CHAPTER 6

DEVELOPMENT OF NOVEL TREND APPROACH: VARIABLE SIZED CLUSTER ANALYSIS (VSCA)

6.1 General

The purpose of the Mann-Kendall (MK) test (Mann 1945, Kendall 1975, Gilbert 1987) is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (downward) trend means that the variable consistently increases (decreases) through time, but the trend may or may not be linear. The MK test can be used in place of a parametric linear regression analysis, which can test if the slope of the estimated linear regression line is different from zero. The regression analysis requires that the residuals from the fitted regression line be normally distributed, an assumption not required by the MK test; that is, the MK test is a non-parametric (distribution-free) test. To determine the intermediate behavior of trend and characterization of trend pattern, the Novel tool VSCA has been developed. The tool also can determine the change point detection.

6.2 Variable Sized Cluster Analysis

In fact, time-series data usually do not display a monotonic behavior but show non-periodic alternating behavior during the intermediate time span. Thus, the traditional parametric/non-parametric tests for aggregated datasets may not unravel real-time information, which may differ from the aggregated character of the datasets. Changes in the mean precipitation, frequencies of occurrences of extremes, the magnitude of evapotranspiration, and river discharges result from non-stationarity in hydrologic processes. Complex interdependence and interactive dynamic mechanisms of climatic variables continuously update hydrometeorological processes and add non-stationarity to climatic variables. Anthropogenic activities (Zhang et al. 2010) also perturb the interactive mechanisms by impacting the forcings, impacting the state of the climate.

In order to capture the intermediate character of precipitation trend which swings during a study period, the present study proposes to apply repeated test protocol of (Zhang et al. 2010), which was implemented for stream flow dataset of west-branch Susquehanna River in Bower watershed. By varying the size of the flow dataset of the west-branch Susquehanna River in the Bower watershed, they repeatedly applied the MK test to discern the pattern of the trend of river flow through color variation for abrupt or gradual changes. The methodology involved different data clusters of varying sizes with changed beginning and ending times of the data series. The pattern of precipitation trend over Uttar Pradesh (UP) that falls in the middle range of the Ganges River basin and state of Jharkhand of India is analyzed and plotted for ascertaining consistency/ monotonicity of the trend. Most of the work pertaining to the analysis of precipitation trends in the Indian sub-continent accounted for the spatially averaged dataset from various states or grid dataset from a large segment of India by assuming rainfall to be uniformly distributed over the large region of India, which appears to be unrealistic; hence station (district) scale data was applied for a more realistic analysis of precipitation trends. Thus, rainfall data of eighteen divisional regions of Uttar Pradesh and five division regions of state Jharkhand, India, were considered to obtain the overall view of the trend pattern. The approach involves the repetitive application of the Mann-Kendall test with moving window, with varied beginning and ending times, i.e., varying size of the data cluster with a moving window was applied to yield 3-D pictorial representation for the easy realization of the pattern of rainfall trends over the state area.

The Mann-Kendall test with varying beginning (T_b) and ending (T_e) points was repeatedly applied to the precipitation time series dataset for 18 synoptic stations of Uttar Pradesh. The minimum size of the data cluster (window size) was initially taken as 5, 10, 15, and 20 for the data of district Agra to infer if it affects results of analysis; however, it was later fixed as 10 for repeated application of Mann-Kendall trend test for all districts. The total number of computations to be carried out for repeated application of the Mann-Kendall test for a given district was calculated as:

$$N = [(n-8).\ (n-9)]/\ 2$$

The Z values obtained from each application of the MK test were plotted on 2-D and 3-D planes.

The 3-D visualization of Z values adds a novel feature to this study. The numbers along the abscissa of Figures (7.1, 7.4, 7.7 … 7.51) represent the beginning time, whereas the ordinate axis locates the ending time of the data clusters picked up to determine the Z values. The top view of Figures (7.1, 7.4, 7.7 … 7.51) can be imagined to appear as Figures (7.2, 7.5, 7.8 … 7.53). The Z-statistic is displayed in different shades of colors in the upper left triangular portion of the figures. Figures (7.1, 7.4, 7.7 … 7.52) show the values of Z statistic corresponding to variable-sized data clusters with varying beginning and ending times, i.e., the size of the data cluster adopted for the computation of Z-statistic varies as one moves around the graphical plane. These variable-sized data clusters enable one to calculate different values of Z-statistic to be plotted for the assessment of the pattern of a trend of precipitation. In order to represent the magnitude of the Z-statistic, an axis (Z-axis) normal to the X-Y plane was used; thus, these figures show a 3-D pictorial view of an erratic surface spread in consonance with the varying values of Z-statistics.

The Z values determined from the N number of computations are also plotted in 2-D. Figures (7.2, 7.5, 7.8 … 7.53) The graduated color scheme (or shades) shows the beginning time on the X-axis and the ending time on the ordinate axis (Y-axis). The shades of the graduated color denote the underlying Z values of individual tests. Thus, a point on the plot with coordinates (T_b, T_e) shows the Z value obtained from the Mann-Kendall test applied to a data cluster that starts with T_b and ends with T_e. Since the length of the time series dataset decreases with the increasing beginning time due to the last data being fixed as an ending time (118), the resulting plot covers the upper left triangular zone in all Figures (7.2, 7.5, 7.8 … 7.53). Figures (7.1, 7.4, 7.7 … 7.52) are analogous to a geographical map representing the elevation of a point with a shade of color. The lower and upper triangular markers represent increasing and decreasing trends, respectively, with a circle as a marker for no trend. The higher values of Z indicate an increasing trend and are displayed by color shades close to red, whereas a decreasing trend is demonstrated by the negative values of Z shown by blue color and its varying shades. No trends are marked by green color shade with Z values as zero. A trend pattern can then be observed by the changing elevation (±) and alterations of the color shades, a typical feature obtained in this study. However, a pattern is easily visible in Figures (7.1, 7.4, 7.7 … 7.52).

It is important to note that the line parallel to the hypotenuse of the right-angled triangle is shown in Figures (7.1, 7.4, 7.7 … 7.52) depicts the Z-values for a fixed window size (size of data cluster) but increases beginning time. This way, a point on the hypotenuse represents a window size comprising 10 data with successive parallel lines to it signifying window sizes of 11, 12, 13, 14… The points on vertical lines drawn perpendicular to the abscissa represent an increasing window size for a fixed beginning time, i.e., data from 1^{st} to 10^{th}, 1^{st} to 11^{th}, 1^{st} to 12^{th}, and so on.

Figures display the changes in the pattern of trend, and changes in the shades of color are utilized to identify the point of significant change in the trend of precipitation. Results pertaining to the detection of change points were ascertained by a methodology that involved a plot of U_t statistics of PMW with an application of negative sign, i.e., the plot of "$-U_t$"

(referred to as modified U_t). The induction of negative sign with U_t term was intended to add congruity between physical phenomenon that actually occurred and the data. This change transforms the character of PMW U_t-statistics incongruity with the actual pattern of trend. The same is referred to as inverted PMW U_t -statistics and suggested to be applied accordingly for all cases to show the true picture of the change in the pattern of trend. It detects multiple change points rather than the single point detected by the traditional PMW method.

6.3 Strength of Proposed VSCA

A novel analysis procedure referred to as variable-sized cluster analysis (VSCA) was developed to identify trends and change points in precipitation time series. The procedure involved station-scale rainfall data of 118 years from eighteen divisional districts of the state of Uttar Pradesh (UP) and five divisional districts of Jharkhand, India. In contract with the traditional Mann Kendall (MK) test, which yields a monotonic trend for the whole span of time, VSCA enables to detection of multiple change points while characterizing the pattern of precipitation trend over the historical time period. The Pettit Mann Whitney (PMW) test was also modified to graphically represent multiple change points which confirmed the results of VSCA. Thus, VSCA demonstrated the unified strength of MK and PMW tests. The 3-D figures, drawn for visualizing the changing trend of precipitation, utilized a 118-year long time series dataset with the minimum size of data cluster as 10, which resulted in the right triangular shape of the graphs due to the repeated application of MK test to variable-sized data clusters.

The PMW test relies on a single maximum value of U_t for the overall time span, and the canonical Pettitt-Mann-Whitney test cannot detect the multimodal characteristics of the change point. In contrast, the variable-sized cluster analysis enables one to identify the different change points that might have occurred over a century-long time duration. Thus, VSCA owns a unique potential to detect one or even more change points if they occurred over the long duration of the study period. For water resources planning, development, management, and disaster management, it is essential to know the overall trend of precipitation and intermediate outbursts of the changed behavior.

6.4 Summary

In contravention to the canonical way of handling precipitation data for trend analysis that yields unique trend over the long span of time, the proposed methodology involving variable size cluster analysis presents a pictorial view of the pattern of the changing trend along with the clue for the detection of one or more change points and/or alternate trends, if it occurs, without application of Pettitt Mann Whitney method. The modified PMW statistics drawn on a graph display results congruous to the actual pattern of changing trend, and hence proved to be a better option than the traditional PMW methodology that detects only one change point/ point of the sudden jump. 3-D figures are drawn for variation of the Z-score with varying starting and ending time clearly depicted the pattern of the changing trend as an undulated surface for the visual feel of the variation in precipitation trend that was not detectable by the ordinary MK test involving complete historical dataset. Trend analysis using station scale data showed a more realistic result than the grid-based data built upon spatial averaging. The same trend, if obtained by the different clusters created by sequential elimination of starting year data in the application of MK test, enabled one to have more confidence and reliability on the conclusive result.