International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Predictive Model for the Production of Cotton Crop of the State Telangana before and After Division

T. Rajani Devi[1], Dr. V V Hara Gopal[2]

[1]Department of Mathematics & Statistics, Telangana Mahila Viswavidyalayam (Women's University), Hyderabad-95, India

Email Id: *trdevi11[at]yahoo.com*

[2]Department of Statistics, Osmania University, Hyderabad-7, India

Email Id: *haragopal_vajja[at]yahoo.com*

Abstract: ***Instability is an inherent characteristic of agriculture everywhere, being dependent on Weather conditions, area, yield and production of Crops are liable to substantial variations from year to year. Agricultural production has always involved the exploitation of resources such as soil, water, and energy. Increasing production to feed a growing world population while at the same time conserving resources for future generations has led to a search for 'sustainable' agricultural methods. This Study will be a step in identifying the Possible error and see that how much of Volatility risk is involved fitting the proper distribution of the Secondary data. Also, any study is not simply to fit the kind of Distribution fits well or not, but the estimation of the Cotton production is also evaluated by Deep learning methods such as Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). The data is fewer after division of the state various curves are fitted and compared to predict the production of Cotton.***

Keywords: Distribution fitting, Goodness of fit, Kolmogorav-Smirnov test, Anderson-Darling test, Chi square test, MLP, RNN, LSTM, Curve fitting.

1. Introduction

Cotton is one of the most important Fibre and cash crop of India and plays a dominant role in the industrial and agricultural economy of the country India is the largest producer of cotton globally. It is a crop that holds significant importance for the Indian economy and the livelihood of the Indian cotton farmers. The Southern Zone (which comprises of states like Telangana, Andhra Pradesh, Karnataka, and Tamil Nadu) is the second biggest producer of cotton, producing about 30% of the nation's cotton, with Telangana producing the largest in the Southern Zone and the third largest in the country, contributing 6.587 million bales (bales of 170 kg each). Telangana produces about 53 lakh bales of cotton and covers 18.27 lakh hectares in India. Telangana is the largest producer of paddy and the second-largest producer of cotton in India.

In this Paper various distributions has been fitted for the Cotton Production data for the years 1990- 91 to 2014-15 and 2014-15 to 2019-20 i.e. before and after division of the state Telangana. Subsequently the Optimal Parameters. Here all the Probability distributions are ranked with each Goodness of fit Test. Applied Goodness of Fit Tests are Kolmogorov Smirnov, Anderson Darling and Chi Square Test for Goodness of Fit. The Production Data before and after division is assessed with twenty different Probability distributions . Also, any study is not simply to fit the kind of Distribution fits well or not, but the estimation of the total Food grains are also evaluated by Deep learning methods such as MLP, RNN and LSTM. The data is fewer after bifurcation various curves are fitted and using the best fit curve predictions are given for Cotton Production.

2. Data Source and Description of the Data

Here the Secondary Data is collected from the Source www.rbi.org.in. The Data for Cotton Production is collected for the years 1990-91 to 2014-15 and 2014-15 to 2019-20 before and after division of the state Telangana

2.1 Distribution fitting for Cotton Production before division of Telangana

In this section the behaviour of the underlying Production data using the best-fitted distribution for a dataset is described. Usually, more than one distribution would be of interest in the matching process when fitting the data with a distribution. Raw data is gathered for the years 1990-91 to 2013-14 in order to discover the distributions that would be fitted onto the data.

For the data of Cotton production 20 distributions are fitted and the corresponding Parameters are shown in the Table 2.1below.The Parameters are obtained using Maximum Likelihood Estimation (MLE) with Iterative Parameter Algorithm.

Table 2.1: Fitted Probability Distributions with Parameters for Cotton Crop before Bifurcation

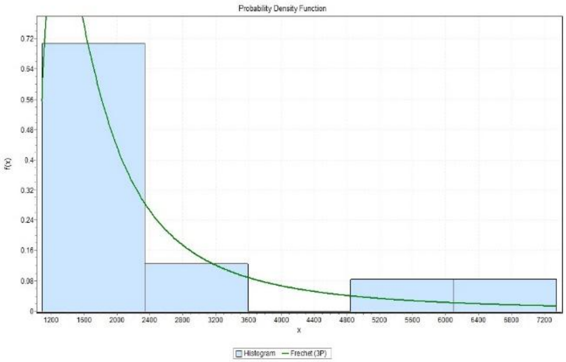
	Distribution	Parameters
1	Fatigue Life	a=0.58 b=2.2E+3
2	Fatigue Life (3P)	a=1.3 b=8.3E+2 g=1.0E+3
3	Frechet	a=2.1 b=1.6E+3
4	Frechet (3P)	a=1.3 b=7.6E+2 g=8.0E+2
5	Gamma	a=2.0 b=1.3E+3
6	Gen. Extreme Value	k=0.4 s=7.6E+2 m=1.7E+3
7	Gen. Pareto	k=0.25 s=1.2E+3 m=9.9E+2
8	Inv. Gaussian	l=5.3E+3 m=2.6E+3
9	Inv. Gaussian (3P)	l=8.7E+2 m=1.6E+3 g=9.5E+2
10	Levy (2P)	s=3.9E+2 g=1.0E+3
11	Log-Gamma	a=1.8E+2 b=0.04
12	Log-Logistic	a=2.8 b=2.0E+3
13	Log-Logistic (3P)	a=1.2 b=7.4E+2 g=1.1E+3
14	Log-Pearson 3	a=5.0 b=0.26 g=6.4
15	Lognormal	s=0.56 m=7.7
16	Lognormal (3P)	s=1.3 m=6.7 g=1.0E+3
17	Pearson 5	a=3.9 b=7.4E+3
18	Pearson 5 (3P)	a=1.5 b=1.1E+3 g=8.4E+2
19	Pearson 6	a_1=1.3E+2 a_2=4.0 b=60.0
20	Pearson 6 (4P)	a_1=38.0 a_2=1.5
		b=29.0 g=8.6E+2

After determining the probability density function to be used to fit the data, goodness of fit (GOF) tests would be performed to quantitatively select the best fitting distribution. The results of evaluation of Probability density functions and their rankings are displayed in the following Table.

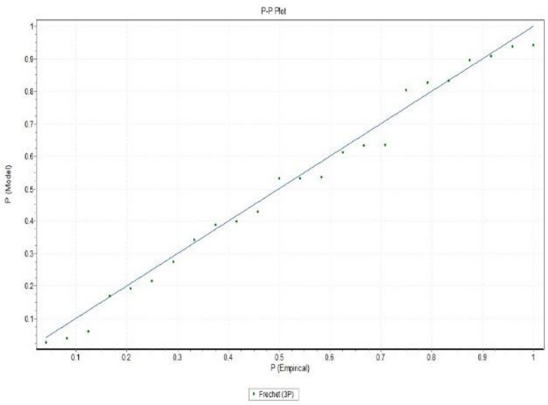
Table 2.2: Goodness of fit summary

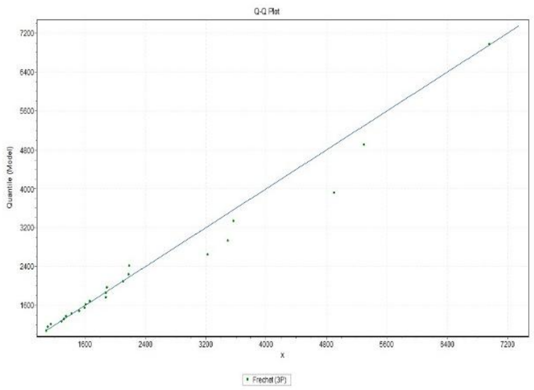
#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Fatigue Life	0.22	20	1.1	19	1.7	14
2	Fatigue Life (3P)	0.1	7	0.24	6	0.71	6
3	Frechet	0.11	8	0.43	10	1.7	15
4	Frechet (3P)	0.1	4	0.23	1	0.01	4
5	Gamma	0.21	19	1.2	20	5	20
6	Gen. Extreme Value	0.12	10	0.44	11	0.87	8
7	Gen. Pareto	0.12	9	0.33	8	1.4	11
8	Inv. Gaussian	0.18	15	0.82	15	1.6	12
9	Inv. Gaussian (3P)	0.09	2	0.24	4	1.40E-04	1
10	Levy (2P)	0.2	17	1.1	18	1.1	10
11	Log-Gamma	0.19	16	0.83	16	2.3	16
12	Log-Logistic	0.16	13	0.78	14	1.6	13
13	Log-Logistic (3P)	0.09	3	0.29	7	0.71	7
14	Log-Pearson 3	0.14	11	0.43	9	0.91	9
15	Lognormal	0.2	18	0.97	17	2.4	17
16	Lognormal (3P)	0.09	1	0.24	5	0.71	5
17	Pearson 5	0.16	14	0.68	13	2.6	19
18	Pearson 5 (3P)	0.1	5	0.23	3	0	2
19	Pearson 6	0.16	12	0.67	12	2.6	18
20	Pearson 6 (4P)	0.1	6	0.23	2	0	3

From the table above it is clear that Frechet (3p) distribution ranked 1 with the Anderson Darling test while with Kolmogorov Smirnov test Log normal (3p) secures rank 1 Chi Square fits the data well and Inv. Guassian (3p) distribution ranks first .Among all these three tests the best one can choose is Anderson Darling test and automatically say that Frechet (3p) is the best distribution represents the data considered. Here are the visualisations of pdf, P-P Plot and Q-Q Plot for the best fit Frechet (3p) distribution.



International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942





2.2 Distribution fitting for Cotton Production after division of Telangana

In this section Cotton production data is gathered for the years 2013-14 to 2019-20 in order to discover the distributions that would be fitted onto the data. 20 distributions are fitted and the corresponding Parameters are shown in the Table 2.3 below. The Parameters are obtained using Maximum Likelihood Estimation (MLE) with Iterative Parameter Algorithm.

Table 2.3: Fitted Probability distributions with Parameters for Cotton Crop after bifurcatio

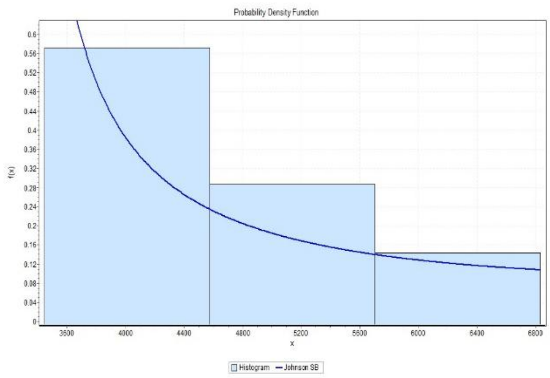
#	Distribution	Parameters
1	Erlang	m=13 b=3.5E+2
2	Error	k=2.3 s=1.3E+3 m=4.6E+3
3	Fatigue Life (3P)	a=1.5 b=5.7E+2 g=3.4E+3
4	Frechet (3P)	a=1.2 b=5.8E+2 g=3.2E+3
5	Gamma	a=13.0 b=3.5E+
6	Gen. Extreme Value	k=0.27 s=7.7E+2 m=3.9E+3
7	Gen. Pareto	k=0.04 s=1.4E+3 m=3.2E+3
8	Gumbel Max	s=9.9E+2 m=4.0E+3
9	Inv. Gaussian	l=6.1E+4 m=4.6E+3
10	Inv. Gaussian (3P)	l=7.2E+2 m=1.3E+3 g=3.3E+3
11	Johnson SB	g=0.76 d=0.49
		l=4.6E+3 x=3.3E+3
12	Levy (2P)	s=2.8E+2 g=3.4E+3
13	Log-Gamma	a=1.0E+3 b=0.01
14	Log-Pearson 3	a=8.0 b=0.09 g=7.7
15	Lognormal (3P)	s=1.4 m=6.4 g=3.4E+3
16	Nakagami	m=3.1 W=2.3E+7
17	Normal	s=1.3E+3 m=4.6E+3
18	Pearson 5 (3P)	a=1.4 b=8.4E+2 g=3.2E+3
19	Pearson 6	a_1=3.9E+2 a_2=19.0 b=2.2E+2
20	Rice	n=4.5E+3 s=1.2E+3

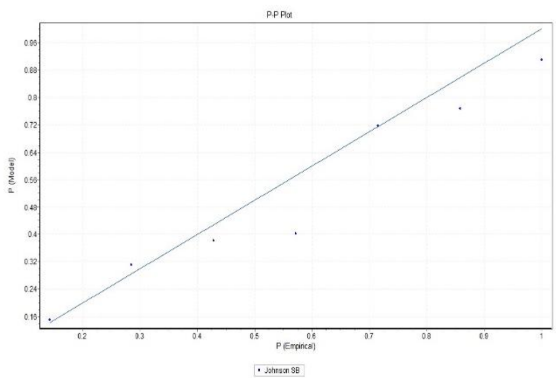
After determining the probability density function to be used to fit the data, goodness of fit (GOF) tests would be performed to quantitatively select the best fitting distribution. The results of evaluation of Probability density functions and their rankings are displayed in the following Table.

Table 2.4: Goodness of fit summary

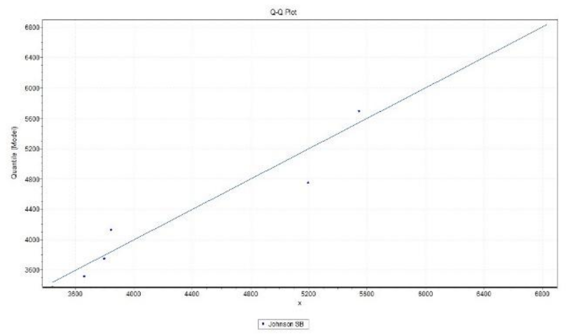
Distribution	Kolmogorov Smirnov		Anderson Darling	
	Statistic	Rank	Statistic	Rank
Erlang	0.26	10	0.44	13
Error	0.29	17	0.49	17
Fatigue Life (3P)	0.22	5	0.34	6
Frechet (3P)	0.22	4	0.31	2
Gamma	0.28	16	0.45	15
Gen. Extreme Value	0.23	8	0.39	8
Gen. Pareto	0.19	2	0.35	7
Gumbel Max	0.28	15	0.43	12
Inv. Gaussian	0.28	14	0.43	11
Inv. Gaussian (3P)	0.23	9	0.34	5
Johnson SB	0.17	1	0.28	1
Levy (2P)	0.22	6	0.42	10
Log-Gamma	0.29	18	0.45	16
Log-Pearson 3	0.27	13	0.4	9
Lognormal (3P)	0.21	3	0.33	4
Nakagami	0.27	12	0.44	14
Normal	0.3	19	0.5	19
Pearson 5 (3P)	0.23	7	0.32	3
Pearson 6	0.3	20	0.51	20
Rice	0.27	11	0.5	18

The Production Data after division is assessed with twenty different Probability distributions as described in the Table 2.4 Johnson SB distribution secures rank 1 with K-S Test as well as with A-D Test, it is considered to be best fit distribution for production data of Cotton after division. Here are given pdf, P-P Plot and Q-Q Plot for best fitting distribution Johnson SB.



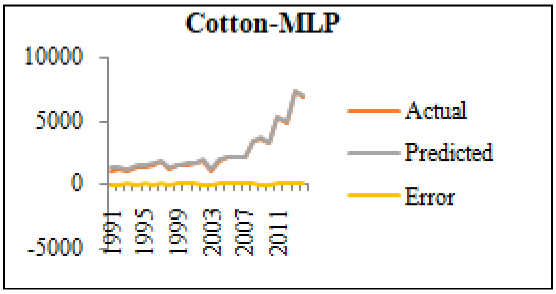


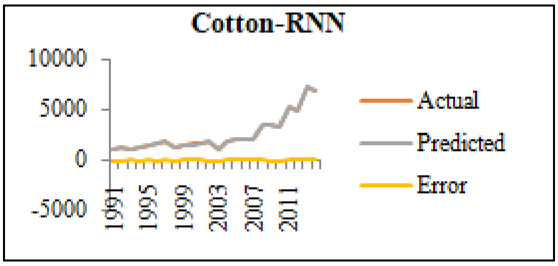
International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

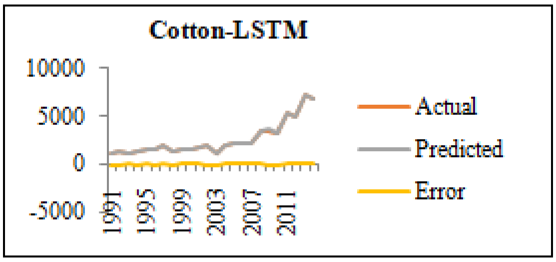


2.3 Forecasting the Production of Cotton for the State Andhra Pradesh using Deep Learning Techniques

In the following figures the predictions of the original and fitted data are visually shown for MLP, RNN and LSTM.







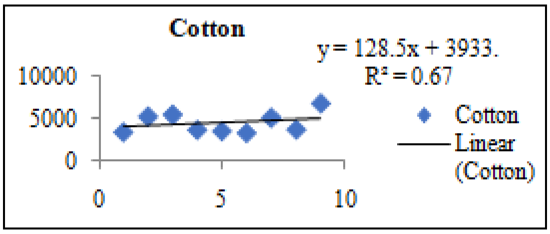
The accuracy of a classifier's predictions is verified using a variety of measures, includes Mean Square Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are explained below.

Table 2.5: Comparisons of MLP, RNN and LSTM

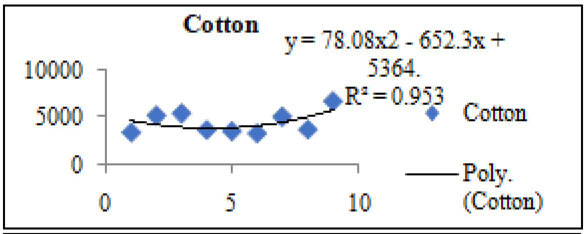
Cotton	MLP	RNN	LSTM
MSE	2108.207	527.0519	131.763
RMSE	45.91522	22.95761	11.4788
MAPE	0.01859	0.009322	0.00467

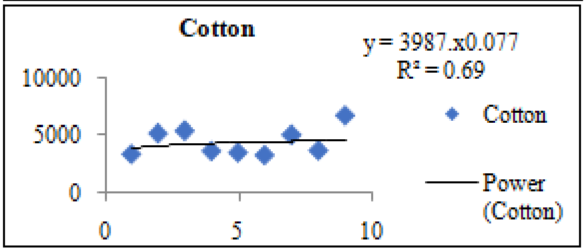
The metrics MSE, RMSE and MAPE determine the model identity and best model fitted to the data used. Obviously from all these models we found that for the LSTM model MSE, RMSE and MAPE is optimum as compared to the other models fitted that is MLP, RNN. Here LSTM out performs with least MSE (131.763), RMSE (11.4788) and MAPE (0.00467) compared to other models.

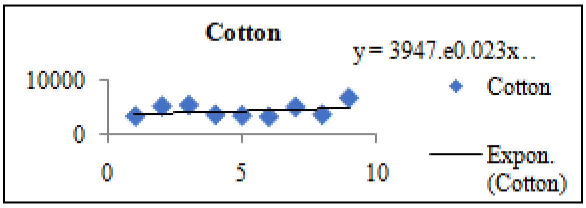
2.4 Forecasting the Production of Cotton for the State Telangana



Year	Production of Cotton
2020-21	6649
2021-22	7636.38
2022-23	8779.92
2023-24	10079.62
2024-25	11535.48







The accuracy of a classifier's predictions is verified using a variety of measures, including R Square (R^2), Mean Square Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are explained below.

Table 2.6: Comparisons of various fitted Curves

	Straight Line	Second degree polynomial	Power Curve	Exponential
R^2	0.67	0.953	0.69	0.69
MSE	1147364	938693	1199350	1165297
RMSE	1071.151	968.8617	1095.149	1079.489
MAPE	0.215599	0.191047	0.220484	0.21967

The metrics R Square, MSE, RMSE and MAPE determine the model identity and best model fitted to the data used. Obviously from all these models we found that for Second degree Polynomial R Square is closer to 1, MSE, RMSE and MAPE is minimum as compare to other models Straight line, Power Curve and Exponential Curves. The following are the predictions of Cotton Production using Second degree Parabolic Equation which is the best fit for Prediction.

Table 2.7: Predictions for production of Cotton

3. Conclusions and Future Study

From the factual data we conclude that the Cotton production in the states AP&TS before and after seems to be different. For combined state of AP i.e before the division Frechet (3p) distribution fits well and after division i.e for the state Telangana Johnson SB distribution fits well. LSTM performs well for predictions of Cotton for the combined state Andhra Pradesh. As the data is less after bifurcation second degree polynomial fitted well with least MSE, RMSE and MAPE..It is observed that production of Cotton increased from 2020-21 to 2024-25 using the best curve fit parabolic equation. For the agriculture stake holders, this study offers a broad and perfect scope for how the data behaves. It is possible to examine the agriculture production for the entire country of India in detail and determine whether the distribution is the same as in the states of Andhra Pradesh and Telangana. This will make it possible to forecast when there is a lack of cotton production.

References

[1] Kumar V, & Bala S. Best fit probability distribution analysis of precipitation and potential evapotranspiration of India's highly dense population state-Bihar. *MAUSAM*. 2022 73(1), 139-150.

[2] Chaudhari RH, Khokhar AN, Paramr DJ, Patel HV, Kumar P, & Kumar R. Fitting of the distribution for CV value of the cotton and tobacco experiment. *Journal of Pharmacognosy and Phytochemistry*. 2020 (5S), 884-890.

[3] Amin MT, Rizwan M, &Alazba AA. A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open Life Sciences*.2016 11(1), 432-440.

[4] Beckman RJ, & Tiet jen GL. Maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*. 1978 7(3-4), 253-258.

[5] Mayank Champaneri; Darpan Chachpara; Chaitanya Chandvidkar; Mansing Rathod. Crop Yield Prediction Using Machine Learning2020, 4, 645-648, https://www.ijsr.net/get_abstract.php?paper_id=SR20402185927

[6] Palanivel, K., &Surianarayanan, C. An approach for prediction of crop yield using machine learning and big data techniques 2019, 3, 110-118, https://iaeme.com/Home/article_id/IJCET_10_03_013.

[7] Ahmar, A.S.; Singh, P.K.; Ruliana, R.; Pandey, A.K.; Gupta, S. Comparison of ARIMA, SutteARIMA, and Holt-Winters, and NNAR Models to Predict Food Grain in India. Forecasting 2023, 5, 138–152. https://doi.org/10.3390/ forecast5010006

[8] R. Hande, A. Ahuja, A. Watwani, P. Chichriya and S. Shamnani, "Krishi Manch: Disease Detection in Rice Crops using CNN, " 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2021, pp. 1-6, doi: 10.1109/ICNTE51185.2021.9487695.