



HAL
open science

**Control Charts for Monitoring
Time-Between-Events-and-Amplitude Data**
Philippe Castagliola, Giovanni Celano, Dorra Rahali, Shu Wu

► **To cite this version:**

Philippe Castagliola, Giovanni Celano, Dorra Rahali, Shu Wu. Control Charts for Monitoring Time-Between-Events-and-Amplitude Data. Control Charts and Machine Learning for Anomaly Detection in Manufacturing, pp.43-76, 2022, 10.1007/978-3-030-83819-5_3 . hal-03328534

HAL Id: hal-03328534

<https://hal.science/hal-03328534>

Submitted on 30 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Control Charts for Monitoring Time-Between-Events-and-Amplitude Data

Philippe Castagliola¹, Giovanni Celano², Dorra Rahali³, and Shu Wu⁴

¹ Université de Nantes & LS2N UMR CNRS 6004, Nantes, France
`philippe.castagliola@univ-nantes.fr`

² Università di Catania, Catania, Italy `giovanni.celano@unict.it`

³ Centre de Recherche en Informatique, Signal et Automatique de Lille, Lille,
France `dorra.rahali@centralelille.fr`

⁴ School of Logistics Engineering, Wuhan University of Technology, Wuhan, China
`wushu0208@hotmail.com`

1 Introduction

Today, due to the large availability of data, various kinds of processes can (and have to) be monitored using Statistical Process Monitoring (SPM) techniques based on advanced control charts. These kinds of processes can be of course industrial ones but, they can also be non industrial ones like in the biological/health-care (diseases, like the Covid-19 for instance), the geological (earthquakes or volcanic eruptions) or the accidental (traffic accidents, forest fires) fields. In all of these situations, people are usually focusing on two characteristics:

1. the time T between two consecutive specific (usually, adverse) events of interest E ,
2. the amplitude X of each of these events.

The characteristics T and X defined above are the key factors to be monitored for an event and they are usually referred to as the TBEA (Time Between Events and Amplitude) characteristics. In general, a decrease in T and/or an increase in X can result in a negative, hazardous or disastrous situation that needs to be efficiently monitored with control charts.

The first TBE (i.e. without taking into account the amplitude characteristic) type of control chart goes back to Calvin (1983), who proposed to monitor the cumulative number of conforming items between two non-conforming ones. The initial idea was to find a method to improve the traditional attribute control charts that are known to be ineffective in the case of high-quality processes in which the occurrence of non-conforming products is very rare. This initial idea has then been investigated by Lucas (1985) and Vardeman

and Ray (1985) and, subsequently, many other researchers started to contribute to this area. Radaelli (1998) proposed to design and implement one- and two-sided Shewhart-type TBE control charts assuming that the counts can be modeled as a homogeneous Poisson process. Gan (1998) developed an EWMA (Exponentially Weighted Moving Average) control chart monitoring the rate of occurrences of rare events based on the inter-arrival times of these events. Benneyan (2001) used the geometric (“g” chart) and the negative binomial (called “h” chart) distributions in order to monitor the number of cases between hospital-acquired infections. Xie et al. (2002) proposed a control chart for TBE data based on the exponential distribution while Borrer et al. (2003) extended it using a CUSUM (Cumulative Sum) scheme and evaluated its robustness in the case of the Weibull and lognormal distributions. Liu et al. (2006) compared the ATS (Average Time to Signal) performance of several continuous TBE charts including the CQC, CQC-r, exponential EWMA and exponential CUSUM charts. Zhang et al. (2007) investigated the case of gamma distributed TBE data and they developed a control chart based on a random-shift model to compute the out-of-control ATS. In the case of multistage manufacturing processes, Shamsuzzaman et al. (2009) developed a control chart for TBE data and designed it using a statistical oriented approach while Zhang et al. (2011a) designed it using a first economic oriented approach and Zhang et al. (2011b) developed it using a second economic oriented approach assuming random process shifts. The use of supplementary runs rules has also been proposed for monitoring TBE data by Cheng and Chen (2011). Qu et al. (2014) studied some TBE control charts that can be used for sampling inspection. Shafae et al. (2015) evaluated the performance of three TBE CUSUM charts and Fang et al. (2016) proposed a generalized group runs TBE chart for a homogenous Poisson failure process.

The first paper that proposed a combined scheme for monitoring the time interval T of an event E as well as its amplitude X has been introduced by Wu et al. (2009) who referred it to as a TBEA (Time Between Events and Amplitude) chart. After this paper, several single TBEA charts have been developed, see for instance Qu et al. (2013), Cheng et al. (2017), Ali and Pievatolo (2018), Qu et al. (2018) and, very recently, Sanusi et al. (2020).

As it can be noticed, this stream of research is rather recent and few publications have already been devoted to. Therefore, the goal of this chapter is to further investigate it and to hopefully open new research directions. More specifically, this chapter will be splitted into three parts:

1. In section 2 we will introduce and compare three different statistics, denoted as Z_1 , Z_2 and Z_3 , suitable for monitoring TBEA data, in the case of four distributions (gamma, lognormal, normal and Weibull), when the time T and the amplitude X are considered as *independent* random variables.

2. In section 3, we will compare the three statistics introduced in section 2, for the same distributions, but considering that the time T and the amplitude X are *dependent* random variables. A model based on three types of Copulas will be used to define the dependence between T and X .
3. Finally, in section 4, in order to overcome the “distribution choice” dilemma, we will introduce a distribution-free approach coupled with an upper-sided EWMA scheme. In addition, a specific technique called “continuousify” will be presented in order to compute the Run Length properties of the proposed upper-sided EWMA TBEA control chart in a reliable way.

2 TBEA charts for independent Times and Amplitudes

2.1 Model

Let $D_0 = 0, D_1, D_2, \dots$ be the dates of occurrence of a specific negative event E, let $T_1 = D_1 - D_0, T_2 = D_2 - D_1, \dots$ be the time intervals between two consecutive occurrences of the event E and let X_1, X_2, \dots be the corresponding magnitudes of this event occurring at times D_1, D_2, \dots (see Figure 1). It must be noted that $D_0 = 0$ is the date of a “virtual” event which has no amplitude associated with.

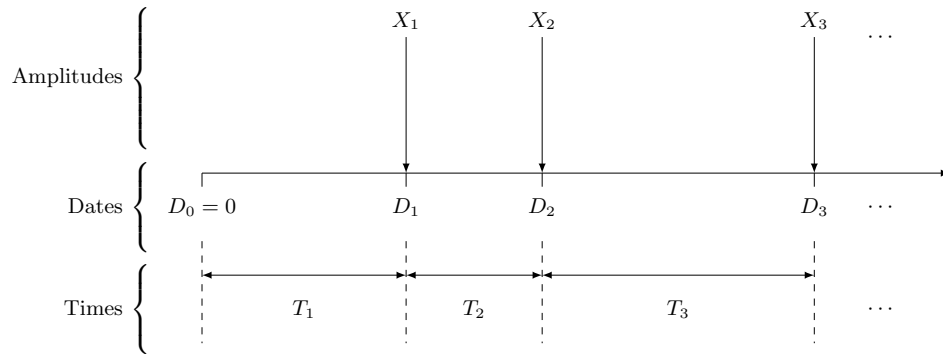


Fig. 1. Dates of occurrence $D_0 = 0, D_1, D_2, \dots$, time intervals $T_1 = D_1 - D_0, T_2 = D_2 - D_1, \dots$ and amplitudes X_1, X_2, \dots of a negative event E

In this section, we assume that T and X are two mutually *independent* continuous random variables, both defined on $[0, +\infty)$. Let $F_T(t|\theta_T)$ and $F_X(x|\theta_X)$ be the c.d.f. (cumulative distribution function) of T and X , respectively, and let $f_T(t|\theta_T)$ and $f_X(x|\theta_X)$ be the p.d.f. (probability distribution function) of

T and X , respectively, where $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_X$ are the corresponding vector of parameters. Let also define $\mu_T = E(T)$, $\mu_X = E(X)$, $\sigma_T = \sigma(T)$ and $\sigma_X = \sigma(X)$ be the expectation and standard-deviation of T and X , respectively. By definition, when the process is *in-control*, we have $\boldsymbol{\theta}_T = \boldsymbol{\theta}_{T_0}$, $\boldsymbol{\theta}_X = \boldsymbol{\theta}_{X_0}$, $\mu_T = \mu_{T_0}$, $\mu_X = \mu_{X_0}$, $\sigma_T = \sigma_{T_0}$, $\sigma_X = \sigma_{X_0}$ and, when the process is *out-of-control*, we have $\boldsymbol{\theta}_T = \boldsymbol{\theta}_{T_1}$, $\boldsymbol{\theta}_X = \boldsymbol{\theta}_{X_1}$, $\mu_T = \mu_{T_1}$, $\mu_X = \mu_{X_1}$, $\sigma_T = \sigma_{T_1}$, $\sigma_X = \sigma_{X_1}$.

Because the reference scales for the random variables T and X can be very different and, in order to not favour one random variable over the other one, we suggest to define (and work with) the “normalized to the mean” new random variables T' and X' as the in-control standardized counterparts of T and X , i.e.

$$T' = \frac{T}{\mu_{T_0}},$$

$$X' = \frac{X}{\mu_{X_0}}.$$

Clearly, when the process is in-control we have $E(T') = E(X') = 1$.

2.2 Statistics to be monitored

In order to simultaneously monitor the time T between an event E and its amplitude X , we suggest to define several dedicated statistics $Z = Z(T', X')$, functions of the random variables T' and X' , satisfying the following two properties:

$$Z \uparrow \text{ if either } T' \downarrow \text{ or } X' \uparrow, \quad (1)$$

$$Z \downarrow \text{ if either } T' \uparrow \text{ or } X' \downarrow. \quad (2)$$

Of course, there are many possible choices for the statistic Z . A first possible choice for the statistic Z (denoted as the Z_1 statistic) satisfying properties (1) and (2) is simply

$$Z_1 = X' - T'. \quad (3)$$

This random variable is defined on $(-\infty, +\infty)$ and its c.d.f. $F_{Z_1}(z|\boldsymbol{\theta}_Z)$ and p.d.f. $f_{Z_1}(z|\boldsymbol{\theta}_Z)$ can be obtained by integrating (see Figure 2 (a) and (b)) over all the couples $(X', T') \in \mathbb{R}^{+2}$ satisfying $Z_1 = X' - T' \leq z$, and they are equal to

$$F_{Z_1}(z|\boldsymbol{\theta}_Z) = 1 - \mu_{X_0} \int_0^{+\infty} F_T((x-z)\mu_{T_0}|\boldsymbol{\theta}_T) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx, \quad (4)$$

$$f_{Z_1}(z|\boldsymbol{\theta}_Z) = \mu_{T_0}\mu_{X_0} \int_0^{+\infty} f_T((x-z)\mu_{T_0}|\boldsymbol{\theta}_T) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx, \quad (5)$$

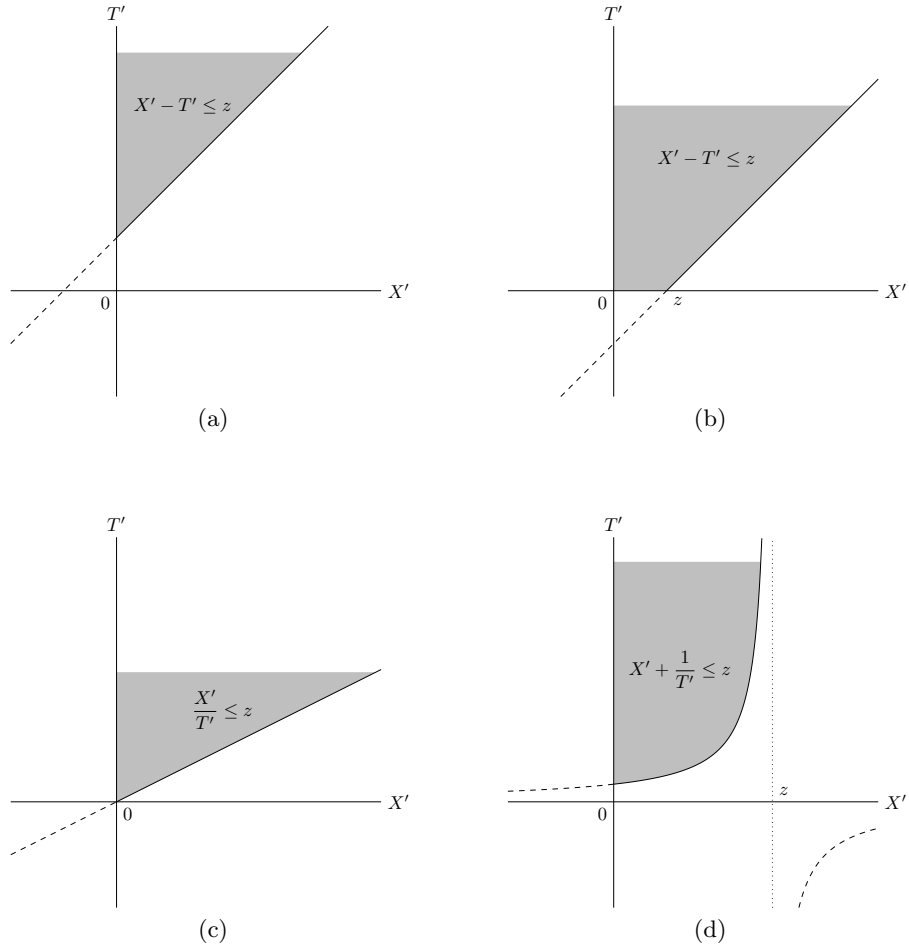


Fig. 2. Integration areas used for statistics (a) and (b) Z_1 , (c) Z_2 and (d) Z_3

where $\theta_Z = (\theta_T, \theta_X)$ is the corresponding combined vector of parameters.

A second possible choice for the statistic Z (denoted as the Z_2 statistic) satisfying properties (1) and (2) is defining it as the ratio between the two characteristics of an event E :

$$Z_2 = \frac{X'}{T'}. \tag{6}$$

This random variable is defined on $[0, +\infty)$ and its c.d.f. $F_{Z_2}(z|\theta_Z)$ and p.d.f. $f_{Z_2}(z|\theta_Z)$ can be obtained by integrating (see Figure 2 (c)) over all the couples $(X', T') \in \mathbb{R}^{+2}$ satisfying $Z_2 = \frac{X'}{T'} \leq z$, and they are equal to

$$F_{Z_2}(z|\boldsymbol{\theta}_Z) = 1 - \mu_{X_0} \int_0^{+\infty} F_T\left(\frac{x\mu_{T_0}}{z}|\boldsymbol{\theta}_T\right) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx, \quad (7)$$

$$f_{Z_2}(z|\boldsymbol{\theta}_Z) = \frac{\mu_{T_0}\mu_{X_0}}{z^2} \int_0^{+\infty} x f_T\left(\frac{x\mu_{T_0}}{z}|\boldsymbol{\theta}_T\right) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx. \quad (8)$$

Finally, a third possible choice for the statistic Z (denoted as the Z_2 statistic) satisfying properties (1) and (2) is

$$Z_3 = X' + \frac{1}{T'}. \quad (9)$$

This random variable which should be considered as a hybrid of the two previous ones is also defined on $[0, +\infty)$ and its c.d.f. $F_{Z_3}(z|\boldsymbol{\theta}_Z)$ and p.d.f. $f_{Z_3}(z|\boldsymbol{\theta}_Z)$ can be obtained by integrating (see Figure 2 (d)) over all the couples $(X', T') \in \mathbb{R}^{+2}$ satisfying $Z_3 = X' + \frac{1}{T'} \leq z$, and they are equal to

$$F_{Z_3}(z|\boldsymbol{\theta}_Z) = F_X(z\mu_{X_0}|\boldsymbol{\theta}_X) - \mu_{X_0} \int_0^z F_T\left(\frac{\mu_{T_0}}{z-x}|\boldsymbol{\theta}_T\right) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx, \quad (10)$$

$$f_{Z_3}(z|\boldsymbol{\theta}_Z) = \mu_{T_0}\mu_{X_0} \int_0^z \frac{1}{(z-x)^2} f_T\left(\frac{\mu_{T_0}}{z-x}|\boldsymbol{\theta}_T\right) f_X(x\mu_{X_0}|\boldsymbol{\theta}_X) dx. \quad (11)$$

More details on how to derive the c.d.f. and p.d.f. of statistics Z_1 , Z_2 and Z_3 provided above can be found in the Appendix section of Rahali et al. (2019). Concerning these c.d.f. and p.d.f. it has to be noted that it is generally not possible to obtain a closed form solution for them and the only solution is to numerically compute these ones by using quadrature techniques.

2.3 Control limit

As it is more important to detect an increase in Z (in order to avoid more damages or injuries, for instance) rather than a decrease, we suggest to only define an upper control limit UCL_Z for the TBEA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$ as

$$\text{UCL}_Z = F_Z^{-1}(1 - \alpha|\boldsymbol{\theta}_{Z_0}), \quad (12)$$

where α is the type I error, $\boldsymbol{\theta}_{Z_0} = (\boldsymbol{\theta}_{T_0}, \boldsymbol{\theta}_{X_0})$ and $F_Z^{-1}(\dots|\boldsymbol{\theta}_{Z_0})$ is the inverse c.d.f. of Z numerically obtained by solving equation $F_Z(z|\boldsymbol{\theta}_{Z_0}) = \alpha$ for z using a one dimension root finder.

2.4 Time to Signal properties

The type II error β of the upper-sided TBEA charts based on statistic $Z \in \{Z_1, Z_2, Z_3\}$ is equal to

$$\beta = F_Z(\text{UCL}_Z|\boldsymbol{\theta}_{Z_1}), \quad (13)$$

where $\boldsymbol{\theta}_{Z_1} = (\boldsymbol{\theta}_{T_1}, \boldsymbol{\theta}_{X_1})$. The out-of-control ATS (Average Time to Signal) and SDTS (Standard Deviation Time to Signal) of the upper-sided TBEA charts based on statistic $Z \in \{Z_1, Z_2, Z_3\}$ can be obtained using the expectation and variance of compound random variables (see also Rahali et al. (2019) for more details) and they are equal to

$$\text{ATS}_1 = \frac{\mu_{T_1}}{1 - \beta}, \quad (14)$$

$$\text{SDTS}_1 = \sqrt{\frac{\sigma_{T_1}^2}{1 - \beta} + \frac{\mu_{T_1}^2 \beta}{(1 - \beta)^2}}. \quad (15)$$

When the process is in-control, we have $1 - \beta = \alpha$ and, consequently, we have the following equivalence for the in-control ATS

$$\text{ATS}_0 = \frac{\mu_{T_0}}{\alpha} \Leftrightarrow \alpha = \frac{\mu_{T_0}}{\text{ATS}_0}.$$

2.5 Comparative studies

As in Rahali et al. (2019), in order to compare the three TBEA charts defined in sub-section 2.2 and based on the statistics $Z \in \{Z_1, Z_2, Z_3\}$ we have chosen to investigate four different types of distribution that are only dependent on two parameters a and b . **The choice of the Gamma, Lognormal, Normal and Weibull distributions is driven by the fact that these ones are very often selected to model time oriented random variables. For this reason, the two parameters beta distribution has been excluded from the benchmark as it is rarely selected for representing time oriented variables. Of course, more complex distributions could have been considered (like the four parameters Beta or Johnson's distributions) but, due to the fact that only the nominal mean μ_0 and standard-deviation σ_0 are assumed to be known, we restricted our choice to a selection of two parameters distributions.** These distributions are summarized in Table 1 with their names, parameter settings and p.d.f. $f(x|a, b)$. In this table, $f_{\text{Nor}}(\dots)$ stands for the p.d.f. of the normal (0, 1) distribution.

Table 1. Distributions used for the comparison of the 3 TBEA charts

Names	Parameters	$f(x a, b)$
Gamma	$a > 0, b > 0$	$\frac{\exp(-\frac{x}{b})x^{a-1}}{b^a \Gamma(a)}$
Lognormal	$a, b > 0$	$(\frac{b}{x}) f_{\text{Nor}}(a + b \ln(x))$
Normal	$a, b > 0$	$\frac{1}{b} f_{\text{Nor}}(\frac{x-a}{b})$
Weibull	$a > 0, b > 0$	$\frac{a}{b} (\frac{x}{b})^{a-1} \exp(-(\frac{x}{b})^a)$

More specifically, we have selected 9 different in-control configurations to be investigated for T (the normal distribution has not been considered as a possible choice for the *time* between events) and 11 in-control configurations to be investigated for X , i.e. a total of 99 scenarios for (T, X) . All of them (see Table 2) are such that the in-control mean is $\mu_0 = 10$ and the in-control standard-deviation is $\sigma_0 \in \{1, 2, 5\}$ (except for the normal distribution, where $\sigma_0 \in \{1, 2\}$ only). In addition, Table 2 also provides the values of the in-control parameters a_0 and b_0 and the corresponding skewness coefficient γ_0 .

Table 2. In-control configurations to be investigated

Distributions	T	X	a_0	b_0	μ_0	σ_0	γ_0
Gamma	• •		100	0.1	10	1	0.2
	• •		25	0.4	10	2	0.4
	• •		4	2.5	10	5	1
Lognormal	• •		-23.0334	10.0249	10	1	0.3010
	• •		-11.5277	5.0494	10	2	0.6080
	• •		-4.6382	2.1169	10	5	1.6250
Normal	○ •		10	1	10	1	0
	○ •		10	2	10	2	0
Weibull	• •		12.1534	10.4304	10	1	-0.7155
	• •		5.7974	10.7998	10	2	-0.3519
	• •		2.1013	11.2906	10	5	0.5664

The upper control limits UCL_Z for the three TBFA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$, satisfying $ATS_0 = 370.4$, have been obtained in Table 3 for the 99 possible scenarios defined in Table 2. As it can be seen, regardless of the statistic Z and the distribution of X , these upper control limits tend to be similar if σ_0 is small (say $\sigma_0 = 1$) and the distribution of T is either gamma or lognormal. But, when T follows a Weibull distribution, the control limits are larger than those of the gamma or lognormal distributions.

When an out-of-control situation occurs in a TBFA process (corresponding to an upper shift in Z), it can be due to i) a mean shift *only in the time* T from μ_{T_0} to $\mu_{T_1} = \delta_T \mu_{T_0}$, ii) a mean shift *only in the amplitude* X from μ_{X_0} to $\mu_{X_1} = \delta_X \mu_{X_0}$, or iii) a combination of the two previous cases, where $\delta_T \leq 1$ and $\delta_X \geq 1$ are the parameters quantifying the change in the time and amplitude, respectively. But, as the actual values of δ_T and δ_X are usually unknown by the practitioner, it is therefore difficult to evaluate the three TBFA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$ using the ATS_1 criterion

Table 3. Upper control limits UCL_Z for the three TBEA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$, satisfying $ATS_0 = 370.4$

		Statistic Z_1										
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	0.273	0.458	1.177	0.275	0.471	1.234	0.268	0.429	0.252	0.400	1.096
	2	0.404	0.547	1.213	0.405	0.556	1.266	0.402	0.527	0.394	0.509	1.138
	5	0.755	0.852	1.395	0.755	0.855	1.431	0.754	0.844	0.752	0.836	1.341
Lognormal	1	0.271	0.457	1.177	0.273	0.471	1.234	0.266	0.428	0.249	0.399	1.096
	2	0.391	0.540	1.212	0.392	0.550	1.264	0.389	0.520	0.380	0.500	1.136
	5	0.682	0.794	1.369	0.682	0.799	1.408	0.681	0.783	0.678	0.772	1.312
Weibull	1	0.293	0.465	1.178	0.295	0.478	1.235	0.289	0.438	0.277	0.410	1.097
	2	0.460	0.578	1.219	0.460	0.587	1.271	0.458	0.562	0.454	0.547	1.145
	5	0.823	0.908	1.418	0.823	0.910	1.452	0.823	0.902	0.822	0.896	1.369

		Statistic Z_2										
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	1.314	1.500	2.226	1.316	1.513	2.280	1.310	1.474	1.296	1.448	2.148
	2	1.590	1.735	2.422	1.591	1.742	2.463	1.588	1.720	1.583	1.706	2.359
	5	3.615	3.713	4.315	3.615	3.713	4.308	3.615	3.712	3.615	3.711	4.309
Lognormal	1	1.310	1.498	2.225	1.312	1.511	2.279	1.306	1.472	1.291	1.445	2.147
	2	1.555	1.708	2.405	1.556	1.715	2.447	1.553	1.692	1.547	1.677	2.341
	5	2.820	2.941	3.623	2.820	2.942	3.623	2.820	2.937	2.819	2.934	3.603
Weibull	1	1.356	1.524	2.238	1.358	1.537	2.291	1.353	1.500	1.343	1.476	2.160
	2	1.762	1.875	2.507	1.762	1.881	2.550	1.761	1.865	1.758	1.856	2.447
	5	4.934	5.010	5.502	4.934	5.010	5.492	4.934	5.010	4.934	5.010	5.506

		Statistic Z_3										
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	2.299	2.474	3.188	2.301	2.488	3.245	2.295	2.447	2.282	2.419	3.107
	2	2.566	2.668	3.277	2.567	2.676	3.328	2.565	2.653	2.561	2.640	3.204
	5	4.587	4.604	4.764	4.587	4.605	4.807	4.587	4.604	4.587	4.603	4.738
Lognormal	1	2.295	2.472	3.188	2.297	2.486	3.244	2.291	2.445	2.277	2.416	3.107
	2	2.530	2.641	3.266	2.531	2.649	3.317	2.529	2.625	2.524	2.611	3.193
	5	3.787	3.817	4.084	3.787	3.818	4.127	3.787	3.815	3.787	3.813	4.043
Weibull	1	2.342	2.499	3.196	2.344	2.512	3.252	2.339	2.472	2.329	2.445	3.115
	2	2.742	2.812	3.357	2.743	2.819	3.413	2.742	2.801	2.740	2.793	3.282
	5	5.912	5.921	6.002	5.912	5.922	6.024	5.912	5.921	5.912	5.921	5.994

defined in (14) that depends on a specific values for δ_T and/or δ_X . For this reason, it is therefore preferable to use the following more general criterion denoted as $EATS_1$ (Expected Average Time to Signal) and defined as

$$EATS_1 = \sum_{\delta_T \in \Omega_T} \sum_{\delta_X \in \Omega_X} f_{\delta_T}(\delta_T) f_{\delta_X}(\delta_X) ATS_1(\delta_T, \delta_X),$$

where Ω_T and Ω_X are the sets of the potential shifts for δ_T and δ_X , respectively, and $f_{\delta_X}(\delta_X)$ and $f_{\delta_T}(\delta_T)$ are the probability mass functions of the shifts δ_T and δ_X over the sets Ω_T and Ω_X . In this chapter, we adopt the classical assumption that confines $f_{\delta_T}(\delta_T)$ and $f_{\delta_X}(\delta_X)$ to be discrete uniform distributions over Ω_T and Ω_X , respectively. If we want to investigate i) a mean shift only due to the time T then we suggest to fix $\Omega_T = \{0.5, 0.55, \dots, 0.9, 0.95\}$ and $\Omega_X = \{1\}$, ii) a mean shift only due to the amplitude X then we suggest to fix $\Omega_T = \{1\}$ and $\Omega_X = \{1.1, 1.2, \dots, 1.9, 2\}$ and iii) a mean shift due to the time T and the amplitude X then we suggest to fix $\Omega_T = \{0.5, 0.55, \dots, 0.9, 0.95\}$ and $\Omega_X = \{1.1, 1.2, \dots, 1.9, 2\}$.

For the 99 possible scenarios defined in Table 2, Table 4 gives the $EATS_1$ values of the three TBEA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$ when $\Omega_T = \{0.5, 0.55, \dots, 0.9, 0.95\}$ and $\Omega_X = \{1.1, 1.2, \dots, 1.9, 2\}$. Values of $EATS_1$ in bold characters are the smallest ones among statistics Z_1 , Z_2 or Z_3 . From Table 4, it can be deduced that when there is a shift in both T and X , the most efficient statistic is Z_1 (in 56% of the cases with an average $EATS_1$ value $\bar{EATS}_1 = 14.91$), followed by statistic Z_2 (in 35% of the cases with $\bar{EATS}_1 = 29.15$) and, finally, statistic Z_3 (in only 5% of the cases with $\bar{EATS}_1 = 11.74$).

The cases where the mean shift is only due to the time T or the mean shift is only due to the amplitude X have both been investigated in Rahali et al. (2019) (see Tables 3 and 4, pages 245–246). In these cases, the conclusions are

- if the mean shift is only due to the time T , then the most efficient statistic is Z_3 (in 71% of the cases with $\bar{EATS}_1 = 59.88$) followed by Z_2 (in 29% of the cases with $\bar{EATS}_1 = 71.82$) while the statistic Z_1 never provides the smallest $EATS_1$ and should not be considered here as an efficient monitoring statistic.
- if the mean shift is only due to the amplitude X , then the statistic Z_1 is the best option as it always gives the smallest $EATS_1$ values, regardless of the combination under consideration. In this case, Z_2 and Z_3 should not be considered as potential efficient monitoring statistics.

2.6 Illustrative example

This illustrative example has been detailed for the first time in Rahali et al. (2019) and it is based on a real data set concerning the time (T in days)

Table 4. EATS₁ values when $\Omega_T = \{0.5, 0.55, \dots, 0.9, 0.95\}$ and $\Omega_X = \{1.1, 1.2, \dots, 1.9, 2\}$ for the three TBEA charts based on statistics $Z \in \{Z_1, Z_2, Z_3\}$

Statistic Z_1												
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	8.4	12.4	47.8	8.5	13.0	55.9	8.4	11.3	8.1	10.4	38.6
	2	10.5	14.5	48.7	10.5	15.0	56.2	10.4	13.6	10.2	12.8	40.1
	5	16.9	21.2	52.5	16.9	21.5	58.4	16.9	20.5	16.8	19.9	45.6
Lognormal	1	8.4	12.4	47.8	8.5	13.0	55.9	8.3	11.3	8.1	10.3	38.6
	2	10.2	14.3	48.6	10.2	14.8	56.1	10.1	13.3	9.9	12.5	39.9
	5	14.7	19.3	52.1	14.7	19.6	58.4	14.7	18.5	14.5	17.8	44.7
Weibull	1	8.8	12.7	47.9	8.9	13.3	55.9	8.8	11.6	8.5	10.7	38.7
	2	12.4	16.0	49.0	12.4	16.5	56.4	12.3	15.1	12.1	14.4	40.6
	5	19.5	23.7	54.1	19.5	23.9	59.6	19.5	23.1	19.4	22.6	47.5

Statistic Z_2												
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	8.5	11.3	30.4	8.5	11.6	34.4	8.5	10.6	8.3	10.1	25.7
	2	12.1	15.0	32.2	12.1	15.2	34.9	12.1	14.5	12.0	14.1	28.9
	5	33.6	34.9	43.0	33.6	34.9	43.1	33.6	34.9	33.6	34.9	42.6
Lognormal	1	8.5	11.2	30.4	8.5	11.6	34.3	8.4	10.6	8.2	10.0	25.7
	2	11.4	14.3	31.9	11.4	14.5	34.7	11.4	13.8	11.2	13.4	28.4
	5	24.5	26.5	38.2	24.5	26.5	38.6	24.5	26.4	24.4	26.3	37.1
Weibull	1	9.2	11.9	30.8	9.2	12.3	34.7	9.1	11.3	9.0	10.7	26.3
	2	16.7	19.1	34.9	16.7	19.3	37.3	16.6	18.8	16.6	18.5	32.1
	5	47.6	48.4	53.7	47.6	48.4	53.6	47.6	48.4	47.6	48.4	53.6

Statistic Z_3												
T	$X \rightarrow$	Gamma			Lognormal			Normal		Weibull		
\downarrow	σ_0	1	2	5	1	2	5	1	2	1	2	5
Gamma	1	8.5	11.4	38.6	8.5	11.8	44.7	8.4	10.6	8.2	9.9	31.7
	2	12.6	14.8	35.7	12.6	15.1	40.0	12.6	14.3	12.5	13.9	30.7
	5	50.9	51.0	52.4	50.9	51.0	53.5	50.9	51.0	50.9	51.0	52.1
Lognormal	1	8.4	11.3	38.7	8.5	11.8	44.8	8.4	10.6	8.2	9.9	31.7
	2	11.7	14.0	35.7	11.7	14.3	40.3	11.7	13.5	11.6	13.1	30.5
	5	35.1	35.5	40.5	35.1	35.5	42.1	35.1	35.4	35.1	35.4	39.4
Weibull	1	9.3	12.1	38.5	9.3	12.6	44.4	9.2	11.3	9.0	10.6	31.8
	2	19.0	20.3	38.3	19.0	20.6	42.4	19.0	19.9	18.9	19.7	33.7
	5	73.3	73.3	73.5	73.3	73.3	73.9	73.3	73.3	73.3	73.3	73.5

between fires in forests of the region “Provence - Alpes - Côte D’Azur” in the south-east of France and their amplitudes (X measured as the burned surface in $ha = 10000m^2$). This data set reports a total of 92 *significant* fires from 2016/10 to 2017/9: the data set has been split into $m = 47$ fires occurring during the “low season” from 2016/10 to mid 2017/6 (used as Phase 1 data) and $n = 45$ fires occurring during the “high season” from mid 2017/6 to 2017/9 (used as Phase 2 data). The values of T and X are recorded in Table 5 (as well as the values of the statistics Z_1 , Z_2 and Z_3) and they are also plotted in Figure 3 where it is clear that the values of T during the “high season” are smaller than those during the “low season” and the values of X during the “high season” are larger than those during the “low season”.

The use of the Kendall’s and Spearman’s rank correlation tests on the whole data set yields p -values larger than the significance level of 0.05 (0.2 for the Kendall’s test and 0.19 for the Spearman’s test) validating the fact that the random variables T and X are uncorrelated (a key assumption in this section). Among the four distributions considered in Table 1, the use of the Kolmogorov-Smirnov’s test shows that the best fit for both T and X is the lognormal distribution with parameters ($a_0 = -1.2648, b_0 = 1.0302$) for T and ($a_0 = -1.6697, b_0 = 0.8624$) for X .

The three TBEA charts, corresponding to the statistics $Z \in \{Z_1, Z_2, Z_3\}$ are plotted in Figure 4 along with their upper control limits $UCL_{Z_1} = 6.0306$, $UCL_{Z_2} = 28.1209$ and $UCL_{Z_3} = 19.3885$ (assuming $ATS_0 = 730$, i.e. 2 years). As it can be seen, these charts detect several out-of-control situations during the “high season” confirming that a decrease in the time between fires occurred as well as an increase in the amplitude of these fires.

3 TBEA charts for dependent Times and Amplitudes

3.1 Motivation

In the previous section, the time T between events and their amplitudes X have been considered as *independent* random variables. But, in practice, this is not always the case. For example, there are natural situations for which the amplitudes tend to become larger when the times between events become shorter (i.e. a negative correlation). Such kind of situations is likely to occur for example in the case of earthquakes for which, in a first phase, small amplitude earthquakes may occur with a low frequency (large time between events) and, suddenly, in a second phase, the occurrence frequency of these earthquakes may increase (shorter time between events) with a negatively correlated increase in their amplitudes. The same kind of situations may also arise in the case of forest fires occurring, in a first phase during the “humid season”, with a low frequency and small amplitudes (surfaces burned) and

Table 5. Phase 1 and 2 data sets corresponding to time (T in days) between fires, amplitudes (X as the burned surface in ha) and the values of the statistics Z_1 , Z_2 and Z_3 .

Phase 1						Phase 2					
i	T	X	Z_1	Z_2	Z_3	i	T	X	Z_1	Z_2	Z_3
1	9	3.68	-1.37	0.16	0.88	1	1	1.00	-0.11	0.40	5.54
2	17	1.99	-2.96	0.05	0.47	2	2	3.70	-0.09	0.75	3.01
3	34	6.00	-5.78	0.07	0.60	3	2	3.17	-0.13	0.64	2.97
4	7	1.19	-1.19	0.07	0.87	4	3	18.40	0.81	2.47	3.18
5	3	135.80	9.45	18.23	11.82	5	3	1.00	-0.47	0.13	1.90
6	2	14.37	0.69	2.89	3.79	6	1	2.22	-0.02	0.89	5.63
7	14	8.10	-1.96	0.23	0.99	7	2	19.09	1.04	3.84	4.14
8	2	32.31	2.01	6.51	5.11	8	1	2.00	-0.04	0.81	5.62
9	6	3.07	-0.87	0.21	1.14	9	2	34.28	2.16	6.90	5.26
10	1	10.03	0.56	4.04	6.21	10	2	3.00	-0.14	0.60	2.95
11	1	7.93	0.40	3.19	6.05	11	1	6.63	0.31	2.67	5.96
12	1	1.50	-0.07	0.60	5.58	12	1	4.47	0.15	1.80	5.80
13	6	23.30	0.62	1.56	2.63	13	7	8.24	-0.67	0.47	1.39
14	3	3.73	-0.27	0.50	2.10	14	1	769.45	56.49	309.87	62.14
15	3	4.73	-0.20	0.63	2.17	15	1	4.37	0.14	1.76	5.79
16	2	3.19	-0.13	0.64	2.97	16	1	90.70	6.50	36.53	12.15
17	2	6.25	0.09	1.26	3.19	17	1	11.49	0.66	4.63	6.31
18	1	3.60	0.08	1.45	5.73	18	6	3590.78	263.36	241.01	265.73
19	1	6.12	0.27	2.46	5.92	19	1	1427.92	104.98	575.04	110.63
20	3	1.50	-0.44	0.20	1.93	20	1	255.96	18.67	103.08	24.32
21	4	1.33	-0.63	0.13	1.46	21	1	1.00	-0.11	0.40	5.54
22	12	1.42	-2.09	0.05	0.56	22	4	13.88	0.29	1.40	2.39
23	3	5.75	-0.13	0.77	2.25	23	1	138.28	10.00	55.69	15.65
24	3	3.47	-0.29	0.47	2.08	24	2	8.90	0.29	1.79	3.39
25	2	13.31	0.61	2.68	3.71	25	3	1.50	-0.44	0.20	1.93
26	1	26.31	1.75	10.60	7.41	26	4	34.63	1.82	3.49	3.92
27	1	18.54	1.18	7.47	6.83	27	1	82.56	5.90	33.25	11.55
28	2	66.17	4.51	13.32	7.61	28	1	2.00	-0.04	0.81	5.62
29	1	9.90	0.55	3.99	6.20	29	1	162.08	11.75	65.27	17.40
30	3	4.22	-0.24	0.57	2.13	30	4	3.26	-0.49	0.33	1.61
31	7	34.28	1.24	1.97	3.31	31	2	285.91	20.69	57.57	23.79
32	4	2.23	-0.57	0.22	1.53	32	1	2.00	-0.04	0.81	5.62
33	1	1.84	-0.05	0.74	5.60	33	3	11.57	0.30	1.55	2.67
34	1	2.88	0.03	1.16	5.68	34	9	34.70	0.91	1.55	3.16
35	1	21.46	1.40	8.64	7.05	35	1	431.00	31.56	173.57	37.21
36	1	4.46	0.15	1.80	5.80	36	1	10.89	0.62	4.39	6.27
37	1	58.27	4.11	23.47	9.76	37	4	1.00	-0.66	0.10	1.44
38	1	8.84	0.47	3.56	6.12	38	6	1.50	-0.99	0.10	1.02
39	13	1.03	-2.30	0.03	0.50	39	1	1.17	-0.10	0.47	5.55
40	7	16.57	-0.06	0.95	2.00	40	2	1.27	-0.27	0.26	2.83
41	14	4.96	-2.20	0.14	0.76	41	1	26.25	1.75	10.57	7.40
42	1	1.37	-0.08	0.55	5.57	42	3	11.66	0.31	1.57	2.68
43	3	23.39	1.17	3.14	3.55	43	1	3.03	0.04	1.22	5.69
44	20	1.70	-3.53	0.03	0.40	44	1	12.00	0.70	4.83	6.35
45	22	5.30	-3.63	0.10	0.64	45	1	1.00	-0.11	0.40	5.54
46	1	15.64	0.97	6.30	6.62						
47	9	5.14	-1.27	0.23	0.99						

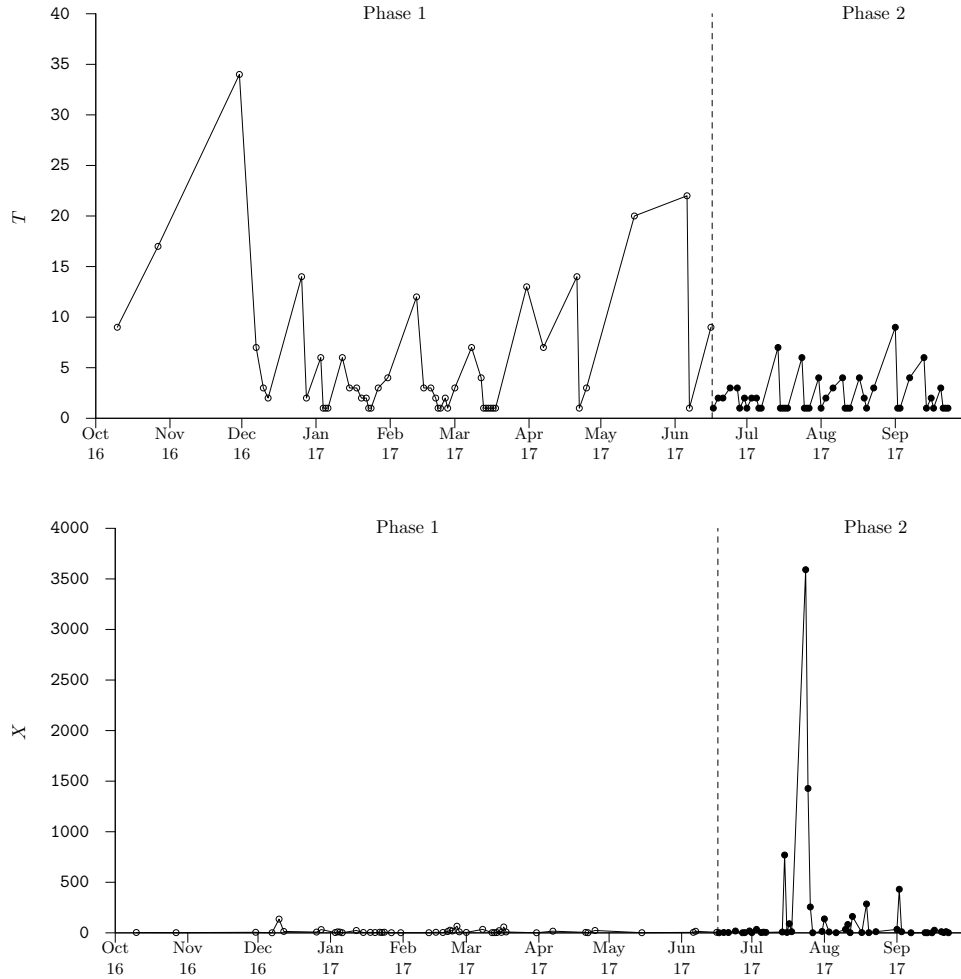


Fig. 3. Time (T in days) between fires and amplitudes (X as the burned surface in ha) corresponding to the data set in Table 5

becoming more disastrous during the “dry season” with shorter time between the occurrence of forest fires and larger amplitudes (see for instance the 2019 forest fires in Amazonia or Siberia or the 2020 forest fires in Australia and USA). Positive correlation between T and X (i.e. the time between events becomes shorter and the amplitude becomes smaller) is also possible as the the forthcoming illustrative example in this section will depict it.

Until now, very few research papers have investigated TBEA control charts by considering the potential dependence between the two variables T and X .

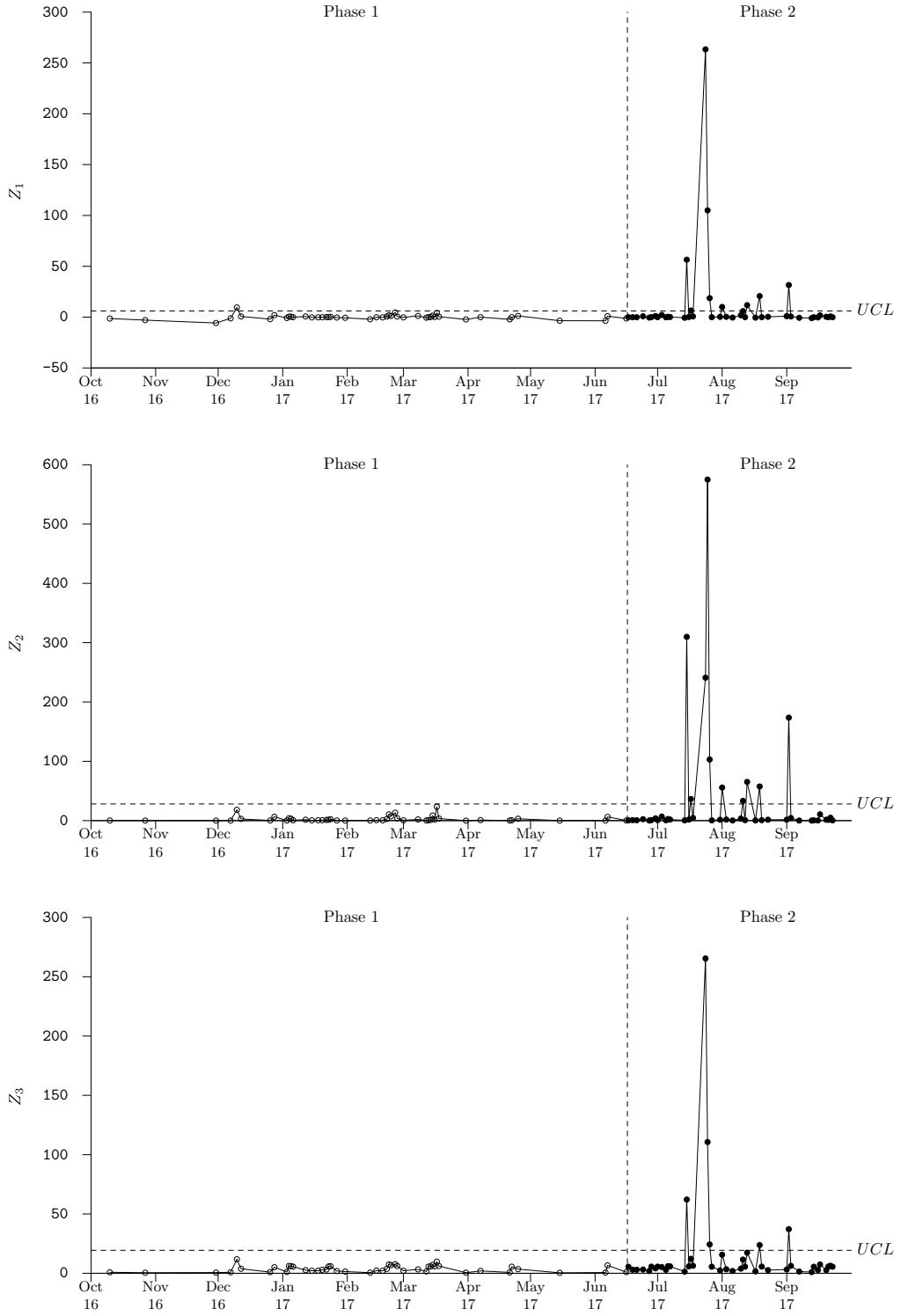


Fig. 4. Statistics Z_1 , Z_2 and Z_3 corresponding to the data set in Table 5

Cheng and Mukherjee (2014) were the first to investigate a T^2 TBEA control chart by using a bivariate SAT (Smith-Adelfang-Tubbs) Gamma distribution to model the joint probability of T and X . This work has been extended later by Cheng et al. (2017) who developed a similar approach based on a MEWMA (multivariate exponentially weighted moving average) procedure.

In this section, instead of specifying a particular bivariate joint distribution for (T, X) , like the SAT distribution for instance, we will assume that the marginal distributions of T and X are both known (and they can be almost anything) and we will use the Copulas mechanism (popularized by Sklar (1959)) in order to model the dependence between the time T and the amplitude X . The use of Copulas in the Statistical Process Monitoring field is not so common. We can cite for instance Fatahi et al. (2011), Dokouhaki and Noorossana (2013), Busababodhin and Amphanthong (2016) and Sukparungsee et al. (2018) who all proposed various types of control charts based on Copulas.

3.2 Model

In this section, we assume that $(X, T) \in \mathbb{R}_+^2$ and their joint continuous c.d.f. is equal to

$$F_{(T,X)}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta) = C(F_T(t|\boldsymbol{\theta}_T), F_X(x|\boldsymbol{\theta}_X)|\theta), \quad (16)$$

where $F_T(t|\boldsymbol{\theta}_T)$ and $F_X(x|\boldsymbol{\theta}_X)$, as defined in section 2.1, are the marginal c.d.f. of T and X , respectively, $C(u, v|\theta)$ is a Copula containing all information on the dependence structure between T and X , and θ is a dependence parameter quantifying the dependence between the marginals. In addition, let

$$f_{(T,X)}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta) = c(F_T(t|\boldsymbol{\theta}_T), F_X(x|\boldsymbol{\theta}_X)|\theta)f_T(t|\boldsymbol{\theta}_T)f_X(x|\boldsymbol{\theta}_X) \quad (17)$$

be the joint p.d.f. of (X, T) where $f_T(t|\boldsymbol{\theta}_T)$ and $f_X(x|\boldsymbol{\theta}_X)$ are the marginal p.d.f.'s of T and X , respectively, and $c(u, v|\theta) = \frac{\partial C(u, v|\theta)}{\partial u \partial v}$ is the Copula density. As explained in section 2.1, in order to not favor one random variable over the other one, the new random variables $T' = \frac{T}{\mu_{T_0}}$ and $X' = \frac{X}{\mu_{X_0}}$ are introduced as the in-control standardized counterparts of T and X , where μ_{T_0} and μ_{X_0} are the (known) in-control expectation of T and X , respectively. It is easy to show that the joint c.d.f. $F_{(T', X')}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta)$ and joint p.d.f. $f_{(T', X')}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta)$ of $(X', T') \in \mathbb{R}_+^2$ are equal to

$$\begin{aligned} F_{(T', X')}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta) &= C(F_T(t\mu_{T_0}|\boldsymbol{\theta}_T), F_X(x\mu_{X_0}|\boldsymbol{\theta}_X)|\theta), \\ f_{(T', X')}(t, x|\boldsymbol{\theta}_T, \boldsymbol{\theta}_X, \theta) &= \mu_{T_0}\mu_{X_0}c(F_T(t\mu_{T_0}|\boldsymbol{\theta}_T), F_X(x\mu_{X_0}|\boldsymbol{\theta}_X)|\theta) \\ &\quad f_T(t\mu_{T_0}|\boldsymbol{\theta}_T)f_X(x\mu_{X_0}|\boldsymbol{\theta}_X). \end{aligned}$$

The closed form formulas for the c.d.f. and p.d.f. of the statistics Z_1 , Z_2 and Z_3 defined in section 2.2 are provided in Rahali et al. (2021). The definition

of the upper control limit UCL_Z of the TBEA charts with dependent times T and amplitudes X is similar to the one in equation (12) and it just requires the addition of the dependence parameter θ , i.e. $UCL_Z = F_Z^{-1}(1 - \alpha | \boldsymbol{\theta}_{Z_0}, \theta)$. The formulas for computing ATS_1 and $SDTS_1$ are the same as in equations (14) and (15), respectively.

3.3 Comparative studies

In order to compare the three TBEA charts for dependent times T and amplitudes X based on statistic $Z \in \{Z_1, Z_2, Z_3\}$, the same distributions listed in Table 1 have been chosen and the same 99 possible scenarios defined in Table 2 have been investigated. The Archimedean bivariate Copulas of Gumbel (1960), Clayton (1978) and Frank (1979) have been chosen in this section to model the dependence between T and X . The Gumbel's (also called Gumbel-Hougaard's) and Clayton's Copulas are two asymmetric Copulas exhibiting a larger dependence in the positive tail than in the negative one (for the Gumbel's Copula) and in the negative tail than in the positive one (for the Clayton's Copulas). The Frank's Copula is a symmetric one that can be used to model dependence structures with either positive or negative correlation. Other Archimedean bivariate Copulas could have also been investigated, like for instance the Ali et al. (1978) and Joe (1993) Copulas but, for simplicity and also due to their popularity, we only confined our investigations to the Gumbel's, Clayton's and Frank's Copulas. Details concerning the definition of each of these Copulas $C(u, v | \theta)$, the domain of definition for θ and the relationship between the Kendall's rank correlation coefficient τ and the dependence parameter θ are provided in Table 6. In order to facilitate the use of these Copulas, Table 7 simply provides pre-computed values of θ for several selected values of the Kendall's rank correlation coefficient $\tau \in \{0, 0.1, 0.2, \dots, 0.9\}$.

When it is not possible to model a negative dependence with the Copulas defined above, it is possible to use 90 or 270 degrees rotated versions C_{90} or C_{270} of these Copulas using the following transformations (see Brechmann and Schepsmeier (2013))

$$\begin{aligned} C_{90}(u_1, u_2) &= u_2 - C(1 - u_1, u_2), \\ C_{270}(u_1, u_2) &= u_1 - C(u_1, 1 - u_2). \end{aligned}$$

Similarly to Table 3 (for independent times T and amplitudes X) and assuming $ATS_0 = 370.4$, the upper control limits UCL_Z of the three TBEA charts with dependent times T and amplitudes X are reported in Tables 3–5 of Rahali et al. (2021) for the 99 scenarios defined in Table 2 and for the 3 Copulas defined above. From Tables 3–5 in Rahali et al. (2021), the following conclusions can be drawn

Table 6. Details concerning Gumbel’s, Clayton’s and Frank’s Copulas

Name	$C(u, v \theta)$	Domain for θ	τ and θ
Gumbel	$\exp\left(-\left((-\ln(u))^\theta + (-\ln(v))^\theta\right)^{\frac{1}{\theta}}\right)$	$[1, \infty)$	$\tau = 1 - \frac{1}{\theta} \Leftrightarrow \theta = \frac{1}{1-\tau}$
Clayton	$\max(0, u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$	$[-1, \infty) \setminus \{0\}$	$\tau = \frac{\theta}{\theta+2} \Leftrightarrow \theta = \frac{2\tau}{1-\tau}$
Frank	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\mathbb{R} \setminus \{0\}$	$\tau = 1 + \frac{4(D_1(\theta)-1)}{\theta}$

Note: $D_1(\theta)$ is the Debye function of the first kind defined as $D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t-1} dt$.

Table 7. Pre-computed values of θ for several selected values of $\tau \in \{0, 0.1, 0.2, \dots, 0.9\}$

τ	θ		
	Frank	Clayton	Gumbel
0.0	0.00	0.00	1.00
0.1	0.91	0.22	1.11
0.2	1.86	0.50	1.25
0.3	2.92	0.86	1.43
0.4	4.16	1.33	1.67
0.5	5.74	2.00	2.00
0.6	7.93	3.00	2.50
0.7	11.41	4.67	3.33
0.8	18.19	8.00	5.00
0.9	38.28	18.00	10.00

- For a fixed statistic $Z \in \{Z_1, Z_2, Z_3\}$, scenario in Table 2 and type of Copula, the larger τ , the smaller the control limit UCL_Z .
- For a fixed scenario in Table 2, value of τ and type of Copula, the upper control limits of the statistic $Z \in \{Z_1, Z_2, Z_3\}$ always satisfy $UCL_{Z_1} < UCL_{Z_2} < UCL_{Z_3}$.
- For a fixed scenario in Table 2, value of τ and statistic $Z \in \{Z_1, Z_2, Z_3\}$, the values of UCL_Z are more or less the same no matter the type of Copula considered.

As in Section 2.5 for independent times T and amplitudes X , $EATS_1$ values (for shifts in both T and X) have been reported in Tables 7–9 of Rahali et al. (2021) in the case of dependent times T and amplitudes X . These Tables only consider the Frank’s Copula and $\tau \in \{0.2, 0.5, 0.8\}$. These results clearly show that, irrespective of the level of dependence, when a shift in both T and X is likely to occur, the best option is to use the statistics Z_1 or, eventually, Z_2 , but not the statistic Z_3 which is never considered as an efficient one. These conclusions are similar to the ones obtained in Section 2.5 and they remain

identical for other Copulas like the Clayton's or Gumbel's ones. More details concerning these aspects can be found in Rahali et al. (2021).

3.4 Illustrative example

The following illustrative example has been detailed for the first time in Rahali et al. (2021) and it is related to a company that has recorded for one of its bottleneck machine, during about 6 years (from 08/01/2012 to 27/12/2018) all the breakdown dates (D_i in days) as well as the estimated corresponding incurred costs (X_i , in euros) which include all the repair and restart costs (spare parts, manpower) and the cost of manufacturing disruption, see Table 8. This data set of 44 dates is divided into two parts

- 30 breakdowns recorded during 5 years (2012 to 2016) and used in this example as a Phase 1 data set.
- 14 breakdowns recorded during 2 years (2017 to 2018) and used in this example as a Phase 2 data set.

The times T_i and amplitude X_i , $i = 1, \dots, 44$ in Table 8 have also been plotted in Figure 5 with \circ (for Phase 1) and \bullet (for Phase 2), respectively. From the scatterplot shown in Figure 5, it can be noted that when the time T_i between consecutive breakdowns is smaller (larger), the corresponding cost X_i seems to be also smaller (larger), indicating a potential slight positive correlation between T and X . Investigations (during the period 2012 to 2016) about this phenomena have clarified why such a positive correlation between T and X may occur in this situation. After the occurrence of a breakdown, if the next one occurs after a *short* period of time, it is often due to the same problem occurring to the process: consequently, the time to looking for the breakdown causes is reduced. The spare parts costs are also reduced as they have already been purchased for the previous breakdown. On the contrary, when the next breakdown occurs after a *long* period of time, the causes are usually different from the previous breakdown and need more time to be revealed; new spare parts need to be purchased, thus increasing the cost. In order to evaluate if a positive correlation significantly exists between T and X , the Kendall's and Spearman's rank correlation coefficients have been estimated to $\hat{\tau} = 0.4657$ and $\hat{\rho} = 0.6129$ as well as their corresponding p -values 0.00035 and 0.00032, respectively, thus confirming a positive correlation between T and X . In this example, a Frank's Copula has been chosen to model the dependence between T and X . As explained in Rahali et al. (2021), if $\hat{\tau} = 0.4657$ then an estimation of the dependence parameter is $\hat{\theta} = 5.14$.

Concerning the marginal distributions of T and X , the best fit using the Kolmogorov-Smirnov's test is to choose a Gamma distribution for the time T with parameters ($a_0 = 11.6488, b_0 = 5.0562$) and the Weibull distribution for the cost X with parameters ($a_0 = 4.8472, b_0 = 5396.4958$).

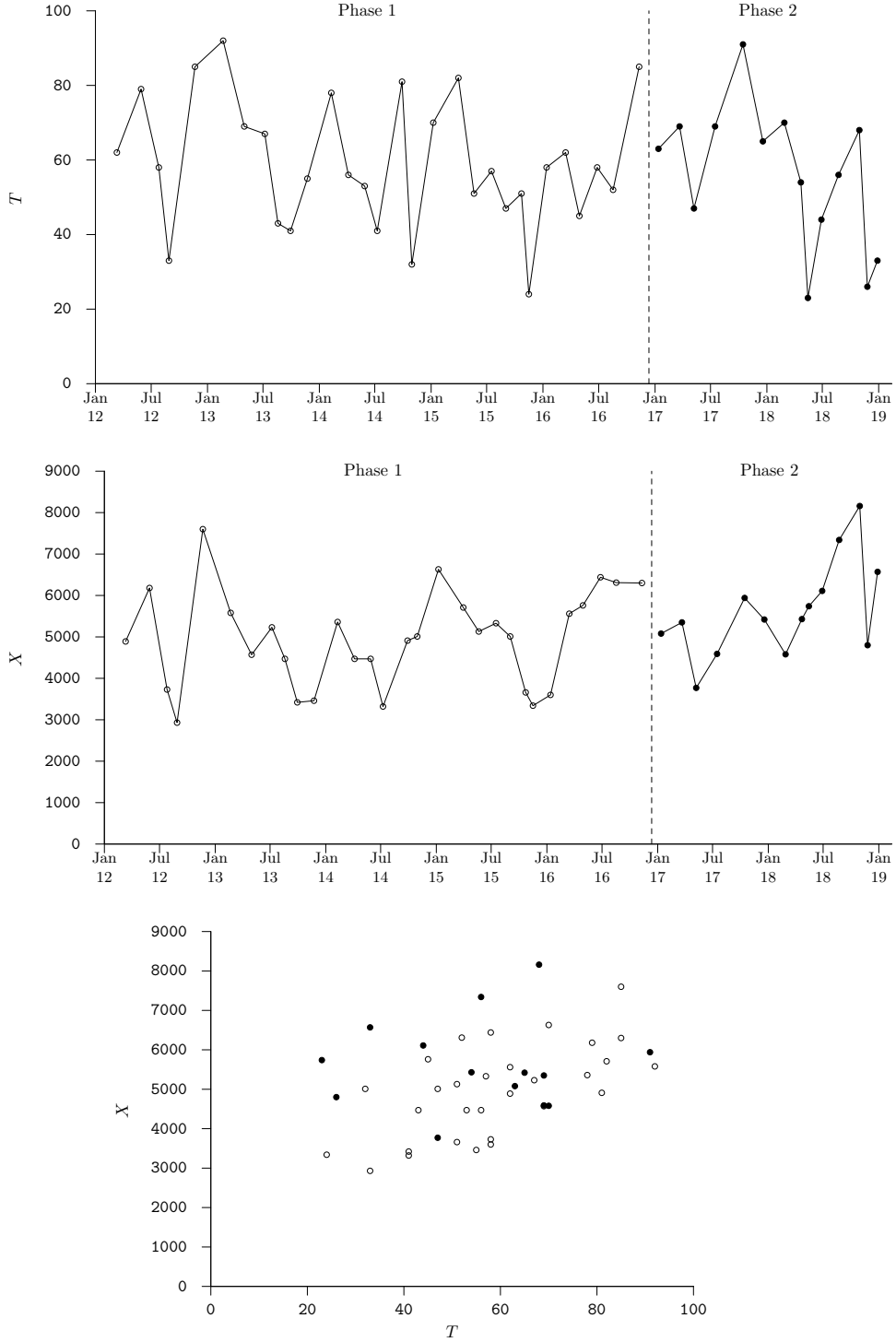


Fig. 5. Phase 1 (○) and 2 (●) data corresponding to the time (T in days) between breakdowns and amplitudes (X in euros) corresponding to the data set in Table 8.

Table 8. Phase 1 and 2 data sets corresponding to the time (T_i in days) between two consecutive breakdowns, amplitudes (X_i cost in euros) and the values of the statistics Z_1 , Z_2 and Z_3 .

Phase 1							Phase 2						
Date	i	T_i	X_i	$Z_{1,i}$	$Z_{2,i}$	$Z_{3,i}$	Date	i	T_i	X_i	$Z_{1,i}$	$Z_{2,i}$	$Z_{3,i}$
10/03/12	1	62	4890	-0.064	0.939	1.939	11/01/17	1	63	5080	-0.043	0.960	1.962
28/05/12	2	79	6180	-0.092	0.932	1.995	21/03/17	2	69	5350	-0.090	0.923	1.935
25/07/12	3	58	3730	-0.231	0.766	1.770	07/05/17	3	47	3770	-0.036	0.955	2.015
27/08/12	4	33	2930	0.032	1.057	2.377	15/07/17	4	69	4590	-0.243	0.792	1.782
20/11/12	5	85	7600	0.093	1.065	2.230	14/10/17	5	91	5940	-0.344	0.777	1.848
20/02/13	6	92	5580	-0.434	0.722	1.768	18/12/17	6	65	5420	-0.008	0.993	2.002
30/04/13	7	69	4570	-0.247	0.789	1.778	26/02/18	7	70	4580	-0.262	0.779	1.767
06/07/13	8	67	5230	-0.080	0.930	1.937	21/04/18	8	54	5430	0.181	1.197	2.189
18/08/13	9	43	4470	0.174	1.238	2.274	14/05/18	9	23	5740	0.770	2.972	3.721
28/09/13	10	41	3420	-0.005	0.993	2.128	27/06/18	10	44	6110	0.488	1.654	2.574
22/11/13	11	55	3460	-0.234	0.749	1.770	22/08/18	11	56	7340	0.533	1.561	2.536
08/02/14	12	78	5360	-0.241	0.818	1.839	29/10/18	12	68	8160	0.495	1.429	2.516
05/04/14	13	56	4470	-0.047	0.951	1.956	24/11/18	13	26	4800	0.529	2.199	3.236
28/05/14	14	53	4470	0.004	1.004	2.015	27/12/18	14	33	6570	0.768	2.371	3.113
08/07/14	15	41	3320	-0.025	0.964	2.108							
27/09/14	16	81	4910	-0.382	0.722	1.720							
29/10/14	17	32	5010	0.470	1.864	2.854							
07/01/15	18	70	6630	0.152	1.128	2.182							
30/03/15	19	82	5710	-0.238	0.829	1.873							
20/05/15	20	51	5130	0.171	1.198	2.192							
16/07/15	21	57	5330	0.110	1.114	2.111							
01/09/15	22	47	5010	0.215	1.269	2.266							
22/10/15	23	51	3660	-0.126	0.855	1.895							
15/11/15	24	24	3340	0.268	1.657	3.129							
12/01/16	25	58	3600	-0.257	0.739	1.743							
14/03/16	26	62	5560	0.072	1.068	2.074							
28/04/16	27	45	5760	0.401	1.524	2.473							
25/06/16	28	58	6440	0.317	1.322	2.318							
16/08/16	29	52	6310	0.393	1.445	2.408							
09/11/16	30	85	6300	-0.169	0.883	1.967							

Assuming an in-control ATS value, $ATS_0 = 9125$ days (i.e. 25 years), the upper control limits of the three TBEA charts based on statistics Z_1 , Z_2 and Z_3 are found to be $UCL_{Z_1} = 0.57$, $UCL_{Z_2} = 2.06$ and $UCL_{Z_3} = 3.18$, respectively, see Rahali et al. (2021) for more details. The three TBEA charts, corresponding to the statistics Z_1 , Z_2 and Z_3 , are plotted in Figure 6 along with their upper control limits. As it can be seen, the Phase 1 part of these charts seems to confirm the fact that from 2012 to 2016, the time between consecutive breakdowns and their corresponding costs were in stable state. But, from

2017, things seems to have changed as several out-of-control situations (see values in bold in Table 8) have been detected by all the TBEA charts due to more frequent breakdowns and an increasing maintenance cost (also due to an aging machine). Every time an out-of-control situation is detected, the production has been stopped, the root causes of the breakdown have been searched for, analyzed and repaired. Then, the machine has been restarted.

4 A distribution-free TBEA chart

4.1 Motivation

The TBEA control charts developed in the previous sections are *parametric* ones, i.e. they assume that the distributions of the Times X and their Amplitudes X are perfectly known. However, as it is mentioned in Qiu (2014), in many practical situations, the distributions of these random variables are unknown (or their parameters cannot be reliably estimated by means of a Phase 1 retrospective study) making the implementation of traditional parametric control charts to be an incorrect approach. For instance, the parametric distributions that have been investigated in Sections 2 and 3, are the Gamma, Lognormal, Normal and Weibull, but are these distributions really appropriate in these cases? To overcome this problem, *distribution-free* control charts have been proposed and investigated in the literature. Among the most recent ones, we can cite for instance Celano et al. (2016), Abid et al. (2016, 2017a,b, 2018), Castagliola et al. (2019) and Chakraborti and Graham (2019b). Most of these approaches use nonparametric statistics like the Sign or the Wilcoxon signed-rank statistics. Practical guidelines for the implementation of such distribution-free control charts can be found in Chakraborti and Graham (2019a).

In this Section, we present an upper-sided distribution-free EWMA control chart for monitoring TBEA data. Moreover, as evaluating the Run Length properties of any EWMA scheme based on discrete data is a challenging problem, we will also introduce a specific method called “continuousify” which allows to obtain much better and replicable results.

4.2 Model

In this Section, we assume that $F_T(t|\theta_T)$ and $F_X(x|\theta_X)$ are the *unknown* continuous c.d.f. of T_i and X_i , $i = 1, 2, \dots$, where θ_T and θ_X are *known* quantiles, respectively. As for the previous sections, when the process is in-control, we have $\theta_T = \theta_{T_0}$, $\theta_X = \theta_{X_0}$ and, when the process is out-of-control, we have $\theta_T = \theta_{T_1}$, $\theta_X = \theta_{X_1}$. Without loss of generality, we will consider that θ_T and θ_X are the median values of T_i and X_i , respectively. Of course, if necessary,

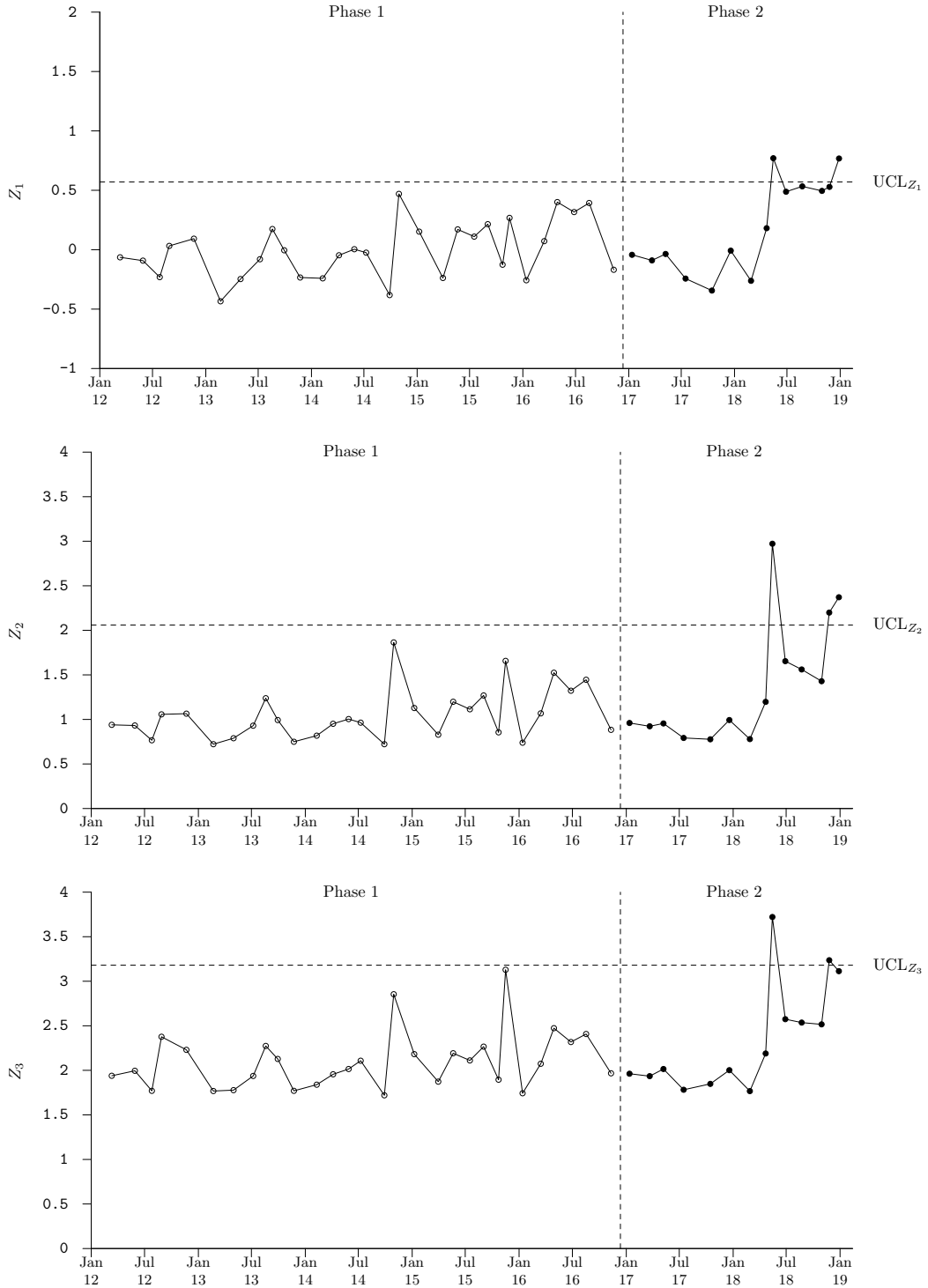


Fig. 6. Statistics Z_1 , Z_2 and Z_3 corresponding to the data set in Table 8.

other quantiles can also be considered.

Let $p_T = P(T_i > \theta_{T_0} | \theta_T) = 1 - F_T(\theta_{T_0} | \theta_T)$ and $p_X = P(X_i > \theta_{X_0} | \theta_X) = 1 - F_X(\theta_{X_0} | \theta_X)$, $i = 1, 2, \dots$, be the probabilities that T_i and X_i are larger than θ_{T_0} and θ_{X_0} assuming that the actual median values are θ_T and θ_X , respectively. Let us also define $q_T = 1 - p_T$ and $q_X = 1 - p_X$. If the process is in-control, we have $p_T = p_{T_0} = 1 - F_T(\theta_{T_0} | \theta_{T_0}) = 0.5$, $p_X = p_{X_0} = 1 - F_X(\theta_{X_0} | \theta_{X_0}) = 0.5$ and, when the process is out-of-control, we have $p_T = p_{T_1} = 1 - F_T(\theta_{T_0} | \theta_{T_1})$, $p_X = p_{X_1} = 1 - F_X(\theta_{X_0} | \theta_{X_1})$.

The upper-sided distribution-free EWMA TBEA control chart that will be introduced in this section is based on the following sign statistics ST_i and SX_i , for $i = 1, 2, \dots$ as

$$\begin{aligned} ST_i &= \text{sign}(T_i - \theta_{T_0}), \\ SX_i &= \text{sign}(X_i - \theta_{X_0}), \end{aligned}$$

where $\text{sign}(x) = -1$ if $x < 0$ and $\text{sign}(x) = +1$ if $x > 0$. The case $x = 0$ will not be considered here (even if it can happen in practice) because the random variables T_i and X_i are assumed to be continuous ones. In order to simultaneously monitor (T_i, X_i) , $i = 1, 2, \dots$, we suggest to define the statistic S_i , for $i = 1, 2, \dots$ as

$$S_i = \frac{SX_i - ST_i}{2}.$$

Since $ST_i \in \{-1, +1\}$ and $SX_i \in \{-1, +1\}$ we have $S_i \in \{-1, 0, +1\}$ and, more precisely, we have:

- $S_i = -1$ when T_i increases ($ST_i = +1$) and, at the same time, X_i decreases ($SX_i = -1$). In this case, the process seems to be in an *acceptable* situation.
- $S_i = +1$ when T_i decreases ($ST_i = -1$) and, at the same time, X_i increases ($SX_i = +1$). In this case, the process seems to be in an *unacceptable* situation.
- $S_i = 0$ when both T_i and X_i increase or when both T_i and X_i decrease. In this case, the process seems to be in an *intermediate* situation.

It can be easily proven that the p.m.f. (probability mass function) $f_{S_i}(s|p_T, p_X) = P(S_i = s|p_T, p_X)$ and the c.d.f. $F_{S_i}(s|p_T, p_X) = P(S_i \leq s|p_T, p_X)$ of S_i are equal to

$$f_{S_i}(s|p_T, p_X) = \begin{cases} p_T q_X & \text{if } s = -1 \\ p_T p_X + q_T q_X & \text{if } s = 0 \\ q_T p_X & \text{if } s = +1 \\ 0 & \text{if } s \notin \{-1, 0, 1\} \end{cases},$$

and

$$F_{S_i}(s|p_T, p_X) = \begin{cases} 0 & \text{if } s \in (-\infty, -1) \\ p_T q_X & \text{if } s \in [-1, 0) \\ p_T + q_T q_X & \text{if } s \in [0, 1) \\ 1 & \text{if } s \in [1, +\infty) \end{cases}.$$

If we define an EWMA TBEA type control chart directly monitoring the statistic S_i , it is known (see Wu et al. (2021) for details) that, due to the strong discrete nature of this statistic, an accurate computation using Markov chain or integral equation methods, for instance, of the corresponding Run Length properties (ARL, SDRL, ...) is impossible. Consequently, it is actually possible to define an EWMA TBEA type control chart based on S_i but it is unfortunately impossible to correctly evaluate its properties and, therefore, it is impossible to design it in a reliable way. In order to overcome this problem and before any implementation of an EWMA scheme, Wu et al. (2021) suggested to define an extra parameter $\sigma \in [0.1, 0.2]$ and to transform the discrete random variable S_i into a new continuous one, denoted as S_i^* defined as a mixture of 3 normal random variables $Y_{i,-1} \sim \text{Nor}(-1, \sigma)$, $Y_{i,0} \sim \text{Nor}(0, \sigma)$ and $Y_{i,+1} \sim \text{Nor}(+1, \sigma)$, with weights $w_{-1} = p_T q_X$, $w_0 = p_T p_X + q_T q_X$ and $w_{+1} = q_T p_X$ (corresponding to the probabilities $f_{S_i}(s|p_T, p_X)$, $s \in \{-1, 0, +1\}$), respectively, i.e.

$$S_i^* = \begin{cases} Y_{i,-1} & \text{if } S_i = -1, \\ Y_{i,0} & \text{if } S_i = 0, \\ Y_{i,+1} & \text{if } S_i = +1. \end{cases}$$

This strategy has been called *continuousify* by Wu et al. (2021). It is easy to prove that the c.d.f. $F_{S_i^*}(s|p_T, p_X) = P(S_i^* \leq s|p_T, p_X)$ of S_i^* is equal to

$$F_{S_i^*}(s|p_T, p_X) = p_T q_X F_{\text{Nor}}(s|-1, \sigma) + (p_T p_X + q_T q_X) F_{\text{Nor}}(s|0, \sigma) + q_T p_X F_{\text{Nor}}(s|+1, \sigma), \tag{18}$$

and its expectation $E(S_i^*)$ and variance $V(S_i^*)$ are equal to

$$E(S_i^*) = p_X - p_T, \\ V(S_i^*) = \sigma^2 + p_T q_T + p_X q_X.$$

When the process is in-control, we have $p_{T_0} = q_{T_0} = 0.5$, $p_{X_0} = q_{X_0} = 0.5$ and the expectation and variance of S_i^* simplify to $E(S_i^*) = 0$ and $V(S_i^*) = \sigma^2 + 0.5$. As the main goal is to detect an increase in S_i (in order to avoid, for instance, more damages or injuries/costs) rather than a decrease, the following upper-sided EWMA TBEA control chart based on the statistic Z_i^* is proposed

$$Z_i^* = \max(0, \lambda S_i^* + (1 - \lambda) Z_{i-1}^*), \tag{19}$$

with the following upper asymptotic control limit UCL defined as

$$\text{UCL} = \underbrace{E(S_i^*)}_{=0} + K \sqrt{\frac{\lambda}{2 - \lambda}} \times \underbrace{\sqrt{V(S_i^*)}}_{=\sqrt{\sigma^2 + 0.5}} = K \sqrt{\frac{\lambda(\sigma^2 + 0.5)}{2 - \lambda}}, \tag{20}$$

where $\lambda \in [0, 1]$ and $K > 0$ are the control chart parameters to be fixed and the initial value $Z_0^* = 0$. The zero-state ARL and SDRL of the proposed

distribution-free upper-sided EWMA TBEA control chart can be obtained using the standard approach of Brook and Evans (1972) which assumes that the behavior of this control chart can be well represented by a discrete-time Markov chain with $m+2$ states, where states $i = 0, 1, \dots, m$ are transient and state $m+1$ is an absorbing one. The transition probability matrix \mathbf{P} of this discrete-time Markov chain is

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{r} \\ \mathbf{0}^\top & 1 \end{pmatrix} = \begin{pmatrix} Q_{0,0} & Q_{0,1} & \cdots & Q_{0,m} & r_0 \\ Q_{1,0} & Q_{1,1} & \cdots & Q_{1,m} & r_1 \\ \vdots & \vdots & & \vdots & \vdots \\ Q_{m,0} & Q_{m,1} & \cdots & Q_{m,m} & r_m \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where \mathbf{Q} is the $(m+1, m+1)$ matrix of transient probabilities, where $\mathbf{0} = (0, 0, \dots, 0)^\top$ and where the $(m+1, 1)$ vector \mathbf{r} satisfies $\mathbf{r} = \mathbf{1} - \mathbf{Q}\mathbf{1}$ (i.e. row probabilities must sum to 1) with $\mathbf{1} = (1, 1, \dots, 1)^\top$. The transient states $i = 1, \dots, m$ are obtained by dividing the interval $[0, \text{UCL}]$ into m subintervals of width 2Δ , where $\Delta = \frac{\text{UCL}}{2m}$. By definition, the midpoint of the i -th subinterval (representing state i) is equal to $H_i = (2i-1)\Delta$. The transient state $i = 0$ corresponds to the “restart state” feature of our chart and it is represented by the value $H_0 = 0$. Concerning the proposed upper-sided EWMA TBEA control chart, it can be easily shown that the generic element $Q_{i,j}$, $i = 0, 1, \dots, m$, of the matrix \mathbf{Q} is equal to

- if $j = 0$,

$$Q_{i,0} = F_{S_i^*} \left(-\frac{(1-\lambda)H_i}{\lambda} \middle| p_T, p_X \right), \quad (21)$$

- if $j = 1, 2, \dots, m$,

$$Q_{i,j} = F_{S_i^*} \left(\frac{H_j + \Delta - (1-\lambda)H_i}{\lambda} \middle| p_T, p_X \right) - F_{S_i^*} \left(\frac{H_j - \Delta - (1-\lambda)H_i}{\lambda} \middle| p_T, p_X \right) \quad (22)$$

Let $\mathbf{q} = (q_0, q_1, \dots, q_m)^\top$ be the $(m+1, 1)$ vector of initial probabilities associated with the $m+1$ transient states. In our case, we assume $\mathbf{q} = (1, 0, \dots, 0)^\top$, i.e. the initial state corresponds to the “restart state”. When the number m of subintervals is sufficiently large (say $m = 300$), this finite approach provides an effective method that allows the ARL and SDRL to be accurately evaluated using the following classical formulas

$$\text{ARL} = \mathbf{q}^\top (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{1}, \quad (23)$$

$$\text{SDRL} = \sqrt{2\mathbf{q}^\top (\mathbf{I} - \mathbf{Q})^{-2} \mathbf{Q} \mathbf{1} + \text{ARL}(1 - \text{ARL})}. \quad (24)$$

In order to illustrate the “power” of the *continuousify* technique on the upper-sided EWMA TBEA control chart (in the case $K = 3$ and $\lambda = 0.2$), Table 9 presents some ARL values obtained without (left side) and with (right side)

this technique, corresponding to 3 combinations for (p_X, p_T) , a number of subintervals $m \in \{100, 120, \dots, 400\}$ and $\sigma = 0.125$. When the *continuousify* technique is not used, the random variable S_i^* in (19) is replaced by S_i or, equivalently, the parameter σ is set to 0. For comparison purpose, the last row of Table 9 also provides the ARL values obtained by simulations. As it can be seen, if the *continuousify* technique is not used (left side), the ARL values obtained using the Markov chain method can have a large variability without any clear monotonic convergence when m increases. It turns out that with these unstable ARL values, it can be really difficult to design and optimize the upper-sided EWMA TBEA control chart. On the other side, if the *continuousify* technique is used (right side), the ARL values obtained using the Markov chain method are clearly very stable, even for small values of m , but (and this is the price to pay for this stability) they are *a bit* larger than those obtained by simulations (compare 26.08, 12.23, 27.88 vs. 24.71, 11.66, 26.46). This property is due to the extra term $\sigma > 0$ in (20).

Table 9. ARL for the distribution-free EWMA TBEA chart computed with and without the *continuousify* technique.

m	without continuousify			with continuousify ($\sigma = 0.125$)		
	$p_T = 0.3$ $p_X = 0.8$	$p_T = 0.2$ $p_X = 0.9$	$p_T = 0.1$ $p_X = 0.6$	$p_T = 0.3$ $p_X = 0.8$	$p_T = 0.2$ $p_X = 0.9$	$p_T = 0.1$ $p_X = 0.6$
100	32.96	12.18	40.16	26.08	12.23	27.87
120	18.57	10.89	18.55	26.08	12.23	27.87
140	28.88	11.91	31.56	26.08	12.23	27.87
160	20.31	11.08	20.81	26.08	12.23	27.87
180	28.36	11.82	31.36	26.08	12.23	27.87
200	24.47	11.55	26.42	26.08	12.23	27.87
220	17.97	10.05	18.39	26.08	12.23	27.87
240	27.98	11.88	31.36	26.08	12.23	27.87
260	57.68	16.41	77.81	26.08	12.23	27.87
280	21.17	11.41	21.75	26.08	12.23	27.87
300	26.75	11.75	29.94	26.08	12.23	27.88
320	21.69	11.43	22.43	26.08	12.23	27.88
340	26.07	11.93	27.39	26.08	12.23	27.88
360	16.68	10.43	16.5	26.08	12.23	27.88
380	29.2	13.09	29.83	26.08	12.23	27.88
400	20.33	11.02	20.7	26.08	12.23	27.88
Simu	24.71	11.66	26.46	26.09	12.23	27.87

4.3 Comparative studies

Since the *continuousify* technique is demonstrated to provide reliable ARL values, it is therefore possible to compute the optimal chart parameters (λ^*, K^*) for the upper-sided EWMA TBEA control chart minimizing the out-of-control $\text{ARL}(\lambda^*, K^*, \sigma, p_T, p_X)$ for $p_T \neq 0.5$ and $p_X \neq 0.5$ under the constraint $\text{ARL}(\lambda^*, K^*, \sigma, 0.5, 0.5) = \text{ARL}_0$, where ARL_0 is a predefined value for the in-control ARL. These optimal values are listed in Table 10 with the corresponding out-of-control (ARL, SDRL) values for $p_T \in \{0.1, 0.2, \dots, 0.4\}$ (only considering a decrease in T), $p_X \in \{0.5, 0.6, \dots, 0.9\}$ (only considering an increase in X), for four possible choices for $\sigma \in \{0.1, 0.125, 0.15, 0.2\}$ and assuming $\text{ARL}_0 = 370.4$. These values of (λ^*, K^*) can be freely be used by practitioners who need to optimally detect a specific shift in the times and/or in the amplitudes.

A comparison between the upper-sided EWMA TBEA control chart introduced in this Section and the three parametric TBEA control charts presented in Subsection 2.1 has been investigated in Wu et al. (2021) using the EARL_1 (instead of the EATS_1) for the following two scenarios

- Scenario #1: a Normal distribution for X with in-control mean $\mu_{X_0} = 10$ and standard-deviation $\sigma_{X_0} = 1$ and a gamma distribution for T with in-control mean $\mu_{T_0} = 10$ and standard-deviation $\sigma_{T_0} = 2$, i.e. $X \sim \text{Nor}(10, 1)$ and $T \sim \text{Gam}(25, 0.4)$.
- Scenario #2: a Normal distribution for X with in-control mean $\mu_{X_0} = 10$ and standard-deviation $\sigma_{X_0} = 2$ and a Weibull distribution for T with in-control mean $\mu_{T_0} = 10$ and standard-deviation $\sigma_{T_0} = 1$, i.e. $X \sim \text{Nor}(10, 2)$ and $T \sim \text{Wei}(12.1534, 10.4304)$.

Based on the results in Tables 3 and 4 in Wu et al. (2021), the conclusion is that no matter the scenario (#1 or #2) or the statistic considered $Z \in \{Z_1, Z_2, Z_3\}$, the out-of-control EARL_1 values obtained for the distribution-free upper-sided EWMA TBEA control chart are always smaller than the ones obtained for the parametric Shewhart control charts introduced in Subsection 2.1, thus showing the advantage of using the proposed distribution-free control chart in situations where the distributions for T and X are unknown.

4.4 Illustrative example

We consider here the same illustrative example as the one already presented in Section 2.6 concerning the days T_i between fires in forests of the french region “Provence - Alpes - Côte D’Azur” and their amplitudes X_i (burned surface in $ha = 10000m^2$). In order to compute the control limit UCL of the distribution-free upper-sided EWMA TBEA chart, the following values have been fixed: $p_T = 0.3$, $p_X = 0.7$, $\sigma = 0.125$ and $\text{ARL}_0 = 370.4$. The corresponding optimal values for λ and K are found to be $\lambda^* = 0.07$ and

Table 10. Optimal values for (λ^*, K^*) with the corresponding out-of-control values of (ARL,SDRL) for $p_T \in \{0.1, 0.2, \dots, 0.4\}$, $p_X \in \{0.5, 0.6, \dots, 0.9\}$ and $\sigma \in \{0.1, 0.125, 0.15, 0.2\}$

$\sigma = 0.1$					
p_T	0.5	0.6	p_X 0.7	0.8	0.9
0.5	(-, -)				
	(370.40, -)				
0.4	(0.010, 1.773)	(0.025, 2.174)			
	(105.66, 74.04)	(50.77, 32.32)			
0.3	(0.025, 2.174)	(0.045, 2.387)	(0.070, 2.515)		
	(51.54, 32.55)	(30.55, 18.04)	(20.50, 11.38)		
0.2	(0.040, 2.348)	(0.070, 2.515)	(0.100, 2.591)	(0.145, 2.639)	
	(31.30, 17.51)	(20.74, 11.40)	(14.85, 7.67)	(11.19, 5.55)	
0.1	(0.060, 2.474)	(0.090, 2.571)	(0.135, 2.634)	(0.180, 2.645)	(0.240, 2.627)
	(21.40, 10.76)	(15.16, 7.37)	(11.32, 5.40)	(8.76, 3.84)	(6.99, 2.74)

$\sigma = 0.125$					
p_T	0.5	0.6	p_X 0.7	0.8	0.9
0.5	(-, -)				
	(370.40, -)				
0.4	(0.010, 1.774)	(0.025, 2.174)			
	(106.19, 74.55)	(51.11, 32.63)			
0.3	(0.025, 2.174)	(0.045, 2.387)	(0.070, 2.515)		
	(51.88, 32.87)	(30.79, 18.25)	(20.68, 11.53)		
0.2	(0.040, 2.348)	(0.065, 2.496)	(0.100, 2.592)	(0.140, 2.638)	
	(31.53, 17.72)	(20.91, 11.27)	(14.99, 7.80)	(11.32, 5.57)	
0.1	(0.060, 2.474)	(0.090, 2.572)	(0.135, 2.634)	(0.175, 2.648)	(0.225, 2.639)
	(21.57, 10.92)	(15.30, 7.50)	(11.44, 5.49)	(8.88, 3.89)	(7.10, 2.75)

$\sigma = 0.15$					
p_T	0.5	0.6	p_X 0.7	0.8	0.9
0.5	(-, -)				
	(370.40, -)				
0.4	(0.010, 1.775)	(0.025, 2.175)			
	(106.83, 75.16)	(51.53, 33.01)			
0.3	(0.025, 2.175)	(0.045, 2.387)	(0.070, 2.515)		
	(52.30, 33.26)	(31.08, 18.51)	(20.90, 11.71)		
0.2	(0.040, 2.348)	(0.065, 2.496)	(0.095, 2.584)	(0.135, 2.636)	
	(31.82, 17.97)	(21.13, 11.45)	(15.17, 7.81)	(11.47, 5.61)	
0.1	(0.055, 2.449)	(0.090, 2.573)	(0.130, 2.632)	(0.170, 2.651)	(0.215, 2.646)
	(21.79, 10.76)	(15.47, 7.64)	(11.59, 5.53)	(9.02, 3.96)	(7.23, 2.80)

$\sigma = 0.2$					
p_T	0.5	0.6	p_X 0.7	0.8	0.9
0.5	(-, -)				
	(370.40, -)				
0.4	(0.010, 1.777)	(0.025, 2.176)			
	(108.43, 76.68)	(52.57, 33.96)			
0.3	(0.020, 2.085)	(0.045, 2.387)	(0.065, 2.496)		
	(53.33, 32.51)	(31.81, 19.15)	(21.44, 11.90)		
0.2	(0.040, 2.348)	(0.065, 2.496)	(0.090, 2.574)	(0.125, 2.630)	
	(32.53, 18.61)	(21.66, 11.92)	(15.60, 8.01)	(11.84, 5.74)	
0.1	(0.055, 2.449)	(0.085, 2.562)	(0.120, 2.624)	(0.155, 2.652)	(0.195, 2.658)
	(22.31, 11.21)	(15.89, 7.83)	(11.95, 5.64)	(9.34, 4.06)	(7.53, 2.91)

$K^* = 2.515$ (see results in Table 10) and the upper control limit UCL is equal to

$$\text{UCL} = 2.515 \times \sqrt{\frac{0.07 \times (0.125^2 + 0.5)}{2 - 0.07}} = 0.344.$$

The in-control median values for T and X have been estimated from the Phase 1 data set and they are equal to $\theta_{T_0} = 3$ days and $\theta_{X_0} = 5.3$ ha. These values are used to compute the values ST_i , SX_i , S_i and S_i^* in Table 11. As it can be noticed, some dates are such that $T_i - \theta_{T_0} = 0$. Of course, this situation is not supposed to happen as the times T_i are supposed to be continuous random variables but, due to the measurement scale (days), this situation actually happens. When this situation occurs, a possible simple strategy consists in assigning $ST_i = 0$ (instead of -1 or $+1$). For this reason, in Table 11, some values of $S_i = s = \pm 0.5$ and the corresponding values for S_i^* are obtained by randomly generating a $\text{Nor}(s, \sigma)$ random variable, as it is already the case for values $s \in \{-1, 0, +1\}$. For instance, in Table 11, when $D_i = 70$ we have $S_i = 0.5$ and the corresponding value for S_i^* has been randomly generated from a $\text{Nor}(0.5, 0.125)$ distribution ($S_i^* = 0.552$). The values Z_i^* have been computed using eqn.(19), for both Phase 1 and 2 data sets, recorded in Table 11 and plotted in Figure 7 along with the distribution-free EWMA TBEA upper control limit $\text{UCL} = 0.344$. If the distribution-free upper-sided EWMA TBEA chart does not detect any out-of-control situations during the Phase 1 (validating the in-control state of this phase), it nevertheless detects several out-of-control situations during the period mid-June 2017 – end of September 2017, (see also the bold values in Table 11), confirming that a decrease in the time between fires occurred with a concurrent increase in the amplitude of these fires. This conclusion is consistent with the one obtained in Section 2.6 in which a parametric approach assuming a lognormal distribution for both T_i and X_i was used.

5 Conclusions

The three contributive Sections of this Chapter have clearly demonstrated that efficient solutions do exist when the aim is to simultaneously monitor the time interval T of an event E as well as its amplitude X . These solutions can be either parametric, for independent or dependent situations, and they can also be distribution-free if there is no a priori knowledge about the distributions associated with T and X .

In our opinion, future researches on the monitoring of TBEA data can be undertaken toward several directions:

- In the proposed parametric approaches, the estimation of the parameters (for the distributions or the Copulas) is not taken into account at all in the design and evaluation of the TBEA control charts. The impact of the

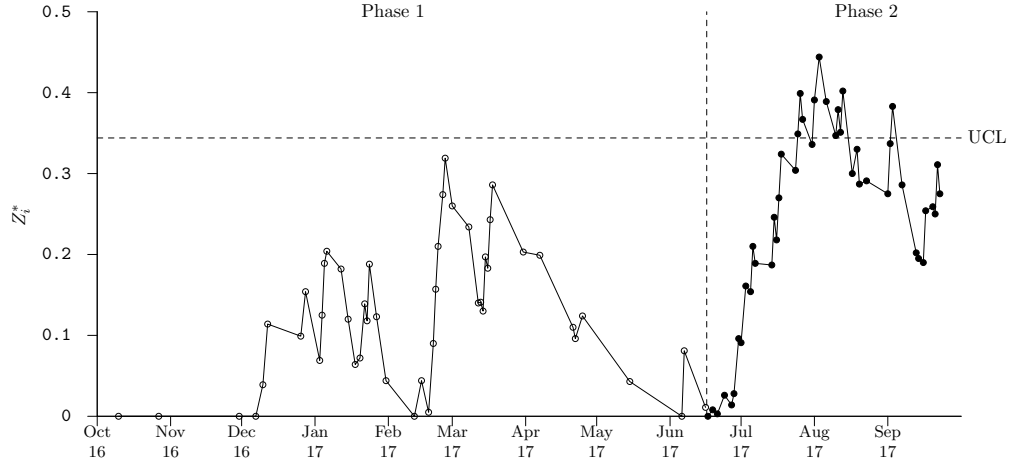


Fig. 7. Distribution-free EWMA TBEA chart with statistic Z_i^* corresponding to the data set in Table 11

parameter estimation is known to strongly influence the efficiency of any control chart and, therefore, researches on this topic should be done.

- Measures like times or amplitudes are obviously subject to measurement errors. These kinds of error are also known to negatively impact the efficiency of any control chart. Researches investigating the impact of the measurement errors on T and/or X should also be undertaken in order to evaluate how much they actually impact the performance of parametric TBEA control charts.
- Instead of considering X as univariate random variable, it could be considered in some cases as a p -variate random vector $\mathbf{X} = (X_1, \dots, X_p)$ where each X_k is the amplitude of a specific characteristic and the goal would be to simultaneously monitor (T, \mathbf{X}) . For instance, in the forest fires example, the amplitude could be considered as a bivariate vector $\mathbf{X} = (X_1, X_2)$ where X_1 would be the burned surface and X_2 would be the cost related to the fires.
- Often, historical data availability in monitoring of adverse events is limited to a few records. Thus, knowledge about the frequency distribution of these events is too restricted to fit a reliable statistical model. With these scenarios, there is room for approaching the monitoring problem with distribution-free approaches, which, therefore, deserve a lot of attention by researchers.

References

- M. Abid, H.Z. Nazir, M. Riaz, and Z. Lin. Use of Ranked Set Sampling in Non-parametric Control Charts. *Journal of the Chinese Institute of Engineers*, 39(5):627–636, 2016.
- M. Abid, H.Z. Nazir, M. Riaz, and Z. Lin. An Efficient Nonparametric EWMA Wilcoxon Signed-Rank Chart for Monitoring Location. *Quality and Reliability Engineering International*, 33(3):669–685, 2017a.
- M. Abid, H.Z. Nazir, M. Riaz, and Z. Lin. Investigating the Impact of Ranked Set Sampling in Nonparametric CUSUM Control Charts. *Quality and Reliability Engineering International*, 33(1):203–214, 2017b.
- M. Abid, H.Z. Nazir, M. Tahir, and M. Riaz. On Designing a New Cumulative Sum Wilcoxon Signed Rank Chart for Monitoring Process Location. *PloS one*, 13(4):e0195762, 2018.
- M.M Ali, N.N. Mikhail, and M.S. Haq. A Class of Bivariate Distributions including the Bivariate Logistic. *Journal of Multivariate Analysis*, 8(3):405–412, 1978.
- S. Ali and A. Pievatolo. Time and Magnitude Monitoring based on the Renewal Reward Process. *Reliability Engineering & System Safety*, 179:97–107, 2018.
- J.C. Benneyan. Number-Between g -Type Statistical Quality Control Charts for Monitoring Adverse Events. *Health care management science*, 4(4):305–318, 2001.
- C.M. Borrór, J.B. Keats, and D.C. Montgomery. Robustness of the Time Between Events CUSUM. *International Journal of Production Research*, 41(15):3435–3444, 2003.
- E.C. Brechmann and U. Schepsmeier. Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine. *Journal of Statistical software*, 52(3):1–27, 2013.
- D. Brook and D.A. Evans. An Approach to the Probability Distribution of CUSUM Run Length. *Biometrika*, 59(3):539–549, 1972.
- P. Busababodhin and P. Amphanthong. Copula Modelling for Multivariate Statistical Process Control: A Review. *Communications for Statistical Applications and Methods*, 23(6):497–515, 2016.
- T. Calvin. Quality Control Techniques for “Zero Defects”. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, 6(3):323–328, 1983.
- P. Castagliola, K.P. Tran, G. Celano, A.C. Rakitzis, and P.E. Maravelakis. An EWMA-type Sign Chart with Exact Run Length Properties. *Journal of Quality Technology*, 51(1):51–63, 2019.
- G. Celano, P. Castagliola, S. Chakraborti, and G. Nenes. The Performance of the Shewhart Sign Control Chart for Finite Horizon Processes. *International Journal of Advanced Manufacturing Technology*, 84(5):1497–1512, 2016.

- S. Chakraborti and M.A. Graham. *Nonparametric Statistical Process Control*. Wiley, 2019a.
- S. Chakraborti and M.A. Graham. Nonparametric (Distribution-Free) Control Charts: an Updated Overview and Some Results. *Quality Engineering*, 31(4):523–544, 2019b.
- C.S. Cheng and P.W. Chen. An ARL-Unbiased Design of Time-Between-Events Control Charts with Runs Rules. *Journal of Statistical Computation and Simulation*, 81(7):857–871, 2011.
- Y. Cheng and A. Mukherjee. One Hotelling T^2 Chart Based on Transformed Data for Simultaneous Monitoring the Frequency and Magnitude of an Event. In *2014 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 764–768, 2014.
- Y. Cheng, A. Mukherjee, and M. Xie. Simultaneously Monitoring Frequency and Magnitude of Events Based on Bivariate Gamma Distribution. *Journal of Statistical Computation and Simulation*, 87(9):1723–1741, 2017.
- D.G. Clayton. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, 65(1):141–151, 1978.
- P. Dokouhaki and R. Noorossana. A Copula Markov CUSUM Chart for Monitoring the Bivariate Auto-Correlated Binary Observations. *Quality and Reliability Engineering International*, 29(6):911–919, 2013.
- Y.Y. Fang, M.B.C Khoo, S.Y. Teh, and M. Xie. Monitoring of Time-Between-Events with a Generalized Group Runs Control Chart. *Quality and Reliability Engineering International*, 32(3):767–781, 2016.
- A.A. Fatahi, P. Dokouhak, and B.F. Moghaddam. A Bivariate Control Chart Based on Copula Function. In *Quality and Reliability (ICQR), 2011 IEEE International Conference on*, pages 292–296. IEEE, 2011.
- M.J. Frank. On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes mathematicae*, 19(1):194–226, 1979.
- F.F. Gan. Designs of One- and Two-Sided Exponential EWMA Charts. *Journal of Quality Technology*, 30(1):55, 1998.
- E.J. Gumbel. Distributions des Valeurs Extremes en Plusieurs Dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9:171–173, 1960.
- H. Joe. Parametric Families of Multivariate Distributions with Given Margins. *Journal of Multivariate Analysis*, 46(2):262–282, 1993.
- J.Y. Liu, M. Xie, T.N. Goh, and P.R. Sharma. A Comparative Study of Exponential Time Between Events Charts. *Quality Technology & Quantitative Management*, 3(3):347–359, 2006.
- J.M. Lucas. Counted Data CUSUM's. *Technometrics*, 27(2):129–144, 1985.
- P. Qiu. *Introduction to Statistical Process Control*. Chapman and Hall, Boca Raton(USA), 2014.
- L. Qu, Z. Wu, M.B.C. Khoo, and P. Castagliola. A CUSUM Scheme for Event Monitoring. *International Journal of Production Economics*, 145(1):268–280, 2013.

- L. Qu, Z. Wu, M.B.C. Khoo, and A. Rahim. Time-Between-Event Control Charts for Sampling Inspection. *Technometrics*, 56(3):336–346, 2014.
- L. Qu, S. He, M.B.C. Khoo, and P. Castagliola. A CUSUM Chart for Detecting the Intensity Ratio of Negative Events. *International Journal of Production Research*, 56(19):6553–6567, 2018.
- G. Radaelli. Planning Time-Between-Events Shewhart Control Charts. *Total Quality Management*, 9(1):133–140, 1998.
- D. Rahali, P. Castagliola, H. Taleb, and M.B.C. Khoo. Evaluation of Shewhart Time-Between-Events-and-Amplitude Control Charts for Several Distributions. *Quality Engineering*, 31(2):240–254, 2019. doi: 10.1080/08982112.2018.1479036.
- D. Rahali, P. Castagliola, H. Taleb, and M.B.C. Khoo. Evaluation of Shewhart Time-Between-Events-and-Amplitude Control Charts for Correlated Data. *Quality and Reliability Engineering International*, 37(1):219–241, 2021. doi: 10.1002/qre.2731.
- R.A. Sanusi, S.Y. Teh, and M.B.C. Khoo. Simultaneous Monitoring of Magnitude and Time-Between-Events Data with a Max-EWMA Control Chart. *Computers & Industrial Engineering*, 142:106378, 2020.
- M.S. Shafae, R.M. Dickinson, W.H. Woodall, and J.A. Camelio. Cumulative Sum Control Charts for Monitoring Weibull-Distributed Time Between Events. *Quality and Reliability Engineering International*, 31(5):839–849, 2015.
- M. Shamsuzzaman, M. Xie, T.N. Goh, and H.Y. Zhang. Integrated Control Chart System for Time-Between-Events Monitoring in a Multistage Manufacturing System. *The International Journal of Advanced Manufacturing Technology*, 40(3):373–381, 2009.
- A. Sklar. Fonctions de Répartition à n Dimensions et leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.
- S. Sukparungsee, S. Kuvattana, P. Busababodhin, and Y. Areepong. Bivariate Copulas on the Hotelling's T^2 Control Chart. *Communications in Statistics—Simulation and Computation*, 47(2):413–419, 2018.
- S. Vardeman and D.O. Ray. Average Run Lengths for CUSUM Schemes when Observations are Exponentially Distributed. *Technometrics*, 27(2):145–150, 1985.
- S. Wu, P. Castagliola, and G. Celano. A Distribution-Free EWMA Control Chart for Monitoring Time-Between-Events-and-Amplitude Data. *Journal of Applied Statistics*, 48(3):434–454, 2021. doi: 10.1080/02664763.2020.1729347.
- Z. Wu, J. Jiao, and H. Zhen. A Control Scheme for Monitoring the Frequency and Magnitude of an Event. *International Journal of Production Research*, 47(11):2887–2902, 2009.
- M. Xie, T.N. Goh, and P. Ranjan. Some Effective Control Chart Procedures for Reliability Monitoring. *Reliability Engineering & System Safety*, 77(2):143–150, 2002.

- C.W. Zhang, M. Xie, J.Y. Liu, and T.N. Goh. A Control Chart for the Gamma Distribution as a Model of Time Between Events. *International Journal of Production Research*, 45(23):5649–5666, 2007.
- H.Y. Zhang, M. Shamsuzzaman, M. Xie, and T.N. Goh. Design and Application of Exponential Chart for Monitoring Time-Between-Events Data under Random Process Shift. *The International Journal of Advanced Manufacturing Technology*, 57(9):849–857, 2011a.
- H.Y. Zhang, M. Xie, T.N. Goh, and M. Shamsuzzaman. Economic Design of Time-Between-Events Control Chart System. *Computers & Industrial Engineering*, 60(4):485–492, 2011b.