

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2025): 4.49

© 2026 Stats & Maths

Maths 2026; SP-11(3): 13-19

www.mathsjournal.com

Received: 09-01-2026

Accepted: 13-02-2026

Pooja

Research Scholar, Department of
Agricultural Statistics, Shri
Durga Ji Post Graduate College,
Chandeshwar, Azamgarh, Uttar
Pradesh, India

Rajendra Prasad Kaushal

Assistant Professor, Department
of Agricultural Statistics, Shri
Durga Ji Post Graduate College,
Chandeshwar, Azamgarh, Uttar
Pradesh, India

Abhay Kumar Yadav

Research Scholar, Department of
Agronomy, Shri Durga Ji Post
Graduate College, Chandeshwar,
Azamgarh, Uttar Pradesh, India

Forecasting pearl millet area and production in India using regression models

Pooja, Rajendra Prasad Kaushal and Abhay Kumar Yadav

DOI: <https://www.doi.org/10.22271/math.2026.v11.i3Sa.2279>

Abstract

Estimating crop production and forecasting crop area are the key to taking the necessary actions to preserve food security and solve environmental issues. Crop production forecasting is the skill of predicting a crop's future values before it is harvested. The primary goal of this study was to get a precise picture of India's pearl millet output and area. One of the four major cereals rice, maize, sorghum, and millets grown in tropical semi-arid regions of the world, mostly in Africa and Asia, is pearl millet (*Pennisetum glaucum*), sometimes referred to as bajra. In northwest India, pearl millet is a significant cereal crop. The key to taking the required steps to maintain food security and address environmental challenges is to estimate crop production and forecast crop area. The ability to anticipate the future values of cultivated crops prior to their harvest is known as crop production forecasting. The primary goal of this study was to provide a detailed picture of the region and production of pearl millet in India. India is the world's top producer of pearl millet, accounting for over 40.51 percent of all millet produced worldwide. Rajasthan is India's top producer of pearl millet, accounting for over 44% of the nation's total millet production. Using linear, logarithmic, quadratic, cubic, power, and exponential models, regression modeling was used for the area and production of pearl millet in India over a 71-year period (1951-2021). The results showed that the cubic model outperformed other models in terms of accurately forecasting India's pearl millet production and area.

Keywords: Pearl millet, area, production, regression study, simulating, forecasting

Introduction

Bajra, or pearl millet (*Pennisetum glaucum*), is a grain crop that is produced mostly in Asia and Africa in tropical semi-arid regions of the world. Bajra can be cultivated in regions where other cereal crops, like wheat or maize, would not thrive since it is well suited to production systems with little rainfall (200-600 mm), low soil fertility, and high temperatures. It is an important fact that our country, India, is termed as the original home of nutria cereals (shree Anna millet). Pearl millet requires minimal fertilizer and pesticide inputs and is resistant to drought and other harsh weather. India is the world's top producer of pearl millet, accounting for over 44% of global production. Nearly 51% of India's gross cultivated area is made up of pearl millet. In India, pearl millet crops were grown on 7.41 million hectares of land in 2019-20. With almost 28 million tons of pearl millet produced, or 51% of India's total production, Rajasthan leads the country in this regard. The world's production of pearl millet has increased over the past 15 years mostly due to the availability of improved genetics, the adoption of hybrids in India, and the extension of the area under pearl millet cultivation in West Africa. The production of pearl millet is hampered by a number of biotic and abiotic stressors, which significantly reduce yields. India is the world's greatest producer of pearl millet, covering 7.14 million hectares and producing 8.06 million tons with an average yield of 1132 kg/ha (<http://www.aicpmip.res.in/pmnews 2017. pdf>) and continues to play a critical role in food and livelihood security of millions of people in India. It is the main source of calorie intake for the 500 million people of Asia and Africa. It has a brief life cycle, tiny seeds, and a C4 panicoid, highly cross-pollinated annual crop. Pearl millet, a warm-season coarse grain, may flourish in barren soils with little water-holding capacity, limited rainfall, high temperatures, and droughty conditions when other cereals would not. Food, feed, fodder

Corresponding Author:**Abhay Kumar Yadav**

Research Scholar, Department of
Agronomy, Shri Durga Ji Post
Graduate College, Chandeshwar,
Azamgarh, Uttar Pradesh, India

forage, brewing, biofuel, building materials, and fuel for cooking in dryland environments are just a few of the many use for this crop. In data science, regression is a crucial technique for predicting a range of continuous targets, which helps to prevent this work from being overly demanding and

from being seen as comprehensive in the actual world. Regression-related studies are still thoroughly assessed by scholars today since they receive the most attention. The area and production of pearl millet were fitted and forecasted in the current study using regression modeling.

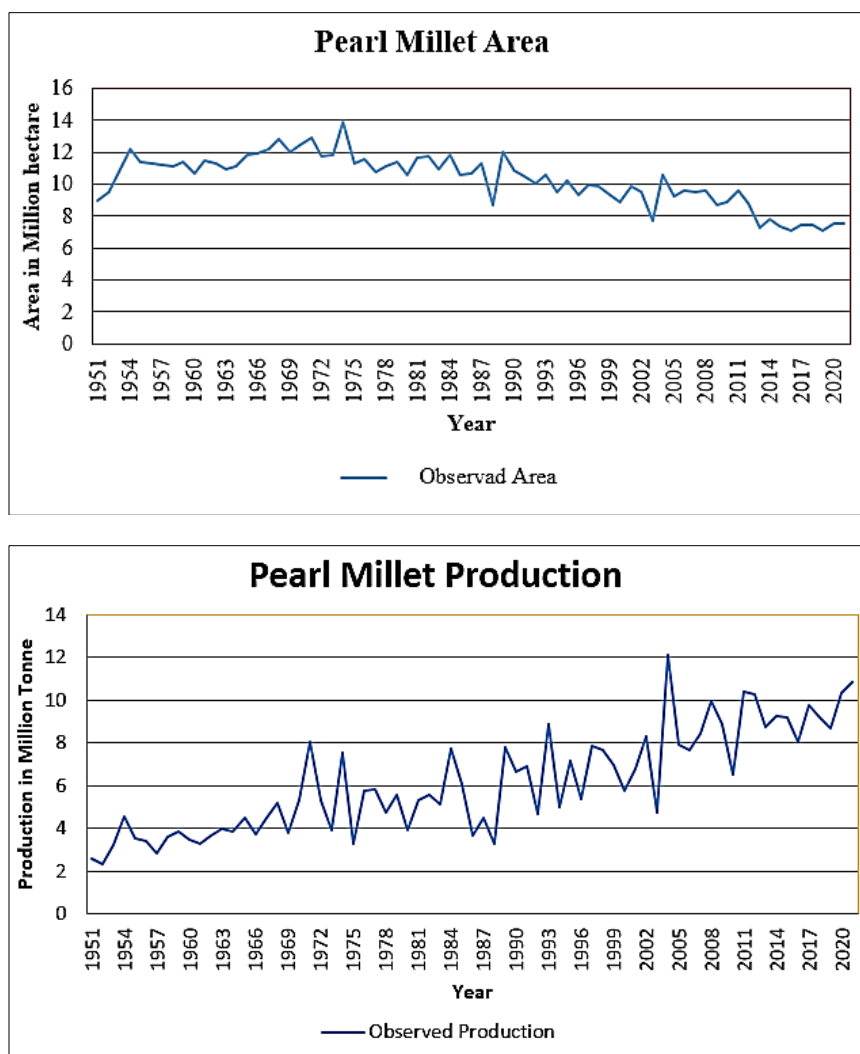


Fig 1: Area and production observed

Data and Methodology

Collection of data: The Agricultural Statistics at a Glance 2022, Government of India (GOI), provided the secondary

data needed for the study on the area, production, and yield of pearl millet in India over a 71-year period (1951-2021).

Table 1: Yearly data of Pearl Millet Area and Production Area - Million Hectares Production - Million Tonnes

Year	Area	Production	Year	Area	Production	Year	Area	Production
1951	9.02	2.6	1975	3.27	26.13	1999	9.3	6.96
1952	9.52	2.35	1976	5.74	30.41	2000	8.9	5.78
1953	10.77	3.19	1977	5.85	28.88	2001	9.83	6.76
1954	12.2	4.55	1978	4.73	30.02	2002	9.53	8.28
1955	11.37	3.52	1979	5.57	30.44	2003	7.74	4.72
1956	11.34	3.43	1980	3.95	26.97	2004	10.61	12.11
1957	11.25	2.87	1981	5.34	29.02	2005	9.23	7.93
1958	11.17	3.62	1982	5.54	31.09	2006	9.58	7.68
1959	11.43	3.87	1983	5.13	27.75	2007	9.51	8.42
1960	10.7	3.49	1984	7.72	33.90	2008	9.57	9.97
1961	11.47	3.28	1985	6.05	31.17	2009	8.75	8.89
1962	11.28	3.65	1986	3.66	26.20	2010	8.9	6.51
1963	10.96	3.96	1987	4.51	26.83	2011	9.61	10.37
1964	11.1	3.88	1988	3.3	26.36	2012	8.78	10.28
1965	11.83	4.52	1989	7.78	31.47	2013	7.3	8.74
1966	11.97	3.75	1990	6.65	34.76	2014	7.81	9.25
1967	12.24	4.47	1991	6.89	32.70	2015	7.32	9.18
1968	12.81	5.19	1992	4.67	25.99	2016	7.13	8.07
1969	12.05	3.8	1993	8.88	36.59	2017	7.46	9.73

1970	12.49	5.33	1994	4.97	30.82	2018	7.48	9.21
1971	12.91	8.03	1995	7.16	29.88	2019	7.11	8.66
1972	11.77	5.32	1996	5.38	29.03	2020	7.54	10.36
1973	11.82	3.93	1997	7.87	34.11	2021	7.57*	10.86*
1974	13.93	7.52	1998	7.64	30.40			

Source of Data: Directorate of Economics & Statistics, DAC&FW, New Delhi. *5th advanced estimate.

Regression Study

Regression analysis uses data analysis to determine the link between response variables and explanatory variables. The variables chosen for examination in this study are production, area, and year. For this study, year is acknowledged as an explanatory variable, while pearl millet production and area are regarded as response variables.

The different steps undertaken for conducting this study is illustrated as a flow chart in Fig. 2. Data on pearl millet production and area was gathered from Agricultural Statistics

at a Glance 2022, Government of India, during a 71-year period (1951-2021). Simple regression analysis was used to model the area and production of pearl millet. The association between an independent and a dependent variable was used to forecast the response variable's future observation. As a result, the study was carried out with the presumption that the projected future values for pearl millet production and area would continue the previous pattern. The primary advantage of this strategy was described as its simplicity and capacity to forecast any available data.

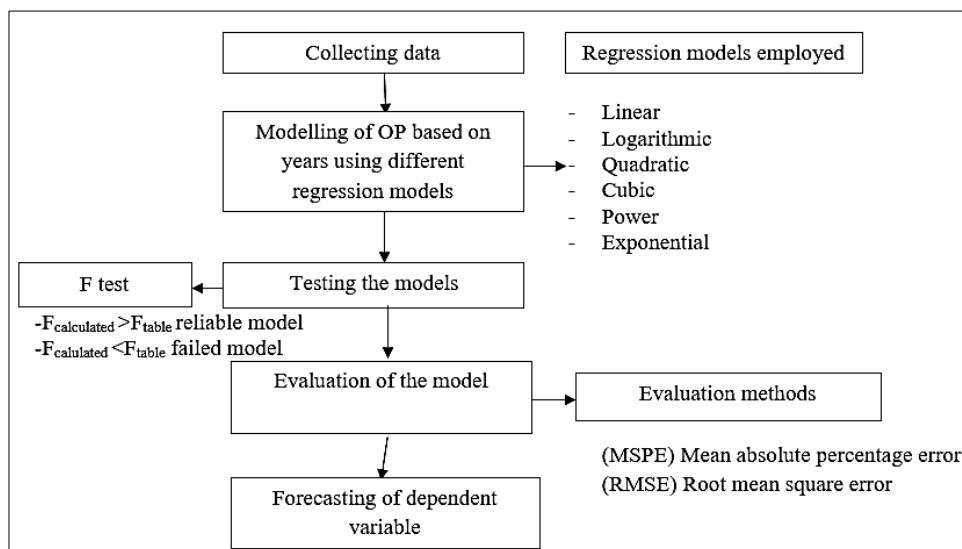


Fig 2: The study's flow chart

Regression analysis methods, such as linear, logarithmic, power, exponential, quadratic, and cubic regressions, were used in this study to create appropriate models for area and production. It is observed that a linear model has straight-line relationships, while a non-linear model has curved ones. The data was divided into training and testing datasets. While the

testing dataset comprises data from the last five years (2017-2021) used to assess the fitted regression model, the training dataset contains data over 66 years (1951-2016) used to develop the model.

The mathematical expression for the regression models both linear and non-linear used in this study is as follows:

$$\text{Linear regression model: } Y_t = a + bt + \varepsilon_t \quad \text{--- (1)}$$

$$\text{Quadratic regression model: } Y_t = a + bt + ct^2 + \varepsilon_t \quad \text{--- (2)}$$

$$\text{Cubic regression model: } Y_t = a + bt + ct^2 + dt^3 + \varepsilon_t \quad \text{--- (3)}$$

$$\text{Logarithmic regression model: } Y_t = a + b \ln(t) + \varepsilon_t \quad \text{--- (4)}$$

$$\text{Power regression model: } Y_t = at^b + \varepsilon_t \quad \text{--- (5)}$$

$$\text{Exponential regression model: } Y_t = ae^{bt} + \varepsilon_t \quad \text{--- (6)}$$

Were;

' Y_t ' indicates the dependent variable i.e., area or production, ' t ' denotes the independent variables, time in years, ' a ' is the intercept, ' b ', ' c ' and ' d ' are representing the regression coefficients and ' ε_t ' is the error term, $\varepsilon \sim N(0, \sigma^2)$.

By calculating coefficient of determination (R^2) and performing a F test, the fitted model is verified. R^2 was employed to determine how much a response variable

changed overall as a result of the regression model. The efficacy of model fitted was indicated by a higher R^2 value. Acceptance of the fitted model and the model's advanced independent variable requires, F test was performed. The fitted model's significance was determined using the F test. The model is accepted or rejected depending on whether the F value if the F value obtained after undergoing F test is higher than the F table value (from F distribution table), the model is accepted or otherwise rejected.

By examining the performance of regression models developed for the data, the optimal forecasting model was

found. The chosen criterion for verifying was the lowest value for RMSE and MAPE. Fitted models' work (Eqs. 7 and 8). Better predictions are indicated by smaller RMSE and MAPE values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (7)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (8)$$

Where, y_t is the observed value and $\hat{y}_t$ is the predicated value, n is the total number of measurements.

Table 2: Linear and Non-linear models of Pearl Millet Area in India

Area Model	Parameter				Criteria			
	a	B	c	d	R^2	Adj. R^2	RMSE	MAPE
Linear	126.915*	-0.058*			0.5754*	0.569*	1.0337	0.080
Logarithmic	893.435*	-116.29*			0.572*	0.566*	1.0374	0.080
Quadratic	-7164.9*	7.285	-0.001		0.767*	0.760*	0.7658	0.026
Cubic	-329641*	494.472*	-0.247*	4.12*	0.798*	0.789*	0.7135	0.053
Power	93.425*	-11.997*			0.581*	0.575*	1.0898	0.085
Exponential	14.351*	-0.006*			0.584*	0.578*	1.086	0.085

Table 3: Linear and Non-linear models of Pearl Millet Production in India

Production Model	Parameter				Criteria			
	a	b	C	d	R^2	Adj. R^2	RMSE	MAPE
Linear	-187.284*	0.097*			0.702*	0.697*	1.301	0.166
Logarithmic	-1461.91*	193.324*			0.701*	0.697*	1.302	0.166
Quadratic	2155.819*	-2.262	0.0005		0.710*	0.702*	1.282	0.165
Cubic	-146883*	222.899*	-0.113*	1.9*	0.713*	0.700*	1.275	0.162
Power	-247.362*	32.803*			0.716*	0.712*	1.288	0.162
Exponential	-31.061*	0.0165*			0.716*	0.712*	1.288	0.162

1. Source: Secondary data collected from Directorate of Economics & Statistics, DAC&FW, New Delhi and <https://eands.dacnet.nic.in/>
2. When a model's criteria value is marked up it means that it performs better than other models in that particular Criteria.
3. *Show significance at the 5% probability level.

Table 4: Findings from the Validation Tests of the Developed Models

Validation test	Regression type					
	Linear	Logarithmic	Quadratic	Cubic	Power	Exponential
t (area)	93.51	92.34	111.92	88.08	95.74	97.04
t(production)	162.39	161.82	83.47	55.61	174.34	174.18

Table 5: MAPE Model Evaluation Criteria In accordance with Lewis (1982)

MAPE	Evaluation
$MAPE \leq 10\%$	High accuracy prediction
$10\% < MAPE \leq 20\%$	Good prediction
$20\% < MAPE \leq 50\%$	Reasonable prediction
$MAPE > 50\%$	Inaccurate prediction

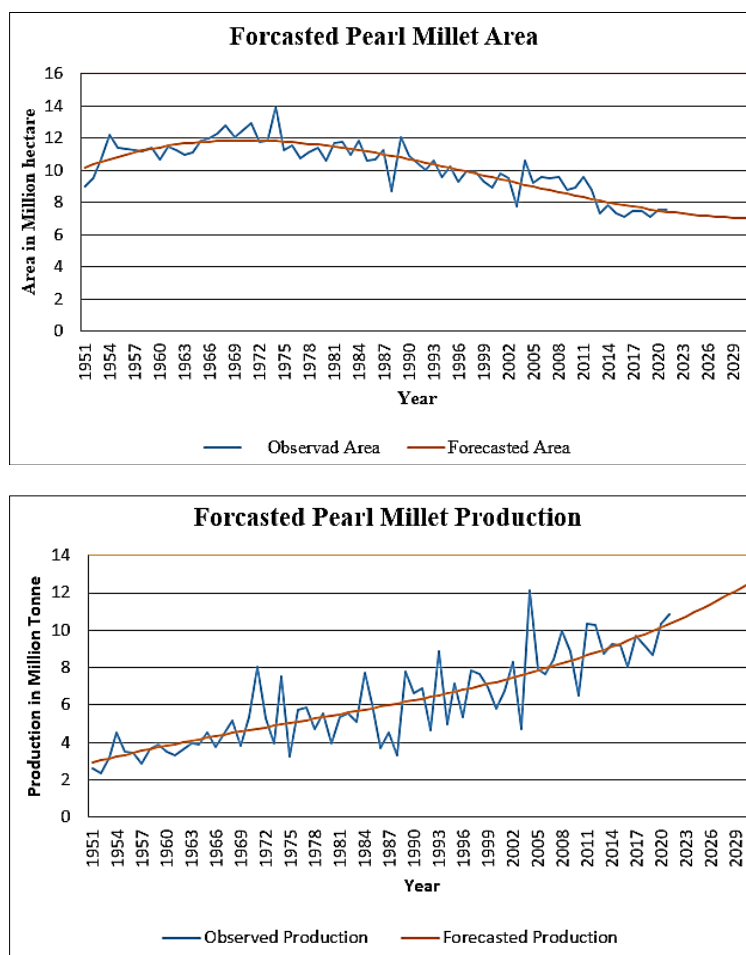
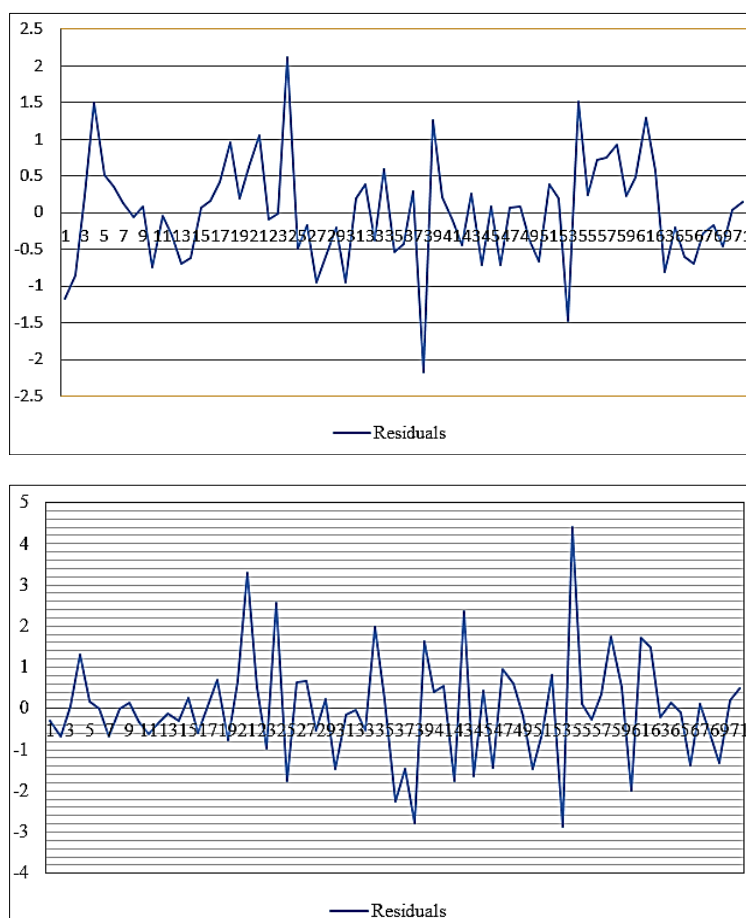
Modeling the production and area of pearl millet

The regression analysis's findings using various techniques are explained in Table 2. The R^2 values lie above 0.969 and 0.870 for the fitted regression models, shown a significant degree of correlation between area and years, production and year, respectively. Additionally, the determination

coefficients showed that only a small percentage (0.011-0.010%) of the variance in area and production may be attributed to factors other than the predictors (see Eqns. 1 to 6). The cubic model is the most suitable for predicting the area and production of pearl millet, according to the model evaluation criteria (Eqns. 9 and 10).

$$\text{Forecasted Cubic model is area } \hat{Y}_{Area} = -329641 + 494.472t - 0.247t^2 + 4.12t^3 \quad \text{-- (9)}$$

$$\text{Forecasted Cubic model is Production } \hat{Y}_{Production} = -146883 + 222.899t - 0.113t^2 + 1.9t^3 \quad \text{-- (10)}$$

**Fig 3:** Actual and Predicted production and area values**Fig 4:** Area and production residuals predicted

The above describes the fitted models' validation outcomes. The calculated value of F is higher than the F table which confirmed that Eqns. (7) to (8) are significant at 95% confidence level. Therefore, it was determined that the explanatory variables used in this study to create the regression models for pearl millet production and area are significant.

Forecasting Pearl Millet

The cubic model built for area and production yielded MAPE values of 0.053 and 0.162, respectively (see Table 2, 3). These results showed that pearl millet production and area could be effectively estimated using the cubic model, which was selected as the best forming model.

The MAPE values of the cubic models showed that they are both highly accurate prediction models for the area and production of pearl millet (see Table 2, 3). Furthermore, it is

clear from Fig. 3 that the actual and expected values for the area and output of pearl millet are very close to each other.

Table 6: Pearl Millet Area and Production Forecasts Using the Cubic Model

Year	Forecasted Area	Forecasted Production
2022	7.366	10.55
2023	7.30	10.76
2024	7.25	10.96
2025	7.20	11.18
2026	7.16	11.14
2027	7.12	11.62
2028	7.09	11.85
2029	7.07	12.09
2030	7.05	12.33
2031	7.04	12.58

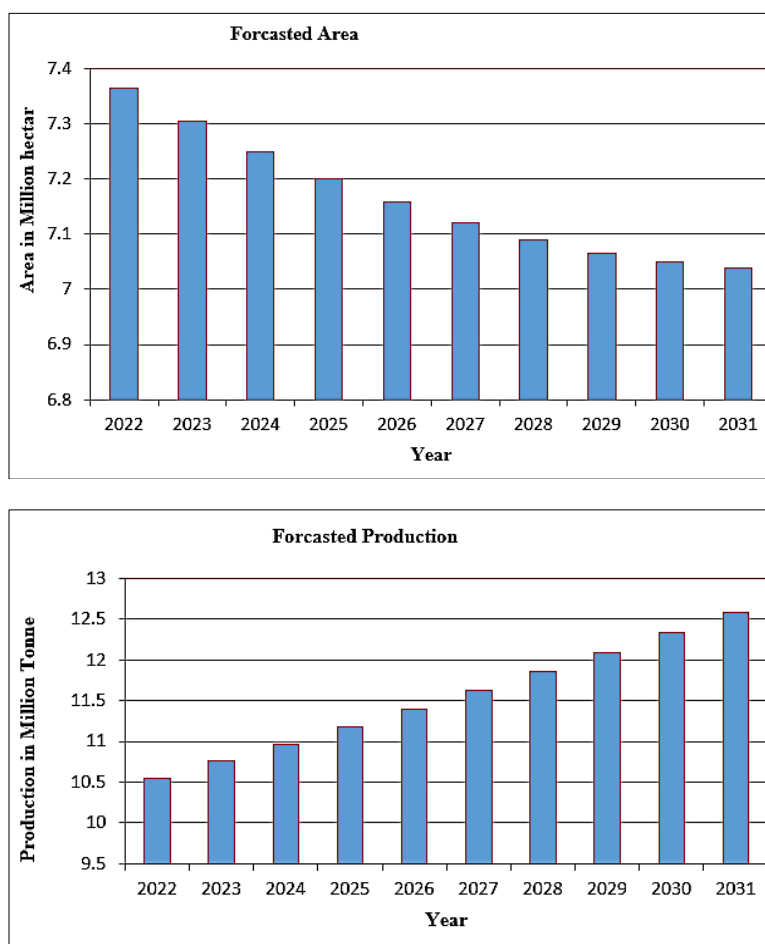


Fig 5: Bar graph to show production and area forecasts for 10 years

The area and production of pearl millet predicting using cubic model for next 10 years are given in Table no.6 and illustrated in Fig. 3 and Fig. 5. According to the findings, pearl millet production and area might reach 12.58 million tonnes and 7.04 million hectares in 2031.

Conclusions

Regression modeling was used in this study to forecast pearl millet production and area, which are important for farmers' decision-making and for the government's agricultural policy-making. The study's primary findings are as follows:

- According to the study's findings, the cubic regression model performed better than any other model.
- The determination coefficients for the pearl millet production and area training data sets are 0.870 and

0.969, respectively, indicating the good reliability of the fitted models.

- The cubic model fitted for the area and production of pearl millet provided a statistically significant result according to the F test. The best cubic model was found to have great accuracy based on the small values of MAPE and RMSE for the fitted models.

In order to model the area and production of pearl millet, the researchers will need to apply other models such as artificial intelligence, grey theory forecasting, hybrid approach, multiple linear regression, and other econometric methods in the future.

Future research findings can be evaluated by comparing them to the current study.

References

1. Nutri Cereals Portal. Available from: <https://nutricereals.dac.gov.in/>
2. USDA Foreign Agricultural Service. Crop Explorer: Commodity View (Millet). Available from: https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=0459100&sel_year=2018&rankby=Production
3. Altman N, Krzywinski M. Crafting 10 years of statistics explanations: Points of significance. *Annu Rev Stat Appl.* 2025;12(1):69-87.
4. India. Ministry of Agriculture and Farmers Welfare. Department of Agriculture. Agricultural statistics at a glance 2022. New Delhi: Government of India; 2022. p.72.
5. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci.* 2021;7:e623.
6. Yadav RS, Mehta V, Tiwari A. An application of time series ARIMA forecasting model for predicting nutri cereals area in India. *Pharma Innov J.* 2022;11(3):1260-1267.
7. Yadav RS, Mehta V, Krishna G. Regression models for forecasting nutri cereals area and production in India. *Pharma Innov J.* 2023;SP-11(3):1260-1267.
8. Dhankar N, Routroy S, Sharma SK. A multiple regression model to analyze the impact of endogenous factors on pearl millet and stover yield in India. *J Model Manag.* 2026;21(2):584-612.
9. Kutty VK. Regression model. In: *Global millets production for a sustainable future.* 2025. p.231.