Change points in space-time, methodology and applications

Puntos de cambio en espacio- tiempo, metodología y aplicaciones

MUÑIZ-MERINO, Lucila†*, JUÁREZ-HERNANDEZ, Bulmaro and CRUZ-SUARES, Hugo Adan

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Fisico Matemáticas

ID 1st Author: *Lucila, Muñiz –Merino* / **ORC ID**: 0000-0001-7732-5514, **CVU CONACYT ID**: 385558

ID 1st Coauthor: *Bulmaro, Juárez-Hernandez* / **ORC ID**: 0000-0002-6260-1296, **CVU CONACYT ID**:78065

ID 2nd Coauthor: *Hugo Adan, Cruz-Suares* / **ORC ID**: 0000-0002-0732-4943, **CVU CONACYT ID**:202875

DOI: 10.35429/JQSA.2019.19.6.17.28 Received September 10, 2019; Accepted December 15, 2019

Abstract

In this work, we review publications which analyze, develop and apply concepts of change points, in general, the formulation of the problem of the change point, and different problems, including some of its applications are presented. Applications include temporal, spatial and temporal-space change points, parametric and non-parametric methods are used.

Change points, Parametric, Non-Parametric

Resumen

En este trabajo se hace una revisión de publicaciones que analizan, desarrollan y aplican conceptos de puntos de cambio, en general se presenta la formulación del problema del punto de cambio, y diferentes problemas del mismo, incluidas algunas de sus aplicaciones. Las aplicaciones incluyen puntos de cambio temporal, espacial y espacio temporales, se utilizan métodos paramétricos y no paramétricos

Puntos de cambio, Paramétricos, No paramétricos

Citación: MUÑIZ-MERINO, Lucila, JUÁREZ-HERNANDEZ, Bulmaro and CRUZ-SUARES, Hugo Adan. Change points in space-time, methodology and applications. Journal of Quantitative and Statistical Analysis. 2019. 6-19: 17-28

* Correspondence to Author (email: lucymerino74@hotmail.com)
† Investigador contribuyendo como primer autor.

Introduction

The change point is one of the central problems of statistical inference, as it relates to the theory of statistical control, hypothesis tests (when detecting whether there is any change in the succession of random variables observed) and estimation theory (when estimating the number of changes and their corresponding locations). Change point problems originally arose in quality control and can generally be found in various experimental and mathematical disciplines such as environmental sciences, epidemiology, seismic signal processes, economics, finance, geology, medicine, biology, physics, etc. (Chen & Gupta, 2012). The change points are presented abruptly and gradually (Brodsky & Darkhovsky, 1993), and (Brodsky, Brodsky, & Darkhovsky, 2000) carry out their analysis on independent and dependent random variables over time and space.

The general objective of this paper was to investigate and analyze the methodologies developed in the study of change points and their different applications in the space-time problem. The methodologies consist in finding the test statistics and by means of it, when contrasting the hypotheses, deciding whether there are change points.

Most spatial-temporal analysis approaches are developed in three separate stages (separate spatial analysis for each time point, separate temporal analysis for each location and analysis of spatial-temporal data with methods for random fields in R^{d+1} (Schabenberger & Gotway, 2004). The first two approaches can be considered conditional methods because they isolate a particular point or time location and apply standard techniques for the type of resulting data. A two-stage variation on the topic is to combine the results of the conditional analyses into a second stage. Two-stage approaches are common in statistical application, in which multiple sources of variation exist.

Analyses of space-time, spatial and temporal change points are performed under different approaches: nonparametric and parametric. In this work, different approaches are included regarding the change points, both spatial, temporal and space-time; parametric and non-parametric.

The detection of change points is important, because when estimating future events, we want accurate models and to be able to predict with greater precision. This is achieved using some methodology of change points; therefore, we analyze and include methodologies and some works that apply them, in addition to some works in which some other methodologies not presented in this article were used.

In general, the problem of change points in space and time treats different change points according to (Xun, Shashi, & Reem, 2014), the changes in space and time are classified in the following way: changes in statistical parameters, change in the value, change in the model adjusted to the data, this is reflected in the change of behavior of the trend which can be linear or polynomial, change in the attributes of the slope. In space, the changes are: raster-based, vector-based and image-based. In spacetime the change also refers to volume.

This paper is structured as follows: the second section presents the formulation of the problem of change points with respect to parameters, their classification and diagnostic methods; section 3 presents change point methods in parametric and nonparametric form; section 4 summarizes in tables some applications with different parametric and nonparametric methods; section 5 includes the results; section 6 presents the conclusions; and section 7 provides the bibliographic references. The detection of change points is important, because when estimating future events, we want accurate models and to be able to predict with greater precision. This is achieved using some methodology of change points; therefore, we analyze and include methodologies and some works that apply them, in addition to some works in which some other methodologies not presented in this article were used.

Change Point Formulation

When a change point is mentioned, the first question that comes to mind is: what is a change point? (Chen & Gupta, 2012). It is defined as the site, or point in time t, in a succession of data $\{x_{t_i}\}$ $i = 1,..,n$ observed and ordered with respect to time, such that these observations follow a distribution F_1, before a point, and in another point after it, the distribution is F_2.

From the statistical point of view, the succession of observations shows a non-homogeneous behaviour. In general, the problem of change points according to (Chen & Gupta, 2012) is visualised as follows:

Let $X_1, X_2, \dots, X_n$ be a succession of independent random vectors (or variables) with probability distribution functions $F_1, F_2, \dots, F_n$, respectively. Then the problem of change points is to test the null hypothesis H_0 of the non-existence of change against the alternative H_a that there is at least one change point:

$$H_0: F_1 = F_2 = \cdots = F_n \quad vs$$
$$H_a: F_1 = \cdots = F_{(k_1)} \neq F_{(k_{1+1})} = \cdots = F_{(k_q)} \neq F_{(k_{q+1})} = \cdots = F_n$$

Where $1 < k_1 < k_2 < \cdots < k_q < n$; q is the unknown number of change points and $k_1, k_2, \dots, k_q$ are the respective unknown positions that have to be estimated. If the distributions $F_1, F_2, \dots, F_n$ become a common parametric family $F(\theta)$, where $\theta \in R^p$, then the problem of change points is to test the null hypothesis H_0 on the non-existence of change in the parameters $\theta_i, i = 1, \dots, n$ of the population against the alternative H_a that there is at least one change point:

$$H_0: \theta_1 = \theta_2 = \cdots = \theta_n = \theta, \quad unknown \quad vs$$
$$H_a: \theta_1 = \cdots = \theta_{(k_1)} \neq \theta_{(k_{1+1})} = \cdots = \theta_{(k_q)} \neq \theta_{(k_{q+1})} = \cdots = \theta_n$$

where q and $k_1, k_2, \dots, k_q$ must be estimated. These hypotheses together reveal the inference aspects of change points to determine if any change point exists in the process, estimate their number and their respective positions.

In several cases it is assumed that the observations are independent and identically distributed (i.i.d.), since the analysis is more complex if there is dependence between the observations. In the case of time series, the dependence occurs among the observations within each time segment; in the case of space-time data, the dependence occurs over space and time.

According to (Brodsky & Darkhovsky, 1993) and (Brodsky, Brodsky, & Darkhovsky, 2000) the problems and methods of diagnosing change points can be classified as follows:

By the character of the information on the diagnostic object: Retrospective analysis (a posteriori) and sequential analysis; by the character of statistical diagnostic methods: Parametric, nonparametric and semiparametric methods; by the character of the diagnostic object: Statistical diagnostic problems for random processes (in discrete or continuous time) and statistical diagnostic problems for random fields; by the character of statistical dependence between observations: change point problems can be formulated for random sequences with independent observations, and change point problems for dependent observations in time or space, in a one-dimensional and multidimensional form, a single change point or multiple change points; by the mechanism of change in the state of the diagnostic object: Detection of abrupt change (change point problems), detection of gradual change, detection in regression relations; by the mechanism of change in the state of the diagnostic object: Detection of abrupt change (change point problems), detection of gradual change, detection in regression relations. These changes are observed in the probabilistic characteristics of the observations.

Change Point Methods

This section presents parametric and non-parametric methods of change points. Then, they will be described in detail.

Parametric methods

The parametric methods are: the standard normal homogeneity test, Von Newman, Buishand Ranges, t Motion, the Cumulative Anomaly Test and one more based on kernel. The standard normal homogeneity test was developed by (Alexandersson, 1986) to compare the mean of the first *k* years of recording with that of the last *n* - *k* years, the hypothesis contrast is:

$$H_0: \mu_1 = \mu_2 \quad vs \quad H_a: \mu_1 \neq \mu_2$$

where μ_1 is the average of the first *k* years and μ_2 is the average of the last *n* - *k* years. The $T(k)$ statistic is calculated as equation (1) based on equations (2) and (3):

$$T(k) = k\,\overline{z_1}^2 + (n-k)\overline{z_2}^2, \quad k = 1, \dots, n, \quad (1)$$

where

$$\overline{z_1} = \frac{1}{k}\sum_{i=1}^{k} z_i \quad \text{, and} \quad (2)$$

Y $\bar{z_2} = \frac{1}{n-k}\sum_{i=k+1}^{n} z_i.$ (3)

If a change occurs in year k, then $T(k)$ peaks near year $k = k_0$. Test statistics T_0 is given in equation (4)

$$T_0 = \max\big(T(k)\big)\, for \quad 1 \le k \le n\,. \qquad (4)$$

The null hypothesis is rejected when T_0 is larger than a certain critical value, i.e. there is a change point. Another proof is the relation of (Von Newman, 1941), which is related to the first-order serial correlation coeficiente. The relation, N, of Von Neumann is defined as the ratio of the difference of the successive mean square (year to year) and the sample mean square (Von Newman, 1941). The test statistic for change point detection in the time series $x_1, x_2, \ldots, x_n$, is described in equation (5):

$$N = \frac{\sum_{i=1}^{n}(x_i - x_{i-1})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (5)$$

The X_i are normally distributed with mean μ and variance σ^2. The hypotheses regarding the change point with respect to the mean are the following:

H_0: $E(N), i = 1, \ldots, m,$ (i.e., the average is constant)

Vs

H_a: $E(N) + \Delta, i = m+1, \ldots, n$ (there is a change of size Δ).

For homogeneous series, the expected value under the null hypothesis is constant $E(N) = 2$. Non-homogeneous series or samples with a change will have a value of N less than 2; any other value implies that the time series has a rapid variation in its mean.

One more parametric test is the Buishand range test. (Buishand, 1982) developed this statistical test. The adjusted partial sum, which is the cumulative deviation of the mean for the observation k of a series $x_1, x_2, \ldots, x_n$ with mean μ, can be calculated using the equations (6) and (7):

$$S_{0^*} = 0, \quad \text{and} \qquad (6)$$

$$S_{k^*} = \sum_{i=1}^{k}(x_i - \bar{x}),\ k = 1, \ldots, n \qquad (7)$$

Where X_i have normal distribution. For homogeneous series, the values of S_{k^*} fluctuate around zero, since in the random time series the deviation from its mean is generally distributed on both sides of the mean of the series. If the series breaks in year K, then S_{k^*} reaches a maximum (negative change) or a minimum (positive displacement) near year k = K.

The readjusted partial sums are obtained by dividing the S_{k^*} by the sample standard deviation, as shown in the equation (8):

$$S_{k^{**}} = \frac{S_{k^*}}{D_y}, \qquad k = 0, \ldots, n, \qquad (8)$$

With (10)

$$D_y^2 = \sum_{i=1}^{n}\frac{(y_i - \bar{y})^2}{n} \qquad (10)$$

The homogeneity test is based on the adjusted rescalated partial sums $S_{k^{**}}$. The statistic for homogeneity developments is (11):

$$Q = \max_{0 \le k \le n} |S_{k^{**}}|\,, \qquad (11)$$

high Q values are an indication of a change.

Another test is the t Motion test, which is used by (Yin, Liu, Yi, & Liu , 2015), who used it to detect possible abrupt change points. The t Motion test is the t-test of two samples (Snedecor & Cochran, 1989) which is used to determine if two population means are equal. The data can be paired or unpaired. If they are paired it means that there is a one-to-one correspondence between the values in the two samples, for paired samples the difference is usually calculated. For unpaired samples, the sample sizes for two samples may not be the same. Sample variances can be assumed to be the same or different.

The hypothesis contrast to the mean to detect a change is:

$$H_0: \mu = \mu_2 \quad vs \quad H_a: \mu_1 \ne \mu_2,$$

When the variances are the same, the test statistic is (12):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (12)$$

ISSN 2410-3438

Where, $s = \sqrt{\frac{n_1 s_1^2 + n_1 s_2^2}{n_1 + n_2 - 2}}$.

This test statistic is used to detect a change point, for a given time series x_n which has n random samples, the reference point is set for a given time. The samples of two subsucessions before and after a point data are n_1 and n_2. $\bar{x}_1$ and $\bar{x}_2$ are the average of two subsucessions and s_1^2 and s_2^2 are the variances of two subsucessions.

The t Motion test according to (Yin, Liu, Yi, & Liu , 2015) is carried out in three steps to detect abrupt change. First, the same length of two subsections before or after the point data is defined; normally, $n_1 = n_2$. Second, according to the mathematical expression in (2), the statistical value of two subsections is successively calculated using the t Motion method for a set of point data. Third, the average values of two samples are compared at a given significance level to detect the change. If $|t_1| < t_\alpha$, the analyzed variable has abrupt change in the point data.

A test analyzed by (Lishan, Suiji, & Xiaoli, 2010) is the cumulative anomaly method for revealing abrupt changes in water discharge and suspended sediment concentration (SSC). According to (Lishan, Suiji, & Xiaoli, 2010), mathematically cumulative anomaly is a method for distinguishing a trend change from discrete data and is widely used in meteorology to analyze variations in precipitation and temperature. For a discrete x_i series, the cumulative anomaly (X_t) for point data x_t can be expressed as (13):

$$\hat{X}_t = \sum_{i=1}^{t}(X_i - \bar{X}), \ \ t = 1,2,\dots,n \quad (13)$$

$with\, \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and $\bar{X}$ is the mean value of the x_i series, and n is the number of discrete points. As the equation suggests, the cumulative anomaly can be used to analyze the magnitude of the fluctuation of a series of discrete data, usually the increase in the cumulative anomaly value indicates the point data involved that are larger than the average, otherwise lower than the average. In the study, the variable x represents the average annual discharge and the average annual suspended sediment concentration (SSC), respectively.

This is a test that has been used by some other researchers besides Ran et al. to analyze changes: (Suiji, Xunxia, Ming, & Zhao, 2012) and for changes in space-time (Xiujing, Shifeng, Yongyong, & Cuicui, 2013).

Another method is Single Spectrum Analysis (SSA) ((Morkvina, 2001) and (Morskvina & Zhiglijavsky, 2003)), this is a powerful technique of time series analysis, used to detect change points.

Algorithm to detect change points:

Let $x_1, x_2, \dots, x_T$, a time series with T<∞. Select: from the window width N(N≤T) the lag parameter ($M \leq \frac{N}{2}$, k = N – M + 1 and 0≤p<q, I=(1,..,l) where l denotes the number of eigenvectors that form the base of the subspace.

1. The base matrix $X^{(n)}$, called the trajectory matrix, $X^{(n)}$ has multivariate data with M characteristics and k observations. Columns $X_j^{(n)}$ with j=1,...,n of $X^{(n)}$ are considered as vectors falling into dimensional M space.

2. The covariance lag matrix.. $R_n = X^{(n)}(X^{(n)})^T$.

3. The M eigenvalues and eigenvectors of R_n.

4. $D_{n,I,p,q}$ is the matrix of the sum of the differences for the vectors $X_j^{(n)}$, the matrix with columns $X_j^{(n)}$ $(j = p + 1, \dots, q)$ is called the test matrix in (14),

$$D_{n,I,p,q} = \sum_{j=p+1}^{q}((X_j^{(n)})^T X_j^{(n)} - (X_j^{(n)})^T U U^{T X_j^{(n)}}. \quad (14)$$

And $U_{i1}, \dots, U_{iI}$ denotes the eigenvectors that form the basis of subspace $L_{n,I}, l < M$.

5. Calculate the normalized square distance S_n, like in 15

$$S_n = \frac{\tilde{D}_{n,I,p,q}}{v_n} \quad (15)$$

ISSN 2410-3438

$and\ \widetilde{D}_{n,I,p,q} = \frac{1}{M(q-p)} D_{n,I,p,q}$ which is the normalized sum of square distances (normalization is made with respect to the number of elements in the test matrix). v_j is an estimator of the normalized sum of square distances $\widetilde{D}_{n,I,p,q}$ in the time intervals [j + 1, j + N] where the no change hypothesis can be accepted. It is suggested to use $\widetilde{D}_{\tilde{n},I,p,q}$, where $\tilde{n}$ is the largest value of j<n such that the null hypothesis of no change in the interval [j + 1, j + N] has been accepted.

The decision rule in the proposed algorithm is to declare a change if for any n, $S_n \geq H$, where H is a fixed threshold, i.e. large values of $D_{n,I,p,q}$ and S_n indicate that there is a change in the structure of the series.

Note: A general recommendation is to select p ≥ k; this makes the base columns and the test matrix consist of different elements. In this case, the algorithm is more sensitive to changes than its economic version (in the sense of the number of x_t involved in each iteration n) when p < k and so, some of the base columns and test matrices match.

To obtain a smoother behaviour of the test statistic $D_{n,I,p,q}$, q needs to be selected slightly larger than p. If the difference q-p is also large, then the behaviour of $D_{n,I,p,q}$ becomes larger; this perhaps for example when p = 0 and q = p (that is, the base and the test matrix coincide).

Regarding the kernel-based space-time change points test is the work of (Jansenberger & Steinnocher , 2014) who made a contribution which focuses on space-time changes. For the spatial quantification of such changes, the dual kernel density estimation method was used. For such method, two different data sets were related to each other. Changes in spatial concentration of grocery stores of two retail groups in the province of Austria were analyzed for 1998 and 2001.

The researchers use a quartic kernel function. However, since the result of an analysis is not strongly influenced by the selected function, there are no rules concerning the selection of an appropriate function. There are several different kernel functions, the most common kernel function is the normal distribution function in (16):

$$k\left(\frac{d_i}{n}\right) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(\frac{d_i}{\mu}\right)^2\right] \quad (16)$$

With $\mu = \left(\frac{s-s_i}{b}\right)$, based on this function the density estimate is expressed in (17):

$$\hat{f}(s) = \frac{1}{b^2} \sum_{i=1}^{n} k\left(\frac{d_i}{b}\right) \quad (17)$$

where d_i is the distance between the s points and the location of the observed point. Because bandwidth b is the standard deviation of the normal distribution, this function extends to infinity in all directions, i.e. it was applied to each of the points in the region.

In this analysis, the quartic Kernel function is used, which has a circumscribed radius, which is also the bandwidth. The quartic kernel function is applied to a limited area around each location and has the functional form in (18)

$$K(\mu) = \begin{cases} \frac{3}{4}(1-\mu^2\mu)^2 & for\ \mu^2\mu \leq 1 \\ 0 & d.o.f. \end{cases} \quad (18)$$

Kernel density is also estimated, which is an interpolation technique that relates individual point locations or points for an entire area and provides estimates of density λ(s) (19), at a location within the study region R.

$$\lambda(\hat{s}) = \sum_{i=1}^{n} k\left(\frac{d_i}{b}\right) \quad (19)$$

If the dual kernel is applied to a variable, this is known as a singular density estimate. If it is applied to two variables, this is called a dual density estimate. In the latter case, a kernel density is estimated for each of the variables individually and then the two density estimates are related to each other by a simple algebraic operation such as a sum, difference and quotient. The most commonly used operation is the quotient. In this study, however, the difference in absolute value of the densities was used and with it the space-time changes in the region of study were visualized.

Non-parametric tests

Within the non-parametric tests, there are: Pettit, Mann Kendall and Lepage. The nonparametric Pettitt test (Pettitt, 1979) detects a change in an unknown time, he uses a version of the test of two Mann-Whitney samples, and calculates their statistical significance.

He considers a time series $x_i(1 \le i \le n)$ and uses a $U_{t,n}$ statistic, which is equivalent to a Mann-Withney statistic, which is used to prove that the two samples $x_1, \ldots, x_t$ and $x_{t+1}, \ldots, x_n$, are from the same population. The null hypothesis H_0 of the Pettitt test is the absence of change points, while the alternative hypothesis is the existence of a change point. It is considered to (20)

$$D_{ij} = sgn(X_i - X_j) \quad (20)$$

where $sgn(x) = \begin{cases} 1, & x > 0 \\ 0\ , & x = 0 \\ -1 & x < 0. \end{cases}$

The statistic test is given in (21):

$$U_{t,n} = \sum_{i=1}^{t} \sum_{j=1+1}^{n} D_{ij}, \quad t = 2, \ldots, n. \quad (21)$$

The Pettitt test uses the test statistics given in (22),

$$k_t = \max_{1 \le t \le n} |U_{t,n}| \quad (22)$$

For testing one tail and for changing directions, the statistics given in (23) are used,

$$k_1^+ = \max_{1 \le t \le n} |U_{t,n}|, \quad k_1^- = \min_{1 \le t \le n} |U_{t,n}|. \quad (23)$$

Obviously $k_t = \max(k_t^+, k_t^-)$. It should be noted that in the null hypothesis H_0, $E(D_{ij}) = 0$ and the distribution of $U_{t,n}$ is symmetrical around zero for each t. So, k_t^+ and k_t^- have the same null distribution. The statistics k_t^+ are k_t^- are from a tail, and use the theory of Mann-Whitney, k_t^+ can be expected to be large if there has been a downward change in the level of the beginning of the series. k_t^+ can be large if $F_1(x) \le F_2(x)$, with strict inequality for at least some x. Similarly, k_t^- can be expected to be large if there has been an upward change or $F_1(x) \ge F_2(x)$.
The significant change point is at the maximum $U_{t,n}$ value and the level of significance associated with k_t^+ and k_t^- is determined approximately by (24)

$$\rho = 2 \exp\left(-\frac{6k_t^2}{n^3+n^2}\right), \quad (24)$$

if ρ is smaller than the specific significance level, e.g. 0.05, the null hypothesis is rejected. In other words, if there is a significant change point, the time series is divided into 2 parts at the location of the change point t.

The probability of approximate significance for a change point is defined as p = 1 - ρ. The Mann-Kendall test is a non-parametric method for trend detection and change points due to its robustness and simplicity. The Mann-Kendall test has been widely used to evaluate the monotonic trend significance of hydrometeorological variables. (Liu, Xu, & Huang, 2012) cites this test in their analysis and mention the Mann-kendall- Sneyers test; while (Sneyers, 1990) calls this statistic the Mann statistic. For time series, the magnitudes $x_i(i = 1,2, \ldots, n)$ mean time series that are compared with $x_j(i = 1,2, \ldots, i-1)$. For each comparison, the number of cases is counted $x_j > x_i$ which is denoted by r_i. The null hypothesis H_0 indicates the existence of no trend in the time series, while the alternative hypothesis H_1 establishes that there is a trend in the data set. Under the null hypothesis (no trend), the range series is (25):

$$S_k = \sum_{i=1}^{k} r_i, \quad (25)$$

Where

$$r_i = \begin{cases} +1 & x_i > x_j \\ & j = 1,2,3, \ldots, i \\ 0 & x_i \le x_j \end{cases}$$

has normal distribution with mean and variance given by: $E(S_k) = \frac{k(k+1)}{4}$, $var(S_k) = \frac{k(k-1)(2k+5)}{72}$.

Forward Sequential Statistics in (26)

$$U_{F_k} = \frac{[S_k - E(S_k)]}{\sqrt{var(S_k)}} \quad k = 1,2, \ldots, n \quad (26)$$

is a standardized normal variable. The backward sequence U_{B_k} is estimated using the same equation but with an inverted series of data. In a 2 tail trend test, the null hypothesis is accepted with a level of significance α if $|U_{F_k}| \le (U_{F_k})_{1-\frac{\alpha}{2}}$ where $(U_{F_k})_{1-\frac{\alpha}{2}}$ is the critical value of the standard normal distribution with a probability α. U_{F_k}>0 denotes an upward trend, while the opposite denotes a downward trend (i.e U_{B_k} is similar to U_{F_k}).

The sequential version of the test used allows the detection of the approximate time of occurrence of the change of trend by localizing the intersection of the forward and backward curves of the test statistic.

A point of intersection within the range of confidence indicates a change point. Another nonparametric test is the Lepage test. This is a two-sample test for location and dispersion (Lepage, 1971), which has been widely used to detect changes such as long-term trends, cyclical variations and staggered changes for precipitation. Lepage assumes that the size of the study series is equal to or greater than ten and the Lepage statistic (HK) follows the Chi-square distribution ($\chi 2$) with two degrees of freedom. It is assumed that the samples come from continuous distributions and are independent. The Lepage (HK) statistic is given in (27):

$$HK = \frac{[W-E(W)]^2}{V(W)} + \frac{[A-E(A)]^2}{V(A)} \quad (27)$$

Let $x = (x_1, x_2, \ldots, x_{n1})$ and $y = (y_1, y_2, \ldots, y_{n2})$ two independent samples of size n_1 and n_2. It is assumed that $\mu_i = 1$ if the smallest i-th observation in a combined sample size (n_1, n_2) belongs to x and $\mu_i = 0$ if it belongs to y.

The null hypothesis H_0 of the Lepage test assumes that the distributions from which the two samples come are equal, contrasting against the alternative H_a in which they are considered to be different. If the HK test statistic exceeds 5.99, the difference between two samples is judged as significant at the confidence level of 95 percent (significance level of 5 percent), i.e. the null hypothesis that the distributions are equal is rejected, therefore there is a change point.

The terms in Eq. (27) can be derived from equations (28), (29), (30) and (31).

$$W = \sum_{i=1}^{n_1+n_2} i\mu_i, \quad (28)$$

$$E(W) = \frac{n_1(n_1+n_2+1)}{2} \quad (29)$$

$$V(W) = \frac{n_1 n_2(n_1+n_2+1)}{2} \quad (30)$$

$$A = \frac{1}{2}n_1\big((n_1+n_2+1)+1\big) + \sum_{i=1}^{n_1+n_2} \left|i - \frac{1}{2}(n_1+n_2+1)\right|\mu_i \quad (31)$$

If $(n_1 + n_2)$ is even $E(A)$ and $V(A)$ will be calculated as in (32) and (33):

$$E[A] = \frac{n_1(n_1+n_2+2)}{4}, \text{ and} \quad (32)$$

$$V(A) = \frac{n_1 n_2 (n_1+n_2-2)(n_1+n_2+2)}{48(n_1+n_2-2)}. \quad (33)$$

The statistical characteristics of the segments divided by the change points are detected by the mean and the coefficient of variation (Cv). Thus, $\mu_x = E[x] = \mu_1'$ and $Cv = \frac{S_x}{x}$, where $S_x = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2}{(n-1)}}$.

Application problems

Three tables are presented below, which summarize the different application problems of the change points studied and the methodology used for their detection, as well as the researchers who have used the methodologies. The three tables consist of three columns, which contain the problem or application, the author or authors and the model or methodology used to detect the change point or points. In the first table there are applications in which the change point analysis is carried out by means of the methods listed in column 3, superscript is placed, the models range from 1 to 4. In the second table, superscript is placed for researchers who use the model listed in column 3, the models range from 1 to 15. In table 3 the same procedure is performed, models from 1 to 9 are listed and the researchers who used it are placed as superscripts.

Problem	Authors	Models
Lake temperature	(Chavaillaz, Joussaume, Bony, & Braconnot, 2015)[1]	1. Regression
Precipitation and temperature	(Skliris, y otros, 2014)[1]	
Contaminant concentration	(Abdel, El, Sean, Rong, & Yalin, 2011)[1]	
Precipitation and temperature	(Chengjing, y otros, 2012)[1]	
Change in ocean aerosols	(Cermak, Wild, Knutti, Mishchencko, & Heidinger, 2010)[1]	
Growth of Chinese spruce	(Ma, Shi, Wang, & He, 2006)[1]	
Land use change	(Bollinger, Kienast, Soliva, & Rhuterford, 2007)[1]	
Sea surface temperature	(González Taboada & Andón , 2012)[1]	
Vegetation index	(Luan, y otros, 2018)[1]	
Spatial concentration of grocery stores	(Jansenberger & Steinnocher , 2014)[2]	2.Kernel
Hydrology (flow)	(Yang, Chen, Xu, & Zhang, 2009) [3]	3. Single spectrum analysis
Climate (Precipitation)	(Xiujing, Shifeng, Yongyong, & Cuicui, 2013) [4]	4.Pettit

Table 1 Regression, Kernel, Single Spectrum Analysis, Pettit, t Motion, Buishand Ranges, Normal Standard Homogeneity and Von Newman

Problem	Authors	Models
Climate (temperature)	(Yin, Liu, Yi, & Liu , 2015)[1]	1.t Motion 2.Pettit
Climate (temperature)	(Malekian & Kazemzadeh, 2015)[1,2,3]	3. Ranges of buishand
Change in air temperature	(Chakraborty, y otros, 2017)[1,2,4,5]	4. Homogeneity Normal standard
Zink contamination	(Tlili, y otros, 2011)[6]	5.Von Neuman
Growth of Chinese spruce	(Ma, Shi, Wang, & He, 2006)[7]	6. Multivariate Statistics
Exotic Vegetation	(Tierney & Cushman, 2005)[8,10]	7. Gini and Lorentz Coefficient
Change of land use	(Bollinger, Kienast, Soliva, & Rhuterford, 2007)[9]	8.MANOVA 9. Logistic Regression
Soil breathing	(Akburak & Makineci, 2012)[10]	10.ANOVA
Detection of telecommunication fraud	(Hilas, Rekanos , & Mastorocostas, 2013)[11]	11.ARIMA 12.MEDIAN
Climate (Temperature)	(Luo, Bryan, Bellotti, & Williams , 2005)[12]	13.ANCOVA
Soil water content	(Cubera & Moreno, 2007)[13]	14.t Motion
Precipitation rates	(Chavaillaz, Joussaume, Bony, & Braconnot, 2015)[14,15]	15.Mann Kendall
Water temperature	(Peter, 2017)[2]	

Table 2 MANOVA, ARIMA, Bayesian space, Mann Kendall, Median and ANOVA

Problem	Authors	Models
Climate (precipitation)	(Chen, Kimb, & Kimc, 2016)[1]	1. Bayesian Space
Climate (temperature)	(Yin, Liu, Yi, & Liu , 2015)[2]	2.Mann Kendall
Climate (precipitation)	(Biana, y otros, 2017)[2,3]	3.Pettit
Change in air temperature	(Chakraborty, y otros, 2017)[3]	
Temperature in the lake	(Yankova, Villiger, Pernthaler, Schanz, & Posch , 2017)[2]	
Precipitation	(Adeyeri, Lamptey, Lawin, & Sanda, 2017)[3]	
Climate (temperature)	(Malekian & Kazemzadeh, 2015)[2,3]	
Climate (precipitation, flow)	(Gebremicael, Mohamed, Zaag, & Hagos, 2017)[2,3]	
Vegetation index, habitat change in estuary	(Marcoe & Pilson, 2017)[4]	4. Data comparison
Change of the beach	(Michalowska, Glowienka, & Pekala, 2016)[5]	5.Images 6.Bayesian space
Intellectual coefficient	(Cai, Lawson, McDermott, & Aelion, 2016)[6]	7. Yamamoto Method 8. Wavelet Method
Climate (Precipitation, temperature)	(Huiying, y otros, 2016)[7,8,9]	9.Trend rate

Table 3 Mann Kendall, Pettit, Lepage, Bayesian, spline, comparative data and images

Results

As a result of this research, the different applications, their authors and applied models are presented summarized in tables. It can be observed that the studies of hydrometeorological variables are of main interest.

As can be observed in the introduced methodologies, both parametric and nonparametric, what was done was to determine the test statistic and, by means of this, to contrast the hypotheses in order to decide whether there are change points. Therefore, the objective is to determine the test statistic in any proposed methodology, often having the need to obtain limit distributions in order to make the contrast.

In terms of applications, for example in Bayesian models, in order to find the aposteriori, both informative and non-informative aprioris can be worked on.

Conclusion

Many of the works on both parametric and non-parametric change points apply the methodology of change points in precipitation, temperature, change in air temperature, water temperature and water flow, to name a few. Although these application problems occur frequently, there are also some others that have been summarized in the tables of the application section, however the hydrometeorological variables as can be observed are of main interest. In the same way, it is of interest to identify the change points in the inference to obtain better accuracy.

References

Abdel, H., El, S., Sean, B., Rong, Z., & Yalin, C. (2011). Modelling temporal and spatial changes of PCBs in fish tissue from Lake Huron. *Environ Monit Assess, 173*, 611-623.

Adeyeri, Lamptey, Lawin, & Sanda. (2017). Spatio-Temporal Precipitation trend and homogeneity analysis in komadugu yobe Basin, Lake Chad region graduate research programme on climate change and water resources. *Climatol Weather Forecasting, 5*(3).

Akburak, S., & Makineci, E. (2012). Temporal changes of soil respiration under different tree species. *Environ. Monit Assess, 185*, 3349-3358.

Alexandersson, H. (1986). A homogeneity test applied to precipitation. *Journal of Climatology, 6*, 661-675.

Biana, G., Dua, J., Songa, M., Xua, Y., Xiea, S., Zhenga, W., & Xuc, C. (2017). A procedure for quantifying runoff response to spatial and temporal changes of impervious surface in Qinhuai River basin of southeastern China. *Catena, 157*, 268-278.

Bollinger, J., Kienast, F., Soliva, R., & Rhuterford, G. (2007). Spatial sensitivity of species habitat patterns to scenarios of land use change (Switxerland). *Landscape Ecol., 22*, 773-789.

Brodsky, B. E., & Darkhovsky, B. S. (1993). *Mathematics and its Applications Noparametrics Methods in Change-Point Problems* (Vol. 243). Kluer Academic Publishers.

Brodsky, Brodsky, B. E., & Darkhovsky, B. S. (2000). *Non Parametric Diagnostic Problems and Methods. Mathematics and Its Aplications.* Kluwer Academic, Dordrecht.

Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hidrology, 58*, 11-27.

Cai, B., Lawson, A., McDermott, S., & Aelion, C. (2016). A Bayesian semiparametric approach with change points for spatial ordinal data. *Stat. Methods Med. Res., 25*(2), 644-658.

Cermak, J., Wild, M., Knutti, R., Mishchencko, M., & Heidinger, K. (2010). Consistency of global derived aerosol and cloud data sets with recent brightening observations. *Geophysical research Letters, 37*.

Chakraborty, D., Saha, S., Sink, R., Sethy , B., Kumar , A., Saikia, U., . . . Daschaudhuri, D. (2017). Trend Analysis and Change Point Detection of North Eastern India. *Environ. Process*.

Chavaillaz, Y., Joussaume, S., Bony, S., & Braconnot, P. (2015). Spatial stabilization and intensification of moistering and drying rate patterns under future climate change. *Clim Dyn, 47*, 951-965.

Chen, J., & Gupta, A. (2012). *Parametric Statistical Changepoint Analysis.* Bogota: Birkhauser.

Chen, S., Kimb, S., & Kimc, T.-W. (2016). Spatial change Assessment of the observed and proyected extreme precipitation via change point detection. *Procedia Engineering, 154*, 1359-1367.

Chengjing, N., Hairong, L., Linsheng, Y., Ye, C., Erfu, D., Shaohong, W., . . . Yongfeng, L. (2012). Spatial and temporal changes in extreme temperature and extreme precipitation in Guangxi. *Quaternary International, 263*, 162-171.

Cubera, E., & Moreno, G. (2007). Effect of single Quercus ilex trees upon spatial and seasonal changes in soil water content in deshesas of central western Spain. *Ann. For. Sci., 64*, 355-364.

Gebremicael, T., Mohamed, Y., Zaag, P., & Hagos, E. (2017). Temporal and spatial changes of rainfall and streamflow in the Upper Tekeze-Atbara river basin Ethiopia. *Hydrol. Earth Syst Sci., 21*, 2127-2142.

González Taboada, F., & Andón , R. (2012). Patterns of change in sea superface temperature in the North Atlantic during the last three decades: beyond mean trends. *Climatic Changes, 115*, 419-431.

Hilas, C., Rekanos , L., & Mastorocostas, P. (2013). *Change Point Detection in Time Series Using Higher-Order Statistics: A Heuristic Approach.* Mathematical Problems in Engineering.

Huiying, Z., Lijuan , G., Huihui, Q., Haixia, Z., Xiufen, L., & Heilongjiang, Z. (2016). The climate change variations in the northern Greater Kningan Mountains during the past centuries, Institute of Meteorological Science, Harbin 150030, China. *J. Geogr. Sci., 26*(5), 585-602.

Jansenberger, E., & Steinnocher , S. (2014). Dual Kernel Density Estimation as a Method Describing Spatio Temporal Changes in the Upper Austrian Food Retailling Market. *7th AGILE Conference on Geographic Information Science 29 April- 1 Mayo 2004.*

Lepage, Y. (1971). A Combination of Wilcoxon´s anAnsari Bradley´s Statistics. *Biometrika, 58*(1), 213-217.

Lishan, R., Suiji, W., & Xiaoli, F. (2010). Channel change at Toudaoguai Station and its responses to the operation of upstream reservoirs in the upper Yellow River. *J. Geogr. Sci.*, 231-247.

Liu, L., Xu, Z., & Huang, J.-X. (2012). Spatio temporal variation and abrupt changes for majer climate variables in the Taihu Basin, China. *Stch Environ Res Risk Assess, 26*, 777-791.

Luan, J., Liu, D., Zhang, L., Huang, Q., Feng, J., Lin, M., & Li, g. (2018). Analysis of the spatial-temporal change of the vegetation index in the upper reach of Han River Basin in 2000-2016. *Proc. IAHS, 379*, 287-292.

Luo, Q., Bryan, B., Bellotti, W., & Williams , M. (2005). Spatial analysis of enviromental change impacts on wheat production in mid-lower north, sourt australian. *Climatic Change, 72*, 213-228.

Ma, Z., Shi, J., Wang, G., & He, Z. (2006). Temporal changes in the inequality of growth of Cunninghamia lanceolata (Lamb) Hook: a novel application of the Gini coefficient and Lorenz asymmetry. *Genetica, 126*, 343-351.

Malekian, A., & Kazemzadeh, M. (2015). Spatio-Temporal Analysis of Regional Trends and Shift Changes of Autocorrelated Temperature Series in Urmia Lake Basin. *Water Recour Manage, 30*, 785-803.

Marcoe, K., & Pilson, S. (2017). Habitat change in the lower Columbia River estuary 1870-2009. *J. Coast Conserv, 21*, 505-525.

Michalowska, K., Glowienka, E., & Pekala, A. (2016). Spatial-temporal detection of changes southern coast of the baltic sea based on multitemporal aerial photographs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciencies*, XLI-B2.

Morkvina, V. (2001). Aplication of the singular spectrum analysis for change-point detection in time series. *Ph. D. Thesi, Cardi University*.

Morskvina, V., & Zhiglijavsky, A. (2003). An on singular based on singular spectrum analysis for change-point detection, comunication in statistics. *Stat. Simul., 32*, 319-352.

Peter, A. (2017). Long-term temporal and spatial changes in the richness and relative abundance of the inshore fish community of the British North Sea Coast. *Journal of Sea Reseach*.

Pettitt, A. (1979). A Non-Parametric Approach to the Change-Point Problem. *Jornal of the Royal Statistical Society, Series C (Applied Statistics), 28*(2), 126-135.

Quant, R. E. (1958). The estimation of parameters of a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc., 50*, 853-880.

Schabenberger, & Gotway. (2004). *Statistical methods for spatial data analysis.* CRC Press.

Skliris, N., Marsh, R., Josey, S. A., Good, S. A., Liu, C., & Allan, R. P. (2014). Salinity changes in the World Ocean since 1950 in relation to changing suface freshwater fluxes. *Clim Dyn, 43*, 709-736.

Tierney, T., & Cushman, J. (2005). Temporal changes Temporal changes in native exotic vegetation and soild characteristics following disturbances by feral pigs in a California grassland. *Biological Invasions, 8*, 1073-1089.

Tlili, A., Corcoll, N., Bonet, B., Moirin, S., Montuelle, B., Bérad, A., & Guasch, H. (2011). In situ spatio-temporal changes in pollution-induced community tolerance to zinc in autotrophic and heterotrophic biofilm communities. *Ecotoxicology, 20*, 1823-1839.

Von Newman, J. (1941). Distribution of the Ratio of the Mean Square Successive Difference to the variance. *The Annals of Mathematical Statistics, 12*(4), 367-395.

Xiujing, M., Shifeng, Z., Yongyong, Z., & Cuicui, W. (2013). Temporal and spatial changes of temperature and precipitation in Hexi Corridor during 1955-2011. *J. Geogr. Sci., 23*(4), 653-667.

Xun, Z., Shashi, S., & Reem, Y. (2014). Spatial temporal change footprint pattern discovery: an inter-disciplinary survey. *WIREs Data Mining Knowl Discov, 4*, 1-23.

ISSN 2410-3438

Yang, T., Chen, X., Xu, C., & Zhang, Z.-C. (2009). Estimation as a Method Describing Spatio Temporal Changes in the Upper Austrian Food Retailling Market. *Stochastic Environmental Research and Risk Assessment, 23*, 1071-1087.

Yankova, Y., Villiger, J., Pernthaler, J., Schanz, F., & Posch , T. (2017). Prolongation, deepening and warming of the metalimnion change habitat conditions of the harmful filamentous cyanobacterium Planktothrix rubencens in a prealpine lake. *Hydrobiologia, 776*, 125-138.

Yin, Y., Liu, H., Yi, X., & Liu , W. (2015). Spatiotem,poral Variation and Abrupt Analysis of Temperature fron 1960 to in the Huang-Huai-Hai Plain, China2012. *Advances in Meteorology*.