Environment and Ecology 41 (4D) : 3087—3092, October—December 2023
Article DOI: https://doi.org/10.60151/envec/CFTP8413
ISSN 0970-0420

# Non Parametric Trend Analysis of Tea Cultivation in India

**Arya V. Chandran, D. S. Dhakre, Debasis Bhattacharya, K. A. Sarkar**

Received 5 June 2023, Accepted 25 October 2023, Published on 29 December 2023

## ABSTRACT

With the second-largest producer and the highest consumer in the world, tea is one of the most important beverage crops in India. Considering the importance of tea industry to India's national economy, a trend analysis of tea cultivation has been done here. Using data on the acreage, production, and productivity of tea in India from 1960 to 2021, and following a non-parametric approach Modified Mann-Kendall test has been used to detect the trend in the data. Addition to this, Sen's slope estimator and innovative trend analysis have been performed to analyze the magnitude of trend whereas the Pettitt Mann-Whitney test has been used to detect the single change point. The findings of tests revealed that the area, production and productivity of tea in India have significantly increasing trend over the years.

**Keywords** Innovative trend analysis, Modified Mann Kendall test, Non parametric, Pettitt Mann-Whitney test, Tea cultivation.

## INTRODUCTION

Tea (*Camellia sinensis*) is one of the most important hot beverage of the world with an apparent global consumption of 6173 million kgs in 2021. India is the second largest producer of tea in the world after China with a production of 1344.40 M.kg from an area of 619773.70 ha for the year 2021-2022 (Official website of Tea board, Government of India). In addition to this, India is the largest consumer of the beverage with domestic consumption of 1168 M.kg in the year 2021-2022. The domestic population consumes around 87% of the total tea produced in India. In the year 2021-2022, India ranks four in the quantity of tea exported the most tea globally after Kenya, China and Sri Lanka (200.79 M.kg) in the year 2021-2022. India with its diverse climatic conditions, produces wide varieties of teas suitable to different tastes and preferences of consumers. World's famous teas like Darjeeling, Assam, Sikkim, Nilgiris and Kangra which are famed for their delicate flavor are produced in India. In India, tea cultivation is mainly spread over Assam, West Bengal, Tamil Nadu, Kerala and Karnataka states where it is an important industry and a way of income and employment generation.

Arya V. Chandran[1]*, D. S. Dhakre[2], Debasis Bhattacharya[3], K. A. Sarkar[4]

[2,4]Assistant Professor, [3]Professor of Head
[1,2,3,4]Department of Agriculture Statistics, Palli Siksha Bhavan Visva-Bharati University, Sriniketan 731226, WB, India

Email :aryavc2926@gmail.com
*Corresponding author

Thus, tea industry is of considerable importance in the national economy of India in terms of income generation, earning foreign exchange, employment generation and contribution to the national exchequer. Hence it is important to study the growth and trend in the data of area, output and yield of tea to determine the performance of the crop.

Trends are the long term movements or patterns that can be seen in time series dataset. Trend analysis is used to indicate whether the data is moving up or down for some or all of the time span. As the majority of India's population depends on agriculture for their livelihood, it signifies a better understanding of trends in area under cultivation, production and productivity of crops. By using the Mann-Kendall (MK) test and Sen's slope approach, Sharma and Singh (2019) studied the trend in area, production and productivity of pulse crops in different districts of Rajasthan and grouped the districts with increasing and decreasing trends. In the Rasuwa district of Nepal, Dawadi *et al.* (2022) discovered an increasing trend in potato production and a decreasing trend in the production of millets and wheat using non-parametric trend analysis. Vaidya *et al.* (2018) used Sen's slope method to quantify the trends in climatic variables as well as the productivity of crops and showed that the productivity of all crops has changed significantly over time except barley. Numerous studies have been carried out using non-parametric methods to analyze the trends in various climatic factors (Sanusi *et al.* 2021, Singh and Kumar 2022). Esit (2023) found that the Innovative Trend Analysis (ITA) with a significant test is more sensitive than the Mann-Kendall test by using monthly and annual hydro metrological data. Using annual and seasonal rainfall data, Wu and Qian (2017) showed that the advantage of ITA over the MK test is that ITA is able to observe sub trends in the series. The present study is confined to the trend analysis of area, production and productivity of tea in India by using different non-parametric procedures. Non-parametric tests have the advantage of being more reliable and adaptable than parametric tests and are no affected by the presence of outliers and autocorrelation.

## MATERIALS AND METHODS

This section contains a description of the materials used, including the various data sets and the technique employed for the study.

### Data

The secondary data regarding the area, production and productivity of tea in India from 1960 to 2021 are obtained from various publications and the website of the tea board, India.

### Methodology

This section gives a brief description of the statistical methods employed in the present study.

### Wallis and Moore phase-frequency test

The Wallis and Moore phase-frequency test is a non-parametric test that identifies the autocorrelation present in the data. It is used for testing the null hypothesis $H_0$, which states that the series comprises of random data, against the alternative hypothesis $H_1$, which states that the series is significantly different from being random. The test statistics of the Wallis and Moore phase-frequency test for n number of observations, higher than 30 is as follows:

$$z = \frac{\left| h - \frac{2n-7}{3} \right|}{\sqrt{\frac{16n-29}{90}}}$$

Where h is the number of phases, excluding the first and last phases. For n $\leq 30$, a continuity correction of -0.5 will be included in the denominator. If the series is free from autocorrelation, trend analysis can be done by the Mann-Kendall (MK) test, otherwise, Modified Mann-Kendall (MMK) test is to be applied.

### Mann-Kendall (MK) test

The Mann-Kendall (MK) test is a non-parametric test to assess whether there is a monotonic trend in the series. A monotonic trend means that the variable under study is consistently increasing or decreasing over time. The monotonic trend may be upward or downward and need not be linear. As it is a non-parametric test, the data does not have to meet the assumptions of normality but should not be serially correlated. The MK tests can detect the presence of significant trend in either direction only, but fails assess the magnitude of change. The test procedure for the MK test is as follows:

$H_0$: There is no monotonic trend present in the data

$H_1$: Monotonic trend is present in the data

For calculating MK statistics, we need to determine the sign of all n(n-1)/2 possible differences $y_j$-$y_k$, where the series is arranged in the order of time and j>k. sgn($y_j$-$y_k$) is an indicator function which takes the values as follows:

$$sgn\,(y_j\text{-}y_k) = \begin{cases} 1, \textit{if}\,(y_j\text{-}y_k) > 0 \\ 0, \textit{if}\,(y_j\text{-}y_k) = 0 \\ \text{-}1, \textit{if}\,(y_j\text{-}y_k) < 0 \end{cases}$$

The MK test statistic, S is calculated as the difference between the number of positive differences and number of negative differences. If S is a positive number, observations in recent time tend to be larger than earlier observation which means that the trend is positive and vice versa.

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn\,(y_j\text{-}y_k)$$

The test is conducted using a normal approximation with the zero mean and the variance as follows:

$$Var\ (S) = \frac{[n\,(n\text{-}1)\,(2n\text{+}5)]}{18}$$

For samples where n >10, the variance term is modified as:

$$Var\,(S) = \frac{1}{18}\left[\, n\,(n\text{-}1)\,(2n\text{+}5)\text{-} \sum_{p\text{-}1}^{g} t_p\,(t_p-1)\,(2t_p+5) \,\right]$$

Where g is the number of tied group, $t_p$ is the tie length of $p^{th}$ group. Then,

$$Z_{-MK} = \begin{cases} \dfrac{S-1}{\sqrt{Var\,(S)}}, \textit{if}\,S > 0 \\ 0, \textit{if}\,S = 0 \\ \dfrac{S+1}{\sqrt{Var\,(S)}}, \textit{if}\,S < 0 \end{cases}$$

A positive value of $Z_{MK}$ indicates that the data tend to increase with time and negative value indicate decreasing trend. By comparing the Z value with the critical value $Z_{\alpha}$ at α % level of significance, the significant presence of the trend can be evaluated.

**Modified Mann-Kendall test**

A modified version of MK test is based on modified variance of S. Yue *et al.* (2002) have proven that the variance of distribution of S will increase by the presence of positive autocorrelation. So in the presence of significant positive auto correlation, a correction factor is used to improve the value of *Var (S)*.

$$\text{Var* (S)} = \eta \times \frac{(n\,(n\text{-}1)\,(2n\text{+}5)}{18}$$

$$\text{where } \eta = 1 + \frac{2}{n\,(n\text{-}1)\,(n\text{-}2)} \times \sum_{i=1}^{n\text{-}1} (n\text{-}1)\,(n - i - 1)\,(n\text{-}i\text{-}2)\,r_i$$

Where $r_i$ is the auto correlation coefficient of ranks at lag i. $r_i$ is determined after removing a suitable non parametric trend estimator from the initial data. The slope of the trend was estimated by using Sen's slope approach.

$$r_i = \frac{\sum_{k=1}^{n\text{-}1} (R_k - \bar{R})\,(R_{k+1} - \bar{R})}{\sum_{k=1}^{n} (R_k - \bar{R})^2}$$

where $\bar{R}$ is the average rank and Rk is the rank of $y_k$. The test statistic is calculated as:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var^*\,(S)}}, \textit{if}\,S > 0 \\ 0, \textit{if}\,S = 0 \\ \dfrac{S+1}{\sqrt{Var^*\,(S)}}, \textit{if}\,S < 0 \end{cases}$$

Like MK test, positive Z value indicates that an increasing trend is present in the data and negative Z value indicates a decreasing trend. Significant trend can be identified by comparing Z value with critical value Zα.

**Sen's slope**

The Sen's slope estimator is a non-parametric procedure developed by Sen which gives the magnitude of a time series trend. Here, the trend is assumed to be linear, and the test is not affected by outliers. Sen's slope estimate for a number of N data sample pairs is calculated as follows:

$$Q_k = \left(\frac{y_j - y_{jk}}{j - k}\right)$$

where $y_j$ and $y_k$ are the observed values at time *j* and k respectively and $j>k$. If there are n values of $y_j$ in the time series, there will be N = n (n-1)/2 slope estimates (Aditya *et al*. 2021). The Sen's slope estimator is calculated as the median of these N slope estimate values.

*Sen's slope estimator, β=Median ($Q_k$)*

If β > 0, the trend is upward and if β <0, the trend is downward.

**Innovative trend analysis (ITA)**

Sen (2012), proposed innovative trend analysis (ITA), a technique used to detect the trend in time series data. It is free from the assumptions of normality and auto correlation. In this technique the entire time series is divided into two equal parts. After arranging the sub series in ascending order, the first half of the series placed on X axis and second half on the Y axis of the Cartesian coordinate system. Data values on the scatter plot may be collected on the $45^0$ -1:1 linear line. If the data points accumulate on 1:1 line, it indicates the absence of trend in the data. While if the data points fall above or below the 1:1 line, this indicates that there is a positive or negative trend respectively. A positive slope of ITA indicates increasing trend and negative slope indicates decreasing trend (Sen 2012). The slope if trend in ITA is calculated as follows:

$$Slope,s = \frac{2\,(\bar{y}_2 - \bar{y}_1)}{n}$$

where $\bar{y}_1$ and $\bar{y}_2$ are the arithmetic mean of first and second half of the series and n is the total number of observations. The pdf of slope abides with the Normal distribution with a zero mean and standard deviation σs. The slope is significant if the calculated slope value is greater than the critical value $s_{cri}$ (Sen 2017). The $z_{\alpha/2}$ value $_{produced}$ from the standard normal distribution is referred to as $s_{cri}$ (Esit 2023).

At α % level of significance, the trend slope confidence interval is obtained as follows:

$$CL_{(1-\alpha)} = 0 \pm s_{cri}\,\sigma_s$$

If the innovative trend slope value falls outside the lower and upper confidence limits, then the alternative hypothesis $H_1$ is adopted, and that indicates the existence of a significant trend (Sen 2017).

ITA trend indicator is used to compare trends in two time series that may have different magnitudes. To allow direct comparison, the indicator is multiplied by 10 to get the same scale as that of the Mann–Kendall test and linear regression analysis. ITA trend indicator, D is calculated as:

$$D = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_{2t} - y_{1t})}{\bar{y}_1}$$

where $y_{1i}$ is the i$^{th}$ value of first half of series and $y_{2i}$ is the $i^{th}$ value of second half series. $\bar{y}_1$ is the mean of first half of the series. Positive value of trend indicator (D) indicates positive trend and negative D value indicates negative trend.

**Pettitt Mann-Whitney (PMW) test**

The PMW test detects the single change point when applied to a time series data. It tests the null hypothesis $H_0$ that there does not exist a significant change point in the series against the alternative hypothesis $H_1$ that states about the existence of change point in the series. Let $y_1, y_2, \ldots, y_n$ be a time series of n observation and let t is the most likely change point. Two samples of time series $y_1, y_2, \ldots, y_t$ and $y_{t+1}, y_{t+2}, \ldots, y_n$ can be derived by dividing the whole series at time $t$. The PMW test statistic is defined as:

$$K_t = max \mid U_t \mid$$

where Ut represents changing point which is obtained as:

$$U_t = \sum_{t=1}^{t} \sum_{j=t+1}^{n} sgn\,(y_t - y_j)$$

$$\text{where } sgn\,(yi - yj) = \begin{cases} +1 \text{ if } (y_i - y_j) > 0 \\ 0 \text{ if } (y_i - y_j) = 0 \\ -1 \text{ if } (y_i - y_j) < 0 \end{cases}$$

To identify the changing point in time series, we plot $U_t$ against t. The $U_t$ will increase up to the change point and then start to decrease. This will happen even it is a local change point. This increase in $U_t$ value followed by a decrease may occur several times in a time series, indicating several local change points. The most significant change point will give by $K_t$ which is the point where $U_t$ value is maximum. The approximate probability for a change point to be significant is calculated as follows:

$$p \cong 2e\left[\frac{-6\,k^2}{(n^3 + n^2)}\right]$$

## RESULTS AND DISCUSSION

The Wallis and Moore phase-frequency test was applied to the data set to find out whether autocorrelation is present in the data. From the test results, it was found that time series data regarding area, production and productivity of tea in India showed significant autocorrelation (Table 1).

**Table 1.** Wallis and Moore phase-frequency test results obtained by using trend package of R software.

| | Area (h=3, n=62) | Production (h=27, n=62) | Productivity (h=30, n=62) |
|---|---|---|---|
| Z statistics | 10.7 | 3.67 | 2.45 |
| p value | <0.01 | <0.01 | <0.05 |

**Table 2.** Trend analysis of data (Results obtained by using modifiedmk and trend package in R).

| | Z value (MMK test) | Sen's slope | Changing point ($K_t$) |
|---|---|---|---|
| Area | 4.29** | 4841.83 | 1990*** |
| Production | 4.70** | 15070.20 | 1990*** |
| Productivity | 4.48** | 18.12 | 1987*** |

As all the data sets show autocorrelation, the modified Mann-Kendall test is used to identify the presence of a monotonic trend in the data. Sen's slope approach is used to find out the slope of the trend. Table 2 shows the results of the trend analysis. From the above table, it is clear that the Z value for the modified Mann-Kendall test is significantly positive. It indicates the presence of an increasing trend in the area, production and productivity of tea in India over the years. Similarly, Sen's slope estimator also indicates the upward trend of the data. From the PMW test, the changing points for area and production were found to be 1990 and 1987 for productivity, respectively.

Further, the trend of the data sets was analyzed using innovative trend analysis results (Fig. 1). In all

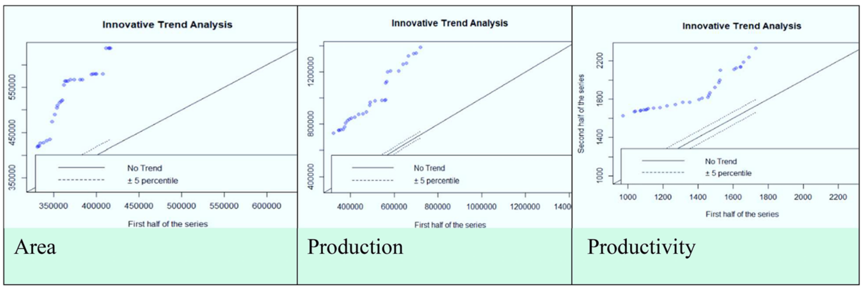


**Fig. 1.** Plot of innovative trend analysis.

**Table 3.** Results of innovative trend analysis (Obtained by using "trendchange" package in R).

| | ITA trend slope | Upper confidence interval at 5% level of significance | Lower confidence interval at 5% level of significance | ITA trend indicator |
|---|---|---|---|---|
| Area | 5221.79 | 285.96 | -285.96 | 4.38 |
| Production | 16038.94 | 528.38 | -528.38 | 9.90 |
| Productivity | 16.63 | 0.990 | -0.990 | 3.81 |

three data sets, all the points accumulate above the 1:1 line, revealing a positive trend in area, production and productivity. The data on area, production and productivity of tea in India were shown to have positive ITA trend slopes, indicating an upward trend (Table 3). All the trend slope values falls above the upper confidence value at 5% level of significance, shows the presence of significant trends in the data. The positive values of ITA trend indicator likewise indicate an upward trend.

## CONCLUSION

The study evaluated the trend in area, production and productivity of tea in India. The Wallis and Moore phase-frequency test revealed that all three time series datasets are serially correlated. Using the Modified Mann-Kendal test for trend analysis, it was discovered that India's acreage, production, and productivity of tea are all significantly growing over time. The same conclusion was drawn from the positive value of Sen's slope estimator. Additionally, the Innovative Trend slope and trend indicator also indicate a positive trend in the area, production, and productivity of tea.

## ACKNOWLEDGMENT

We express our gratitude to the Department of Agricultural Statistics, Visva-Bharati, for all sorts of assistance provided during this study.

## REFERENCES

Aditya F, Gusmayanti E, Sudrajat J (2021) Rainfall trend analysis using Mann-Kendall and Sen's slope estimator test in West Kalimantan. *IOP Conf Ser* : *Earth Environ Sci* 893 (1): In press. https://doi.org/10.1088/1755-1315/893/1/012006

Dawadi B, Shrestha A, Acharya RH, Dhital YP, Devkota R (2022) Impact of climate change on agricultural production: A case of Rasuwa District, Nepal. *Reg Sust* 3(2) : 122–132. https://doi.org/10.1016/j.regsus.2022.07.002

Esit M (2023) Investigation of innovative trend approaches (ITA with significance test and IPTA) comparing to the classical trend method of monthly and annual hydrometeorological variables: A case study of Ankara region, Turkey. *J Water Clim Chang* 14(1) : 305–329. https://doi.org/10.2166/wcc.2022.356

Sanusi W, Abdy M, Sulaiman (2021) Innovative trend analysis of annual maximum precipitation in Gowa regency. *J Phys Conf Ser* 1899 (1): In press. https://doi.org/10.1088/1742-6596/1899/1/012092

Şen Z (2012) Innovative trend analysis methodology *J. Hydrol Eng* 17(9): 1042–1046. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000556

Şen Z (2017) Innovative trend significance test and applications. *Theor Appl Climatol* 127(3–4) : 939–947. https://doi.org/10.1007/s00704-015-1681-x

Sharma S, Singh SP (2019) Production analysis: A non-parametric time series application for pulses in Rajasthan. *Int J Curr Microbiol Appl Sci* 8(01) : 2438–2450. https://doi.org/10.20546/ijcmas.2019.801.257

Singh P, Kumar N (2022) Analysis of trend and temporal variability in rainfall over Northeast India. *MAUSAM* 73 (2) : 307-314. https://doi.org/10.54302/mausam.v73i2.5479.

Tea Board India (2022) Statistics. [Data sets]. https://www.teaboard.gov.in

Vaidya P, Randhawa S, Sharma P, Sharma YP, Satyarthi K, Randhawa SS (2018) Climate variability and crop productivity in Himalayan ecosystem: A case study of Kullu district *MAUSAM* 69(4): 563-570. https://www.researchgate.net/publication/332710340

Wu H, Qian H (2017) Innovative trend analysis of annual and seasonal rainfall and extreme values in Shaanxi, China, since the 1950s. *Int J Climatol* 37(5) : 2582–2592. https://doi.org/10.1002/joc.4866

Yue S, Pilon P, Phinney B, Cavadias G (2002) The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol Process* 16 : 1807-1829. https://doi.org/10.1002/hyp.1095