CHAPTER 4

Methodology

4.1 Overview

This chapter provides a detailed methodology of the potential impact of climate change on the characteristics of drought occurrences across Uttar Pradesh, India. The widely used drought indices Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) are evaluated for 18 stations in Uttar Pradesh state from 1971 to 2018 in this study. Drought characteristics such as intensity, duration, and frequency of different categories are estimated and compared based on SPI and SPEI. Station proportion methodology is adopted at a different timescale, providing a better insight into the temporal variability of spatial extent drought of a specific category. The rank-based Mann-Kendall (MK) and Sens slope methods are used to analyze the temporal variability for meteorological time series and drought severity. Furthermore, this chapter provides the framework for collectively assessing dry and wet events and their transition occurrences from wet-to-dry or dry-to-wet over the study period. And the methodology of projection of precipitation, maximum and minimum temperature for the future period under change scenarios RCP 4.5 and RCP 8.5 climate by statistical downscaling is detailed here.

4.2 Trend Detection Methodology

4.2.1 Mann-Kendall Test

Mann-Kendall (MK) trend is a non-parametric statistical test widely used to determine a monotonic trend at a significance level in a series of the environment. The trend of meteorological variables and drought indices output were determined using the MK trend test in this study. The MK test has advantages over the parametric test. It is a non-parametric test that doesn't require following the assumption of a normal distribution or interfered by the outliers, making this test more robust towards the non-normally distributed time series (Mann 1945; Kendall 1975). This technique is extensively used in trend detection of hydrometeorological variables and climatic extremes (Sah et al., 2021; Tian & Quiring, 2019). In the literature, the researcher focused on the trend assessment of meteorological variables and drought to anticipate future changes (Abhilash et al., 2019; Guhathakurta et al., 2017; Hadi & Tombul, 2018).

The Mann-Kendall (MK) test determines whether a variable has a monotonic upward or downward trend over time. A monotonic upward or downward trend depict the consistent increase or decrease in the value of a variable over a time period. The MK test can be used instead of parametric linear regression analysis to determine if the slope of the predicted linear regression line is greater than zero. In regression analysis, the residuals from the fitted regression line must have a normal distribution. However, the MK test is a non-parametric test that does not follow the assumption of normal distribution. The MK test is best regarded as an exploratory study and should be used to discover stations with significant magnitude changes and quantify these results.

The MK test is based on the following assumptions:

- When no trend exists, the measurements (observations or data) collected over time are independent and identically distributed. The condition of independence implies that the observations are not serially connected over time.
- The procedures for collecting, processing, and measuring samples provide unbiased and representative observations of the underlying populations over time.

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.

The time-series data for a hydrologic variable, x_i, are ranked according to their rank (r_1, r_2,... r_n). Each data set's rank (r_i) is compared to the rest of the data that ranks lower than r_j, *where* $j = i+1, i+2,...n$.

S as the test statistic determined by

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

where,

$$sgn(r_j - r_i) = \begin{cases} +1, & if\ r_j > r_i \\ -1, & if r_j < r_i \end{cases} \tag{4.2}$$

Each of the values *n*, *r*, *i*, and *j* stands for a different time series variable: the number of time-series data, the rank of those data, their magnitude, a specific data value, and the number of data points left in the series. Positive or negative *S* values indicate an increasing or declining trend in the dataset, whereas a larger magnitude of *S* indicate a more consistent trend in the dataset's direction.

The mean and variance of the Mann-Kendall statistic *S* are determined as:

$$E(S) = 0 \tag{4.3}$$

$$var(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} t_p.(t_p - 1).(2t_p + 5)\right] \tag{4.4}$$

where t_p is the cluster size of *pth* tied groups and *q* is the number of tied groups.

The Z_{MK} statistic determined as:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{var(S)}}, & if \quad S > 0 \\ \\ \frac{S+1}{\sqrt{var(S)}}, & if \quad S < 0 \end{cases} \tag{4.5}$$

At a given significance level α, if $|Z|> Z_{1-\alpha/2}$, the null hypothesis (H_0) is rejected, which implies there is no significant trend present in the considered time series and accepts the alternative hypothesis (H_1), which implies that there is a notable trend at significance level α. Positive and negative signs of sing of Z_{MK} value denote the upward and downward trend of the meteorological series. Although the level of significance decision is still subjective, analysts have used both 0.05 (often referred to as 5% of significance) and 0.01 (1% of significance). The lower threshold of significance indicate that more data should differ substantially from the zero hypotheses. In the current analysis of rainfall variability, 0.05 of significance has been seen as more meticulous than 0.01 thresholds. The MK test used in this study to determine long-term changes in meteorological variables and the meteorological drought characteristics at different timescales.

4.2.2 Sen's Slope Estimator

Sen's slope is a non-parametric test for determining the magnitude of a linear trend represented by the median value of the time series slope values. Sen's slope estimator determines the magnitude of the trend in time series of hydrologic variables after determining the nature of the trend. The slope (T_k) for each of the data sets/pairs is calculated using the following equation proposed by Sen (1968) as shown below.

$$T_k = \frac{y_j - y_i}{j - i} \qquad \text{for k = 1, 2, …N} \tag{4.6}$$

y_j is a *j-th* data point, y_i is an *i-th* data point in time series, and j should be greater than i, i.e., $j > i$. The main advantage of Sen's slope estimator is the use of a global median value, which makes it more resistant to the effects of extreme values in the time series. The median of these slope values (T_k) is calculated as follows:

$$\beta = \left\{ \begin{matrix} T_{\frac{N+1}{2}}, & if\ N\ is\ odd \\ \frac{1}{2}\left(T_{\frac{N}{2}} + T_{\frac{N+2}{2}}\right), & if\ N\ is\ even \end{matrix} \right\} \tag{4.7}$$

The positive β value indicate an increasing trend in the time series, while the negative β value indicate a decreasing trend. Having determined T_k as above, Sen's slope estimator β is confirmed by a two-sided test at 100*(1 – α) % confidence interval.

4.3 Pearson's Correlation Analysis

Pearson's Correlation method is used in this study to assess the consistency of the drought indices SPI and SPEI at timescales of 3, 6, 9, and 12-month for all synoptic locations in the study area. The Pearson correlation coefficient (r) is the test statistic that indicate how strong a linear relationship exists between two variables. This is based on the method of covariance and is considered as the best method of measuring the association between variables of interest. It estimates the magnitude of the association or correlation, as well as the relationship's direction. It assigns a value between -1 and 1, with 0 indicating no correlation, 1 representing total positive correlation, and -1 meaning complete negative correlation. This is translated as follows: A correlation value of 0.8 between two variables indicate a significant and robust positive relationship between the two variables. A positive correlation represents that if variable x rises, y variable is also increasing, whereas a negative correlation indicate that if variable X rises, variable Y falls.

Pearson correlation coefficient (r) between two variables, x, and y, calculated using the following equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (4.8)$$

Where, x_i = value of time series of variable x

$\bar{x}$ = mean of time series variable x

y_i = value of time series variable y

$\bar{y}$ = mean of observed variable y.

4.4 Drought Assessment Tools

Drought Indices have been the most widely used method of assessing drought conditions in the past, as they are more useful in decision making than raw data in assessing the severity, duration, and Intensity of drought events around the world. Drought indices are statistical indices calculated by assimilating several drought indicators into a single numerical value, which provides a quantitative assessment of drought characteristics based on the severity, location, and duration of drought episodes (Zargar et al., 2011). Precipitation, temperature, soil moisture, streamflow, river discharge, vegetation state, and ecosystem reactions are critical drought indicators that transform into drought indices that portray a comprehensive picture for drought evaluation and decision making (Alsafadi et al., 2020; Narasimhan & Srinivasan, 2005; Vicente-Serrano et al., 2010). It is possible to identify short-term wet periods within long-term droughts or short-term dry spells within long-term wet periods by monitoring the climate at various timescales. Indices can help people communicate more effectively by simplifying complex relationships. The hazard event and the Indices may also play a historical reference for planners or decision-makers. This gives users a chance of drought occurrence or recurrence varying in severity. But, crucially, climate change will alter historical patterns, and indices can help plan and design applications (Svoboda & Fuchs, 2017). The use of these drought indices triggered drought relief programs and prompted the government to take action and determine the amount of water being drained to gauge the severity of the drought. It is also used to monitor the progress and the propagation of drought. They aim to assess the impact of drought on the landscape over time. In most cases, drought indices developed for one region may not be directly applicable to other regions because of the complexity of drought phenomena, different hydro-climatic conditions, and different catchment characteristics (Keyantash, 2002; Smakhtin & Hughes, 2004). In this study, two existing drought indicators based on meteorological perspectives were analysed. They included: Standardized Precipitation Index (SPI) and Standardised Evapotranspiration Index (SPEI) are the most frequently used drought indices in meteorology. Following the presentation of the drought indices, the evaluation of the drought indices is presented in the upcoming section. Finally, the findings of the investigation are summarised in the chapter's conclusion.

4.4.1 Standardized Precipitation Index (SPI)

Standardized precipitation index (SPI) is a multiscalar drought index developed by Mckee et al. (1993), which requires only precipitation for drought assessment. SPI is used to calculate precipitation deficits at different timescales. Its main advantage is that it can be calculated over multiple timescales. Its multiscalar characteristics enable it to monitor short-term water supplies such as soil moisture, which is critical for agricultural productivity as well as long-term water resources such as groundwater supplies, streamflow, and lake and reservoir levels. The SPI is the transformation of a time series of precipitation into a standardized normal distribution (z distribution). It is estimated by fitting long-term monthly precipitation data to the gamma probability distribution function and converting standard normal distribution with mean zero and standard deviation one using equations (4.17 and 4.18) using classical estimation (Abramowitz and Stegun, 1965; Ghosh, 2019).

(i) Computation of SPI

A long-term monthly precipitation database with at least 30 years of data is required to calculate the SPI for a specific time period at any location. As previously mentioned, the SPI is calculated by fitting a probability density function to the observed precipitation frequency distribution over the time period under consideration. The gamma distribution is considered the best fit for long-term monthly precipitation records (Angelidis et al., 2012). A separate analysis is done for each month and each location in space based on the observed precipitation data. When a monthly precipitation time series is transformed into a standardized normal distribution, the mean of each probability density function is set to zero and the standard deviation to one, ensuring that the SPI values are accurately stated in standard deviations (Edwards, 1997). Since the SPI is a standardization of location and timescale, it shares the advantages of standardization described for the PDSI. The strength of the anomaly is then classified, as shown in Table 4.1 once it has been standardized. During a drought, the SPI value must remain negative for an extended period before it can be considered a drought event. SPI has been used extensively to determine meteorological drought characteristics (Dikici, 2020; Guhathakurta et al., 2017; Łabędzki, 2007; Mekonen et al., 2020; Paulo et al., 2016; Saharwardi et al., 2021).

Methods for calculating SPI are as follows:

First, gamma probability distribution function fitted to t the long-term time series of rainfall defined by the probability density function is

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \; for \; x > 0 \tag{4.9}$$

where α >0 is shape factor

β>0 is scale factor

x>0 is the precipitation.

Γ(α) is the gamma function defined as

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy \tag{4.10}$$

In order to fit the distribution to the data, α and β must be estimated. Using Thom (1958) maximum likelihood approximations, Edwards (1997) proposed the following method for estimating these parameters:

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \tag{4.11}$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \tag{4.12}$$

where,

$$A = ln(\bar{x}) - \frac{\sum_{i=1}^{n} ln(x)}{n} \tag{4.13}$$

For n observations

The calculated parameters are then used to calculate the cumulative probability of an observed over a given month or timescale.

$$G(x) = \int_0^{x} g(x)dx = \frac{1}{\beta^{\hat{\alpha}}(\hat{\alpha})\Gamma(\hat{\alpha})} \int_0^{x} x^{\hat{\alpha}-1}\; e^{\frac{-x}{\hat{\beta}}} dx \tag{4.14}$$

The above equation is reduced to an incomplete gamma function when t is substituted for $x/\hat{\beta}$:

$$G(x) = \frac{1}{\Gamma(\hat{\alpha})} \int_0^x t^{\hat{\alpha}-1e^{-1}}\, dt \quad (4.15)$$

Because the gamma function is undefined at x = 0 and a precipitation distribution can contain zeros, the cumulative probability is:

$$H(x) = q + (1-q)G(x) \quad (4.16)$$

where q denotes the likelihood of no precipitation. If m is the number of zeros in a rainfall time series, m/n can be used to calculate q. For each station in this analysis, a small amount of rainfall was substituted for zero rainfall. This substitution does not affect precipitation distribution. The probability of monthly null precipitation is zero at larger timescales (such as 1, 3, 6, 9, 12, 24, and 48 months). As a result, the errors in calculating the parameters and the monthly null precipitation do not affect the distribution at larger timescales.

The cumulative probability, *H(x),* is then transformed into the standard normal random variable *Z,* which has a mean of zero and a variance of one, and yields the SPI value. In this study, an approximate conversion is used, as provided Abramowitz and Stegun (1965), following Edwards and McKee (1997) and Lloyd-Hughes & Saunders (2002).

$$Z = SPI = -\left[t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right] \quad 0 < H(x) \leq 0.5 \quad (4.17)$$

$$Z = SPI = \left[t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right] \quad 0.5 < H(x) \leq 1 \quad (4.18)$$

$$\text{If,} \quad 0 < H(x) \leq 0.5 \quad t = \sqrt{ln\left(\frac{1}{(H(X))^2}\right)} \tag{4.19}$$

and if,

$$0.5 < H(x) \leq 1 \quad t = \sqrt{ln\left(\frac{1}{(1-H(X))^2}\right)} \tag{4.20}$$

The following are the values of the remaining coefficient:

C_0=2.515517, C_1=0.802853, C_2=0.010328, d_1=1.432788, d_2=0.189269, d_3=0.001308

Drought is considered regional in characteristics, spanning large areas and lasting for longer durations, so understanding them in that context is critical. Drought evaluations at the regional level can provide critical reference data for a variety of water-related activities, and they should be included in strategic short and long-term plans for effective water resource management. By analyzing the temporal variation and spatiotemporal characteristics of SPI, drought severity, intensity, areal extent, and frequency can be developed to assess the severity of localized droughts within the study area and the distribution of drought. Temporal variation of SPI was assessed using the regional representative of SPI values at the timescale of 3-month, 6-month 9-month, and 12-month for 48 years from 1971 to 2018 over the 18 synoptic locations of Uttar Pradesh. A spatial analysis of drought was performed using the SPI values estimated for evaluating the most affected areas for a specific drought event.

4.4.2 Standardized Precipitation Evapotranspiration Index (SPEI)

Standardized Precipitation Evapotranspiration Index (SPEI) is SPI based multivariate drought index proposed by Vicente-Serrano et al. (2010), which introduces the impact of temperature change on the severity of drought (Liu et al., 2021). It assimilates both critical climatic indicator precipitation and temperature, which influences the availability and variability of water on the severity of drought characterization (Gond et al., 2019). The fundamental advantage of the SPEI over prior multiscalar drought indicators such as the SPI is that it accounts for evaporative demand in its computation, making it perfect for evaluating the effects of global warming on drought occurrences. The SPEI was created to combine the ease of calculation and multi-temporal aspect of SPI with the sensitivity of PDSI to changes in evaporation requirement caused by temperature fluctuations or changes over the period (Beguería et al., 2010). The SPEI, like the SPI, can be calculated at different timescales, letting the user to determine the timescale at which the system under investigation responds to drought the most. This study investigates the drought characteristics using SPEI estimated for the under historical period (1971 to 2018) and the influence of changing climate scenario in the near future for the period of 32 years (2019 to 2050) at timescale of 3-month, 6-month, 9-month, and 12-months over Uttar Pradesh.

(i) Estimation of PET by Thornthwaite method

Potential Evapotranspiration (PET) indicate that it refers to the maximum amount of water that soils and vegetation would transfer to the atmosphere if there were no water supply deficit. The main differences between these models are the climatic data variables and the data availability in a given region. Vicente-Serrano et al. (2010) employed the Thornthwaite equation to calculate the PET, which uses the study area's geographic location and monthly mean air temperature as input variables. Thornthwaite equation could be a useful balance between consistency and minimal data requirements. As a result, the current study used the Thornthwaite model (Thornthwaite, 1948) to estimate PET, which only considers temperature as an input variable.

Methods for calculating PET are as follows:

$$PET = 16k\left(\frac{10T}{I}\right)^{a} \tag{4.21}$$

Where k is the correction factor, *T* (°C) is the mean monthly temperature, *I* represent the annual thermal index, and i is the monthly thermal index.

$$I = \sum_{j=i}^{12} i \tag{4.22}$$

$$i = \left(\frac{t}{5}\right)^{1.514} \tag{4.23}$$

$$a = 0.49 + 0.018I - 7.7x10^{-5}I^{2} + 6.7x10^{-7}I^{3} \tag{4.24}$$

Where a is location dependent coefficient, which has the same unit as temperature *T* (°C).

Where k is the correction factor depending on the Latitude and month given below:

$$k = \left(\frac{N}{12}\right)\left(\frac{NDM}{30}\right) \tag{4.25}$$

where NDM denotes the number of days in the month and N denotes the maximum number of sun hours, as shown below

$$N = \left(\frac{24}{\Phi}\right)w_s \tag{4.26}$$

w_s is the hourly angle of the rising sun, which can be calculated as follows:

$$w_s = arccos(-tan\psi tan\delta) \tag{4.27}$$

ψ is the latitude measured in radians. If δ is the solar declination in radians and *J* is the average Julian day of the month, then the following can be estimated:

$$\delta = 0.4093\ sen\left(\frac{2\pi J}{365} - 1.405\right) \quad (4.28)$$

(ii) Computation of SPEI

SPEI uses "water balance" (D_i) as an input variable aggregated at different timescales (3, 6, 9, and 12-month), and these resulting values fit the probability distribution function (e.g., log-logistic) then normalized the water balance to obtain the SPEI. Unlike the SPI index, negative values of SPEI define the drought period based on intensity, severity, magnitude, and duration. After estimating the PET, each month's climatic water balance (Di) values were calculated using equation 28. The PET values calculated using the Thornthwaite equation and monthly rainfall are summed to the n-month scale as follows:

The difference (*Di*) between precipitation (*P*) and *PET* for the month (*i*) is given in equation 4.29.

$$D_i = P_i - PET_i \quad (4.29)$$

The calculated *D* values are accumulated at different timescales:

$$D_n^k = \sum_{i=0}^{k-1} P_{n-1} - (PET)_{n-1} \quad (4.30)$$

K is the timescale (months) of accumulation, and n is the calculation month.

The probability density function of a Log-logistic distribution is given as:

$$f(x) = \frac{\beta}{\gamma}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1}\left(1+\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right)^{-2} \quad (4.31)$$

where α, β, and γ are scale, shape, and origin parameters, respectively for $\gamma > D < \infty$.

The probability weighted moments (PWMs) were used by Vicente-Serrano et al. (2010) to calculate the log-logistic $\alpha, \beta,$ and γ distribution parameters, based on the plotting-position approach (Hosking, 1990), in which the PWMs of the order s are calculated as plotting positions.

$$w_s = \frac{1}{N}\sum_{i=1}^{N}(1 - F_i)^s D_i \quad (4.32)$$

F_i is a frequency estimator based on Hosking (1990) approach where N is the number of data. Using the Plotting Position formulae, the standard deviation of the SPEI series changes noticeably as a function of the SPEI timescale, affecting the spatial comparability of the SPEI values. The standard deviation of the series does not change among the different SPEI timescales if the PWMs obtained by unbiased estimator proposed by Hosking (1986). The unbiased PWMs generated using the following procedure:

$$w_s = \frac{1}{N}\sum_{i=1}^{N}\frac{\binom{N-i}{s}D_i}{\binom{N-1}{s}} \quad (4.33)$$

The probability distribution function for the D series is given as:

$$F(x) = \left[1 + \left(\frac{\alpha}{x} - y\right)^{\beta}\right]^{-1} \quad (4.34)$$

With *f(x)* the SPEI can be obtained as the standardized values of *F(x)* according to the method of Abramowitz and Stegan et al. (1965):

$$SPEI = W - \frac{C_0 + C_1W + C_2W^2}{1 + d_1W + d_2W^2 + d_3W^3} \quad (4.35)$$

Where,

$$W = \sqrt{-2ln(P)} \;\; for\ P \leq 0.5 \quad (4.36)$$

$$W = \sqrt{-2\ln(1-P)} \quad \text{for P > 0.5}$$

P is the probability of exceeding a determined D_i value and is given as

$$P = 1 - f(x) \tag{4.37}$$

while the constants are:

C_0=2.515517, C_1=0.802853, C_2=0.0203, d_1=1.432788, d_2=0.189269, d_3=0.001

Table 4.1 Drought and wet severity class according to SPEI/SPI values

SPI/SPEI Value	Category
>2.0	**Extreme Wet**
1.5 to 1.99	**Severely Wet**
1.0 to 1.49	**Moderately Wet**
0.0 to 0.99	**Near Normal**
-0.99 to 0.0	
-1.0 to -1.49	**Moderately Dry**
-1.99 to -1.5	**Severely Dry**
<-2.0	**Extremely Dry**

Meteorological Organization (WMO) 2012 user guide (URL: WMO_1090_EN.pdf and wamis.org)

4.4.3 Drought Characterization and Run Theory

Drought characterization is defined by the severity, intensity, duration, and frequency of a drought event (Hao & Singh, 2015). The run theory refers to the occurrence of another type of event in the course of a continuous occurrence of similar events, such as drought, no drought days, wet event, continuous rainy days, and alternating natural waters (Wu et al., 2019). A run is a segment of time series in which all of the values of the drought indicator are below a specified threshold (Yevjevich, 1969). Researchers have been utilizing various thresholds for drought characteristics estimation. Pandey et al. (2021) selected SPI ≤ -1 as the threshold for drought identification over the Bundelkhand region, India. A drought event identifies when the value of SPI and SPEI is continuously negative and reaches a value of -1 or less (Mesbahzadeh et al., 2020). This study calculated all drought characteristics at specified threshold SPI/SPEI ≤ ±1 to evaluate drought characteristics based on drought indicator SPI and SPEI over Uttar Pradesh, India, from 1971 to 2018. In this method, a drought event is defined when the value of SPI and SPEI is continuously negative and reaches a value of -1 or less, as shown in Figure 4.1.

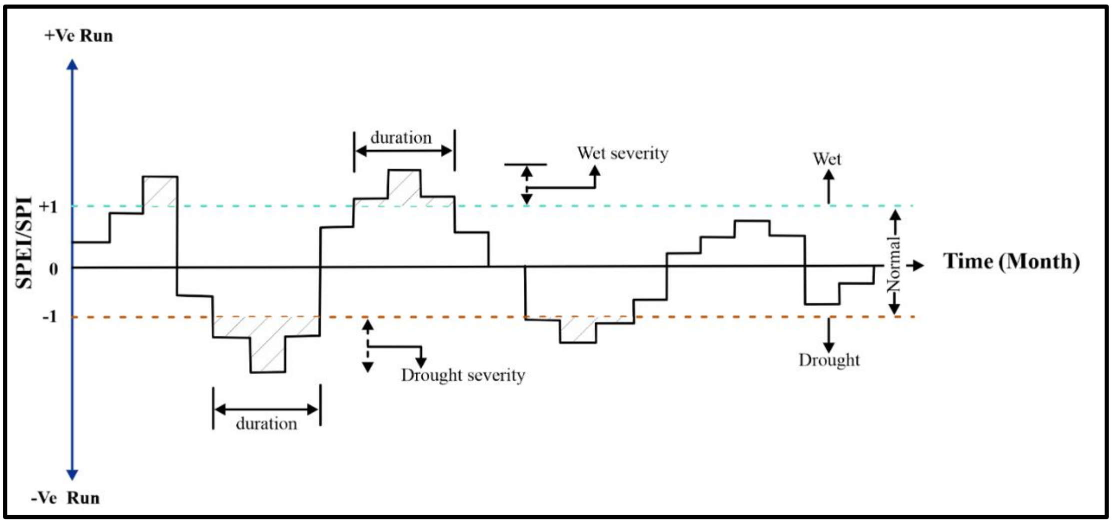

Figure 4.1 Schematic diagram of drought and wet characterization based on run theory.

(a) Drought duration (D*)* is defined as months between the start and end of a drought event when SPI/SPEI ≤ -1.

(b) Drought Severity (*S*) equals absolute indices values during a drought event estimated as drought severity.

$$S = | - \sum_{n=1}^{D} SPI \ / \ SPEI| \ \leq -1 \qquad (4.38)$$

(c) Drought Intensity (*I*) is calculated as the ratio of the absolute value of drought severity to drought duration during a drought event (Wang et al., 2014). It indicate the deviation of the climate index from its normal values.

$$I = \frac{Severity\ (S)}{Duration(D)} \tag{4.39}$$

(d) Drought Frequency (*F*) is used to assess the frequency of occurrence in every 100 years of drought events over the study period. Calculated as

$$F = \frac{D_j}{J.n} \times 100\% \tag{4.40}$$

Where, $F_{j,100}$ = frequency of drought event for timescale *j* in 100 years

D_j = duration of drought event at scale *j*

n= Number of years of the study period

(e) Drought Station Proportion

Drought station proportion estimates the spatial extent of drought occurrence of specific categories in the study region. It is calculated as the percentage of the number of drought stations affected in the study area impacted by the drought of a specific category (Li et al., 2012).

$$P_j = \frac{n_j}{N_j} \times 100\% \tag{4.41}$$

Where n_j is the number of drought-prone stations falling under specific drought categories (say extreme drought categories when SPI/SPEI < -2) occurring at a particular timescale j (say six months). N_j is the total number of studied stations.

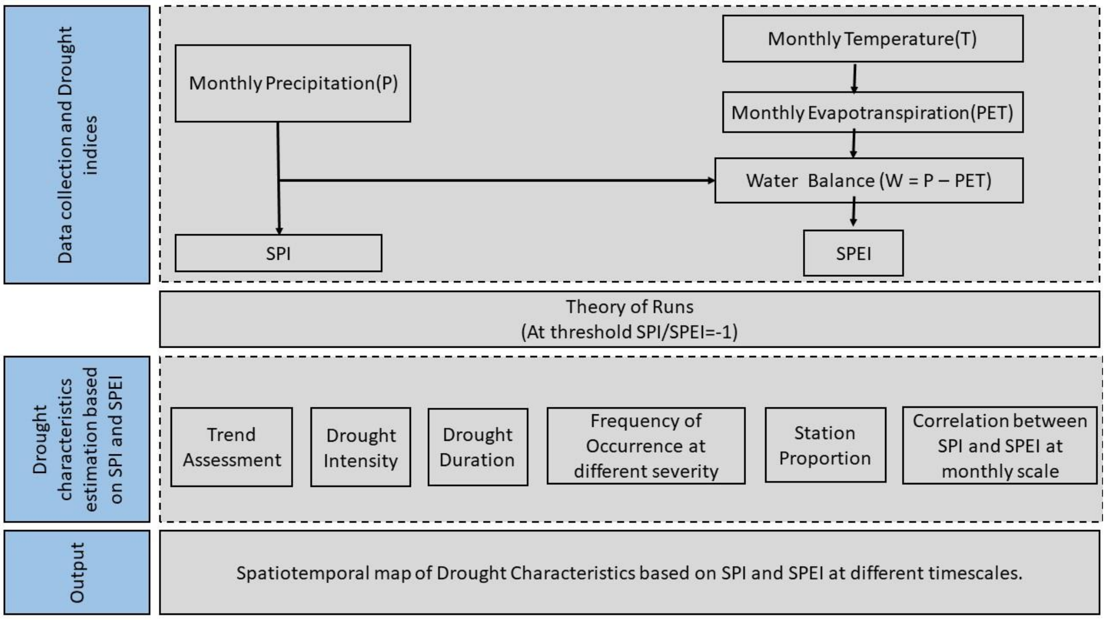

Figure 4.2 The framework of methodology of evaluation drought characteristics under the observed time period

4.5 Evaluations of Concurrent Dry-Wet Event Occurrence

The frequent occurrence of climate extremes (such as droughts and floods) and their rapid transition amplify the related socio-economic repercussions beyond what would be expected from any one particular event. Therefore, it is crucial to understand the spatial and temporal evolution of wet-dry event characteristics and their transition (wet-to-dry and dry-to-wet). This will serve as the basis for identifying and locating the most vulnerable hotspots and informing adaptation and mitigation strategies. We investigated dry and wet events and their transition over the study period to assess the concurrent occurrence of these two climatic extremes over the research region. The dry and wet event characteristics were evaluated using the SPEI, which included the influence of temperature on the estimation of dry and wet characteristics. This study uses SPEI at longer timescales used to measure meteorological drought, but its averaging effect makes it inefficient for estimating flood conditions. We have evaluated the dry and wet characteristics and their transition (dry-to-wet and wet-to-dry) using the SPEI time series at a 1-month timescale from 1971 to 2018 over the 18 synoptic stations of Uttar Pradesh, India.

4.5.1 Dry and wet event characterization

In this study, dry and wet event characteristics are defined based on severity, duration, and intensity. A dry event identifies when the value of SPEI is continuously negative and reaches a value of SPEI≤-1 or less, and a wet event identifies when the value of SPEI is continuously positive and reaches a value of SPEI ≥1. Dry characteristics calculated at specified threshold SPEI ≤ -1 and wet event characteristics at threshold SPEI ≥1 characteristics based on SPEI time series at a 1-month timescale over Uttar Pradesh, India, for the period 1971 to 2018. In the previous section, 4.5.3, the duration, severity, and intensity of dry and wet events are defined and shown in Figure 4.1. The duration, severity, and intensity of dry and wet events are estimated based on run theory, the average wet and dry event duration (AWD and ADD), average wet and dry severity (AWS and ADS), and average wet and dry intensity (AWI and ADI) were calculated for 18 synoptic locations in Uttar Pradesh over 48 years (1971–2018).

4.5.2 Wet-Dry ratio (WD)

The wet-dry ratio (WD) is estimated at each location by taking the natural logarithm of the ratio of the number of wet months (N_w) and the number of dry months (N_d) (De Luca et al., 2020). In this study, the WD ratio was estimated at different severity moderate, severe, and extreme from 1971 to 2018 over the 18 synoptic locations of Uttar Pradesh, India. The WD ratios quantify the predominance of dry or wet events of specific severity over an area. Thus, WD ratio of greater than 0 indicate that wet extremes outweigh dry events, whereas a WD ratio less than 0 indicate that dry extremes predominate over wet events. The natural logarithm was utilized to reduce the range of WD ratio values and to sign-separate wet-dominated from dry-dominated locations.

The WD ratio is defined as:

$$WD\ Ratio = \ln\left(\frac{N_w}{N_d}\right) \tag{4.42}$$

4.5.3 Dry-wet transition time

The average transition time (T_t) is the time interval in months between the dry and wet events of different severity 'moderate', 'severe', and 'extreme' defined by De Luca et al. (2020). The number of transitions and the average transition was calculated separately at different severity levels ('moderate', 'severe', and 'extreme') for every synoptic location for 48 years from 1971 to 2018. The stepwise procedure for determining the wet-to-dry transition time (T_t) is as follows: Wet and dry events were extracted and sorted in ascending order of time (from oldest to most recent); in the case of consecutive dry and wet months, only the first and last month value was retained; the difference in months between wet-to-dry events within the time series was calculated, and the average was determined. Dry-to-wet transition (T_t) was calculated in the same manner, with the exception of storing the first and last months of wet and dry events separately and determining the time gap between dry-to-wet events.

4.5.4 Dry-wet rapid transition time

The consecutive occurrence of wet and dry months is defined as a wet-dry rapid transition event, and it is estimated considering all the severity together (Ansari &

Grossi, 2022). This study determined the geographic hotspot for the concurrent occurrence of climate extreme events by calculating the frequency of wet-to-dry events (wet event followed by dry event) and dry-to-wet (dry event followed by wet event) rapid transition events for each location for the period of 48 years from 1971 to 2018.

4.6 Climate Change Projection

Future climate change projections of precipitation and temperature have been conducted using the Second-Generation Canadian Earth System Model (CanESM2) created by the Canadian Centre for Climate Modeling and Analysis (CCCma). CanESM2 combines the fourth-generation atmospheric general circulation model (CGCM4) and the fourth-generation ocean general circulation model (OGCM4). A climate simulation carried out within the contributes to the IPCC fifth assessment report. CanESM2 climate change scenario is available at a grid size of 2.8125°. Hydrological analysis cannot be performed with the available data at this resolution, highlighting the need for downscaling to make the data usable. Sub-grid and regional climate scenarios cannot be resolved by GCM, which neglects important regional features such as topography, vegetation, and cloudiness that affect local climate. As a result, it is critical to reducing the GCM data from a global to a regional scale. Statistically reducing GCM outputs to a station scale was done in the study. The climate model used in this study is the second-generation Canadian Earth System Model (CanESM2). At ground level, the Earth System Models (ESMs) have a fairly coarse spatial resolution of 200 by 300 km. The data in this CanESM2 model is based on AR5 IPCC climate change scenarios from 2014. The four RCPs (RCP 2.6, RCP 4.5, RCP 6.0, and RCP 8.5) are based on greenhouse gas concentration trajectories rather than emissions, as was the case with SRES earlier. Downscaling can be done directly in

SDSM using the CanESM2 output from the climate station. The CanESM2 model is chosen for this investigation under the RCP 4.5 and RCP 8.5 scenarios due to the aforesaid reasons. Statistical downscaling is a technique in which linear transfer functions between mesoscale atmospheric predictors variables of GCM (e.g., mean sea level pressure, geo-potential height, specific humidity, etc.) and local climatic variables (temperature, precipitation, etc.) are developed for the observed period and used to derive a station scale climate projection for the future period. SDSM version 4.2 (Wilby & Dawson, 2007) was used for statistical downscaling.

SDSM is used to downscale the rainfall from GCM to the regional level. It comprises of stochastic approach as well as Multiple Linear Regression (MLR). The foremost step in SDSM is to develop a quantitative connection between predicted (observed) and predictor. SDSM is further divided into three sub-models i.e., annual, seasonal, and monthly sub-models. All these models drive a regression equation. But the difference is that the annual sub-model derives only a single equation for a whole year, whereas the seasonal derives the equation separately for each season. On the other hand, the monthly sub-model derives the equation for every single month separately. Furthermore, the sub-models may be conditional or unconditional based on the type of parameter to be downscaled.

In the statistical downscaling method, there are seven distinct stages.

1. The quality of the baseline data (daily temperature and rainfall) needs to be reviewed.
2. Screening predictor variables where we look for empirical relationships between the predictors we've chosen and the predictands themselves.

3. A high explained variance value is used to select predictor-predictand pairs for correlation analysis.
4. The selection of predictor variables to be included in the parameter file is determined by the correlation coefficient (r) and significance level (p).
5. Predictand data and a parameter file are used to calibrate the model. The parameter file and NCEP/NCAR reanalysis data were used to create ensembles of synthetic weather series in the weather generator step.
6. In the statistical analysis step, the observed and simulated climate variables were examined to verify the relationship between the predictor and predictand.
7. The ensembles of synthetic weather series are created during the scenario generation step using the parameter file and CanESM2 model data. In the RCP 4.5 and RCP 8.5 scenarios, the downscaled climate variables, respectively, through the creation of scenarios.

Data from 1971 to 1995 were used for calibration, and data from 1996 to 2005 were used for validation. Future climate change projections have been made for the rainfall and temperature for 2019 to 2050 under RCP 4.5 and RCP 8.5. Daily precipitation and temperature projections under RCP 4.5 and RCP 8.5 for all 18 synoptic stations were taken into account for the study region for 30 years (2019 to 2050). SPEI time series estimated for the future period over the study area at timescales of 3-months, 6-months, 9-months, and 12-months. Based on run theory, drought event characteristics (severity, duration, and frequency) are estimated for SPEI time series at a threshold of SPEI≤-1.

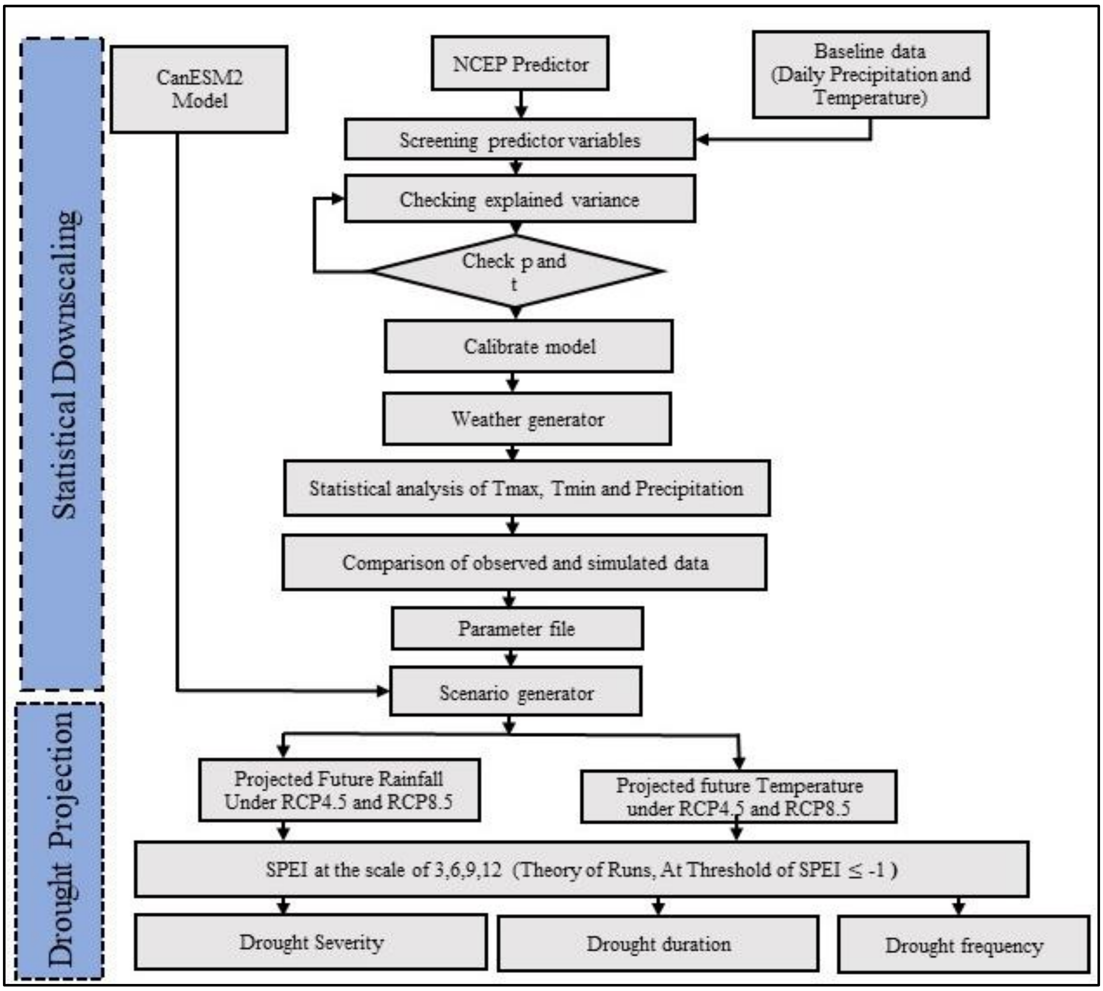

Figure 4.3 Climate change projection and drought characteristics projection under climate change scenarios

4.6.1 Input data for SDSM

The primary input of SDSM is a set of large-scale atmospheric predictor variables, including observed daily precipitation and temperature data from 1971 to 2005. CanESM2 output at the location is downloaded and can be utilized as input directly to SDSM. The present study utilizes the CanESM2 output for RCP 4.5 and RCP 8.5. The 26 standardized large-scale surface and atmospheric predictor variables were generated by the Canadian Centre for Climate Modeling and Analysis (CCCma) for both CanESM2 and NCEP/NCAR from 1971 to 2005. Table 4.2 enlists the 26 standardized large-scale surface and atmospheric predictor variables that were obtained from the SDSM website (http://co-public.lboro.ac.uk/cocwd/SDSM/). Daily climatic variables (precipitation, maximum and minimum temperature) are used as predictand variables in the climate change downscaling data. Daily precipitation and temperature data from the CanESM2 model were collected for RCP 4.5 and RCP 8.5 scenarios and converted to monthly data using the SDSM.

Table 4.2 List of the large-scale climate variables (Daily predictors variable) used for downscaling.

No.	Code	Daily Variable	No.	Code	Daily Variable
1	mslp	Mean sea level pressure	**14**	*P8_f*	*850 hPa geostrophic airflow velocity*
2	*p1_f*	*Surface geostrophic airflow*	**15**	*p8_u*	*850 hPa zonal wind*
3	*p1_u*	*Surface zonal wind*	**16**	*p8_v*	*850 hPa meridional wind*
4	*p1_v*	*Surface meridional wind*	**17**	*p8_z*	*850 hPa vorticity*
5	*p1_z*	*Surface vorticity*	**18**	***p8th***	***850 hPa geostrophic wind direction***
6	***p1th***	***Geostrophic wind direction***	**19**	*p8zh*	*850 hPa divergence*
7	*p1zh*	*Surface divergence*	**20**	p500	500 hPa geopotential height
8	*p5_f*	*500 hPa geostrophic airflow velocity*	**21**	p850	850 hPa geopotential height
9	*p5_u*	*500 hPa zonal wind*	**22**	prcp	Total precipitation
10	*p5_v*	*500 hPa meridional wind*	**23**	s500	500 hPa specific humidity
11	*p5_z*	*500 hPa vorticity*	**24**	s850	850 hPa specific humidity
12	***p5th***	***500 hPa geostrophic wind direction***	**25**	shum	Near surface specific humidity
13	*p5zh*	*500 hPa divergence*	**26**	temp	Mean temperature at 2 m height

Bold indicate variables that have not been normalized. Italics indicate secondary (airflow) variables derived from the pressure field.

Data Outputs

The statistically downscaled daily precipitation (Pr), daily maximum temperature and daily minimum temperature (P, Tmax& Tmin**)** for the current climate (1971–2005) and future scenarios under the RCPs (RCP 4.5 and RCP 8.5) are simulated. The following outputs were obtained during the process;

Daily Precipitation(P)

- Pr.PAR, a list of selected predictors for daily precipitation.
- Pr-syn. OUT, the model output of daily precipitation for the current period (1971–2005) generated using the weather generator.
- Pr -rcp45. OUT, projected daily precipitation under RCP 4.5 (2006–2100).
- Pr -rcp85. OUT, projected daily precipitation under RCP 8.5 (2006–2100).

Temperature Maximum (Tmax)

- Tmax.PAR, a list of selected predictors for daily maximum temperature.
- Tmax-syn. OUT, the model output of daily maximum temperature for the current period(1971–2005) generated using the weather generator.
- Tmax-rcp45. OUT, projected daily maximum temperature under RCP 4.5 (2006–2100).
- Tmax-rcp85. OUT, projected daily maximum temperature under RCP 8.5 (2006–2100).

Temperature Minimum (Tmin)

- Tmin.PAR, a list of selected predictors for daily minimum temperature.
- Tmin-syn. OUT, the model output of daily minimum temperature for the current

period(1971–2005) generated using the weather generator.

- Tmin-rcp45. OUT, projected daily minimum temperature under RCP 4.5 (2006–2100).
- Tmin-rcp85. OUT, projected daily minimum temperature under RCP 8.5 (2006–2100)

4.6.2 Description of statistical downscaling model (SDSM)

The basic principle of statistical downscaling is to identify statistical relationships between an observed small-scale predictand variable and larger-scale predictor variables for a baseline period and then apply these relationships to downscale future climate scenarios using a global climate model (GCM) output predictor. SDSM, developed by Wilby et al. (2002), as an appropriate tool for statistical downscaling is best described as a conditional weather generator because atmospheric circulation indices and regional moisture variables are used to estimate time-varying parameters describing daily weather at individual sites (e.g., daily precipitation or temperatures). SDSM was used in this study for the projection of Pr, Tmax, and Tmin under changing climate scenarios at every location used in this study. The downscaled process is either unconditional (wet day occurrence or air temperature) or conditional on an event (as with rainfall amounts).

For unconditional processes, such as temperature, there is a direct linear relationship between the predictand (U_i) and the chosen predictors (X_{ij}):

$$U_i = \gamma_0 + \sum_{j=1}^{n} \gamma_j X_{ij} + \varepsilon_i \tag{4.43}$$

Where U_i is temperature and X_{ij} is selected NCEP predictors on day i. Υ_j is the regression coefficient estimated for each month using least-square regression. $\mathcal{E}_i$ is assumed to follow a Gaussian distribution and is stochastically generated from normally distributed random numbers and added on a daily basis to the deterministic component model error.

The major steps adopted for downscaling of predictand (Pr, Tmax, and Tmin) are:

(i) Recognition of predictors variable

(ii) Calibration of predictand data and validation,

(iii) Generation of present and future time series data for Pr, Tmax, and Tmin.

(iii) Statistical analysis of downscaled present and projected Pr, Tmax, and Tmin for different scenarios.

(i) Recognition of predictors variable

The most important step in statistical downscaling is the screening/selection of suitable predictors for downscaling predictands. Since the maximum temperature time series is normally distributed so it is considered an unconditional variable. Therefore, to select a multivariate regression model in SDSM, the direct relationship is considered between observational precipitation (Pr), maximum temperature (Tmax), and minimum temperature (Tmin) as the predicted variable and large-scale predictors as independent variables. Regression parameters, along with NCEP and GCM predictors, are used to generate a maximum of 100 daily time series to fit closely with the observed data during validation, and twenty-time series are considered as the standard, a precedent set by other studies as well (Chu et al., 2010).

In general, the consensus is that a combination of variables is superior to individual variables in statistical downscaling, but finding a suitable combination of predictor variables is still a challenge because predictor choice can be complicated by the region, season, and predictands. Various indicators partial correlation, correlation matrix, explained variance, P-value, histograms, and scatterplots could be used to select some suitable predictors from a multitude of atmospheric predictors. Multiple co-linearity, which can misguide results among predictors, must be considered during the selection of predictors. Ordinary Least Squares (OLS) and Dual Simplex (DS) are two optimization methods available for SDSM. OLS is faster than DS and produces comparable results to DS. The selection of the first and the most appropriate predictor (Super Predictor) is relatively easy. Still, the selection of the second, third, fourth, and soon is much more subjective (Mahmood & Babel, 2014). According to Pallant (2007), a correlation coefficient of less than 0.7 between two predictors is acceptable. Percentage Reduction (PRP) in absolute partial correlation is calculated by the absolute correlation coefficient between a predictor and a predictand (R) and the partial correlation coefficient between a predictor and a predictand (Pr) using the following equation:

$$PRP = \left(\frac{Pr-R}{R}\right) \tag{4.44}$$

In the end, the predictor with the least PRP is selected as the second most suitable predictor. Similarly, the third, fourth, and further predictors are chosen by repeating the procedure outlined above.

It can be seen in different studies (Chu et al., 2010; Mahmood & Babel, 2014) that mostly 1-3 large-scale variables are believed to be enough to capture the variation of a predictand during calibration. It is better to use a smaller number of predictors during calibration because as the number of predictors increases in the regression equation, the chances of multiple co-linearity also increase. So, the fewer predictors, the lower the chance of multiple co-linearity during calibration.

There are certain measurement errors for the calibration of prediction outputs when using multivariate regression models that emerged from the validation stage of the SDSM model. Based on the available observed 35 years of data for temperature was divided into two daily datasets, first half (1971– 1995) 25 years of daily data was used for model calibration, and the second half (1995–2005) was used for model validation, In the present study, SDSM, using a monthly, and annually sub-model, was developed with the NCEP predictors that were selected during the screening process at each synoptic site of Uttar Pradesh. Standard Error (SE) of predictions and R square (R^2) were used as performance indicators during the calibration of SDSM, a procedure also followed in studies by Huang et al. (2011) and Mahmood & Babel (2013).

(ii) Weather generator

The weather generator operation produces daily synthetic data for the historical period by using calibration output and observed NCEP reanalysis atmospheric variables. It enables verification of calibrated models (using independent data) and the synthesis of artificial time series for present climate conditions, which means it allows comparison of observed and simulated model outputs. Graphical model outputs were used to plot statistical model outputs and observed data. The statistical outputs were the sum, mean, maximum, minimum, and variance of generated and observed weather data.

(iii) Climate scenarios generation

Since climate change will be highly dependent on human activities in the future, IPCC fifth assessment report (AR5) in 2014 considers various concentration scenarios issued RCPs that are related to the amount of greenhouse gases concentration, not emissions it supersedes Special Report on Emission Scenarios (SRES) projections published in 2000. Four concentration pathways have been selected for climate modeling and research, which describe different climate futures, all of which are considered possible depending on how much greenhouse gases will be emitted in the future. The RCPs, namely RCP 2.6, RCP 4.5, and RCP 8.5, are labeled after a possible range of radiative forcing values in the year 2100 (2.6, 4.5, and 8.5 W/m^2, respectively). Figure 4.4 details the process of scenario generation by employing SDSM.

The climate scenario generation step is based on the assumption that the predictor-predictand relationship is valid in the present as well as in future climate conditions. The synthetic weather data series is created by scenario generation option in the same way as the weather generator. At this time, the CanESM2 model output is provided as an alternative to NCEP/NCAR reanalysis data. Two CanESM2 scenarios, RCP 4.5 and RCP 8.5 are considered in this study. Synthetic daily weather records were generated for each scenario from 2006 to 2100 for all the variables and station data. The average of generated multiple ensembles was utilized as daily weather data for the required period. The future time period, mid-century (upto the 2050s) is considered for the downscaling of precipitation, Tmax, and Tmin.

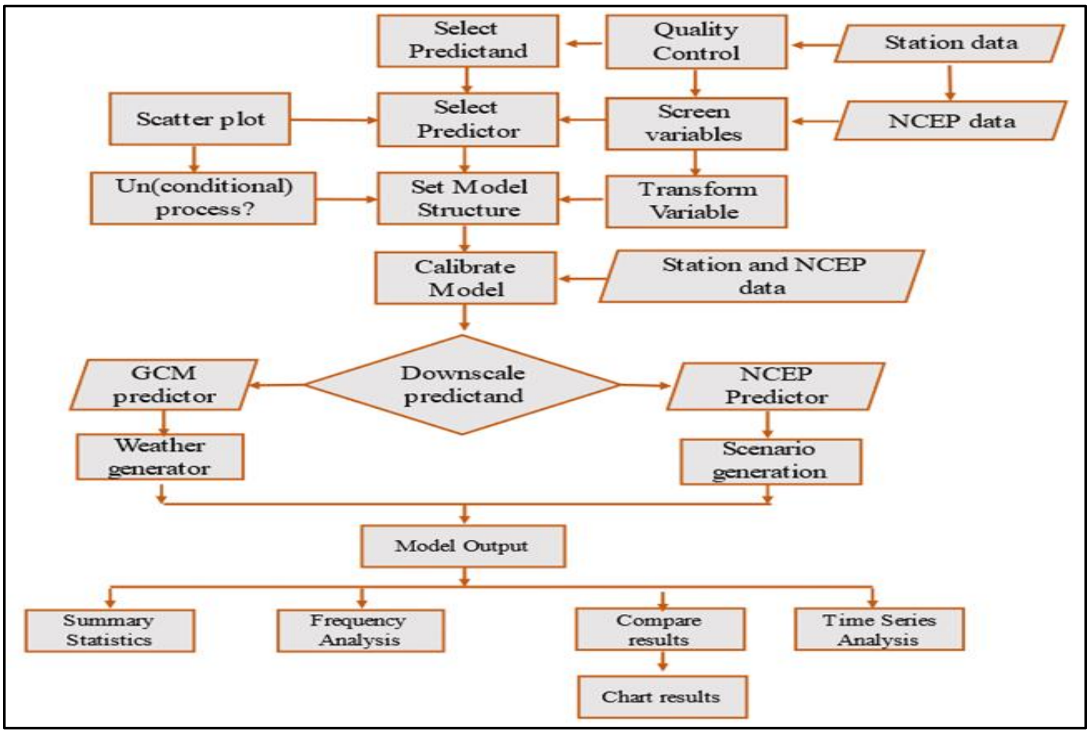

Figure 4.4 Flow chart demonstrates the climate scenario generation by SDSM (Wilby & Dawson, 2007)

4.7 Summary

This chapter describes the detailed methodology adopted in the present study. The first phase of the study provides a detailed methodology to investigate the temporal variability of the meteorological variable using trend assessment. The second phase of the study detailed the methodology to evaluate and compare the variability of drought characteristics estimated based on SPI and SPEI over the study area for the observed period. The next phase of the study detailed the framework to evaluate the spatial and temporal variability of dry and wet event characteristics and their transition characteristics based on SPEI at a one-month scale over the observed period. And in the last phase of this study discusses the development of the statistical relationship between predictors and predictands using the statistical downscaling model. Using the developed statistical relationship and CanESM2 model, the projection of future rainfall and temperature under climate change scenarios RCP 4.5 and RCP 8.5 is discussed. The last phase of this study is the future projection of drought characteristics estimated using SPEI at different timescales under RCP 4.5 and RCP 8.5.