Chapter-5

TREND AND MUTATION POINT DETECTION IN RAPESEED AND MUSTARD CROP FOR SMALL AREA

Bhim Singh*, CK Arya[1], Amar Singh[2], KK Maurya[2], RP Kaushal[3], Sugriv Maurya[4] and AK Sharma[5]

Department of Basic Science, College of Agric ulture, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut -250110 (U.P.), India.

[1]Department of NRM, College of Horticulture and Forestry (AU, Kota), Jhalawar-326023 (Rajasthan) India

[2]Department of Agricultural Statistics, CSSS PG College, Machhra, Meerut-250106 (U.P.), India.

[3]Department of Agricultural Economics, SDJ PG College, Chandeshwar, Azamgarh – 2760128, India.

[4]Department of Agricultural Economics, Janta Mahavidyalaya, Ajeetmal, Auraiya-206121 (U.P.), India.

[5]Department of Agricultural Statistics, AS College, Lakhaoti, Bulandshahr- 203407 (U.P.), India.

*Corresponding author: bhimsingh1@gmail.com

Abstract

Rapeseed-mustard (Brassica species) is the third most important oilseed crop after soybean and groundnut, contributing. The objectives of present study were to identify the mutation point and trend analysis in the time series data of rapeseed and mustard crop for small area. In the present article, we used various well-known non-parametric methods namely Pettitt's test, Standard normal homogeneity test and Buishand's range test. These three techniques were used to capture the mutation point in time series data connected to the area, production and productivity of rapeseed and mustard during the period 1997-2019. Accordingly, Sen's slope estimator was applied to measure the magnitude of the trend of area, production and productivity of rapeseed and mustard and their significances were tested through Mann Kendall (M-K) test. The analysis provided a clear idea about the trends of rapeseed and mustard crops. The results showed the raising trends in production of rapeseed and mustard. The significant ($p< 0.05$) highest rising trend of rapeseed and mustard production was observed in the Meerut district and obtained Sen's slope estimator of 315.67/year in the second-time series (2009-2019) followed by 247.71/year in the overall the study period. However, the production of rapeseed and mustard has been increased significantly in Uttar Pradesh compared to previous years and also it holds the raising trend in the area of 178.75 in the second-time period.

Keywords: Mutation point, MK-trend test, Rapeseed and mustard, Sen's slope, Trend detection

Introduction

Rapeseed-mustard (*Brassica* species) is the third most important oilseed crop after soybean and groundnut, contributing 28.95 % of the total oilseed production in the country (Anonymous, 2020). On an average, India produces around 9.34 million tones of rapeseed and mustard annually (2018-19) on the area of 6.23 million hectare with productivity 1499 kg per hectare. Uttar Pradesh is the second leading rapeseed and mustard producing state in India after Rajasthan. Rajasthan and Uttar Pradesh together produces over 50% of the total rapeseed and mustard production in the demand for the consumption of rapeseed and mustard comes mainly from eastern and northern areas of the country. In Uttar Pradesh, rapeseed and mustard occupies an area of 0.75 million hectares, production of 1.12 million tonnes and productivity of 1483 kg per hectare (2018-19). It has been a great concern that technological change in rapeseed and mustard production has increased variability, which is considered to be one of the important facts. The requirements for small area statistics have been in great demand by public as well as private sectors because the governments are interested in obtaining statistics for smaller domains such as States, Provinces, Districts, Tehsils, Blocks and Village Panchayat etc or different racial and ethnic subgroups for implementing their policies at local. This is also mainly due to the fact that regional planning and the distribution of central funds are often made on a local or regional basis, and these may depend on variables such as crop production/yield, the number of unemployed persons, condition of the housing, use of fertilizers, principal occupation of the households, infrastructural needs, and land and water endowments (Singh *et al.* 2012). These domains are called small area-the term "SMALL" refers to the fact that the sample size in the area or domain from the survey is small (Singh and Sisodia, 2015). In the present study, Meerut district was considered as "small area". Various research workers have made attempt in the past to study the trend and growth analysis of oil seeds productions in the State/country. Notably among them are Rao and Raju (2005), Gangwar and Singh (*2015)*, Viswanatha and Kingsly (2017). Snehdeep *et al*. (2017), etc. However, in the present paper, trend and mutation point detection (MPD) in rapeseed and mustard crop for small area (Meerut district) has been studied. In order to identify trend and MPD for time series data, the more powerful the non-parametric techniques such as Pettitt's, SNH and Buishand's range tests have been studied whereas the trends analysis through Sen's slope estimator and their significances tested with the Mann-Kendall (M-K) test.

Materials and Methods

Meerut district is located in the western part of Uttar Pradesh situated between 28.740 N and 29.270 N latitudes and 77.44^0 and 78.14^0 E longitudes. It has an area more than 2500 km^2 (Annonymus, 2011), majorly dominated by agriculture land use (Fig. 1). It is bounded by Muzaffarnagar district is in the north, Ghaziabad, Bulandsahar and Gautam Budh Nagar districts in the south, Bijnore and Jyotibaphule Nagar in the east and Baghpat in the west (Anonymous, 2013). The time series data on area, production and productivity of rapeseed and mustard for Meerut district of western Uttar Pradesh from 1997–1998 to 2019–2020 (23 years) have been used for the present study. The secondary data have been taken from Area, Production Statistics Information System, Ministry of Agriculture and Farmer Welfare, New Delhi (https://www.aps.dac.gov.in/APY/Public_Report1.aspx). The data obtained were analysed by using Microsoft Office Excel 2007 and XLSTAT 2021 in order to determine the trend and mutation point in rapeseed and mustard crop for Meerut district of western Uttar Pradesh, India.

Methods for mutation point and trend detection

The analysis of mutation point, popularly known as change point tries to find out where significant changes occur in the rapeseed and mustard time series data. Thus, mutation point is a basic required tool in time series analysis. Many different techniques were applied to find out the mutation points in the time series data by the several researchers (Buishand, 1982; Bryson *et al.*, 2012; Li and Lund, 2012; Sharma *et al.*, 2016; Polisetty and Paidipati, 2020). The Pettitt's test, a distribution-free test, developed by Pettitt (1979) which is valuable for evaluating the occurrence of abrupt changes in time series data (Mu *et al.* 2007). Several research workers have been applied different distribution-free statistical methods to find the mutation points in the time series data (Jaiswal *et al.*, 2015; Sharma *et al.*, 2016 and Kalpana and Kiran, 2019; Polisetty and Paidipati, 2020). The details methods of change point detection are given below.

Pettitt's Test (Pettitt, 1979)

This approach was proposed by Pettitt (1979) which is commonly applied to detect significant variations in the average of a time series data when the exact time of the variation is unknown. This test the H_0 (null hypothesis): The times series has no change point, verses H_1 (alternative hypothesis): A change point exists. According to Pettitt's test, if $y_1, y_2, \ldots, y_T$ is a series of observations which has a change point at t in such a way that $y_1, y_2, \ldots, y_t$ has a distribution function $F_1(y)F_1(y)$ which is different from the distribution function $F_2(y)$ of the second part of the series $y_{t+1}, y_{t+2}, y_{t+3} \ldots, y_T$. The statistics $U_{t,T}$ for Pettitt's test is given as:

$$U_{t,T} = \sum_{i=1}^{t} \sum_{j=t+1}^{T} sign(y - y)$$

$$\text{where, } sign(y_i - y_j) = \begin{cases} 1 & if(y_i - y_j) > 0 \\ 0 & if(y_i - y_j) = 0 \\ -1 & if(y_i - y_j) < 0 \end{cases}$$

then test statistic

$$K_T = Max|U_{t,T}|$$

and the significance of K_T is approximated for

$$p = 2\exp\left(\frac{-6K_T^2}{T^2(1+T)}\right) < 0.05$$

Buishand's Range Test (Buishand, 1982)

The Buishand's Range test is given by Buishand (1982) based on the adjusted partial sums or cumulative deviations from the average. The null hypothesis is same as the Pettitt's test. The adjusted partial sum is defined as:

$$S_K = \sum_{i=1}^{k} (y_i - \bar{y})$$

and its test statistic calculated by

$$R_b = \frac{Max(S_k) - Min(S_k)}{\bar{x}}$$

Here, p-value is measured with a Monte Carlo simulation method by m replicates.

Standard Normal Homogeneity (SNH) Test (Alexandersson, 1986)

This method was proposed by Alexandersson (1986) in which the proposed null and alternative hypotheses are same as the Pettitt's test. The test statistic (T_k) of SNH test is given as

$$T_k = kZ_1^2 + (n-k)Z_2^2$$

where

$$Z_1 = \frac{1}{k}\sum_{i=1}^{k}\frac{(y_i - y)}{}, Z_1 = \frac{1}{n-k}\sum_{i=k+1}^{n}\frac{(y - \bar{y})}{}$$

Here, the test statistic is appropriate for compares the average of first n observations and the average of the remaining (n-k) observations with the n data points also, the critical value is T = max T_k and p-value is measured with a Monte Carlo simulation methods by m replicates.

Test for trend Analysis

Non-linear trends in time series of yield data often hinder risk analysis for crop production. The trends in historical time series of sugarcane data have been analysed. The amount of the trend in the time series data was driven by Sen's estimator (Sen, 1968) and significance of the trend in the time series was tested by Mann-Kendall (M-K) test (Mann, 1945; Kendall, 1975). Sen's slope method is a nonparametric approach frequently used for environmental data analysis because it is robust to missing data and outliers (Gilbert, 1987) and also gives a robust estimation of trend (Yue *et al*., 2002). Mann-Kendall test checks the null hypothesis (H_0) of no trend versus alternative hypothesis (H_1) of the presence of increasing or decreasing trend. The Mann-Kendall test plays a crucial role to identify the significant nature of linear trends in the time series data (Polisetty and Paidipati, 2020). The test statistic (S) of M-K test is defined as (Salas, 1993):

where N is number of data points. It is assume, the value of is calculated by the following expressions:

This statistic represents the differences of positive and negative sign of their values for all the differences measured. For large samples (N>10), the test is facilitated using a normal distribution (Helsel and Hirsch, 1992) with the subsequent equations:

Here, N represents number of tied groups in the observations and t_k is data points in the kth tied groups in the observations. The standard normal deviate (Z-statistics) is then calculated as suggested by Hirsch *et al.,* 1993 by the following expression:

If the calculated value of , the null hypothesis (H_0) is rejected at level of significance in a two-tailed test. The sign of 'S' represents upward or down ward trends of time series. If 'S' has positive sign, it represents the upward trend and if 'S' observed negative sign, it indicates down ward trend.

Sen's Slope estimator

A non-parametric or distribution-free test is used to calculate the value of trends in the time series of rapeseed and mustard production. Sen's slope estimator (Sen, 1968) has the sufficient procedure to evaluate the trend in time series data (Polisetty and Paidipati, 2020). Sen's method calculates the slope as a change in measurement in consonance to the change in time. In Sen's slope estimator, the slopes (T_i) (*i = 1, 2, …, n)* of all pairs of data are first calculated by

where y_j and y_k are values at time *j* and *k* (*j* > *k*) respectively. The second quartile of these *n* values of T_i is Sen's slope estimator which is computed as

A positive sign of indicates an upward trend and a negative value indicates downward trend in the time series.

Results and Discussion

The known well-known fact is that rapeseed and mustard is one of the most consumable oilseed crops in north India. Various statistical tools have been used to present the summary statistics for rapeseed and mustard crop in Meerut district during 1997-98 to 2019-20. The highest yield of rapeseed and mustard was recorded in the years of 2013 with the total of 1491.79 kg/ha while the lowest yield recorded in 2003 with the value of 852.52 kg/ha (Table 1 and Fig. 2). The maximum rapeseed and mustard production in Meerut district was 8382.00 tonnes with a standard deviation of 1791.55 tonnes and coefficient of variation of 36.68% (Table 1). The maximum area under rapeseed and mustard was 6263.00 ha with a standard deviation of 1140.36 ha and coefficient of variation of 28.43% along with average productivity of 1197.01 during the study period. The overall average area, production and productivity of rapeseed and mustard were 4011.43 ha, 4883.78 tonnes and 1197.01 kg/ha during 1997-98 to 2019-20. The current study, at the outset, the non-parametric tests Pettitt's, SNH and Buishand's range tests have been used to capture the mutation point for year wise rapeseed and mustard area, production and productivity from 1997-98 to 2019-20. From the Table 2 and Fig. 3, the results exhibits the significant ($p < 0.0001$) mutation points i.e., 2009 year is captured for area and production by two tests such as Pettitt's, Buishand's range test while year 2012 was identified as significant mutation point by the SNH test. In the case of productivity, the identified mutation point was 2007 by Buishand's range test. These turning points play a very crucial role to analyse the monotonic trends in the production of rapeseed and mustard. After, identification of mutation point, the whole time series (1997-2020) were divided into three parts i.e., first-time series (before mutation point), second-time series (after mutation point) and whole-time series (1997-2020). Further, the Sen's slope estimators have been used to compute the degree of monotonic trends. Further, the trend analysis is extended based on segmentation time period. Here, the M-K test has been employed to analysis the significant monotonic trends for indicator wise and segmentation period wise for rapeseed and mustard. A notable point is observed that the raising trends for rapeseed and mustard area, production and productivity in segmentation year wise (Table 3 and Figure 3). The identified all trends are statistically significant at 95% confidence by M-K test except in first and second segmentation period for productivity of rapeseed and mustard. From the Sen's slope estimator analysis, it is observed that statistically significant highest growth of production was observed as 315.67/year during 2009-2019 i.e. second segmentation period and lowest non-significant growth was experienced as 12.99/year in the first sub-time series of productivity (Table 3). The whole-time series results shown the upward significant trends were seen in the all segment of production, area and productivity. Thus, production of rapeseed and mustard showed the highest increasing trend as compared to area and productivity in first & second series and also in whole time period. From the above analysis and discussion, it is clear that the Sen's slope estimator has given accurate results to the researchers. The statistical analysis showed that rapeseed and mustard production is positively increased during the whole-time series (study period). However, at present, the population is increasing continuously along with the production of rapeseed and mustard is successful to meet the necessities of the utilization of rapeseed and mustard.

Table 1: Summary statistics for rapeseed and mustard crop in Meerut district during 1997-98 to 2019-20.

State	Minimum	Maximum	Average	S.D.	C.V.
Area (ha)	2595.00	6263.00	4011.43	1140.36	28.43
Production (tonnes)	2422.00	8382.00	4883.78	1791.55	36.68
Productivity (kg/ha)	852.52	1491.79	1197.01	180.01	852.52

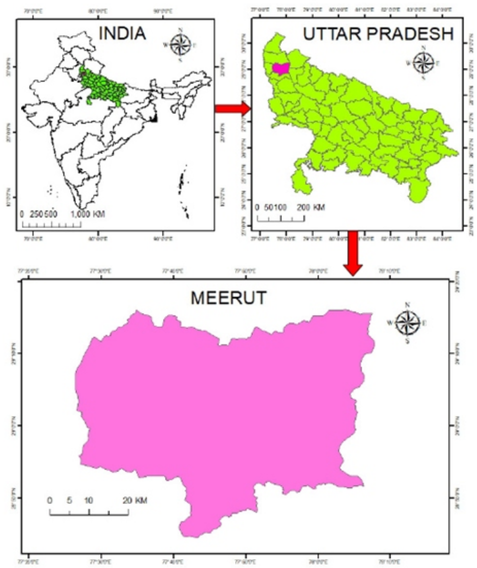
Table 2: Analysis of change point for rapeseed and mustard crop in different states of India.

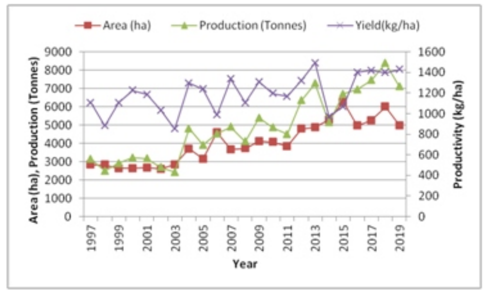
State	Pettitt's Test		Buishand's Range Test		SNH test	
	p-value	Change Point	p-value	Change Point	p-value	Change Point
Area	<0.0001	2009**	<0.0001	2009**	<0.0001	2012**
Production	<0.0001	2009**	<0.0001	2009**	<0.0001	20012**
Productivity	0.117	No trend	0.044	2007*	0.055	No trend

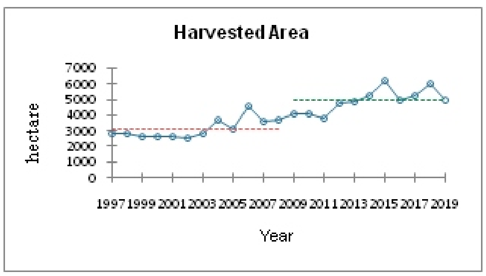
*: Significant @ 5 % level of significance, **: Significant @ 1 % level of significance

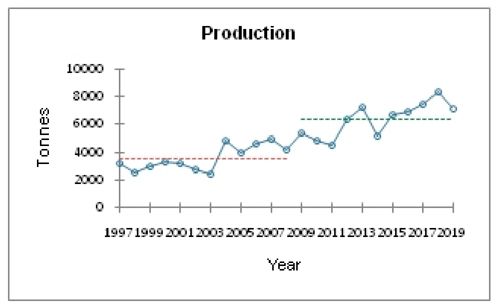
Table 3: Mann-Kendall test and Sen's slope estimators for rapeseed and mustard crop

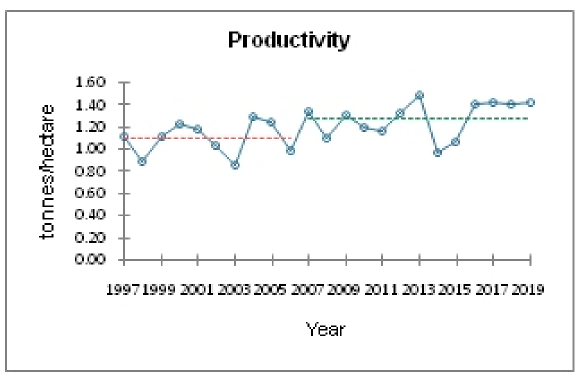
State	Segmentation year	MK-Stat	p-value	Sen's slope
Area	1997-2008	34.00	<0.0236	99.72
	2009-2019	31.00	0.0195	178.75
	1997-2019	191.00	<0.0001	149.50
Production	1997-2008	30.00	0.0467	171.30
	2009-2019	33.00	0.0127	315.67
	1997-2019	187.00	<0.0001	247.71
Productivity	1997-2006	5.00	0.721	12.99
	2007-2019	103.00	0.200	13.40
	1997-2019	22.00	0.007	14.79











References

- Alexandersson, H. (1986). A homogeneity test applied to precipitation data. International Journal of Climatology, 6, 661–675.
- Anonymous (2011). Government of India (2011) Census. Government of India.
- Anonymous (2013). MSME Development Institute (2013) Brief Industrial Profile of District Meerut. Ministry of Micro, Small & Medium Enterprises, Govt. of India, Agra.
- Anonymous 2020. Agriculture Statistics at a Glance (2019), Directorate of Economics and Statistics, Government of India. https://eands.dacnet.nic.in/
- Bryson, C.B., Richard, E.C. and Adrian, W.B. (2012). Trend estimation and change point detection in individual climatic series using flexible regression methods. Journal of Geophysical Research, 117, D16106.
- Buishand, T.A. (1982). Some methods for testing the homogeneity of rainfall records. Journal of Hydrology, 58(1–2),11–27.
- F.A.O. (2019). Crop production. Food and Agriculture Organization of the United Nations. Retrieved 2020-10-09.

- Gangwar, A. and Singh, V. (2015). An analysis of supply response of rapeseed-mustard in different regions of Uttar Pradesh, Journal of Oilseed Brassica, 6 (1), 158-166.
- Gilbert, R.O. (1987). Statistical methods for environmental pollution monitoring. John Wiley & Sons, New York, USA.
- Helsel, D.R. and Hirsch, R.M. (1992). Statistical methods in water resources. Elsevier, New York.
- Hirsch, R.M., Helsel, D.R., Cohn, T.A. and Gilroy, E.J. 1993. Statistical treatment of hydrologic data: In Handbook of Hydrology, Maidment D.R. (ed). McGraw-Hill, New York, pp. 17.1-17.52.
- Jaiswal, R.K., Lohani, A.K. and Tiwari, H.L. 2015. Statistical analysis for change detection and trend assessment in climatological parameters. Environmental Processes, 2(4), 729-749.
- Kalpana, P., Madhavi, K., Venkateswaran, M. and Kiran, K.P. (2019). A Research on change point and trend scrutiny with reference to Castor in India. International Journal of Engineering and Advanced Technology, 9, 240-244.
- Kendall, M.G. (1975). Rank correlation methods. Charles Griffin, London, UK.
- Li, S., Lund, R.B. (2012). Multiple change point detection via genetic algorithms. Journal of Climatology, 25, 674–686.
- Mann, H.B. (1945). Nonparametric tests against trend. Econometrica, 13, 245-259.
- Mu, X, Zhang, L, McVicar, TR, Chille, B. and Gau, P. (2007). Analysis of the impact of conservation measures on stream flow regime in catchments of the Loess Plateau, China. Hydrology Process, 21, 2124–2134.
- Pettitt, A.N. (1979). A non-parametric approach to the change point problem. Journal of Applied Statistics, 28(2), 126–135.
- Polisetty, K. and Paidipati, K.K. (2020). Statistical assessment of trend analysis on production of wheat crop over India. Sarhad Journal of Agriculture, 36(1), 178-184.
- Rao, I.V.Y.R. and V.T. Raju (2005). Growth and instability of groundnut (Arachis hypogaea L.) production in Andhra Pradesh: District wise analysis. Journal of Oilseeds Research, 22(1), 141-149.
- Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of American Statistical Association, 63, 1379-1389.
- Sharma, S., Swayne, D.A. and Obimbo, C. (2016). Trend analysis and change point techniques: a survey. Energy, Ecology and Environment, 1(3), 123–130.
- Singh, B. and Sisodia, B.V.S. (2015). A Note on Estimation of Overall Average for Small Domain. International Journal of Agricultural and Statistical Sciences, 11(1), 245-249.
- Singh, B. and Sisodia, B.V.S., Singh, A. and Kaushal, R.P. (2015). A note on the estimation methods of crop production at the block level. Journal of Applied Statistics, 39 (9), 2015–2027.
- Snehdeep, Sisodia, B.V.S., Mourya, K.K. and Rai, V.N. (2017). Trend and Growth Analysis of Rapeseed and Mustard Production in Uttar Pradesh. International Journal of Agricultural and Statistical Sciences, 13(1), 273-277.
- Viswanatha, R.K. and I.T., Kingsly (2017). Area, Production, Yield Trends and Pattern of Oilseeds Growth in India. Economic Affairs, 62(2), 327-334
- Yue, S., Pilon, P. and Cavadias, G. (2002). Power of the Mann-Kendall and Spearman's Rho tests for detecting monotonic trends in hydrologic series. Journal of Hydrology, 259, 254-271.