RESEARCH REVIEW JOURNALS
Volume-03
Issue-06
June -2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Detecting the Abrupt Change in Hydrological Time Series: A Methodological Review

Rupak Kumar Paul

Assistant Professor, Dewan Abdul Gani College, P.O. – Harirampur, District – Dakshin Dinajpur, W.B., PIN – 733125

ARTICLE DETAILS

Article History
Published Online: 19 June 2018

Keywords
Hypothesis Testing, Non-parametric Test, Step Chang Detection.

ABSTRACT

Hydrological time series particularly the discharge data series carries the signature of the change in basin hydrological phenomena that may occur abruptly or gradually or in more complex manner. Many statistical tests are available for capturing such temporal behaviour of the hydrologic processes but still the task is quite challenging since much care needs to be taken regarding variable selection, data pre-processing and choosing right statistical tests depending upon the nature of the data, geographical condition of the basin, purpose of the study etc. Very often more than one method of change detection is required to be applied and ultimate decision making becomes subjective depending on the geographical condition of the basin. So, one needs to be very cautious about the applicability of the available tests when these are being applied. Here, an attempt has been made to review the assumption and applicability of available statistical tests which are solely designed for detecting step change in hydrological time series.

Introduction

In the present era the basin hydrological phenomena specially the stream flows are getting changed due to several factors like climate change (or climate variability) coupled with human interference through several pathways like land use and land cover change, forest clearing, afforestation, urbanization, dam and reservoir, canal etc. which control the temporal behaviour of river flow at a given point in space and this kind of alteration in stream flows is obviously reflected in discharge time series. As pointed out by Kundzewicz & Robson (2000), two types of changes in stream flow are usually observed: one is the abrupt change (also called step change) which may be caused by such phenomena as dam or reservoir construction, channel diversion etc. and another one is the gradual change (also called trend) which usually occurs as a result of urbanization, deforestation, climate variability or climate change and many other alike interferences. Even a complex combination of these two processes can cause stream flow changes (Kundzewicz, Z. W. & Robson, Alice, 2000 & 2004).

Any water structure is typically designed based on the assumption that hydrological processes do not change with time and if this assumption is violated the existing code of design of the structure has to be likewise revised in order to achieve the target, otherwise the structure would be either under designed i.e., missing the target or overdesigned i.e., becoming overly costly (Radziejewski et al., 2000). So, the issue of change detection in hydrometeorological time series finds its practical essence among water engineering and hydrological community.

Both the kinds of changes in hydrologic time series can be captured by using graphical and statistical methods. The graphical method, in this context, helps us in data visualization but the certainty or uncertainty of the alteration cannot be evaluated by the technique. So, graphical tools can be used as exploratory data analysis in view to diagnose important features of the data set while statistical tests, on the other hand, are the sophisticated devices to detect the change in the concerned parameter with incorporation of uncertainty analysis although there are a lot of limitations of the techniques also.

A number of statistical tests are available in literature and each one is designed for its specific use to detect a specific type of hydrological change. Very often, it is observed that different statistical tools identify different change points for the same time series and the ultimate decision becomes very much subjective depending upon the geographical condition and the nature of change i.e., abrupt or gradual change. Here, an attempt has been made to make a documentation of relevant statistical testes which are popularly used for capturing the abrupt changes occurring in discharge time series data along with a brief discussion on applicability of each.

Data and Data Pre-processing

Much care needs to be taken to ensure the quality of the data before commencing any statistical test, otherwise the poor data quality may result misleading interpretation. It is also important to choose the right variable type, appropriate form and frequency of data which is to be done in accordance to the problem to be resolved and objective of the study. As for example, for evaluating the hydrological processes completer duration series is the right choice while for flood analysis partial duration series or high extreme value series will be useful.

The discharge time series data is usually arranged in four different formats based on frequency of observations: (i) Complete Duration Series which consists of all the available observations, (ii) Partial Duration Series which contains the observations above a specified value, (iii) Annual Exceedance Series wherein observations above a specified value are shown in such a way that number of observations and number of years involved in the time series become equal and (iv) Extreme Value Series which contain the biggest or smallest values within the time step at which data is collected.

After choosing data series of appropriate frequency, it is necessary to check the data for missing value, 'censored' value and any other noise present in the data set. Censored data, as defined by Hipel & McLeod (1994) for water quality data series, refers to that observation which is reported as being less than or greater than detectable level. The defined level below or above which value of observation is reported is said to be *truncation level*. Data series may be *singly censored* if there is only one detection limit and it is *multiple censored* when more than one detection limit is specified for the data series. The management of these issues depends on the variable type and the focus of the study as well (Hipel & McLeod, 1994).

In addition to handling all these issues Exploratory Data Analysis (EDA) is another part of data pre-processing which is essential for detection and visualization of important features of the data, its structure or any anomalies thereby in the dataset by means of graphical and statistical tool. Exploratory data analysis is generally performed at more than one stage. As described by Grubb & Robson (2000), "at each stage, graphs are plotted and then refined so that the important features of the data can be seen clearly. Often patterns or features emerge that need further exploration. These might include seasonal variation, correlation or a problem with some data values. Because it is an exploration of the data, no two analyses will be the same".

Time series plot, multiple time series plot, simple scatter plot, scatter plot with fitted smoothed curve, scatter plot of transformed data or residual value instead of plotting the original data, scatter plot matrix, simple spatial plot of data on map by symbol, contour plot, box and whisker plot, percentile-percentile plot (p-p plot) or quantile-quantile plot (q-q plot) of original data or residual value, histogram, auto correlogram etc. are some of the tools usually used for exploratory data analysis which helps researcher get insight into the behaviour of data series prior to application of statistical test.

An Overview of Hypothesis Testing

Hypothesis test is the mechanism of estimating population properties based on sample dataset which is drawn from the said population and also guessing the precision of the estimation (Pal, 1998). The statistical measures which characterize the population (e.g., population mean, population standard deviation, correlation co-efficient of two different kind of population dataset etc.) are called *"parameters"* whereas the statistical measures worked out based on sample data (e.g., sample mean, sample standard deviation, correlation co-efficient of two different kind of sample dataset etc.) are called *"statistics"*. So, by means of hypothesis test one estimates the parameters based on statistics and judging how precise is such an estimation and this latter one is the issue of significance test and they are inseparable from each other, though very often they are discussed separately. The theory of tests of hypotheses, as pointed out by Hipel & McLeod (1994), was originally developed by Neyman & Pearson (1928, 1933) while significance testing is due largely to Fisher (1973).

This is obviously the problem of *sampling theory* which deals with relationship between sample and population by applying the law of probability (Pal, 1998). What is done in hypothesis testing is that generalization is made based on a particular information i.e., it is a sort of movement from particular (sample) towards general (population) what is known as *"statistical induction"* or *"statistical inference"* and in doing so, a deductive argument is used wherein population property is guessed and then property of sample drawn from that said population is investigated by applying law of probability (Pal, 1998) to know whether or not both the properties i.e., property of population and that of sample do match well with each other.

Hipel & McLeod (1994) have illustrated the concept which is put here. Suppose one wishes to investigate whether or not the discharge of a river possesses certain property. Say, for example, that property is the presence of abrupt change over a certain period of time. Sample annual time series discharge data of that river is available. Now the investigation is done following the steps mentioned below.

First, it is typically assumed that the population from which sample dataset has been drawn does not possesses the property. As per the present example the typical assumption, then, will be that there is no abrupt change in river discharge and this assumption which is made about the population is called *Null Hypothesis* (H_0) since it nullifies the presence of certain property in the population data. According to the logical understanding, there will be no abrupt change in the sample data since this sample has been drawn from the population. It should be kept in mind that whatever assumption or hypothesis is formulated for running the statistical test it is formulated about population and NOT about sample data. Alternatively, it is again assumed that the sample has been drawn from the population which has abrupt change and this assumption is called *Alternative Hypothesis* (H_1).

Second, in order to choose between H_0 and H_1, a test statistic is formulated based on sample dataset in such a way that this test statistic will follow a specific distribution if the null hypothesis is true, otherwise it will follow some other distribution. The distribution which the test statistic follows under the condition that null hypothesis is true (i.e., there is no abrupt change in

discharge as per the present example) is called *distribution under null hypothesis*. This distribution under null hypothesis is known prior to application of the test since the test was designed theoretically based on this distribution.

Third, it is observed whether this computed test statistic falls in the rejection or acceptance zone of the priori known distribution under null for a desired significance level, α (alpha, usually 0.05 or 0.01 is taken as α value). Alpha is the probability of rejecting null hypothesis when it is true. This is technically called *Type-I error*. If this test statistic falls in the acceptance zone, it is said that null hypothesis fails to be rejected, i.e., the sample has been drawn from the population which does not possesses the certain property i.e., abrupt change (as per present example). If, on the other hand, the test statistic falls in the rejection zone of the distribution under null, it is said that the alternative hypothesis is accepted or alternatively it can be said that there are enough evidences to decide that sample data (here, the discharge) has been drawn from some other population where there is abrupt change.

So, the test statistic is considered to follow the distribution under null if it falls in the acceptance zone of that distribution. On the other hand, test statistic is considered not to follow the distribution under null if it falls in its rejection zone; but what distribution does it follow on rejection of null, nothing is talked about, which means investigator only investigates the distribution under null and that is why the null hypothesis is very often called the hypothesis under test (Hipel & McLeod, 1994).

There are all together three distributions involved in hypothesis test: the *first one* is the distribution of the population dataset from which present sample data has been drawn, *second one* is the distribution that test statistic follows if null hypothesis is true what is called distribution under null hypothesis and *third one* is the distribution which test statistic follows if null hypothesis is rejected i.e., this distribution can be said as the distribution under alternative hypothesis.

The first one (i.e., the distribution of population dataset from which sample data in hand has been drawn) is investigated to know whether or not it possesses normal distribution, since there are some tests which assume that the population from which sample data has been drawn follows the normal distribution. It is to be of worth mentioning that normality checking is done using sample data but it is carried out to know the distribution of population data from which this sample has been drawn i.e., normality checking is also a hypothesis test which is a movement from sample to population.

Then the researcher is concerned only about second distribution (i.e., distribution under null) and last one is not investigated at all, since with only investigating the second one (i.e., distribution under null) right decision can be taken. Whether or not it is investigated, the third distribution (i.e., distribution under alternative hypothesis) is important for calculating Type-II error. This Type-II error, usually denoted by β, is the probability of accepting null hypothesis when it is false. Accepting null hypothesis means rejection of alternative hypothesis and the error involved in rejection of alternative hypothesis is the Type-II error which is to be calculated using the distribution under alternative hypothesis (i.e., the distribution which the test statistic follows if the null hypothesis is false). It is just opposite to Type-I error which is calculated using the second distribution i.e., the distribution under null hypothesis (i.e., the distribution which test statistic follows if null hypothesis is true). This is to be of worth mentioning that $(1-\alpha)$ is considered as the confidence level while $(1-\beta)$ is considered as the power of the test.

There are two types of design of statistical tests or to say there are two groups of statistical tests. In one group, the test statistic is computed based on actual data in interval or ratio scale and this group of tests are called *parametric tests*. These parametric tests use mean and standard deviation of the dataset which are essentially tied to normal distribution of the population dataset and that is the reason behind checking the normality of the data under investigation and by the word “parametric” they refer to the characteristics of population (Pal, 1998). In other group of the tests, the test statistic is computed based on either rank (i.e., NOT the original data) of the sample dataset which was originally collected in interval/ratio scale or it is computed based on the frequency of the ordinal or nominal scale data and this group of tests are called *non-parametric tests*. The three most commonly used parametric tests are z-test, Student’s t test, Fisher’s F test (Pal, 1998) while most commonly used non-parametric test is Chi Square test.

In testing the changes in hydrological datasets non-parametric tests are mostly used since this type of time series is very often found to be not normally distributed and the non-parametric tests are applicable here as the tests are distribution-free i.e., these tests do not assume that population from which sample data has been drawn are having normal distribution. So, only two types of distributions are dealt in when one works with non-parametric tests. One is the distribution which the test statistic follows if null hypothesis is true (i.e., distribution under null hypothesis) and second one is the distribution that test statistic follows if null hypothesis is false (i.e., distribution under alternative hypothesis). The distribution of the population dataset from which present sample data has been drawn is not checked since the non-parametric test does not have any assumption regarding this distribution.

The results of statistical tests should be considered as the general check for evidence of the change not the proof since such results are the expression of probability not certainty (Robson et al., 2000). Always there lies a certain amount of probability of detecting change while it is not there in reality. Eventually, the statistical tests do not address the causal explanation for the change

even if it is detected with satisfactory amount of certainty and very often if the data does not properly meet the required test assumption or if the data is very much messy or noisy in nature then the resultant output might be misleading.

The Diagnostic Check

The available statistical tests which are used for change detection in time series include both, *parametric* and *non-parametric* type of test. Non-parametric tests are distribution free i.e., they do not assume any specific distribution of data which are to be tested but parametric tests, if not mentioned otherwise, make assumption that sample data follow normal distribution and/or the data are independent (i.e., there is no autocorrelation in the series). So, prior to the application of statistical tests for change detection, the following diagnostic checks of the data is essential.

1. Kolmogorov-Smirnov Test

This is non-parametric test used for checking normality of the data set by comparing the empirical time series distribution with theoretical normal distribution.

The series is first divided into several classes, then observed relative frequency of all classes is calculated and then summed up to get observed cumulative relative frequency. According for each class theoretical cumulative frequency is calculated using normal distribution. Maximum difference of these two for any class is the test statistic for this test (Pal, 1998). The test statistic is defined as:

$$D = \max\left|F_0(x) - F_r(x)\right|$$

D is the test statistic, $F_0(x)$ is the observed cumulative relative frequency distribution of a random sample of *N* observation while $F_r(x)$ is the theoretical relative cumulative frequency distribution. The critical values for *D* can either be taken from textbooks, or assessed directly using re-sampling (Robson et al., 2000).

The test can also be used to test the trend change between two time periods if change point is known. For that purpose, the series has to be divided into two sub-series based on known change point and then frequency distribution of both the sub-series are compared using this test. So, the test here is the two-sample test. Details about this two sample Kolmogorov-Smirnov test are available in Robson et al. (2000).

Apart from Kolmogorov-Smirnov test, there are other several tests which are used to check the normality of the data, e.g., Anderson-Darling test, Shapiro-Wilk test, Cramer Von-Mises test etc.

Next to normality check, another important aspect that is to be tested for time series data is the test of independence against autocorrelation as there are many tests for change detection assume that data is independent.

Following are the popular ***tests for checking the independence*** (or randomness or homogeneity) ***against autocorrelation*** (or dependence or non-randomness) of the data set.

2. The von Neumann Ratio Test

This test was formulated by John von Neumann in 1941 for testing the independence or homogeneity or randomness or autocorrelation of the time series (Neumann, J. von, 1941). Let $Y_1,\ldots,Y_n$ be the normally distributed annual series (*i* is the year from 1 to *n*) with $\bar{Y}$ being the mean value of the series, von Neumann ratio is defined by:

$$N = \sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2 \Big/ \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

The ratio *N* is closely related to the first-order serial correlation co-efficient and the value is bounded in between 0 and 4 (W.M.O., 1966; Buishand, 1981). Under the null hypothesis of a constant mean, *E(N)* = 2 while for a non-homogeneous record the mean of *N* tends to be smaller than 2 and even its value may exceed 2 if there is much variation in mean (Buishand, 1982).

The null hypothesis (H_0) states that the series is independent (or random or homogeneous) while alternative hypothesis (H_1) states that series is dependent (i.e., autocorrelated or non-random or non-homogeneous). Depending upon the context of the test, the appropriate phrase from the alternative ones given in the parentheses is used to frame the null or alternative hypothesis.

In the context of test of independence, the null is rejected if the value of *N* is less than or greater than 2 and null fails to be rejected if the value of test statistic is equal to or close to 2. On rejection of null, in present context, it is said that data are dependent i.e., series is autocorrelated or it is non-random. The *N* tends toward zero if there is possible positive autocorrelation in the series while possible negative autocorrelation is indicated by larger value of *N* (Neumann, von J. et al., 1941).

In the context of test of homogeneity, null is that the data are homogeneous with a constant mean while alternative hypothesis is that it is non-homogeneous i.e., there is a break somewhere in the data set where the mean value of the series shifts but this test is not capable of detecting the exact change point in the series. A table of percentage points of *N* for normally distributed samples is given by Owen (1962) (Buishand, 1982).

3. Durbin-Watson Test

This test was first introduced in 1950 by James Durbin and Geoffrey Stuart Watson and modified subsequently (Durbin& Watson, 1950; Durbin& Watson, 1951; Durbin & Watson, 1971). The test was originally formulated to capture whether autocorrelation exists in residuals of least square regression, but presently the it is widely used to capture whether serial autocorrelation exists in time series data (Pal., 1998).If $X_t,\ldots,X_n$ is the annual time series (*t* is the year from 1 to *n*), ρ is the serial autocorrelation, the Durbin-Watson statistic, D is formulated as follows:

$$D = \frac{\sum_{t=2}^{n}(X_t - X_{t-1})^2}{\sum_{t=1}^{n} X_t^2}$$

The null hypothesis (H_0) is that there is no autocorrelation in the series (H_0: $\rho = 0$) and alternative hypothesis (H_1) is that the series elements are autocorrelated (H_1: $\rho \neq 0$).The value of D is always bound in between 0 and 4. Since the test statistic D is approximately equal to $2(1-\rho)$, D = 2 it indicates that there is no autocorrelation. On the other hand, there is evidence of positive serial autocorrelation if the value of D is less than 2 and there is evidence of negative serial autocorrelation if the value of D is greater than 2.

This Durbin-Watson test in its D statistic provides the lower and upper limits (critical value), d_l and d_u to which we compare the calculated D value for evaluation of significance of the statistic (Pal., 1998). To test for *positive autocorrelation* at significance α, the test statistic D is compared to the lower and upper critical values (d_l, α and d_u, α) (Pal, 1998) as:

If D< d_l, α, there is statistical evidence that the elements of the series are positively autocorrelated.

If D> d_u, α, there is no statistical evidence that the elements of the series are positively autocorrelated.

If d_l, α <D<d_u, α, the test is inconclusive and this gives a hint that large sample is required.

To test for *negative autocorrelation* at significance α, the test statistic (4 − D) is compared to lower and upper critical values (d_l, α and d_u, α) (Wikipedia, accessed on 21.11.2017):

If (4 − D) < d_l, α, there is statistical evidence that the elements of the series are negatively autocorrelated.

If (4 − D) > d_u, α, there is no statistical evidence that the elements of the series are negatively autocorrelated.

If d_l, α < (4 − D) < d_u, α, the test is inconclusive. this gives a hint that large sample is required.

Statistical Tests for Detection of Abrupt Change

The main objective of the statistical tests discussed below is to evaluate the homogeneity of the dataset i.e., to evaluate whether the mean of the time series is constant or it shifts after a particular point of time. If dataset is proved to be non-homogeneous i.e., if its mean is not constant, rather its value shifts, the test can be used subsequently to detect the exact point of time where the mean of the time series shifts. Let $X_i,\ldots,X_n$ be an annual time series (where *i* is the year from 1 to *n*). There will be a single mean for that series if the series is proved to homogeneous. On the other hand, if the series is proved to be non-homogeneous, there will be more than one mean. If the series has only one change point (say, *k*) there will be two means: one is before mean, say, $\overline{X}_1$ which occurs for first *k* observations and another is after mean, say, $\overline{X}_2$ which occurs

for $(n-k)$ observations. If there are two change points in the series, the series will be divided into three sub-series and there will one mean for each sub-series and soon. There are two groups of homogeneity tests: ***Absolute Homogeneity Test*** and ***Relative Homogeneity Test***.

The first four tests discussed below are used for testing absolute homogeneity while the last one is applied to test for relative homogeneity of the time series data.

1. Distribution-free CUSUM Test

This is a sign based non-parametric test used for detection of the shift in mean of a time series for an unknown time of change. In this test cumulative sum (CUSUM) of the sign of deviation of individual element of the series from its median is used to formulate the test statistic (McGilchrist & Woodyer, 1975). If $X_1,\ldots,X_n$ is a time series with X_{median} being the median, then the test statistic (Robson et al., 2000) is defined by:

$$TS = \frac{2}{n}\max\left|CS_k\right|, \text{ for } k = 1,2\ldots, n; \text{ where, } CS_k = \sum_{i=1}^{k} sign(X_i - X_{median})$$

The sign here follows the sign function i.e., 1 for positive, 0 for zero and -1 for negative. Theseries presents a break at time *k* where max|CS_k|(for $1 \le k \le n$) is observed.

The test statistic (*TS*) follows the Kolmogorov-Smirnov test statistic (*KS*) for the equality of distribution of two random variables: (i) times at which observations greater than the median occur (ii) times at which observations less than the median occur (Robson, Alice et al.,2000) and hence and the standard algorithms for the Kolmogorov-Smirnov test can be used for determination of the percentage points of the test statistic. The critical value of max|V_k| at various significance levels is $1.22\sqrt{n}$ at α = 0.1; $1.36\sqrt{n}$ at α = 0.05 and $1.63\sqrt{n}$ at α = 0.01 (Chiew & McMahon, 1993).

2. Pettitt Test

The Pettitt test (also known as *Median Change Point test*) is a *nonparametric test* formulated by A. N. Pettitt in 1979 (Pettitt, 1979; Robson et al, 2000). The test was derived from the Mann–Whitney test in order to detect a single break point in an annual time series (Pettitt, 1979; Dagnélie, 1970; Pérez, 2011). For the data series $x_i\ldots,x_N$ the variable $U_{t,N}$ is defined as follows:

$$U_{t,N} = \sum_{i=1}^{t}\sum_{j=t+1}^{N} sign(x_i - x_j)$$, where *sign* denotes the sign function i.e., 1 for "positive", 0 for zero and –1 for "negative".

The summation (for $i = 1,2\ldots,t$) used here for computation of variable $U_{t,N}$ is the cumulative sum.

The statistic $U_{t,N}$ is equivalent to a Mann-Whitney statistic for testing that the two samples $x_i\ldots,x_t$ and $x_{t+1}\ldots,x_N$ come from the same population. The statistic $U_{t,N}$ is then considered for values of *t* with $1 \le t < N$ (Pettitt, 1979). Null hypothesis (H_0) states that there is no change in the median of the series while against alternative hypothesis (H_1) states that there is change. The null hypothesis is tested using following test statistic (Pettitt, 1979):

$$K_N = \max_{1\le t<N}\left|U_{t,N}\right|$$

The series presents a single break at time *t* where K_N is observed (Pérez, 2011). For changes in one direction with same null distribution, the statistics (Pettitt, 1979) are defined as follows:

$K_N^+ = \max_{1\le t<N} U_{t,N}$ for positive change and $K_N^- = \min_{1\le t<N} U_{t,N}$ for negative change.

Using rank theory, Pettitt shows that variable $U_{t,N}$ can be computed as a rank statistic. If *k* represents the K_N value of the study series under the null hypothesis, the variable $U_{t,N}$ can be recomputed as:

$U_{t,N} = 2\sum_{i=1}^{k} r_i - k(N+1)$; where *k*=1, 2…, *N* and $\sum r_i$ is the cumulative sum of the rank, r_i of the original value. Here rank is assigned in ascending order i.e., rank 1 is assigned to smallest value of the series while rank r_N is assigned to the N^{th} observation of the series and tie value is addressed in the same way as it is done in computing Mann-Whitney statistic or Spearman's ρ.

The probability of exceeding the *k* value is computed as follows (Pettitt, 1979; Pérezet al., 2011):

$$\text{Prob}(K_N > k) \approx 2\exp\left[\frac{-6k^2}{N^3 - N}\right]$$

This probability calculation is based on Bernoulli experiments (Robson et al. (2000). For a significance level α, the null hypothesis is rejected if α is greater than the estimated probability.

This test has also several modified versions. For example (Robson et al., 2000), the test statistic

$$K_N^* = \max\left|\frac{U_{t,N}}{\sqrt{Nt - t^2}}\right|$$

is an alternative indicator for the change point. For this test statistic the significance levels should be estimated using resampling methods (Robson et al., 2000).

3. Worsley Likelihood Ratio Test

This is parametric test introduced by K. J. Worsley in 1979 (Worsley, 1979) based on original work of P. M. Hawkins (Hawkins., 1977). The test assumes that data is normally distributed and that the change point is unknown (Robson, Alice et al., 2000). If $X_t,\ldots,X_T$ is the time series with $\bar{X}$ being its mean and S_X being the standard deviation, the test statistic Q is then defined as:

$$W = \frac{V\sqrt{T-2}}{\sqrt{1-V^2}}$$

where,

$$V = \max\left|S_k^*\right|;\ S_k^* = \frac{\sqrt{k(T-k)}CS_k}{S_X} \text{ and } CS_k = \sum_{t=1}^{k}(X_t - \bar{X});\ \text{for } k = 0,1\ldots, T$$

Here, CS_k is the cumulative sum. Null hypothesis (H_0) is that data are homogeneous with constant mean while alternative hypothesis (H_1) is that there is a break in the series where series mean shifts. The change point is indicated by the point on time where *V* is observed. Critical values for different significance levels for this test have been derived by Worsley (1979) (Robson et al., 2000).

4. Buishand Test

This is a *parametric test* introduced by T. A. Buishand (Buishand, 1981, 1982 & 1984) for detection of the specific year where the mean of the annual time series shifts. The test assumes that series is independent and normally distributed (Wijngaard et al., 2003; Yozgatligil, Ceylan & Yazici, 2016). The test is formulated based on adjusted partial sum and it is more sensitive to break in the middle of a time series (Hawkins, 1977; Wijngaard et al., 2003). If $Y_1,\ldots,Y_n$ is an annual series, with mean value of $\bar{Y}$ the *adjusted partial sum* (Buishand, 1982; Hänsel et al., 2016) is defined as

$$S_0^* = 0,\ S_k^* = \sum_{i=1}^{k}(Y_i - \bar{Y}),\ k = 1,\ldots,n$$

Summation (for $i = 1,2\ldots,k$) used here for construction of the variable S_k^* is the cumulative sum. When $k = n,\ S_n^* = 0$, then variable S_k^* will float around zero if the series is homogeneous, and there is no systematic pattern in the deviations of the series components from their mean. If there is a break in year *K*, then variable S_k^* reaches a maximum (negative shift) or minimum (positive shift) near the year $k = K$ (Yozgatligil, C. and Yazici, C., 2016). For significance test Buishand (1982) derived three test statistics which are defined as

$$Q = \max_{0\le k\le n}\left|\frac{S_k^*}{D_Y}\right|;\ R = \left[\max_{0\le k\le n} S_k^* - \min_{0\le k\le n} S_k^*\right]\Big/ D_Y \text{ and } U = \frac{1}{n(n+1)}\sum_{k=1}^{n-1}\left(\frac{S_k^*}{D_Y}\right)^2$$

Here D_Y is the standard deviation of the series and variable S_k^* has been standardized by dividing it by standard deviation of the series to compute all these three test statistics. So, *Q*-statistic is the maximum value of *"rescaled adjusted partial sum"* while *R*-statistic is the *"rescaled adjusted range"*. The *U* statistic, on the other hand, is derived from the original formulation by Gardener (1969) which implies that variance is known (Pérez, 2011). If it is unknown the sample variance, D_Y^2 can be used to compute the *U* statistic (Buishand, 1982). The series presents a break at time *k* where *Q* is observed.

Null hypothesis (H_0) states that series is homogeneous while alternative hypothesis (H_1) states that there is break in the series. Large values of these test statistics are an indication for the presence of break in the series (Buishand, 1982). The critical value of $Q/\sqrt{n}$, $R/\sqrt{n}$ and *U* can be found in the study by Buishand (1982). For $n \to \infty$ critical value of *Q* can be obtained from a table of the Kolmogorov-Smirnov goodness-of-fit statistic (Buishand, 1982).

5. Standard Normal Homogeneity Test

This is a parametric test which was originally formulated by Hans Alexandersson in 1986 for detection of a single break in annual precipitation series (Alexandersson, 1986). The test has two assumptions: *first*, the annual series can be described by standard normal distribution and *second*, there is a single possible break in the series and that consists of shift in mean level. The null hypothesis (H_0) states that series is homogeneous with zero mean value and unit standard deviation, while alternative hypothesis (H_1) is states that there is a single break at some point of time (year), *k* (for $1 \le k < n$). Alexandersson (1986) shows that under the alternative hypothesis the mean of the series (μ_1) before break (i.e. for $i \le k$) is different from the mean of the series (μ_2) after break (i.e. for $i > k$), i.e. $\mu_1 \neq \mu_2$ if there is a single break at some point, *k* (for $1 \le k < n$).

If $X_i,\ldots,X_n$ are the observations of an annual series (for $i = 1,2\ldots,n$) with sample mean, $\bar{X}$ and sample standard deviation, S_X, the test statistic for testing the null hypothesis is defined as:

$$T_0 = \max_{1\le k<n}\{T_k\}, \text{ where } T_k = k\bar{z}_1^2 + (n-k)\bar{z}_2^2, \text{ for } k = 1,2\ldots,n;$$

and, $\bar{z}_1 = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_i - \bar{X}}{S_X}\right);\ \bar{z}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\left(\frac{X_i - \bar{X}}{S_X}\right)$. The symbol $\sum$ which has been used in the formula to construct the variable $\bar{z}_1$ and $\bar{z}_2$ is the cumulative sum (for $k = 1,2\ldots,n$).

The null hypothesis is rejected if the T_0 is greater than the critical value. The critical for various sample sizes starting from 10 to 5000 are documented by Khaliq & Ouarda (2007).

In original formulation by Alexandersson (1986) the observed series was replaced by a series of ratios $\left|q_i\right|_{i=1}^{n}$ which was estimated between the observed value of the series to which the test was applied and the value of the reference station and such replacement was done in order to reduce the noise pertaining to the observed series (Alexandersson, 1986; Pérezet al., 2011).

Conclusion

Statistical tests are designed in such a way that their applications are very much specific and purposeful. In other words, it can be said that any test will no more be operable if it is not applied specifically for the purpose for which it was designed. Another important issue that has to be looked into carefully prior to running any statistical test is the format and measurement scale (nominal, ordinal, interval or ratio scale) of the data based on which it is decided which test to apply.

Test results may turn out to be doubtful even when the choice of statistical test is right with respect to the format and the scale of the data. For example, test is capturing the step change in data but there are circumstantial evidences which are strong enough to prove that in reality there is no change at all. This change might be due to change in data collection method, change in accuracy level of the data or events like locational shift of the observatory (for weather data or gauging station of river). So, it is advisable to test the change in causal variable along with testing the variable of interest. For example, along with the change in river flow which is the variable of interest, the change in its causal factor like precipitation is also to be tested simultaneously. The result of the latter will substantiate the former test result. Not only that, it is also advisable to apply more than one test for a single purpose to make it more confirm.

To conclude, one has to proceed very cautiously and systematically, adopting a step-by-step approach to make an accurate decision. These steps might be as follows: *data cleaning* (to check any error in observation), *exploratory data analysis* (to know the nature of the data), *right choice of test tool* depending on data nature, its scale and purpose of test, *running the test* (for change detection), *collection of circumstantial evidences* (i.e., data on other element of the environment like causal factor of the variable of interest) and finally the *testing of the other environmental elements* (causal factor) in terms of their change (to substantiate the test result). Finally, it can be said that it is the intuition, the logical thinking and the experience of the researcher on which the ultimate success of an academic investigation depends.

Acknowledgement: This is to acknowledge that the author has been inspired by two specific works: ***(i)*** Kundzewicz & Robson (eds.) (2000): Detecting Trend and Other Changes in Hydrological Data. World Climate Programme – Water, World Climate Programme Data and Monitoring, WCDMP-45, WMO/TD no. 1013. World Meteorological Organization, Geneva, Switzerland and ***(ii)*** Kundzewicz & Robson (2004): Change detection in hydrological records - a review of the methodology, Hydrological Sciences Journal, 49(1): 7-19, DOI: 10.1623/hysj.49.1.7.53993.

References

1. Alexandersson, H (1986): A homogeneity test applied to precipitation data. Journal of Climatology, 6: 661-675.
2. Buishand, T.A. (1981): The analysis of homogeneity of long-term rainfall records in TheNetherlands. R. Neth. Meteorol. Inst. (K.N.M.I.), De Bilt, Sci. Rep. No. 81-7.
3. Buishand, T.A. (1982): Some methods for testing the homogeneity of rainfall records. J.Hydrol., 58: 11-27.
4. Buishand, T.A. (1984): Tests for detecting a shift in the mean of hydrological time series. J. Hydrol., 73: 51-69.
5. Chiew, F.H.S. & McMahon, T.A. (1993): Detection of trend or change in annual flow of Australian rivers. Int. J. of Climatology., 13: 643-653.
6. Dagnélie, P (1970): Théorie et Méthodes Statistiques. Les Presses Agronomiques de Gembloux 2:463.
7. Durbin, J. & Watson, G.S. (1950): Testing for Serial Correlation in Least Squares Regression: I. Biometrika, 37(3/4): 409-428.
8. Durbin, J. & Watson, G.S. (1951): Testing for Serial Correlation in Least Squares Regression: II. Biometrika, 38(1/2):159-177.
9. Durbin, J. & Watson, G.S. (1971): Testing for Serial Correlation in Least Squares Regression: III. Biometrika, 58(1):1-19.
10. Fisher, R. A. (1 973): Statistical Methods and Scientific Inference. 3rd edition, Oliver and Boyd,Edinburg.
11. Gardner, L.A (1969): On detecting changes in the mean of normal variates. Ann Math Statist 40:116–126.
12. Grubb, H&Robson, Alice (2000): Exploratory/visual analysis. In: Kundzewicz, Z. W. &Robson, Alice (eds) (2000) *Detecting Trend and Other Changes in Hydrological Data*. World Climate Programme – Water, World Climate Programme Data and Monitoring, WCDMP-45, WMO/TD no. 1013. World Meteorological Organization, Geneva, Switzerland.
13. Hänsel, S. et al. (2016): Assessing Homogeneity and Climate Variability of Temperature and Precipitation Series in the Capitals of North-Eastern Brazil. Front. Earth Sci.4:29. doi:10.3389/feart.2016.00029.
14. Hawkins, P.M (1977): Testing a sequence of observations for a shift inrandom location. J. Am. Stat. Assoc. 73: 180–185.
15. Hipel, K.W and McLeod, A.I (1994): *Time series modelling of water resources and environmental systems*. Development in water science – 45, Elsevier, Amsterdam-London-New York, Tokyo.
16. Khaliq, M.N. and Ouarda, T.B.M.J. (2007): On the critical values of the standard normal homogeneity test (SNHT). Int. J.Climatol., 27: 681–687.
17. Kundzewicz, Z. W. & Robson, Alice (eds) (2000): *Detecting Trend and Other Changes in Hydrological Data*. World Climate Programme – Water, World Climate Programme Data and Monitoring, WCDMP-45, WMO/TD no. 1013. World Meteorological Organization, Geneva, Switzerland.

18. Kundzewicz, Zbigniew W. & Robson, Alice J. (2004): *Change detection in hydrological records - a review of the methodology*, Hydrological Sciences Journal, 49(1): 7-19, DOI: 10.1623/hysj.49.1.7.53993.
19. McGilchrist, C.A. and Woodyer, K.D (1975): Note on a Distribution-Free CUSUM Technique. Technometrics, 3: 321-325.
20. Neumann, J. von (1941): Distribution of the ratio of the mean square successive difference to the variance. Annals of Math. Stat., 12: 153-162.
21. Neumann, von J. (1941): The mean square successive difference. The Annals of Mathematical Statistics, 12 (2): 153-162.
22. Owen, D.B., (1962): Handbook of Statistical Tables. Addison-Wesley, Reading, Mass.
23. Pal, Saroj K. (1998): Statistics for Geoscientists, Techniques and Application. Concept Publishing Company, New Delhi, pp. 499-500.
24. Pérez, S. et al. (2011): Abrupt changes in rainfall in the Eastern area of La Pampa Province, Argentina. Theor. Appl. Climatol. 103:159–165.DOI 10.1007/s00704-010-0290-y.
25. Pettitt, A. N (1979): A Non-parametric Approach to the Change-point Problem. Appl. Statist. 28 (2): 126-135.
26. Radziejewski, M. et al. (2000): Phase randomization for change detection in hydrological data. In: Kundzewicz, Z. W. & Robson, Alice (eds) (2000) *Detecting Trend and Other Changes in Hydrological Data*. World Climate Programme – Water, World Climate Programme Data and Monitoring, WCDMP-45, WMO/TD no. 1013. World Meteorological Organization, Geneva, Switzerland.
27. Robson, Alice et al., (2000): Statistical methods for testing for change. In: Kundzewicz, Z. W. & Robson, Alice (eds) (2000) *Detecting Trend and Other Changes in Hydrological Data*. World Climate Programme – Water, World Climate Programme Data and Monitoring, WCDMP-45, WMO/TD no. 1013. World Meteorological Organization, Geneva, Switzerland.
28. Worsley, K.J. (1979) On the Likelihood Ratio Test for a Shift in Location of Normal Populations, Journal of theAmerican Statistical Association, 74:365-367,DOI: 10.1080/01621459.1979.10482519.
29. Yozgatligil, Ceylan and Yazici, Ceyda (2016): Comparison of homogeneity tests for temperature using a simulation study. Int. J. Climatol. 36: 62–81.
30. https://en.wikipedia.org/wiki/Durbin%E2%80%93Watson_statistic (for Durbin-Watson test, accessed on 21.11.2017).

