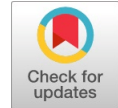# A Research on Change Point and Trend Scrutiny: with Reference to Castor in India

Kalpana P, Madhavi K, Venkateswaran M, Kiran Kumar P

Abstract: Castor is treated as an important non-edible oil crop among Mahua, Karanja and Jatropha. India occupies a leading position in the production of castor followed by China and Brazil. Castor has been used in various purposes from thousands of years ago which has important characteristics. The main objectives of current work are i) Identification of change point and ii) Trend analysis with respect to area, yield and production of castor for the period from 1961 to 2016 in India. The non-parametric statistical methods such as Pettitt's, Standard Normal Homogeneity (SNH) and Buishand's Range tests have conceded to detect the mutation point whereas the magnitude of trend is measured and analysed with the help of Sen's slope estimator and their significance is tested by Mann-Kendall test. The results revealed that the change is identified by mutation point at initially 1988. Interestingly the maximum growth is captured in the second sub-time series on the basis of time scale and area, yield and production. However, the production of castor cannot meet up its annual enormous demand for India.

Keywords: Castor, Change point, Trend analysis, Non-Parametric methods

I. INTRODUCTION

Castor plant was first found in tropical Africa before 1500 B.C. Castor beans produces a non-edible vegetable oil and member of the spurge family, later it is called as scientifically Ricinus communis L. Castor mainly cultivated for procuring the oil from the castor seeds and it is an important renewable resource for the chemical industry and less expensive vegetable oil. In olden days, castor oil is used as facial oil, wick lamps for lighting etc. Recent years, castor oil is using as multi - purposes in many ways for example a lubricant in high speed engines and airplanes; ointments, printing inks, polishes, hair dressings, cosmetic, soaps, waxes, disinfectants etc. It has been utilized in medicines and similarly many veterinary purposes. In worldwide, 3 million hectares of land cultivated for castor and 1.5 million tons of production is produced annually. However, the supply of castor is not proportionate to demand of India. There are numerous factors that influence stability of area, production and yield such as urbanization, floods, environmental factors, fertilizers etc.

Consequently, the requirement of castor is rapidly increasing day by day in industries. To balance the rising demand of castor, the agricultural scientists, farmers have to take more attention to increase the production.

Rajarathinam and Parmar (2011) investigated trends and growth rates of castor crop through parametric, non-parametric tests and time series models. Identified non-parametric models gives best fitted trends among parametric and time series models. Change and instability of area, production, and yield with respect to major crops are studied by various statistical techniques (Hasan et al. 2008; Rakesh 2014; Muhammad et al. 2014; Taj et al. 2015 and Suseela and Chandrasekaran 2016). Ramandeep and Sandhya (2015) analyzed trends of important crops such as cotton, castor and banana of district and state levels of India. Kalpana (2014, 2017) analyzed the growth rates of area, production, yield and seed of groundnut crop and castor crop through different linear and non-liner regression models. In the above mentioned works, the researchers have studied trends, change and instability of major crops - pulse, castor, groundnut, wheat etc. The present study mainly concentrates on identification of change and trend for harvested area, productivity and production of castor during 1961 to 2016 in India. The non-parametric statistical methods such as Pettitt's, Buishand's and SNH tests have been adapted to capture the mutation point whereas the quantified the ratio of trend by Sen's slope estimator and their significance is tested by Mann-Kendall test.

II. MATERIAL AND METHODS

The analysis consists of year wise data for harvested area, productivity and production of castor in India and study period is from 1961 to 2016. The details are presented below sections:

2.1 Tests for Capture Change Point

The change point analysis tries to detect where significant change takes place in time series data and it is fundamental tool in time series analysis. Many researchers have applied various methods to capture the change points for time series related data sets. (Chen and Gupta 2001; Gallagher et al. 2012). In this study, three statistical techniques viz. Pettitt's, Buishand's Range and SNH tests used to identify an abrupt change in the castor data.

2.1.1 Pettitt's Test

This test was introduced first by Pettitt in the early 1979's. It is extensively adapted to examine the abrupt changes (Winingaard et al. 2003; Verstraeten et al. 2006; Dhorde and Zarenistanak 2013).The test statistic is,

$$U_{t,T} = \sum_{i=1}^{t} \sum_{j=t+1}^{T} \text{sign}\left(x_i - x_j\right)$$

$$\text{Where, sign}\left(x_i - x_j\right) = \begin{cases} 1, & \text{if } \left(x_i - x_j\right) > 0 \\ 0, & \text{if } \left(x_i - x_j\right) = 0 \\ -1, & \text{if } \left(x_i - x_j\right) < 0 \end{cases}$$

Then consider, $K_T = \max \left|U_{t,T}\right|$ and its significance is tested at 5% level.

2.1.2 Buishand's Range Test

Since 1982, the test is called as Buishand's range test before it is named as Buishand and the test statistics is calculated by $R_b = \dfrac{\max S_k - \min S_k}{\sigma}$ where, $S_k = \sum_{i=1}^{k}\left(x_i - \hat{x}\right)$. Here, 'p' value is considered with a Monte Carlo Simulation by 'm' replicates.

2.1.3 SNH Test

The test statistic of SNH test T_k is between the mean values of the first n data points and the remaining (n-k) data points (Stepanek et al. 2009).

$$T_k = kz_1^2 + \left(n-k\right)z_2^2, \quad \left(1<k<n\right) \text{ where, } z_1 = \frac{1}{k}\sum_{i=1}^{k}\frac{\left(x_i - \overline{x}\right)}{\sigma} \text{ and } z_2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\frac{\left(x_i - \overline{x}\right)}{\sigma}$$

In the SNH test, the critical value is $T = \max T_k$ and p-value is considered with a Monte Carlo simulation by m replicates.

2.2 Trend Analysis

Generally the magnitude of trend is measured in terms of ratio, this ratio can give an idea about trend i.e., either falling or rising or remaining relatively constant. With the help of trend ratio, it is possible to detect the poor or good signs of management. The non-parametric tests are prominent and well suited for non-normally distributed data.

The M-K test plays a vital role to find the significant nature of monotonic trends for time series data (Tabari et al. 2014; Khalili et al. 2015; Gavrilov et al. 2016; Pohlert 2018). Researchers of World Meteorological Organization (WMO) suggested that M–K test is effective and adaptable for the areas of climatic, hydrological and water resources (Tabari and Marofi 2011). The test statistic of Mann-Kendall (S) is,

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{sign}\left(x_j - x_k\right) \text{ where } \text{sign}\left(x_j - x_k\right) = \begin{cases} 1, & \text{if } \left(x_j - x_k\right) > 0 \\ 0, & \text{if } \left(x_j - x_k\right) = 0 \\ -1, & \text{if } \left(x_j - x_k\right) < 0 \end{cases}$$

Where $E(S) = 0$ and $V(S) = \left\{n\left(n-1\right)\left(2n+5\right) - \sum_{j=1}^{p} t_j\left(t_j - 1\right)\left(2t_j + 5\right)\right\}/18$ and 'p' is the number of tied groups in the data set, 't_j' group and n is the number of data in the time series. For n>10, the standardized statistic for the M-K test is

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\text{Var}(S)}}, & \text{if } S > 0 \\[3mm] \dfrac{S+1}{\sqrt{\text{Var}(S)}}, & \text{if } S < 0 \\[3mm] 0, & \text{if } S = 0 \end{cases}$$

If 'S' is positive, it indicates increasing trend otherwise decreasing trend.

2.3 Sen's Slope Estimator

It is an adequate tool to determine the magnitude of trend in hydro-meteorological time Series (Huang et al. 2013; Talaee 2014; Da Silva et al. 2015; Liuzzo et al. 2016; Amirataee et al. 2016; Zamani et al. 2016). The Sen's slope estimator is defined as the median of these N values of T_i and test statistics is computed by

$$T_i = \frac{y_j - y_k}{j - k} \quad \text{for i=1, 2, ..., N}$$

where y_j and y_k are data values at time j and k (j>k) respectively.

$$\beta = \begin{cases} T_{\frac{N+1}{2}} & \text{N is odd,} \\ \frac{1}{2}\left(T_{\frac{N}{2}} + T_{\frac{N+2}{2}} \right) & \text{N is even} \end{cases}$$

If β is positive, it indicates increasing trend otherwise decreasing trend

III. EXPERIMENTAL STUDY

Table 1: Descriptive statistics for Castor crop from 1961-2016

Item	Minimum	Maximum	Mean	S.D.
Area harvested(ha)	374900.000	1471000.000	667236.964	237322.433
Yield(hg/ha)	1970.000	16923.000	7906.036	4548.101
Production(tonnes)	79800.000	2295000.000	618262.500	542460.716

It is evident from Table 1, the results revealed that the lowest harvested area (i.e., 3,74,900 ha) in the year 1975 and highest harvested area (14,71,000 ha) in the year 2011. The minimum and maximum yield of castor is identified as 1,970 hg/ha in the year 1965 and 16,923 hg/ha in the year 2014. Whereas in the case of production, the minimum and maximum are 79, 800 tonnes at the year1965 and 22,95,000 tonnes at the year 2011. The overall harvested area, yield and production of castor are 667236.964 ha, 7906.036 hg/ha and 618262 tonnes during 196-2016.

Table 2: Change point analysis on Area, Yield and Production of Castor

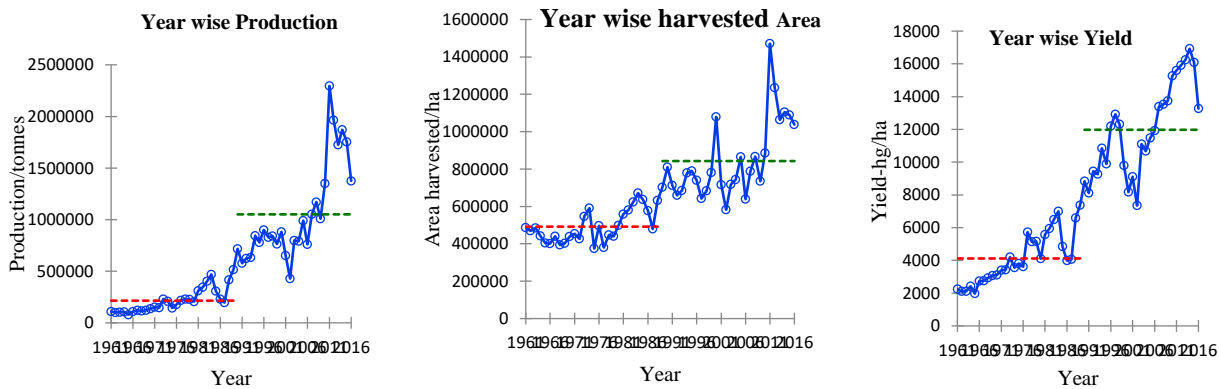| Indicators	Annual Rainfall					
	Pettitt's Test		Buishand's Test		SNH test	
	p-Value	Change Point	p-Value	Change Point	p-Value	Change Point
Harvested Area (ha)	0.0001	1988	0.0001	1988	0.0001	1988
Yield (hg/ha)	0.0001	1989	0.0001	1989	0.0001	1989
Production (tonnes)	0.0001	1989	0.0001	2006	0.0001	1989

Fig.1 Identification of change points for production, area and yield of castor

From the Table 2 and Fig. 1, the results exhibits the significant change points i.e.,1988 and 1989 years are captured for area and yield by three tests such as Pettitt's, Buishand's and SNH test. In the case of production the identified change point is 1989 by Pettitt's and SNH test. A notable observation from the change point analysis is the change is occurred from the year 1988 onwards. Based on the change point results, the time period (1961-2016) is splits into three time frames i.e., first segmentation period (before mutation point), second segmentation period (after mutation point) and whole-time series (1901-2015). Consequently, the trend analysis is extended based on segmentation time period. Here, the M-K test have been employed to analysis the significant monotonic trends for indicator wise and segmentation period wise for castor.

Table 3: Trend analysis results by M-K test and Sen's slope estimator

Indicators	Segmentation Period	Kendall's tau	p-value	Sen's slope
Area harvested	1961-1988	0.381	0.004	6270.833
	1989-2016	0.434	**0.001**	13379.000
	1961-2016	0.699	**0.0001**	11089.610
Yield	1961-1989	0.773	**0.0001**	160.977
	1990-2016	0.641	**0.0001**	310.045
	1961-2016	0.857	**0.0001**	258.744
Production	1961-1989	0.773	**0.0001**	10412.222
	1990-2016	0.601	**0.0001**	35987.500
	1961-2016	0.847	**0.0001**	24525.672

Fig. 2 Sen's slope estimators for area, yield and production of castor

The trend analysis results shows all the indicators i.e., area, yield and production of castor trend values behaves significant upward trend viz., segmentation year wise. On average the highest magnitude of trend for area is 13379 ha, yield is 310.045 hg/ha and production is 35987.5 tonnes. An important observation from second segmentation period is the magnitude ratio of trend is highly raised followed by whole time series and first sub-time series on the basis of area, yield and production and these results are presented in the Table 3 and Fig. 2.

IV. CONCLUSION

The change point analysis captured the change is started from the year 1988 for yearly data of castor. The trend analysis explores all the indicators of castor followed good raising trends and moreover the maximum growth is observed in second segmentation period. A notable point from the results is growth rates of castor is good and it is necessary to maintain for future demand

REFERENCES

1. Chen J, Gupta A (2001) On change point detection and estimation, Communications in Statistics. Simul Comput 30(3):665–697.
2. Dhorde AG, Zarenistanak M (2013) Three-way approach to test data homogeneity: an analysis of temperature and precipitation series over southwestern Islamic Republic of Iran. J Ind Geophys Union 17(3):233–242.
3. Gallagher C, Lund R, Robbins M (2012) Change point detection in climate time series with long-term trends. J Clim 26:4994–5006.
4. Gavrilov MB, Tosic I, Markovic SB, Petrovic P et al. (2016) Analysis of annual and seasonal temperature trends using the Mann-Kendall test in Vojvodina, Serbia. Idojaras 120:183–198.
5. Hasan MN, Monayem MM, Islam MS, Alam QM, Hossain MI (2008) Change and Instability in Area and Production of Wheat and Maize in Bangladesh. Bangladesh J. Agril 33: 409-417.
6. Kalpana P (2014) Statistical Modeling on Growth Rates of Groundnut Crop from 1990 to 2014 in India. Int j Eng Sci Research 4: 1-7.
7. Kalpana P (2017) A Case Study on Change and Instability of Castor. Int J Adv Infor Sci Tech 6: 13-15.
8. Khalili K, Tahoudi MN, Mirabbasi R, Ahmadi F (2015) Investigation of spatial and temporal variability of precipitation in Iran over the last half century. Stoch Environ Res Risk Assess 1–17.
9. Muhammad ABC, Taj U, Jamal U (2014) Oil Seeds Area and Production Variability in Bangladesh. J Appl quan Methods 9.
10. Pohlert T (2018) Non-parametric trend tests and change-point detection. https://cran.r-project.org/web/packages/trend/vignettes/trend.pdf. Accessed on 25 September 2018.
11. Rajarathinam A, Parmar RS, (2011) Application of Parametric and Nonparametric Regression Models for Area, Production and Productivity Trends of Castor (Ricinus communis L.) Crop. Asian J Appl Sci 4: 42-52.
12. Rakesh S (2014) Growth & Instability in Agricultural Production in Haryana: A District level Analysis. Int J Sci Res Publication 4.
13. Ramandeep KM M, Sandhya KG (2015) Analysis of trends in area, production and yield of important crops of India. Int J Agron Agricu Res 7: 86-92.
14. Stepanek P, Zahradnek P, Skalak P (2009) Data quality control and homogenization of air temperature and precipitation series in the area of Czech Republic in the period of 1961-2007. Adv Sci Res 3:23-26.
15. Suseela K, Chandrasekaran M (2016) Growth And Instability in Dry Land Agriculture of Andhra Pradesh Int J Agri, Sci Rese 6: 285-294.
16. Tabari H, AghaKouchak A, Willems P (2014) A perturbation approach for assessing trends in precipitation extremes across Iran. J Hydrol 519:1420–1427
17. Tabari H, Marofi S (2011) Changes of pan evaporation in the west of Iran. Water Resour Manag 25:97–111.
18. Taj UM, Mohammad KH, Ohid MU (2015) Growth and Instability In Area And Production of Pulse In Bangladesh. J Agri Sci Eng 1: 163-167.
19. Verstraeten G, Poesen J, Demarée G, Salles C (2006) Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): implications for assessing soil erosion rates. J Geophys Res 111:1–11.
20. Winingaard JB, Kleink Tank AMG, Konnen GP (2003) Homogeneity of 20th Century European Daily Temperature and Precipitation Series. Int J Climatol 23:679–692.