A Change Point Model Approach for Monitoring Time between Events and Amplitude Data

Published Version:
Scagliarini, M. (2025). A Change Point Model Approach for Monitoring Time between Events and Amplitude Data. Springer [10.1007/978-3-031-96033-8_53].

Availability:
This version is available at: https://hdl.handle.net/11585/1018652 since: 2025-06-30

Published:
DOI: http://doi.org/10.1007/978-3-031-96033-8_53

A Change Point Model Approach for Monitoring Time Between Events and Amplitude Data

Michele Scagliarini

Department of Statistical Sciences, University of Bologna, Italy
michele.scagliarini@unibo.it

Abstract. This study proposes change-point control charts for monitoring Time Between Events and Amplitude (TBEA) data. Using both a real-world case study and a simulation experiment, we evaluate the in-control statistical properties of two change-point control charts. We then compare the results obtained from these charts with those of a control chart specifically designed for monitoring TBEA data. The findings suggest that change-point control charts can be effectively used within the TBEA framework, providing valuable insights into the monitored process.

Keywords: Control Charts; Distribution Free; Climate Change.

1 Introduction and Methodological Note

In practical applications, monitoring the occurrence of unwanted events, such as equipment failures, quality issues, or extreme natural phenomena, is often crucial. In such contexts, two key characteristics are typically of interest: the time interval T between consecutive occurrences of the negative event E and the associated magnitude X. Time Between Events and Amplitude (TBEA) control charts have been specifically developed to simultaneously monitor both the time interval between successive events and their magnitude. Generally, TBEA control charts use, as statistics to be monitored, some functions of the random variables T and X. Most contributions in literature study the statistical properties of these control charts under the assumption that T and X follow specific probability distributions. However, in practice, identifying the true distribution of these variables can be difficult. To address this issue, several distribution-free TBEA control charts have been proposed [1]. Furthermore, in addition to signaling an alarm (i.e., identifying the detection time), it is often useful to estimate the change-point time. Identifying when a shift occurred in a TBEA statistic can aid in tracing potential root causes and enhance overall process understanding. To the best of our knowledge no prior research has addressed this specific issue. Therefore, this study proposes change-point control charts for monitoring TBEA data. Specifically, to provide a comprehensive overview, we consider two non-parametric control charts with distinct characteristics: one based on the Mann-Whitney (MW) statistic, designed to detect shifts in location, and another based on the Kolmogorov-Smirnov (KS) statistic, capable of detecting arbitrary changes in the process distribution. Through a simulation experiment and a case study, we evaluate their statistical properties and compare the results with those obtained from a distribution-free Exponentially Weighted Moving Average (EWMA) TBEA control chart [1].

Let T_i (i=1,2,…) represent the time intervals between two consecutive occurrences of a given event E and let X_i denote the corresponding magnitudes of the event of interest. To monitor T_i and X_i simultaneously, it is assumed that T_i and X_i are continuous random variables, both defined on [0;+∞), with unknown distribution functions $F_T\left(t|\theta_T\right)$ and $F_X\left(x|\theta_X\right)$ respectively, where θ_T and θ_X are known α-quantiles [1]. Without loss of generality, it is considered that θ_T and θ_X are the median values of T_i and X_i, respectively, and when the process is in-control it follows that $\theta_T = \theta_{T_0}$ and $\theta_X = \theta_{X_0}$. The distribution-free EWMA TBEA control chart is based on the sign statistics $\mathrm{ST}_i = \mathrm{sign}\left(T_i - \theta_{T_0}\right)$ and $\mathrm{SX}_i = \mathrm{sign}\left(X_i - \theta_{X_0}\right)$, for i=1,2,…, where $\mathrm{sign}(x) = -1$ if x<0 and sign(x) = +1 if x>0. Let us now define the statistic $S_i = \left(\mathrm{SX}_i - \mathrm{ST}_i\right)/2$ (i=1,2,…), which has the following behaviour: S_i = –1 when T_i increases and, at the same time, X_i decreases (positive situation); S_i = +1 when T_i decreases and, at the same time, X_i increases (negative situation); S_i = 0 when both T_i and X_i increase or when both T_i and X_i decrease (intermediate situation). Note that S_i is a discrete random variable; therefore, it would be impossible to accurately compute the run length properties of a control chart based on this statistic. To solve this problem the authors [1] developed the "continuousify" method. They propose to define an extra parameter $\sigma \in$ [0.1, 0.2] and to convert S_i into a continuous random variable, S_i^*, defined as a mixture of 3 normal random variables $Y_{i,-1} \sim N\left(-1,\sigma\right)$, $Y_{i,0} \sim N\left(0,\sigma\right)$ and $Y_{i,+1} \sim N\left(1,\sigma\right)$, respectively. More precisely S_i^* is defined as: $S_i^* = Y_{i,-1}$ if $S_i = -1$; $S_i^* = Y_{i,0}$ if $S_i = 0$; $S_i^* = Y_{i,+1}$ if $S_i = +1$. Note that the parameter σ has to be fixed. However, the authors demonstrated that this parameter has no impact on the performance of the control chart if it falls within the suggested range. The upper-sided EWMA TBEA control chart uses the statistic $Z_i^* = \max\left(0, \lambda S_i^* + \left(1-\lambda\right) Z_{i-1}^*\right)$ with an upper asymptotic control limit given by $UCL = K\sqrt{\lambda\left(\sigma^2 + 0.5\right)/\left(2-\lambda\right)}$, where $\lambda \in$[0, 1] and K>0 are the control chart parameters and the initial value is $Z_0^* = 0$. The optimal design parameters λ and K can be obtained by studying the statistical properties of the control chart by means of a Markov chain approach [1].

To effectively monitor time between events and amplitude data using change points control charts, it is essential to consider that the variables T and X may have very different scales. Therefore, to avoid favoring one variable over the other, it is appropriate to define and use the normalized variables $T' = T/\mu_{T_0}$ and $X' = X/\mu_{X_0}$, where μ_{T_0} and μ_{X_0} are the in-control means of T and X, respectively. Among the possible choices for the TBEA statistic to be monitored, we consider the ratio $Z_R = T'/X'$ and propose to monitor this statistic using non-parametric change-point control charts. In the change point model framework, it is assumed that the values of the TBEA statistic $z_1, z_2,...$ are generated by the random variables $Z_1, Z_2,...$ with unknown distribution functions F_1, $F_2,...$, respectively. A change point occurs at instant τ when $F_\tau \neq F_{\tau+1}$ and it is usually assumed that the observations are independent and identically distributed between every pair of change points. Therefore, the distribution of the sequence can be described by the following model: $Z_i \sim F_0$ if $0 < i \leq \tau_1$, $Z_i \sim F_1$ if $\tau_1 < i \leq \tau_2, \ldots,$ $Z_i \sim F_k$ if $\tau_k < i \leq m$, where $0 < i < \tau_1 < \tau_2 < \cdots < \tau_k < m$ denote k unknown change points. Within this framework, it is of interest to test H_0: k=0, versus H_1: $k\neq0$, to estimate the number of change points, k, and to estimate their locations $\hat{\tau}_1, \hat{\tau}_2, ..., \hat{\tau}_k$. For testing the aforementioned hypothesis system, [2] proposed a change point control chart, designed to detect changes in the location of the process distribution, based on the Mann-Whitney (MW) U-statistic: $U_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} D_{ij}$, $1 \leq k \leq n-1$, where $D_{ij} = \text{sign}\left(Z_i - Z_j\right)$ with D_{ij} = 1 if $Z_i > Z_j$, D_{ij}=0 if $Z_i = Z_j$, D_{ij}= –1 if $Z_i > Z_j$. The variance of $U_{k,n}$ depends on the split point k, so for this reason it is suitably standardized, thus obtaining the statistic $T_{k,n}$. Therefore, the test for the presence of a change point and the estimate of its time of occurrence are given by maximizing $T_{k,n}$ over k: $T_{\max,n} = \max_{1 \leq k \leq n-1} \left|T_{k,n}\right|$. To detect arbitrary changes in the process distribution, [3] proposed a control chart based on the Kolmogorov-Smirnov (KS) statistic. This control chart tests for a change point immediately following any observation z_k by partitioning the observations into two samples, $S_1=(z_1,\ldots,z_k)$ and $S_2=(z_{k+1},\ldots,z_t)$, and subsequently comparing the corresponding empirical distribution functions $\hat{F}_{S_1}\left(z\right)$ and $\hat{F}_{S_2}\left(z\right)$. The test statistic is based on the maximum difference between the empirical distributions $D_{k,t} = \sup_z \left|\hat{F}_{S_1}\left(z_i\right) - \hat{F}_{S_2}\left(z_i\right)\right|$. Further details can be found in [2] and [3].

2 The Case Study, Results and Concluding Remarks

For effective drought management and resilience policies design, it is crucial to monitor the evolution of drought conditions and detect changes in their characteristics. As the relevant drought characteristics are severity, duration and frequency, a suitable methodology for monitoring drought events should consider the time interval T between two occurrences and the magnitude X of each event. In this context, TBEA control charts may prove useful. The Standardized Precipitation Evapotranspiration Index (SPEI) is a widely used metric for assessing drought conditions [4]. It is based on precipitation, temperature and potential evapotranspiration and can be calculated for accumulation periods ranging from 1 to 48 months. For our analysis, we use a 12-month accumulation period (SPEI-12), which is commonly employed to assess hydrological drought [4]. Positive SPEI values indicate above-average moisture conditions, while negative values indicate drier conditions. The aim is to monitor the characteristics of drought episodes. Hence, the occurrences of SPEI-12≤-1, corresponding to drought conditions, were considered as the events (E) of interest. As the TBEA methodology assumes that the magnitudes of the events are defined on [0; +∞), the absolute values of SPEI-12 at the dates of occurrence were considered as the corresponding magnitudes X_i. Given the monthly frequency of SPEI-12, the time T_i between two events is computed in months. The SPEI-12 data were downloaded from the Global SPEI database for the coordinates of Bologna (44.4949° N, 11.3426° E), where observations are available from December, 1901 to December, 2022 (1453 observations). The dataset includes 226 events E (values of SPEI-12≤-1). For Phase-1 analysis, we selected the time period from July 1946 to July 1989, comprising 517 SPEI-12 observations with 47 events E. This period, was used to estimate in-control median (for the EWMA chart) and mean (for the change-point charts) values: $\hat{\theta}_{T_0} = 1$, $\hat{\theta}_{X_0} = 1.2263$, $\hat{\mu}_{T_0} = 10.9149$ and $\hat{\mu}_{X_0} = 1.2511$, respectively. To ensure comparability, all control charts were configured with the same in-control theoretical statistical properties (ARL0=370 or with a theoretical false alarms rate equal to $\alpha_0 = 0.0027$). Figure 1 shows the results obtained with the TBEA EWMA and KS control charts, respectively. The results for the MW control chart are omitted, as they closely resemble those of the KS control chart.

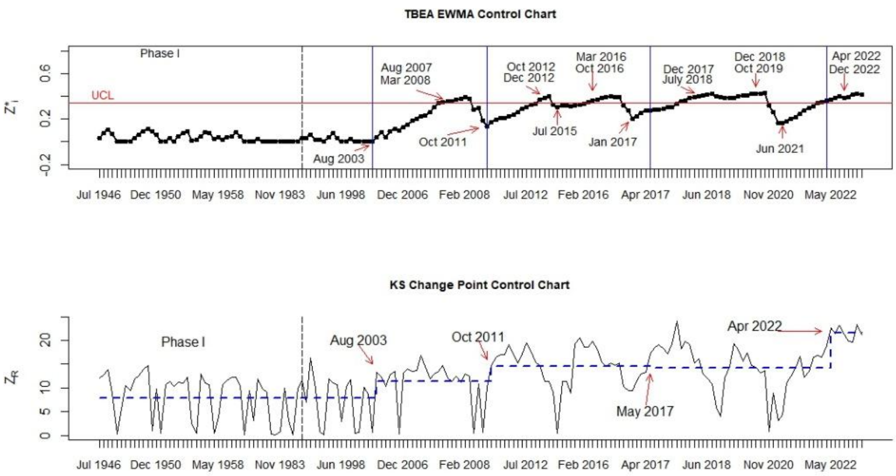

Fig. 1. TBEA EWMA (upper panel, where the consecutive out-of-control events in the original time scale are indicated to facilitate the discussion), Kolmogorov-Smirnov (lower panel, where the estimated change points are also indicated) control charts

Before discussing the results, some key considerations must be addressed. First, no alarms were detected in the Phase 1 dataset, indicating that the estimated medians and means are reliable estimates of the corresponding in-control parameters. Second, it is important to remember that the EWMA control chart is tailored for monitoring TBEA data, with well-documented statistical properties [1]. In contrast, the change-point control charts are not specifically designed for monitoring time-between-events and amplitude statistics. Therefore, a potential concern is that their false alarm rates may differ from theoretical values. To assess the performance of the change-point control charts, we conducted a simulation study in which, to make the comparisons as consistent as possible, we generate data similar to those observed in the present case. Phase-1 was performed on 517 SPEI-12 observations, normally distributed and with mean and standard deviation $\mu_{SPEI-12}$ =0.2082 and $\sigma_{SPEI-12}$ =0.9213, respectively. In these observations occurred 47 events *E*. To reproduce a similar scenario, 500 observations were generated, following a normal distribution with mean $\mu_{SPEI-12}$ and standard deviation $\sigma_{SPEI-12}$. Events *E* were identified, and their corresponding magnitudes X_i and time between two events T_i were determined. The mean values of X_i and T_i were then estimated to configure the KS and MW change-point control charts. Once set up the KS and MW control charts, 1000 additional normally distributed observations with mean $\mu_{SPEI-12}$ and standard deviation $\sigma_{SPEI-12}$ were generated. In these observations the events *E* were identified as the corresponding amplitudes and time between events. The values of the statistic Z_R were computed and the possible change points estimated. This procedure was repeated 10^6 and the false alarm rates were calculated by dividing the total number of estimated change points by the total number of events *E* identified. Table 1 summarizes the results showing the empirical false alarm rates.

Table 1. Empirical false alarm rates obtained from the simulation experiment.

	KS	MW
$\hat{\alpha}_0$	0.00230	0.00180

The results indicate that the empirical false alarm rates do not exceed the theoretical value and that the KS chart's empirical rate closely matches the theoretical value, warranting greater focus on this control chart in subsequent analyses. We now turn to discussing the results. In this work, different methodologies were employed. However, the examined control charts consistently suggest worsening drought conditions over time. The TBEA statistics show significant increases beginning in August 2003 (KS control charts) and August 2007 (EWMA control chart). Consistency among results becomes evident when the change points estimated by the KS control chart are superimposed on the EWMA chart (the continuous vertical lines in Figure 1 upper panel). The KS control chart frequently estimates change points close to the lower turning points of the EWMA statistic (August 2003, October 2011, July 2015, January 2017, June 2021 see also Figure 1 upper panel). The amount of agreement in the results can be objectively quantified by performing an Event Coincidence Analysis (ECA) [5]. ECA evaluates empirical co-occurrence frequencies of distinct event types against those expected under the assumption of uncoupled Poisson processes. In the present case, there are two sequences of events of distinct types: events A (the estimated change points) that occur at times t_i^A; events B (the lower turning point of the EWMA statistics) that occur at times t_i^B. A coincidence occurs if two events are closer in time than a temporal tolerance or coincidence interval ΔT , i.e. $t_i^A - t_i^B \le \Delta T$. To quantify the coincidence intensity between Type A and Type B event sequences, we computed the precursor coincidence rate: $r_p\left(\Delta T,\tau\right)=\frac{1}{N_A}\sum_{i=1}^{N_A}\Theta\left[\sum_{j=1}^{N_B}1_{[0,\Delta T]}\left(\left(t_i^A-\tau\right)-t_j^B\right)\right]$, where $\Theta\left(x\right)=0$ for $x\leq 0$ and $\Theta\left(x\right)=1$ otherwise, $1_I\left(\cdot\right)$ is the indicator function of the interval I and τ is a time lag parameter to adjust for possible lagged relationships. The precursor coincidence rate measures the fraction of A-type events that are preceded by at least one B-type event within the tolerance interval ΔT . It is also possible to test the null hypothesis that the observed number of coincidences can be explained by two independent series of randomly Poisson-distributed events. In our case with ΔT =0 and considering only the instantaneous coincidences within the coincidence interval (τ=0) we obtain $r_p\left(\Delta T,\tau\right)$ =0.5: the 50% of the estimated change points were preceded by an EWMA lower turning point within the same month (ΔT =0). The hypothesis that these coincidences occur by chance can be rejected (*p*-value=0.000), demonstrating substantial agreement between methods. Therefore, these methods might be effectively combined to provide complementary insights. For the EWMA chart, post-signal analyses can be conducted to identify consecutive out-of-control events in the original time scale. For instance, a long period of consecutive out-of-control values occurred from December 2018 to October 2019. Regarding the change-point control charts, it is important to remember that they monitor Z_R, which simultaneously accounts for both time between drought events and their magnitudes. This means that the estimation of a change point corresponds to detecting a change in the drought characteristics. For example, the KS control chart reveals that increases in the TBEA statistic observed in August 2003 and October 2011 indicate an increase in drought severity. In summary, change-point control charts can serve as a complementary tool and be effectively used in conjunction with TBEA control charts. Having an estimate of the time when the shift occurred in a TBEA statistic might allow for a better understanding of the monitored process, optimize response strategies, and enhance overall system resilience, making them a valuable tool across multiple domains.

References

1. Wu, S., Castagliola, P., Celano G.: A Distribution-Free EWMA Control Chart for Monitoring Time-Between-Events-and-Amplitude Data. J. Appl. Stat. 48(3), 434-454 (2021).
2. Hawkins, D.M., Deng, Q.: A Nonparametric Change-Point Control Chart. J. Qual. Technol. 42(2), 165-173 (2010).
3. Ross, G.J., Adams, N.M.: Two Nonparametric Control Charts for Detecting Arbitrary Distribution Changes. J. Qual. Technol. 44(12), 102-116 (2012).
4. Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J. Clim. 23(7), 1696–1718 (2010).
5. Donges, J. F., Schleussner, C-F., Siegmund, J. F., Donner, R. V.: Event Coincidence Analysis for Quantifying Statistical Interrelationships Between Event Time Series. Europ. Phys. J.: Spec. Top. 225(3), 471-487 (2016).