Chapter 4

METHODOLOGY

4.1 General

The preferred techniques for analyzing trends of multiple climatic extremes are nonparametric methods, for they are distribution-free, robust against outliers, and possess a relatively higher potential to tackle non-normally distributed data (4) observed a monotonic tendency with rising precipitation in the state's mountain and coastal areas while analyzing spatio-temporal trend and concentration of monsoon precipitation times series in Damodar River basin region. Furthermore, the monsoon arrives sooner. This chapter provides a detailed methodology for assessing multiple climatic extremes over the Damodar River basin in India.

The widely used multiple climatic extremes, based on maximum and minimum temperatures and precipitation, are evaluated for 15 stations in DRB from 1923 to 2022 in this study. The extreme indices based on precipitation and maxima-minima temperature were analyzed for spatiotemporal research, 3-D pattern characterization of precipitation trends, and multiple change point detection. The rank-based Mann-Kendall (MK) and Sen's slope methods analyze the temporal variability for meteorological time series. Furthermore, this chapter provides the framework for collectively assessing the trend pattern of multiple climatic extremes over a basin. Statistical tests were performed to determine whether there was change or not in chosen climatic variables and the flowchart of the methodology is illustrated in Figure 4.1.

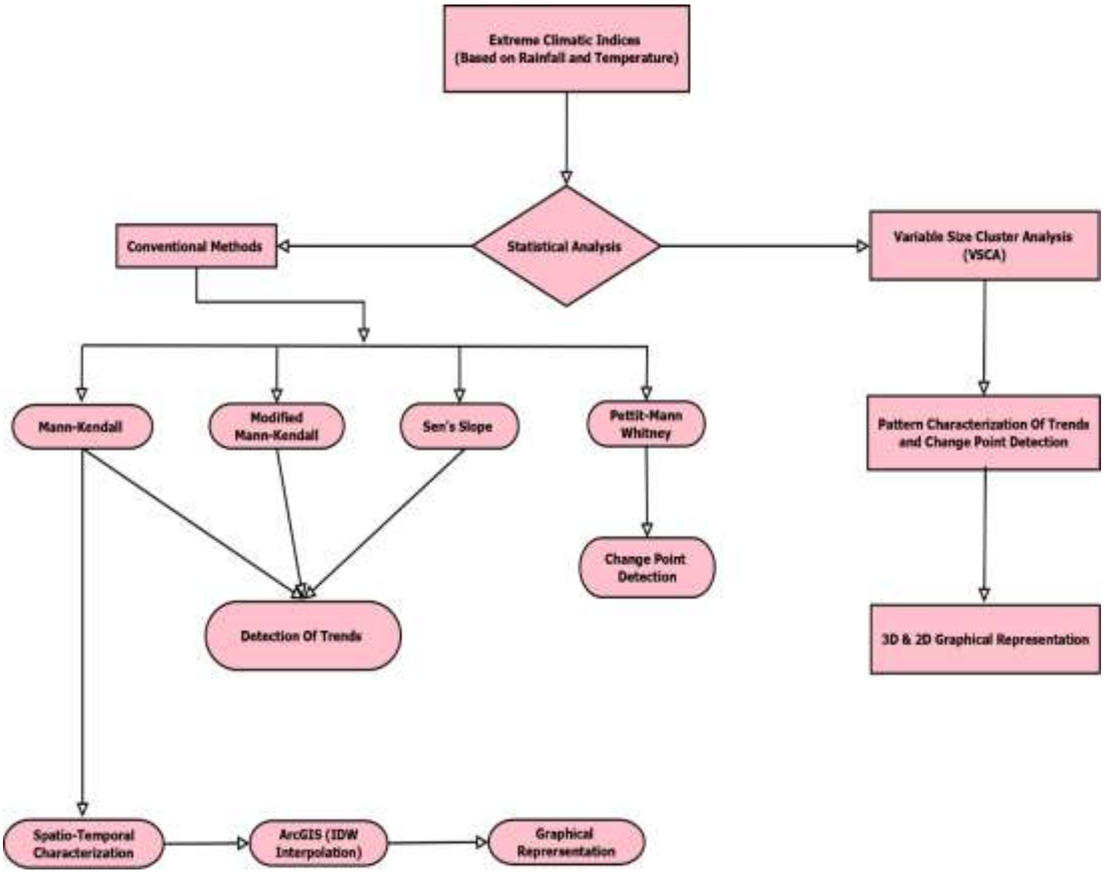

Figure 4.1 Flowchart diagram of methodology

4.2 Statistical Techniques

Statistical analysis involves the process of collecting, organizing, interpreting, and presenting data to uncover patterns, trends, and relationships. It includes descriptive statistics, such as measures of central tendency (mean, median, mode) and variability (range, variance, standard deviation), and inferential statistics, which allow for hypothesis testing, confidence intervals, and predictions through methods like t-tests, ANOVA, and regression analysis. Statistical tools help assess probabilities, correlations, and patterns in time-series data. These techniques are crucial for making informed decisions, validating hypotheses, and drawing meaningful conclusions from data in fields like science, business, and engineering.

4.2.1 Autocorrelation (Serial dependency check and removal)

Autocorrelation checks for serial dependency in long-term time series climatic data. A positive autocorrelation signifies persistence, meaning the data points are dependent on each other. In hydroclimatic time series studies, autocorrelation can bias the MK test results, leading to underestimated test statistics. Many studies, such as Kundu et al. (2015) and Turkes et al. (2002), have identified and addressed this issue. Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between *lagged values* of a time series.

There are several autocorrelation coefficients corresponding to each panel in the lag plot. For example, $r1$ measures the relationship between yt and $yt-1$, $r2$ measures the relationship between yt and $yt-2$, and so on.

The value of rk can be written as shown in equation 4.1

$$r_k = \frac{\sum_{t=1}^{T}(y_t-\bar{y})^2}{\sum_{t=k}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})} \quad 4.1$$

Where T is the length of the time series.

4.2.2 Pre-whitening

Before implementing the M-K test, removing the effect of serial correlation using the pre-whitening procedure (Dash et al., 2009) was done. Trend tests with the pre-whitening method (Yue et al., 2001) ;(Arora et al., 2005; Kishore et al., 2016; Sharma et al., 2019; Suryavanshi et al., 2014) have been implemented to identify a significant trend and persistence in the study. Time-series data is de-trended, Yt, by using the normalized dataset Yt, and the median values of the slope, β. Before analyzing the trend, the time series data is divided by the sample mean, Yt, such that the properties remain unchanged with a mean equal to unity. If the slope tends to zero, there is no requirement for trend analysis.

The process of pre-whitening is used to remove the effect of autocorrelation from a time series before applying statistical tests. This is done by applying a linear model to the time series. The general formula for pre-whitening is shown in equation 4.2:

$$Z_t = X_t - \rho X_{t-1} \quad 4.2$$

Where:

Z_t is the pre-whitened time series at time t,

X_t the original time series value at time t

X_{t-1} is the time series value at time $t-1$ (previous time step),

ρ^ is the estimated autocorrelation coefficient at lag 1.

4.3 Trend Detection Methods

Trend detection of climate variables, such as temperature, precipitation, and sea level, is crucial for understanding long-term climate change. Methods like linear regression are often used to identify gradual increases or decreases in variables over time, such as global warming trends in average temperatures. Moving averages help smooth out short-term variability, revealing clearer trends amidst natural fluctuations. More advanced techniques like Seasonal Decomposition of Time Series (STL) separate the climate data into trend, seasonal, and random components, isolating long-term trends from seasonal effects like monsoon or winter cycles. These methods allow scientists to detect persistent shifts in climate patterns, providing insights into the impacts of global warming and guiding climate action policies.

4.3.1 A brief description of the M-K test

Hydrologic variables, represented by time series data, X ($x_i \in X$ for all i = 1, 2…,n) are ranked (r_1, r_2……, r_n) to increase magnitude. Each data point of rank (r_i) is compared ranked below data (r_j for all j = i+1, i+2…. n), and during the comparison, the sign is noted to calculate the test statistic S as shown in equation 4.3:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(r_j - r_i) \qquad 4.1$$

Where,

$$sgn(rj-ri) = \begin{cases} +1 \; if \; rj > ri \\ 0 \; if \; rj = ri \\ -1 \; if \; rj < ri \end{cases}$$

Where n, r, x i, and j are aspects of a series of data. "n" represents the number of data points in the series, "r" signifies the rank or position of a data point, "x" denotes the magnitude or value of that particular data point "i" represents any given data point in the series and finally "j" refers to any remaining data points appearing after that specific one. The value of S depends on whether the trend in the dataset is increasing or decreasing. The positive and negative values of S indicate the direction of the trend. Considering the assumption that the data is independent and identically distributed, determine the mean and variance of the Mann-Kendall S statistic in the manner as shown in equation 4.4, 4.5 & 4.6:

$$E(S) = 0 \quad 4.2$$

$$Var(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum q_{tp}\left(t_{tp}-1\right)\left(2t_{tp}+5\right)\right] \quad 4.3$$

$$Z = \begin{cases} \frac{s-1}{\sqrt{Var(s)}} & if\ s > 0 \\ 0 & if\ s = 0 \\ \frac{s+1}{\sqrt{Var(s)}} & if\ s < 0 \end{cases} \quad 4.4$$

The null hypothesis (H_O) representing no trend is rejected if $|Z| > Z_{1-\alpha/2}$, where Z is obtained from the standard regular distribution table, and α is the significance level. The Mann-Kendall test uses the variance of S to determine Z, allowing it to identify trends in data over time. Researchers typically choose a significance level of 5% or 10%, although this decision is subjective. A lower significance level indicates a deviation from the hypothesis. In this study, a level of significance of 0.05 has been adopted.

Assumptions

Unlike parametric tests, the Mann-Kendall test does not require any assumptions about the underlying distribution of the data. However, there are some assumptions and considerations to keep in mind when applying the MK test:

1. Independence: The observations in the time series should be independent of each other. Autocorrelation within the data can lead to biased trend detection. If there is evidence of autocorrelation, pre-processing techniques such as detrending or autocorrelation corrections may be necessary.

2. Random Sampling: The data should be collected through random sampling or represent a random sample of the population of interest. This assumption ensures that the observed trends represent the larger population and are not artifacts of biased sampling.

3. Linearity: The Mann-Kendall test detects monotonic trends, meaning the data either consistently increases or consistently decreases over time. It does not assume that the trend follows a linear pattern. However, if the underlying trend is non-monotonic or exhibits sudden changes in direction, the Mann-Kendall test may not be the most appropriate method for trend detection.

4. Homogeneity of variance: The variability of the data should be consistent across the entire time series. Significant changes in variance over time can affect the sensitivity of the Mann-Kendall test to detect trends accurately. Data transformation techniques or robust trend tests may be considered if heteroscedasticity is present.

5. Complete and Continuous Data: The Mann-Kendall test assumes that the time series data are complete and continuous without missing or irregularly spaced observations. Gaps or irregularities in the data can impact the reliability of trend detection and may require imputation or interpolation techniques.

6. Sufficient Data Length: The effectiveness of trend detection with the Mann-Kendall test improves with more extended time series datasets. Short time series may lack statistical power to detect trends reliably, especially if the variability within the data is high.MK can be calculated even when missing data and values are below one or more detection limits, but the test's performance will suffer due to these factors. There must be enough time between samples to ensure no correlation between data collected at different points in time.

4.3.2 Modified Mann-Kendall test

In autocorrelation, pre-whitening has been used to identify trends in a time series. On the other hand, pre-whitening has decreased the detection rate of significant trends in the MK test (Yue et al. 2001). As a result, the MMK test was used to identify trends in an autocorrelated dataset. After removing a non-parametric trend estimate such as Theil and Sen’s median slope from the data, the autocorrelated across ranks of the observations, k is assessed. Because the variance of S is overestimated when the data are positively autocorrelated, only significant values of k are utilized to compute the variance correction factor n/n S:

N is the actual number of observations, n*s is considered an adequate number of observations to account for autocorrelation in the data and is the autocorrelation function of the ranks of the observations. Several for significant autocorrelation in data. The corrected variance is computed as shown in equation 4.7:

$$\sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(s_i - \bar{s})^2 - 2\sum_{k=1}^{n-1}\frac{1}{n-k}\mathrm{cov}(s_i, s_{i+k})\right] \quad 4.5$$

Where:

$\sigma^2$is the$_c$corrected variance.

Si is the rank sum for the $\boldsymbol{i}$-th time point.

$\bar{S}$ is the mean of the rank sums

$cov\ (Si, Si + k)\ i$s the covariance between SiS_iSi and Si+kS_{i+k}Si+k.

n is the number of observations

4.3.3 Sen’s slope estimation method

Sen's slope estimation method, also known as the Theil-Sen estimator, is a non-parametric technique used to determine the magnitude of a trend in time series data. It is instrumental in environmental and hydrological studies for detecting trends in climate data, such as temperature, precipitation, and river flow, as it is resistant to outliers and does not assume any specific distribution for the data. Theil first proposed the method in 1950 and later extended by Sen in 1968.

Sen’s slope is calculated by taking the median of all possible pairwise slopes between data points over time. The formula for the slope between any two data points is as shown in equation 4.8:

$$Q = \frac{X_j - X_i}{j-i} \quad 4.8$$

where Xj and Xi the data values at times $j\ and\ i$, respective, and $j > i$. Sen’s slope estimator is the median of all such slopes, providing a robust measure of the central trend in the data. One of the strengths of this method is its ability to handle missing or irregular data points and its robustness against non-normality or heteroscedasticity in the dataset. It is often used in combination with the Mann-Kendall test to assess the statistical significance of the detected trends.

4.3.4 Pettitt Mann Whitney test

Pettitt -Mann -Whitney (Pettitt, 1979) This method was used to identify the changes in the mean or median of a time series data when there were no clear indications of change points. It is commonly used in climate studies with data to detect a single change point. The test compares the hypothesis (H_O), which assumes that the variable.

(T) follows one or more distributions with the average value (no change), against the alternative hypothesis (Ha), which suggests that a change point exists. The nonparametric statistic KT is defined as shown in equation 4.9 & 4.10;

$$K_T = \max U_{t,T} \qquad 4.6$$

Where,

$$U_{t,T} = \sum_{i=1}^{T} \sum_{j=t+1}^{T} \operatorname{sgn}(X_j - X_i) \qquad 4.7$$

Where,

$U\ t,\ T;$ and the series change point is located at K_T when the statistic is significant. Furthermore,

$$sgn\,(Xj - Xi) = \begin{cases} +1 \; if\ Xj > Xi \\ 0 \; if\ Xj = Xi \\ -1 \; if\ Xj < Xi \end{cases}$$

The approximate significance probability of KT is calculated for p values than or equal to 0.05 using the method as shown in equation 4.11;

$$P \approx -6K_T^2 + 2\,exp\left(\frac{3}{2}\right)T + T \qquad 4.11$$

4.4 Climate Extremes

The expert team (ET) of Climate Change Detection and Indices (ETCCDI), jointly sponsored by the World Meteorological Organization (WMO) Commission of

Climatology (CCI) and the Climate Variability and Predictictibility (CLIVAR) project, developed 27 indices for monitoring the changes in climate extremes (Peterson, 2005). The computation of these indices requires daily precipitation and temperature (minimum and maximum) datasets. This study used 10 precipitation and 5 temperature indices (listed in Table 4.1). All indices used in this study were calculated annually from 1923-2022.

Table 4.1 List of extreme indices of temperature and precipitation as recommended by the ETCCDI, along with their definitions

S. NO.	Index	Index name	Index definition	Units
1	TN10p	Cool nights	Let TN_{ij} be the daily minimum temperature on the day I in period j, and let Tnin10 be the calendar day 10th percentile centered on a 5-day window. The percentage of time for the base period is determined where $TN_{ij} < TN_{in}10$	%
2	TX10p	Cool days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j*, and let $TX_{in}10$ be the calendar day 10^{th} percentile centered on a 5-day window. The percentage of time for the base period is determined where $TX_{ij} < TX_{in}10$	%
3	TN90p	Warm nights	Let TN_{ij} be the daily minimum temperature on day *i* in period *j*, and let $TN_{in}90$ be the calendar day 90^{th} percentile centered on a 5-day window. The percentage of time for the base period is determined where $TN_{ij} > TN_{in}90$	%
4	TX90p	Warm days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j*, and let $TX_{in}90$ be the calendar day 90^{th} percentile centered on a 5-day window. The percentage of time for the base period is determined where $TX_{ij} > TX_{in}90$	%
5	TXx	Max Tmax	Let TX_x be the daily maximum temperatures in month *k*, period *j*. The maximum daily maximum temperature each month is then TX_{xkj}=max (TX_{xkj})	C
6	TXn	Min Tmax	Let TX_n be the daily maximum temperatures in month *k*, period *j*. The minimum daily maximum temperature	C

			each month is then TX_{nkj}=min (TX_{nkj})	
7	TNx	Max Tmin	Let TN_x be the daily minimum temperatures in month k, period j. The maximum daily minimum temperature each month is then TN_{xkj}=max (TN_{xkj}	C
8	TNn	Min Tmin	Let TN_n be the daily minimum temperatures in month k, period j. The minimum daily minimum temperature each month is then TN_{nkj}=min (TN_{nkj})	C
9	FD	Frost days	Let TN_{ij} be the daily minimum temperature on day i in year j. Count the number of days where TN_{ij} < 0°C.	days
10	ID	Ice days	Let TX_{ij} be daily maximum temperature on day i in year j. Count the number of days where TX_{ij} < 0°C.	days
11	SU	Summer days	Let TX_{ij} be daily maximum temperature on day i in year j. Count the number of days where TX_{ij} > 25°C.	days
12	TR	Tropical nights	Let TN_{ij}be daily minimum temperature on day i in year j. Count the number of days where TN_{ij} > 20°C.	days
13	GSL	Growing season length	Let TG_{ij} be the daily mean temperature on day i in year j. Count the number of days between the first occurrence of at least six consecutive days with TG_{ij} > 5°C and the first occurrence after 1st July (1st Jan. in SH) of at least six consecutive days with TG_{ij} < 5°C.	days
14	DTR	Diurnal temperature range	Let TX_{ij} and TN_{ij} be the daily maximum and minimum temperature respectively on day i in period j. If I represent the number of days in j, then $$DTR_j = \frac{\sum_{i=1}^{I}(Tx_{ij} - Tn_{ij})}{I}$$	
15	RX1day	Max 1-day precipitation amount	Let RR_{ij} be the daily precipitation amount on day i in period j. The maximum 1-day value for period j are $Rx1day_j$ = max (RR_{ij})	mm
16	RX5day	Max 1-day precipitation	Let RR_{kj} be the precipitation amount for the 5- day interval ending $_k$, period j. Then maximum 5-day values for period j are	mm

		amount	$Rx5day_j = \max(RR_{kj})$	
17	SDII	Simple daily intensity index	Let RR_{wj} be the daily precipitation amount on wet days, w $(RR \geq 1mm)$ in period j. If W represents number of wet days in j, then: $SDII_j = \frac{\sum_{w=1}^{W} RR_{wj}}{W}$	mm/day
18	R10mm	Number of heavy precipitation days	Let RR_{ij} be the daily precipitation amount on day i in period j. Count the number of days where $RR_{ij} \geq 10mm$	days
19	R20mm	Number of very heavy precipitation days	Let RR_{ij} be the daily precipitation amount on day i in period j. Count the number of days where $RR_{ij} \geq 20mm$	days
20	CDD	Consecutive dry days	Let RR_{ij} be the daily precipitation amount on day i in period j. Count the largest number of consecutive days where $RR_{ij} < 1mm$	days
21	CWD	Consecutive wet days	Let RR_{ij} be the daily precipitation amount on day i in period j. Count the largest number of consecutive days where $RR_{ij} \geq 1mm$	days
22	R95p	Very wet days	Let RR_{wj} be the daily precipitation amount on a wet day w $(RR \geq 1.0mm)$ in period i and let $RR_{wn}95$ be the 95th percentile of precipitation on wet days in the 1961-1990 period. If W represents the number of wet days in the period, then $R95p_j = \sum_{w=1}^{W} RR_{wj}$ where $RR_{wj} > RR_{wn95}$	mm
23	R99p	Very heavy wet days	Let RR_{wj} be the daily precipitation amount on a wet day w $(RR \geq 1.0mm)$ in period i and let $RR_{wn}99$ be the 99th percentile of precipitation on wet days in the 1961-1990 period. If W represents the number of wet days in the period, then: $R99_{pj} = \sum_{w=1}^{W} RR_{wj}$ where $RR_{wj} > RR_{wn99}$	mm
24	PRCPTOT	Annual total wet-day	Let RR_{ij} be the daily precipitation amount on day i in period j. If I represent the number of days in j, then	mm

		precipitation	$PRCPTOT_j = \sum_{i=1}^{I} RR_{ij}$	
25	CSDI	Coldspell duration index	Let TN_{ij} be the daily maximum temperature on day i in period j , and let $TN_{in}10$ be the calendar day 10th percentile centered on a 5-day window. Then the number of days per period is summed where in intervals of at least 6 consecutive days $TN_{ij} < TN_{in}10$	days
26	WSDI	Warm spell duration index	Let TX_{ij} be the daily maximum temperature on day i in period j , and let $TX_{in}90$ be the calendar day 90th percentile centered on a 5-day window. Then the number of days per period is summed where in intervals of at least 6 consecutive days $TX_{ij} > TX_{in}90$	days
27	Rnnmm	Number of user-defined precipitation days	Let RR_{ij} be the daily precipitation amount on day i in period j. Count the number of days where $RR_{ij} \geq nnmm$	days

Out of 27 extreme climatic indices, 15 indices have been used in this research, mentioned in the next chapters in detail.

4.5 Summary

The trend analysis results would aid in formulating effective water management and developing appropriate mitigation measures to protect water resources. The extreme values trend analysis results and 3-D pattern characterization of rainfall patterns of multiple extreme climatic indices would aid in forecasting the pattern and intensity of future extreme occurrences in further depth. That would make it easier to create efficient safety preventive measures, such as increasing drainage capacity and redesigning stormwater drainage systems for floods, and developing appropriate mitigation methods for drought also, which has a high human cost in the contemporary day.