

Application of Change Point Analysis (CPA) to monthly temperature in Tamil Nadu, India

S. G. PATIL

Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu – 641 003, India

(Received 15 June 2018, Accepted 18 February 2019)

e mail : san.santoshpatil@gmail.com

सार – तापमान और वर्षा पैटर्न में परिवर्तन के आधार पर जलवायु परिवर्तन की विशिष्टता बताई जाती है। CPA का उपयोग यह पता लगाने के लिए किया जाता है कि किसी क्रमबद्ध डेटा में कोई परिवर्तन हुआ है या नहीं। वर्तमान अध्ययन तक की अवधि में तमिलनाडु के प्रत्येक जिले के लिए 2002 से 1901 CPA का उपयोग करते हुए ऐतिहासिक मासिक तापमान डेटा में एक प्रक्रिया के औसत में परिवर्तन का पता लगाने से संबंधित है। CPA के विश्लेषण से पता चलता है कि तमिलनाडु के जिले पिछले कुछ वर्षों से तापमान में परिवर्तन प्रदर्शित कर रहे हैं। वर्णनात्मक आंकड़ों और प्रवृत्ति विश्लेषण से पता चला है कि चेन्नै और लगभग सभी जिलों में प्रेक्षित अधिकतम भिन्नता प्रत्येक महीने के हिसाब से महत्वपूर्ण रुझान दर्शाती है।

ABSTRACT. Climate change is characterized by observing changes in temperature and rainfall pattern. CPA is used order to detect whether any changes have occurred on a series of time ordered data. Present study deals with detection of change in the mean of a process in historical monthly temperature data using CPA for each districts of Tamil Nadu for the period of 1901 to 2002. The analysis of CPA shows that the districts in Tamil Nadu are exhibits changes in temperature over the years. Descriptive statistics and trend analysis revealed that maximum variation observed at Chennai and almost all the districts shows month wise significant trend.

Key words – Change point, Mann Kendall, Sen Slope.

1. Introduction

Climatic variability over years has turned out to be a predominant pattern. The last few years have been marked by many changes in climate un-predictable rainfall, flooding, warmer conditions, heat waves etc. This trend of appears to be merely the continuation of a pattern of variability. The tendency of climate to change relatively suddenly has been one of the most surprising.

Climate change is a long-term change in the statistical distribution of weather patterns over periods of time. Trend and Change point analysis of climatologically time series will help to understand the problem of climate change.

The climate change studies based on long historic climate data which exhibits non-homogeneity. One type of non-homogeneity in long meteorological time series is sudden shifts of the mean level. Such abrupt changes in the immediate environment may shift the farm practices, causes pests & disease incidence may affect the grain size, harvest time. Thus there are situations where identification of such change helps to know shifting of crop managements or pattern.

Change point analysis is the process of detecting distributional changes within time-ordered observations. In present study, change point analysis (CPA) is applied to temperature data of each district of Tamil Nadu for detecting any change in the mean of a process in historical data. CPA assumes that the process (time series) must be distributed identically, and the observations must be independent (at least there is no strong autocorrelation). The study also includes the basic characteristics of Temperature e.g., Descriptive statistics, non- parametric trend analysis (Mann Kendall test) etc.

2. Data base and climatic status of the zone

Tamil Nadu is situated between 8° 5' and 13° 35' North latitude and between 76° 15' and 80° 20' East longitude. The climate of the state ranges from dry sub-humid to semi-arid.

Data source: Monthly maximum temperature data were obtained from the http://indiawaterportal.org/met_data/ for the period 1901-2002 of 29 districts of Tamil Nadu.

Software used: The analysis is done by using R 3.1.0- package.

TABLE 1
Descriptive statistics of monthly maximum temperature

District	Min	Max	Q1	Median	Q3	Mean	SD	CV
Ariyalur	27.66	36.65	30.37	32.72	34.08	32.34	2.17	6.7
Chennai	26.8	39.58	29.83	33.38	35.5	32.96	3.15	9.56
Coimbatore	27.78	34.37	29.34	29.95	31.58	30.43	1.48	4.86
Cuddalore	27.44	37.15	30.19	32.91	34.41	32.46	2.42	7.47
Dharmapuri	27.08	36.5	29.74	30.78	33.09	31.32	2.27	7.26
Dindigul	26.21	32.46	27.73	28.32	29.77	28.73	1.36	4.74
Erode	26.92	34.62	28.61	29.34	31.29	29.93	1.78	5.94
Kanchipuram	26.94	38.68	29.82	33.18	35.12	32.74	2.96	9.03
Karur	25.97	33.34	28.17	28.89	30.38	29.22	1.61	5.52
Madurai	27.33	33.18	28.8	29.35	30.64	29.72	1.22	4.11
Nagapattinam	27.33	36.93	30.02	32.84	34.42	32.39	2.45	7.57
Namakkal	28.27	36.76	30.83	31.75	33.39	32.07	1.91	5.95
Nilgiris	27.72	34.69	29.65	30.43	32.1	30.81	1.62	5.25
Perambalur	27.85	36.4	30.66	32.54	33.84	32.29	2.02	6.26
Pudukkottai	27.76	35.06	30.29	31.7	32.73	31.53	1.61	5.12
Ramanathapuram	27.35	33.44	29.26	29.84	30.92	30.05	1.19	3.96
Salem	28.32	37.34	31.1	32.38	34.07	32.62	2.11	6.45
Sivaganga	28.24	34.81	30.38	31.17	32.2	31.29	1.33	4.24
Thanjavur	27.51	35.86	30.21	32.28	33.42	31.9	1.93	6.06
Theni	27.22	32.9	28.69	29.2	30.67	29.63	1.23	4.14
Thoothukudi	25.72	30.31	26.64	27.11	28.3	27.44	0.99	3.62
Tiruchirappalli	27.45	35.2	29.99	31.21	32.46	31.25	1.74	5.58
Tirunelveli	22.71	26.95	23.69	24.11	25.19	24.4	0.9	3.7
Tiruvallur	26.74	39.47	29.92	33.4	35.21	32.86	3.06	9.33
Tiruvannamalai	26.74	37.19	29.76	32.21	33.83	31.88	2.51	7.89
Tiruvarur	27.69	36.27	30.44	32.62	33.81	32.21	2	6.2
Vellore	26.58	37.28	29.64	31.77	33.62	31.71	2.57	8.11
Viluppuram	27.36	37.3	30.23	32.86	34.39	32.41	2.43	7.49
Virudhunagar	27.93	33.24	29.04	29.59	30.9	29.96	1.15	3.82

3. Methodology

3.1. Descriptive statistics

Descriptive statistics includes of minimum value, maximum value, median, quartiles and coefficient of variation. Box plot is five number summary shows minimum, maximum, median, 1st and 3rd quartiles.

3.2. Trend analysis

Trend analysis of a time series consists of the magnitude of a trend and its statistical significance in a

time series. Obviously, different workers have used different methodologies for trend detection. Kundzewicz (2004) has discussed the change detection methodologies for hydrological data.

In general, the magnitude of trend in a time series is determines either using regression analysis (parametric test) or using Sen's estimator method (non-parametric). Both these methods assume a linear trend in the time series. It checks the null hypothesis of no trend versus the alternative hypothesis of the existence of increasing or decreasing trend.

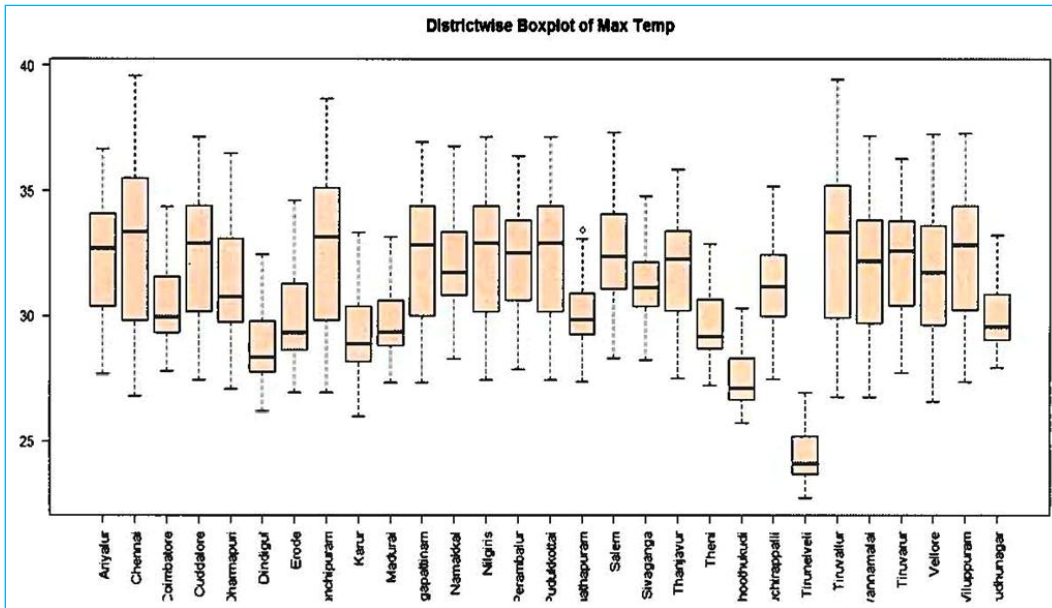


Fig. 1. Box plot of monthly maximum temperature of districts of Tamil Nadu

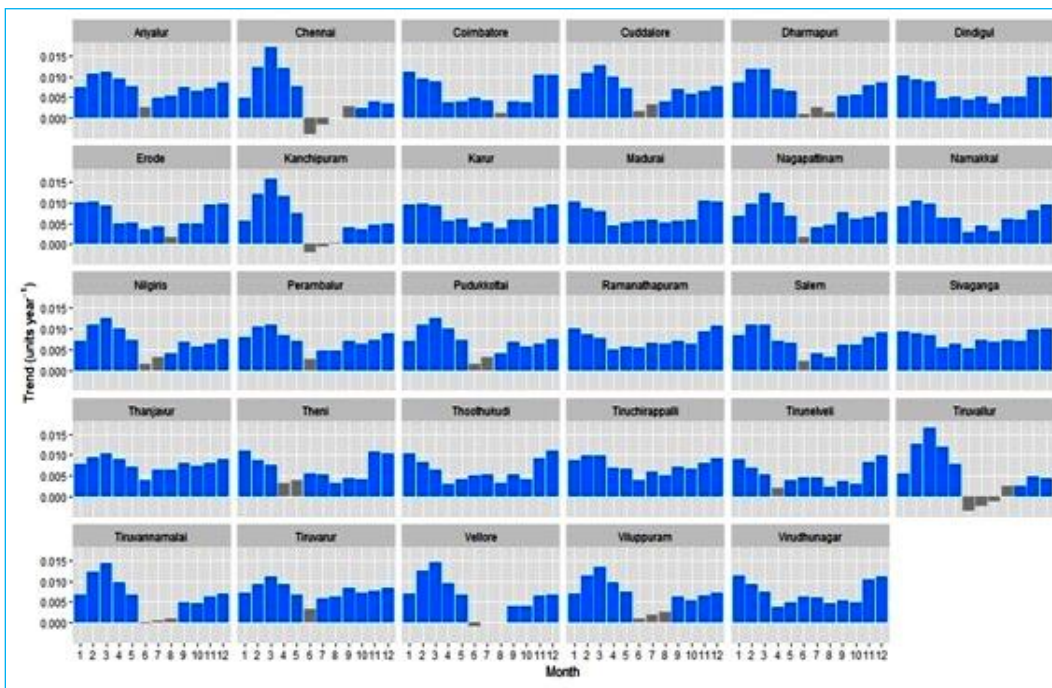


Fig. 2. Mann Kendall test monthly maximum temperature of Districts of Tamil Nadu

3.3. Regression analysis

Trend is one of the important component of time series data. Using simple regression we can model the linear trend at time t as:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

3.4. Sen's method

The Sen's slope estimator is a linear slope estimator that works most effectively on monotonic data. Unlike linear regression, it is not greatly affected by data errors, outliers, or missing data. Sen's slope estimator has been widely used for determining the magnitude of trend. The

TABLE 3

Change point monthly maximum temperature of districts of Tamil Nadu

District	intercept	trend	P	1 st CP	intercept	trend	P	2 nd CP	intercept	trend	P	3 rd CP	intercept	trend	P
Ariyalur	32.14	2.05E-04	0.66	1954(6)	32.01	1.04E-03	0.4	1982(2)	32.98	-5.90E-04	0.77	-	-	-	-
Chennai	32.85	8.32E-05	0.9	1954(6)	32.6	1.53E-03	0.05	-	-	-	-	-	-	-	-
Coimbatore	30.27	7.85E-05	0.64	1982(1)	30.97	-6.37E-04	0.64	-	-	-	-	-	-	-	-
Cuddalore	32.27	1.83E-04	0.72	1954(6)	32.03	1.92E-03	<0.001	-	-	-	-	-	-	-	-
Dharmapuri	31.22	-1.45E-04	0.74	1957(8)	31.23	1.04E-03	0.11	-	-	-	-	-	-	-	-
Dindigul	28.55	6.98E-05	0.65	1982(1)	29.37	-8.55E-04	0.5	-	-	-	-	-	-	-	-
Erode	29.75	1.25E-04	0.54	1982(1)	30.45	-6.10E-04	0.71	-	-	-	-	-	-	-	-
Kanchipuram	32.61	1.06E-04	0.87	1954(6)	32.36	1.63E-03	0.03	-	-	-	-	-	-	-	-
Karur	29.03	1.03E-04	0.57	1982(1)	29.85	-8.21E-04	0.58	-	-	-	-	-	-	-	-
Madurai	29.52	1.14E-04	0.39	1982(12)	30.34	-5.52E-04	0.65	-	-	-	-	-	-	-	-
Nagapattinam	32.19	2.42E-04	0.64	1954(6)	31.92	2.06E-03	<0.001	-	-	-	-	-	-	-	-
Namakkal	31.88	1.36E-04	0.54	1980(4)	32.46	3.52E-04	0.82	-	-	-	-	-	-	-	-
The Nilgiris	32.27	1.83E-04	0.72	1954(6)	32.03	1.92E-03	<0.001	-	-	-	-	-	-	-	-
Perambalur	32.1	8.87E-05	0.7	1982(2)	32.91	-7.01E-04	0.71	-	-	-	-	-	-	-	-
Pudukkottai	32.27	1.83E-04	0.72	1954(6)	32.03	1.92E-03	<0.001	-	-	-	-	-	-	-	-
Ramanathapuram	30.06	-1.99E-03	0.15	1918(9)	29.87	1.88E-04	0.5	1966(12)	29.77	1.94E-03	0.23	1982(3)	30.58	2.18E-04	0.84
Salem	32.43	1.37E-04	0.59	1978(9)	32.87	1.01E-03	0.51	-	-	-	-	-	-	-	-
Sivaganga	31.32	-2.26E-03	0.13	1918(9)	31.07	4.90E-04	0.33	1954(6)	30.99	1.99E-03	0.28	1982(3)	31.89	-1.12E-04	0.93
Thanjavur	31.9	-1.82E-03	0.4	1918(9)	31.68	4.96E-04	0.51	1954(6)	31.65	1.65E-03	0.54	1982(3)	32.5	-1.54E-04	0.93
Theni	29.47	4.19E-05	0.76	1982(3)	30.25	-6.93E-04	0.55	-	-	-	-	-	-	-	-
Thoothukudi	27.26	1.27E-04	0.25	1982(3)	27.91	3.79E-05	0.97	-	-	-	-	-	-	-	-
Tiruchirappalli	31.05	1.06E-04	0.59	1982(2)	31.9	-7.91E-04	0.63	-	-	-	-	-	-	-	-
Tirunelveli	24.25	8.64E-05	0.39	1982(3)	24.8	-1.21E-04	0.88	-	-	-	-	-	-	-	-
Tiruvallur	32.76	5.02E-05	0.94	1954(6)	32.53	1.45E-03	0.06	-	-	-	-	-	-	-	-
Tiruvannamalai	31.74	8.09E-05	0.88	1954(6)	31.55	1.53E-03	0.02	-	-	-	-	-	-	-	-
Tiruvarur	32.25	-2.21E-03	0.31	1918(11)	31.99	5.46E-04	0.49	1954(6)	31.7751	1.60E-03	0.27	1977(10)	32.45	1.50E-03	0.26
Vellore	31.58	3.79E-05	0.94	1954(6)	31.42	1.39E-03	0.03	-	-	-	-	-	-	-	-
Viluppuram	32.24	1.32E-04	0.8	1954(6)	32.03	1.73E-03	<0.001	-	-	-	-	-	-	-	-
Virudhunagar	29.78	6.77E-05	0.59	1981(12)	30.56	-3.45E-04	0.74	-	-	-	-	-	-	-	-

CP : change point; E-0k = 10^{-k}

approach involves computing slopes for all the pairs of ordinal time points using the median of these slopes as an estimate of all the overall slope. Sen's method proceeds by calculating the slope as a change in measurement per change in time.

$$Q = \frac{x_j - x_k}{j - k}$$

where, Q = slope between data points x_j and x_k ; x_j and x_k are datapoints at time j and k ($j > k$) respectively.

3.5. Mann-Kendall test

To ascertain the presence of statistically significant trend in maximum temperature, non-parametric Mann-Kendall (MK) test has been employed, which tested for the significance at 95% level. Since there is fluctuation presents in the weather parameters, non-parametric Mann-

Kendall test is useful because its static is based on the sign of differences, not directly on the values of random variable and therefore trends determined is less affected by the fluctuations. Mann-Kendall test is applicable to the detection of monotonic trend in time series.

The test statistic S is calculated using the formula:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k)$$

$$\text{sgn}(x_j - x_k) = \begin{cases} +1 & \text{if } x_j - x_k > 0 \\ 0 & \text{if } x_j - x_k = 0 \\ -1 & \text{if } x_j - x_k < 0 \end{cases}$$

where, n is the number of observed data series, x_j and x_k are the values in periods j and k respectively,

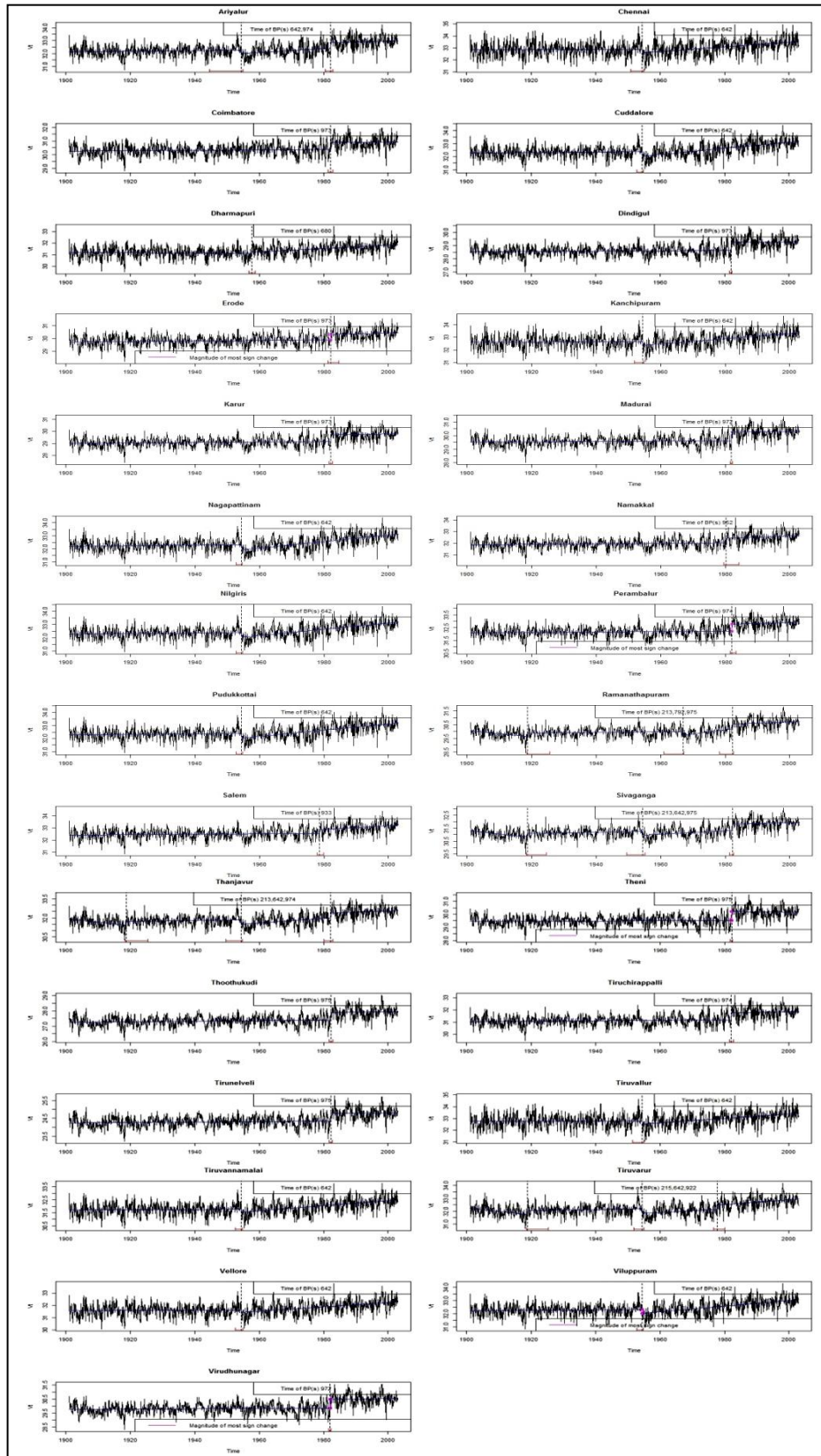


Fig. 3. CPA plots for monthly maximum temperature of districts of Tamil Nadu

$j > k$. For $n \geq 10$, the sampling distribution of S is as follows:

$$Z = \begin{cases} \frac{s - 1}{\sqrt{\text{VAR}(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S + 1}{\sqrt{\text{VAR}(S)}} & \text{if } S < 0 \end{cases}$$

where, $\text{VAR}(S)$ is determined as:

$$\text{VAR}(S) = \frac{1}{18} [n(n - 1)(2n + 5) - \sum_{p=1}^q t_p(t_p - 1)(2t_p + 5)]$$

where, q is the number of tied groups and t_p is the number of data values in the p^{th} group.

3.6. Detecting breakpoints/change points

Most of the existing change detection techniques are unable to account for seasonal variation and analyze time series by aggregating data by year or season or compare specific periods (e.g., summer) between years (Coppin *et al.*, 2004).

Jaxk Reeves *et al.* (2007) review article enumerates, categorizes, and compares many of the methods that have been proposed to detect undocumented change points in climate data series.

Verbesselt *et al.* (2010a&b) investigate the application of a generic change detection method that accounts for seasonality and enables the detection of trend change within the time series.

Breaks For Additive Seasonal and Trend (BFAST) is an iterative algorithm that combines the decomposition of time series into seasonal, trend, and remainder component with methods for detecting changes. An additive decomposition model is used to iteratively fit a piecewise linear trend and a seasonal model (Haywood and Randall, 2008). The general model is of the form:

$$Y_t = T_t + S_t + e_t \quad : t \in T$$

where, t is time period, Y_t is the observed value, T_t is the trend component, S_t the seasonal component and e_t the error term.

Each time before fitting the T_t and S_t components, it is tested whether abrupt changes are occurring. The ordinary least squares residuals-based moving sum (MOSUM) test, is selected for this purpose (Zeileis and

Kleiber, 2005). If the test indicates significant change ($\alpha = 0.05$), the breakpoints are estimated using the method of Bai and Perron (2003). This method minimizes the Bayesian Information Criterion (BIC) to determine the optimal number of breaks and iteratively minimizes the residual sum of squares to estimate the optimal break positions. Change point analysis can detect more stable shift than control charts.

4. Results and discussion

Monthly maximum temperature characteristics of Tamil Nadu reported in Table 1 and visualized in Fig. 1. Average temperature of Tamil Nadu is 30.98 °C, maximum temperature observed in Chennai district (39.58 °C) followed by Tiruvallur (39.47 °C) with highest variation (CV) of 9.56 and 9.33 respectively. Thoothukudi district shows lowest variation of 3.62 with mean temperature of 27.44 °C.

4.1. Trend analysis

Annual and monthly trend for individual month and districts of Tamil Nadu were studied using Mann Kendall test and reported in Table 2 and Fig. 2. Mann Kendall test shows Chennai, Kanchipuram and Tiruvallur exhibits insignificant increasing trend while other districts of Tamil Nadu exhibits significant increasing trend in annual mean and month-wise maximum temperature. Chennai, Kanchipuram and Tiruvallur shows insignificant decreasing trend in June and July.

4.2. Detection of change points

In present study Breaks For Additive Seasonal and Trend (BFAST) performed for each district of Tamil Nadu and plotted trend component of time series to see the change points in the mean of a process in monthly temperature data for the period of 1901 to 2002 and results were presented in Table 3 and Fig. 3. Analysis is done by using R software Table 3 shows that all districts exhibit at least one change point. Around 1954 and 1982, eleven and ten districts show the 1st change in mean process respectively.

Linear trend analysis performed for each district before and after change point observed and results presented in Table 3.

(i) Before 1st change point: All districts of Tamil Nadu shows insignificant trend for monthly maximum temperature.

(ii) After 1st change point: Most of the districts exhibits insignificant trend, but Cuddlore, Kanchipuram,

Nagapattinam, Nilgiris, Pudukkottai, Tiruvannamalai, Vellore and Viluppuram shows positive significant trend.

(iii) Multiple change points: Ariyalur shows two change point, while Ramanathapuram, Sivaganga, Thanjavur and Tiruvarur shows three change points in mean of a process in monthly temperature data over the study period.

5. Conclusions

Long term maximum temperature (1901-2002) over districts of Tamil Nadu were analyzed for monthly trend and changes in mean process is observed over years. Present study revealed that insignificant trend observed during June, July while other month showed significant increasing trend. Around 1954 and 1982 mean process of maximum temperature observed. Before 1st change point the linear trend was non-significant while Cuddlore, Kanchipuram, Nagapattinam, Nilgiris, Pudukkottai, Tiruvannamalai, Vellore and Viluppuram showed significant increasing trend after 1st change point. The analysis showing changes in mean process of maximum temperature thus may be helpful in describing climate change.

Disclaimer

The contents and views expressed in this research paper are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

- Bai, J. and Perron, P., 2003, "Computation and analysis of multiple structural change models", *J. Appl. Econometrics*, **18**, 1-22.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004, "Digital change detection methods in ecosystem monitoring : A review", *International Journal of Remote Sensing*, **25**, 9, 1565-1596.
- Haywood, J. and Randall, J., 2008, "Trending seasonal data with multiple structural breaks. NZ visitor arrivals and the minimal effects of 9/11", *Research report 08/10*, University of Wellington, Victoria, New Zealand, p26.
- Jaxk Reeves, Chen, J., Wang, X. L., Lund, R. B. and Lu, Q., 2007, "A Review and Comparison of Change point Detection Techniques for Climate Data", *Journal of Applied Meteorology and Climatology*, **46**, 900-915.
- Kundzewicz, Z. W. and Robson, A. J., 2004, "Change detection in hydrological records-a review of the methodology", *Hydrol. Sci. J.*, **49**, 1, 7-19
- Verbesselt, J., Hyndman, R., Newnham, G. and Culvenor, D., 2010, "Detecting trend and seasonal changes in satellite image time series", *Remote sensing of Environment*, **114**, 1, 106-115.
- Verbesselt, J., Hyndman, R., Zeileis, A. and Culvenor, D., 2010, "Phenological Change Detection while Accounting for Abrupt and Gradual Trends in Satellite Image Time Series", *Remote Sensing of Environment*, **114**, 12, 2970-2980.
- Zeileis, Achim and Kleiber, Christian, 2005, "Validating multiple structural change models: A case study", *Journal of Applied Econometrics*, **20**, 5, 685-690.