quadratic | 15247.53\*\* [454.42] | -2.68[59.86] | -7.88\*\* [1.67] | - | 92.63\*\* | 0.91\*\* | 1.13 | 794.24 | 539.57 |
| 92.16 |
| **Cubic** | **14653.84\*\* [631.89]** | **187.23****[154.09]** | **-21.25\* [10.15]** | **0.25****[0.19]** | **94.04\*\*** | **0.87\*\*** | **0.58** | **771.67** | **507.23** |
| **93.44** |
| **Non-linear** | Exponential | 17246.09\* [535.90] | -0.021\*\* [0.002] | - | - | 82.27\*\* | 0.93\* | 3.82\*\* | 1231.64 | 1009.57 |
| 81.72 |
| **Logistic** | **15935.63\*\* [571.10]** | **0.03\* [0.02]** | **0.11\*\* [0.02]** | **-** | **92.87\*\*** | **0.91\*\*** | **1.13** | **780.82** | **532.25** |
| **92.41** |

\* Significant at 5% level \*\* Significant at 1% level Values in brackets [ ] indicates standard errors

## Table 4: Statistics of fitted ETS models for area under pearl millet

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Aspect** | **Model** | **Initial states** | **Smoothing parameters** | **AIC** | **BIC** | **S-W TEST** | **B-L TEST** | **RMSE** |
| 𝒍 | 𝒃 | 𝜶 | 𝜷 |
| **Area** | **(A,A,N)** | **15932.99** | **-62.83** | **0.14** | **0.14** | **595.88** | **603.51** | **0.96** | **14.57** | **946.42** |
| (A,N,N) | 15055.93 | - | 0.58 | - | 601.99 | 606.58 | 0.90\*\* | 20.22 | 1098.27 |
| (M,M,N) | 16230.87 | 1.00 | 0.21 | 0.06 | 606.84 | 614.47 | 0.91\*\* | 16.40 | 970.99 |
| (M,A,N) | 16327.65 | -62.75 | 0.23 | 0.03 | 607.56 | 615.19 | 0.92\* | 14.94 | 983.12 |

\* Significant at 5% level \*\* Significant at 1% level

## Fitting of ARIMA Models for Area

The initial step in fitting ARIMA models involves transforming non-stationary time series into a stationary form. In Fig.1, the time series initially appeared non-stationary due to an autocorrelation plot lacking a sharp tail-off towards zero, along with visual evidence of non- constant mean and variance over time. This was further supported by a non-significant result from the Augmented Dickey-Fuller (ADF) test. However, after applying first-order differencing, stationarity was achieved, confirmed by a significant result from the ADF test (Table 5). The Fig. 2 depicts stationary time series after first-order differencing.

Given that d equals 1, the algebraic family of ARIMA models, including ARIMA (0,1,1), ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (2,1,1) and ARIMA (2,1,2), each incorporating a drift parameter, were identified for the fitting process to determine the optimal ARIMA model. Subsequently, to verify whether the inclusion of the drift parameter enhances the model's performance, additional models from the algebraic family, such as ARIMA (2,1,2), ARIMA (3,1,3), ARIMA (3,1,2), ARIMA (1,1,3) and ARIMA (2,1,3) without drift, were tested.

The results of fitting ARIMA models with drift are summarized in Table 6. Among the fitted models, ARIMA (0,1,1) exhibited the lowest values for both AIC and BIC, along with significant moving average (MA) term. Residuals from all fitted ARIMA models met the assumptions of normality and independence, except for ARIMA (1,1,0), which exhibited non- normally distributed residuals, as indicated by the Shapiro-Wilk (S-W) and Box-Ljung (B-L) tests. Based on these findings, ARIMA (0,1,1) with drift was selected as a best fit for the area of pearl millet.

The results of fitting ARIMA models without drift are summarized in Table 7. Among the fitted models, ARIMA (3,1,3) exhibited the lowest value for AIC and RMSE, along with significant autoregressive (AR) and moving average (MA) terms. Despite having significant AR and MA terms and the lowest BIC value, ARIMA (2,1,2) displayed higher RMSE and AIC values compared to ARIMA (3,1,3). Residuals met the assumptions of normality and independence, as indicated by the Shapiro-Wilk and Box-Ljung tests in Table 7.

In a similar study, Bhagyashree and Rajarathinam (2009) employed regression models to analyze pearl millet area trends in Gujarat, concluding that nonparametric regression was the most suitable model. Likewise, Dhekale *et al*. (2017) utilized various parametric and nonparametric regression techniques to examine historical patterns in area of tea across key states in India.



**Fig. 1: ACF and PACF plot of original data series for area under pearl millet**

****

**Fig. 2: ACF and PACF plot of differenced data series for area under pearl millet Table 5: Summary of ADF test**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aspect** | **ADF Statistic at Level** | **P value** | **ADF Statistic at First****difference** | **P value** |
| **Area** | -1.75 | 0.67 | -4.61 | <0.01\*\* |

**Table 6: Statistics of fitted ARIMA models with drift parameter for area under pearl millet**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Aspect** | **Model** | **Drift** | **AR(Φ)** | **MA(θ)** | **AIC** | **BIC** | **S-W TEST** | **B-L TEST** | **RMSE** |
| **Φ1** | **Φ2** | **Φ3** | **θ1** | **θ2** | **θ3** |
| **Area** | **(0,1,1)** | **-265.50\*\* [60.64]** | **-** | **-** | **-** | **0.6648\*\* [0.11]** | **-** | **-** | **550.34** | **558.83** | **0.94** | **6.55** | **795.57** |
| (1,1,0) | -263.13[137.66] | -0.39\* [0.16] | - | - | - | - | - | 561.58 | 566.07 | 0.90\*\* | 11.92 | 1076.49 |
| (1,1,1) | -265.25\*\* [60.81] | -0.06[0.22] | - | - | -0.64\*\* [0.14] | - | - | 556.28 | 562.27 | 0.94 | 6.90 | 957.43 |
| (2,1,1) | -258.26\*\* [57.76] | -0.29[0.22] | -0.40\* [0.19] | - | -0.42\* [0.19] | - | - | 554.81 | 562.29 | 0.96 | 3.76 | 903.91 |
| (2,1,2) | -237.15\*\* [77.44] | 0.39\* [0.16] | -0.57\*\* [0.16] | - | 1.33\*\* [0.15] | 1.00\*\* [0.20] | - | 551.92 | 560.9 | 0.97 | 0.48 | 778.45 |

\* Significant at 5% level \*\* Significant at 1% level Values in brackets [ ] indicates standard errors

## Table 7: Statistics of fitted ARIMA models without drift parameter for area under pearl millet

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Aspect** | **Model** | **AR(Φ)** | **MA(θ)** | **AIC** | **BIC** | **S-W TEST** | **B-L TEST** | **RMSE** |
| **Φ1** | **Φ2** | **Φ3** | **θ1** | **θ2** | **θ3** |
| **Area** | (2,1,2) | 0.53\*\* [0.18] | -0.45\* [0.19] | - | -1.34\*\* [0.13] | 1.00\*\* [0.14] | - | 557.48 | 563.97 | 0.96 | 1.55 | 872.35 |
| **(3,1,3)** | **1.36\*\* [0.21]** | **-0.94\*\* [0.24]** | **0.55\*\* [0.16]** | **-2.22\*\* [0.24]** | **2.18\*\* [0.41]** | **-0.88\*\* [0.26]** | **556.81** | **567.28** | **0.95** | **4.16** | **806.20** |
| (3,1,2) | 0.62\*\* [0.18] | -0.53\*\* [0.18] | 0.24[0.19] | -1.37\*\* [0.13] | 1.00\*\* [0.17] | - | 556.89 | 565.87 | 0.95 | 3.20 | 840.56 |
| (1,1,3) | 0.06[0.27] | - | - | -0.82\*\* [0.23] | 0.19[0.27] | 0.58\*\* [0.21] | 557.19 | 564.67 | 0.97 | 4.226 | 887.98 |
| (2,1,3) | 0.30[0.30] | -0.35[0.24] | - | -1.06\*\* [0.31] | 0.60[0.43] | 0.29[0.30] | 557.62 | 566.6 | 0.96 | 2.87 | 857.40 |

\* Significant at 5% level \*\* Significant at 1% level Values in brackets [ ] indicates standard errors

## Table 8: Observed, predicted and per cent forecast error of best fitted ARIMA models for area under pearl millet

|  |  |  |  |
| --- | --- | --- | --- |
| **Years** | **Observed values** | **ARIMA (3,1,3) without drift** | **ARIMA (0,1,1) with drift** |
| **PV** | **PFE** | **PV** | **PFE** |
| 2014-15 | 4530.16 | 7682.21 | 69.58 | 6521.98 | 43.97 |
| 2015-16 | 3911.68 | 7506.13 | 91.89 | 6256.47 | 59.94 |
| 2016-17 | 4313.71 | 6663.63 | 54.48 | 5990.97 | 38.88 |
| 2017-18 | 3971.52 | 6217.66 | 56.56 | 5725.46 | 44.16 |
| 2018-19 | 3916.00 | 6308.26 | 61.09 | 5459.96 | 39.43 |
| 2019-20 | 4498.31 | 6385.20 | 41.95 | 5194.45 | 15.48 |
| 2020-21 | 4602.60 | 6157.77 | 33.79 | 4928.95 | 7.09 |
| 2021-22 | 4452.10 | 5826.94 | 30.88 | 4663.44 | 4.75 |

\*PV = Predicted values, PFE = Per cent forecast error

## Fitting of nonparametric model for area

The nonparametric regression model was fitted for area under pearl millet crop in the Gujarat using PROC LOESS procedure and the results are given in Table 9. The fitted graph for the area of pearl millet, depicted in Fig. 4, illustrates a decreasing trend over time. The optimum bandwidth (or smoothing parameter) was calculated as 0.691. Residuals were found to meet the assumptions of independence but failed to meet the assumption of normality, as indicated by result of the Shapiro-Wilk and the run test. Furthermore, the values of RMSE (772.919) and MAE (558.485) were relatively lower than those of parametric models, suggesting the superiority of nonparametric models.

## Table 9: Trend in area of pearl millet in Gujarat using nonparametric regression model

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Bandwidth** | **Goodness of fit** |
| **RMSE** | **MAE** | **Run Test** | **S-W Test** |
| Area | 0.691 | 772.919 | 558.485 | 0.752 | 0.91\*\* |

\*\* Significant at 1% level

**Fig.3: Fitting of nonparametric model for area under pearl millet, 1980-81 to 2013-14**

|  |  |
| --- | --- |
|  |  |

**Fig. 4: Observed vs Fitted area under pearl millet using ARIMA (0,1,1) with drift and ARIMA (3,1,3) without drift parameter, 1980-81 to 2013-14**

**Table 10: Observed, predicted and per cent forecast error of best fitted models for area under pearl millet**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Years** | **Observed values** | **Cubic** | **Logistic** | **ARIMA (0,1,1)****with drift** | **ETS (A,A,N)** |
| **PV** | **PFE** | **PV** | **PFE** | **PV** | **PFE** | **PV** | **PFE** |
| 2014-15 | 4530.16 | 6090.62 | 34.45 | 5794.65 | 27.91 | 6521.98 | 43.97 | 6414.48 | 41.59 |
| 2015-16 | 3911.68 | 5731.79 | 46.53 | 5385.62 | 37.68 | 6256.47 | 59.94 | 6219.47 | 58.99 |
| 2016-17 | 4313.71 | 5385.47 | 24.85 | 4991.25 | 15.7 | 5990.97 | 38.88 | 6024.45 | 39.66 |
| 2017-18 | 3971.52 | 5053.19 | 27.24 | 4613.14 | 16.15 | 5725.46 | 44.16 | 5829.45 | 46.78 |
| 2018-19 | 3916.00 | 4736.46 | 20.95 | 4252.54 | 8.59 | 5459.96 | 39.43 | 5634.44 | 43.88 |
| 2019-20 | 4498.31 | 4436.82 | -1.37 | 3910.39 | -13.07 | 5194.45 | 15.48 | 5439.42 | 20.92 |
| 2020-21 | 4602.60 | 4155.80 | -9.71 | 3587.32 | -22.06 | 4928.95 | 7.09 | 5244.41 | 13.94 |
| 2021-22 | 4452.10 | 3894.93 | -12.51 | 3283.65 | -26.24 | 4663.44 | 4.75 | 5049.40 | 13.41 |

\*PV = Predicted values, PFE = Per cent forecast error

* 1. **Forecast and Per Cent Forecast Error**

The observed and fitted values of the best-fitted models from each category for the area of pearl millet are depicted in Fig. 5, Fig. 6, Fig. 7 and Fig. 8. The test data from 2014-15 to 2021-22 were utilized to validate these models, with per cent forecast errors calculated from the predicted values and presented in Table 10, alongside the observed values. From 2014-15 to 2018-19, all four best-fitted models tended to overestimate the actual values, with the logistic model consistently exhibiting lower per cent forecast errors during this period. However, in contrast, the cubic and logistic models tended to underestimate the actual values, while the ARIMA (0,1,1) with the drift and ETS (A,A,N) model goes on overestimating the actual values in the years 2019-20 to 2021-22. Notably, the ARIMA (0,1,1) with the drift model demonstrated greater consistency in the last two years, with the lowest per cent forecast error among the best fitted models.

In the analysis for the area of pearl millet, five linear and non-linear models were evaluated. The cubic model (linear) and logistic model (non-linear) were identified as the best fits based on lower RMSE, MAE and higher adjusted R², with the cubic model being slightly superior. Among ARIMA models, ARIMA (0,1,1) with drift was selected as the best fit due to significant parameters and lower RMSE (795.57) compared to RMSE (806.20) of ARIMA (3,1,3) without drift. From ETS models ETS (A,A,N) model emerged best based on lower AIC, BIC and RMSE (946.42) values. The nonparametric model with a bandwidth of 0.691 produced an RMSE of 772.919, indicating a better fit over the non-linear and ARIMA models. However, the cubic model was ultimately identified as the best fit due to its slightly lower RMSE than nonparametric model. However, ARIMA (0,1,1) with drift excel for forecasting purpose based on lowest per cent forecast error in last two testing years and it states that including a drift term improves the model’s fit and predictive ability.

**CONCLUSION**

Comparison among the best-fitted model from linear, non-linear, time series (ARIMA) and nonparametric models based on RMSE values suggests the order of best fit models as follows: cubic > nonparametric > logistic > ARIMA (0,1,1) with drift > ETS (A,A,N).

In conclusion, we can say the area under pearl millet showed declining trend and had not shown any rise in area under cultivation in recent years despite of National Millet Mission (2018) and IYOM (Internation Year of Millets, 2023) initiatives and it is likely that area under cultivation of pearl millet will continue to fall.

|  |  |
| --- | --- |
|  |  |
| **Fig. 5: Observed vs Fitted area under pearl millet using cubic model, 1980-81 to 2013-14** | **Fig. 6: Observed vs Fitted area under pearl millet using logistic model, 1980-81 to 2013-14** |
|  |  |
| **Fig. 7: Observed vs Fitted area under pearl millet using ARIMA (0,1,1) with drift, 1980 -81 to 2013-14** | **Fig. 8: Observed vs Fitted area under pearl millet using ETS (A,A,N) model, 1980-81 to 2013-14** |

# REFERENCES

Anonymous, (1996). Food and Agriculture Organization of the United Nations. Retrieved from: <http://www.fao.org/3/w1808e/w1808e0c.htmI>.

Anonymous, (2022). *Agricultural Statistics at a Glance- 2022*. Department of Agriculture & Farmers Welfare, MoA & FW, Government of India, India. Retrieved from: [https://agricoop.nic.in/Documents/CWWGDATA/Agricultural\_Statistics\_at\_a\_Glance\_2](https://agricoop.nic.in/Documents/CWWGDATA/Agricultural_Statistics_at_a_Glance_2022_0.pdf) [022\_0.pdf](https://agricoop.nic.in/Documents/CWWGDATA/Agricultural_Statistics_at_a_Glance_2022_0.pdf)

Anonymous, (2023). Food and Agriculture Organization of the United Nations. *International Year of Millets 2023*. Retrieved from <https://www.fao.org/millets-2023/en>

Ashfaq, M., Bashir, M. K., & Khan, A. A. (2014). New approach to forecasting agro-based statistical models. *Advances in Environmental Biology*, 8(1), 252-257

Bhagyashree, S. D., & Rajarathinam, A. (2009). Statistical modelling on area, production and productivity trends of bajra crop grown in Gujarat state. *International Journal of Agricultural Statistics Science,* 27 (1-2), 291-296.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Revised Second Edition, Holden Day, San Francisco.

Chaudhari, M. R., Patel, M. M., & Patel, R. R. (2023). Growth rates of cost and returns in summer pearl millet cultivation in Gujarat. *International Journal of Statistics and Applied Mathematics*, 8(6S), 123-130.

 Cleveland, W. S., & Devlin, S. J. (1988). *Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting*. Journal of the American Statistical Association, 83(403), 596-610.

Cromwell, J. B., Labys, C., & Terraza. M. (1994). *Univariate Test for Time Series models*. Sage Publications, Cromwell.

Dhekale, B. S., Sahu, P. K., Vishwajith, K. P., Mishra, P., & Narsimhaiah, L. (2017). Application of parametric and nonparametric regression models for area, production and productivity trends of tea (*Camellia sinensis*) in India. *Indian Journal of Ecology,* 44 (2), 192-200.

Dickey, D. A., and W. A. Fuller. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica.* 49 (1), 1057–1072.

Dodiya, R. D., and Barad, A. H. (2022). Effectiveness of biopesticides against *Spodoptera litura* infesting groundnut under field condition. The Pharma Innovation Journal, 11(8), 1601.

Dodiya, R. D., Barad, A. H., and Italiya, J. V. (2024). Efficacy of Insecticides against *Spodoptera litura* (Fabricius) Infesting Groundnut (Arachis Hypogaea Linnaeus). Pesticide Research Journal, 36(2), 125-129.

Dodiya, R. D., Barad, A. H., Italiya, J. V., and Prajapati, H. N. (2024). Impact of Weather Parameters on Population Dynamics of Tobacco Leaf Eating Caterpillar, *Spodoptera litura* (F.) Infesting Groundnut. *Environment and Ecology*, 42(1A), 301-306.

Dodiya, R. D., Barad, A. H., Italiya, J. V., Pathan, N. P., and Bhatt, N. A. (2023). Host Suitability of Tobacco Leaf Eating Caterpillar (Spodoptera litura F.) under Laboratory Condition. International Journal of Bio-resource and Stress Management, 14(Aug, 8), 1186-1195.

Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II*. International Journal of Forecasting*, 22(4), 637-666.

Hurvich, C. M., Simonoff, J. S., & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2), 271-293. Retrieved from <https://www.jstor.org/stable/2985940>

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice. (2nd ed.)*.

McGraw-Hill, New York.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis*. John Wiley & Sons, Inc.

Nath, S. K. (2008). *Manual on area and crop production statistics*. Central Statistical Organisation.

OTexts. <https://otexts.org/fpp2/>

Panwar, S., & Dimri, A. K. (2018). Trend analysis of production and productivity of major crops and its sustainability: A case study of Haryana. *Indian Journal of Agricultural Research*, *52*(5), 571-575.

SAS Institute Inc. (2021). *The LOESS Procedure*. SAS/STAT User's Guide. Retrieved from <https://documentation.sas.com/doc/en/statug/latest/statug_loess_syntax01.htm>

Satyavathi, C. T., Ambawat, S., Khandelwal, V., & Srivastava, R. K. (2021). Pearl millet: a climate-resilient nutricereal for mitigating hidden hunger and provide nutritional security. *Frontiers in Plant Science*, *12*, 659938.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika,* 52, 591-611.

Sidney S., & Castellan, N. Jr. (1988). *Nonparametric statistics for the behavioral sciences.*

Yonar, H., Yonar, A., Mishra, P., Abotaleb, M., Alkhatib, A., Makarovskikh, T. and Cam, M. 2022. Modeling and forecasting of milk production in different breeds in Turkey. *Indian Journal of Animal Science*, 92(1):105-111.