Research Article Environmetrics

Received: 2 March 2015, Revised: 2 June 2015, Accepted: 4 June 2015, Published online in Wiley Online Library: 3 August 2015

(wileyonlinelibrary.com) DOI: 10.1002/env.2349

A criterion for environmental assessment using Birnbaum–Saunders attribute control charts

Victor Leiva[a,b], Carolina Marchant[b,c], Fabrizio Ruggeri[d]* and Helton Saulo[e]

Assessing environmental risk is useful for preventing adverse effects on human health in highly polluted cities. We design a criterion for environmental monitoring based on an attribute control chart for the number of dangerous contaminant levels when the concentration to be monitored follows a Birnbaum–Saunders distribution. This distribution is being widely applied to environmental data. We provide a novel justification for its usage in environmental sciences. The control coefficient and the minimum inspection concentration for the designed criterion are determined to yield the specified in-control average run length, whereas the out-of-control case is obtained according to a shift in the target mean. A simulation study is conducted to evaluate the proposed criterion, which reports its performance to provide earlier alerts of out-of-control processes. An application with real-world environmental data is carried out to validate its coherence with what is reported by the health authority. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: average run length; inspection point; moment estimation; Monte Carlo simulation; np control chart; R software

1. INTRODUCTION

Control charts are widely used for quality monitoring in industry; see Montgomery (2005) and Leiva *et al.* (2014). These charts alert when a process is out of control, and then an action must be taken for being in control again. The use of such charts has been expanded to other areas, including environmental and health risk assessment; see Lund and Seymour (1999), Manly and Mackenzie (2000), Grigg and Farewell (2004), Ferreira-Baptista and De Miguel (2005), Woodall (2006), Morrison (2008), and Saulo *et al.* (2015). The power or effectiveness of a control chart is usually measured by the average run length (ARL), which is the average number of inspected samples required to advert an out-of-control condition after it has occurred. On the one hand, practitioners of control charts desire to advert an out-of-control condition as fast as possible, when a process is out of control, that is, the out-of-control ARL (ARL1) must be as small as possible. On the other hand, when the process is in control, these practitioners desire to have a fewest possible number of false alarms, that is, to have a large in-control ARL (ARL0). Control charts are classified by variables or attributes, depending on whether the data used to monitor the quality characteristic are generated by measurements (continuous data) or by attributes (qualitative data), as presence (non-conforming) or absence (conforming) of an event. Thus, for this monitoring, we can use detection tools for variables, such as x-bar (sample mean) or R (range) charts, or for attributes, such as p (proportion of non-conforming) or np (number of non-conforming) charts; see Epprecht *et al.* (2003), Hsu (2004), Wu and Wang (2007), Rodrigues *et al.* (2011), and Joekes and Barbosa (2013).

The first component to develop the criterion stated in the paper title is the np chart used for subgroups with constant size; see Montgomery (2005, pp. 282). International guidelines often utilize exceedance probabilities of pollutant concentrations to establish environmental alerts. Episodes of extreme air contamination need to be monitored, alerted and corrected for protection of human health. An increment of respiratory diseases may be caused, for example, by high concentrations of particulate matter (PM); see Cox (2000), Clyde (2000), Reich *et al.* (2008), Almaskut *et al.* (2012), Marchant *et al.* (2013), and Alencar and Santos (2014). However, monitoring airborne PM pollution is not an easy task, because PM is a complex mixture of substances in solid or liquid states with different properties and factors; see more details in Querol *et al.* (2004). The discussion on accurate properties that need to be monitored for health protection is still an open subject; see WHO (2003, 2006). To facilitate the monitoring, air quality standards have been defined in terms of the PM mass concentration with a diameter below 2.5 (PM2.5) or 10 (PM10) μm. The chemical mass balance has been often used to detect PM sources. In our study, we model PM10 as a dilution and concentration of air masses, without taking into account factors of composition, chemical reactivity and production-degradation

* *Correspondence to: Fabrizio Ruggeri, CNR IMATI, Via Bassini 15, 20133 Milano, Italy. E-mail: fabrizio@mi.imati.cnr.it*

a *Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Viña del Mar, Chile*

b *Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile*

c *Department of Statistics, Universidade Federal de Pernambuco, Recife, Brazil*

d *CNR-IMATI, Milano, Italy*

e *Institute of Mathematics and Statistics, Universidade Federal de Goiás, Goiânia, Brazil*

equilibrium; see Vione *et al.* (2006) and Kessler *et al.* (2010). We follow the existing Chilean air quality guideline, considering PM10 levels which can be harmful for human health according to this guideline (PM10 levels greater than 150 μg/normalized cubic meters –m^{3}N–) and not its composition; see Marchant *et al.* (2013). One can monitor the number of times that dangerous PM10 levels are greater than the regulation in each air quality station. Therefore, environmental risk assessment is important and can be conducted by np charts.

Probability models employed to describe air contaminant concentrations have usually positive asymmetry or skewness (non-normality). Then, the normal or Gaussian model is not suitable for computing the mentioned exceedance probabilities. Several practitioners have used the log-normal (LN) model for describing air contaminant concentrations, mainly due to its physical arguments (Ott, 1990) and its relationship with the normal model. Among others, beta, exponential, extreme values, gamma, inverse Gaussian, Johnson bounded system, log-logistic, Pearson, and Weibull distributions have been also used for modeling concentrations of pollutants, although without the theoretical arguments that the LN model has; see Marchant *et al.* (2013).

The second component to develop our criterion is a distribution with positive skewness that is being widely studied. This is named the Birnbaum–Saunders (BS) model, which has good properties and a close relation with the normal model; see Birnbaum and Saunders (1969). It has been widely applied to diverse fields, including environmental and earth sciences; see, for example, Leiva *et al.* (2008, 2009, 2012, 2015), Ferreira *et al.* (2012), Marchant *et al.* (2013), and Saulo *et al.* (2013, 2015). As in the LN model, and unlike the other mentioned probability models, the physical arguments of the BS model allows us to postulate it as a candidate to describe air contaminant concentrations; see details in Section 2. In addition, the BS model has properties similar to the LN model. Despite the wide use and development of methodologies based on the BS model, there are few studies on quality monitoring tools and nothing on attribute control charts based on it; see Balakrishnan *et al.* (2007), Lio *et al.* (2010), Aslam *et al.* (2011), Leiva *et al.* (2014), and Saulo *et al.* (2015).

The objective of this paper is to propose a criterion based on np charts to assess environmental risk when the pollutant concentration follows a BS model. We formalize the usage of this model for environmental contaminant data. We illustrate our criterion with air contamination data from the city of Santiago, Chile, which is one of the most polluted cities around the world; see Leiva *et al.* (2008), Vilca *et al.* (2010), Marchant *et al.* (2013), and the section of application of the present paper for more details. The criterion allows us to assess environmental risk and could provide information for preventing adverse effects on human health of the population of Santiago, Chile. We show the coherence between our criterion and what is happening in a real-world situation dictated by a Chilean authority. Note that although the criterion is applied in this work to environmental monitoring, it can be used in a more general setting, for example, for monitoring production processes associated with items subject to failure.

In Section 2, we justify the usage of the BS distribution for air contamination data. In Section 3, we propose a criterion for environmental assessment using BS attribute control charts. In Section 4, we provide some numerical results using Monte Carlo simulation, reporting ARLs and discussing a computational implementation in the R software; see www.r-project.org. In Section 5, we validate the proposed criterion with real-world environmental data collected by the Chilean official environmental authority. In Section 6, we present some conclusions and possible future works. In the Appendix, some mathematical and statistical features of the BS distribution are provided.

2. JUSTIFYING THE BS MODEL AS A CONTAMINANT DISTRIBUTION

Assuming a BS model to describe pollutant concentration data based on an empirical fitting can be a reasonable argument. However, this argument may be strengthened if we justify why the BS model might be suitable for such a description. On the one hand, Ott (1990) provided a physical explanation of the LN model as a pollutant concentration distribution, based on the proportionate-effect model; see Aitchison & Brown (1973, p. 22). On the other hand, Desmond (1985) discussed an incorrectness in the usage of Cramér's biological model, linked to the proportionate-effect model, to justify the LN model as a life distribution, which rather conducts to a BS model; see Cramér (1946, p. 219). Relating these works, we justify the BS model as an environmental contaminant distribution.

The Cramér's biological model (or proportionate-effect model) can be expressed as

$$X_{j+1} = X_j + Y_{j+1}\, g(X_j), \quad j = 0, 1, \ldots \tag{1}$$

where Y_{j+1} is a random variable (RV) corresponding to the magnitude of the $(j+1)$th impulse and X_{j+1} the accumulated total amount after application of this impulse. From (1), the LN model is obtained when $g(X) = X$, whereas the BS model is obtained with $g(X) = 1$. The relationship defined in (1) was originally proposed in a biological context, whereas Desmond (1985) put it in a context of fatigue life. We discuss relationship (1) in a context of environmental contamination.

The appropriateness of the model given in (1) in the fatigue life context is supported by Frost and Dugdal's differential-equation model used in engineering (Desmond, 1985) defined as

$$\frac{\mathrm{d}e}{\mathrm{d}m} = \frac{S^3 e}{c} = c_1 f(\Delta K) \tag{2}$$

where e is the crack size, m is the cycle number, S is the applied stress, c is a material constant, ΔK is the range of the stress intensity factor, $f(\cdot)$ is an empirically determined function, and c_1 is an experimental constant. From fracture mechanics, one can relate ΔK in (2) to the crack size e by means of

$$\Delta K = \alpha \Delta S e^{1/2} \tag{3}$$

where α is a material indicator and ΔS the applied stress amplitude. Based on (2) and (3) and using the approximation $f(\alpha \Delta S e^{1/2}) \approx c_2 + c_3 h(e)$, with $h(\cdot)$ being a function of the crack size, we have

$$\frac{\mathrm{d}e}{\mathrm{d}m} \approx c_0 + c_4\, h(e) \tag{4}$$

where c_0 and c_4 contain all stress and material factors included in c_1, c_2, and c_3, with c_0 being considered as an initial value. Equation (4) is treated by engineers as if it were a deterministic relation, but considering c_0, c_4, and e as RVs is closer to reality. Note the similarity between differential-equation model (4) and proportionate-effect model (1).

Now, by retaking (1), applying the central limit theorem, and considering the increment $\Delta X_j = X_{j+1} - X_j$ in the $(j+1)$th impulse is small enough such that summation can be changed by integration, we obtain that

$$\sum_{j=1}^{m} Y_j = \sum_{j=1}^{m} \frac{\Delta X_j}{g(X_j)} \approx \int_{X_0}^{X_m} \frac{\mathrm{d}x}{g(x)} \tag{5}$$

follows approximately a normal distribution, where X_0 is the initial crack size. To obtain the fatigue life distribution, suppose that the mean of Y_j is η and its variance ϱ^2, which generalizes Assumption 2 of Birnbaum and Saunders (1969), conducting to

$$I(X_t) = \int_{X_0}^{X_t} \frac{1}{g(x)}\,\mathrm{d}x \sim \mathrm{N}(t\eta, t\varrho^2) \tag{6}$$

where X_t is the crack size at time t. Assume now that $X_c > X_0$ is the critical crack length at which failure occurs, where X_c and X_0 are non-random. Then, $T = \inf\{t\colon X_t > X_c\}$ is the time to fatigue failure. Thus, from (6) and using the equivalent events $\{T \leqslant t\}$ and $\{X_t > X_c\}$, it follows that the cumulative distribution function (CDF) of T at t is given by

$$F_T(t) = \Phi\left(\frac{t\eta - I(X_c)}{\sqrt{t}\varrho}\right) \tag{7}$$

where $\Phi(\cdot)$ is the standard normal or N(0, 1) CDF. From (7), note that the model given in (1) leads to life distributions in the BS family independently of the form of $g(X)$.

The choice of the function $g(X)$ determines the dependence of the rate of crack extension on previous crack size. Empirical evidence suggests that a power function for $g(X)$ is reasonable, that is, $g(X) = X^{\delta}$, where δ is a material indicator, for $\delta \geqslant 0$, but $\delta \neq 1$; see Desmond (1985). Thus, from (6), we have

$$I(X_t) = \int_{X_0}^{X_t} \frac{\mathrm{d}x}{g(x)} = \frac{1}{\delta - 1}\left[\frac{1}{X_0^{\delta-1}} - \frac{1}{X_t^{\delta-1}}\right] = \frac{X_0^{1-\delta} - X_t^{1-\delta}}{\delta - 1} \sim \mathrm{N}(t\eta, t\varrho^2) \tag{8}$$

From (8) and by symmetry of the normal distribution, $X_t^{1-\delta} \sim \mathrm{N}(X_0^{1-\delta} + [1-\delta]t\eta, [1-\delta]^2 t\varrho^2)$. Then, the CDF of T at t is

$$F_T(t) = \begin{cases} \Phi\left(\frac{X_c^{1-\delta} - X_0^{1-\delta} + [\delta-1]t\eta}{[\delta-1]\sqrt{t}\varrho}\right), & \text{if } \delta > 1; \\ \Phi\left(\frac{X_0^{1-\delta} - X_c^{1-\delta} + [1-\delta]t\eta}{[1-\delta]\sqrt{t}\varrho}\right), & \text{if } \delta < 1 \end{cases} \tag{9}$$

From (9), note that the BS model results for $\delta = 0$. However, the case $\delta = 1$, corresponding to the LN model as a particular case when $g(X) = X$ in the proportionate-effect model given in (1), cannot be obtained from (9). Therefore, the distribution of T with CDF given in (9) is of BS type and not of LN type; see Desmond (1985).

The parallel between biological and fatigue life models discussed in the previous text could be put in a context of contamination, making a parallel between proportionate-effect and pollutant concentration models. This last parallel is also valid as an argument for modeling data of outdoor and indoor atmospheric quality, of water quality in lakes, groundwater and soils, and of trace metals in human tissue, blood, and feces; see Ott (1990) and references therein.

To understand the model given by (1) in a contamination context, consider the following. We first assume a deterministic setting. Suppose that an environment has a pollutant concentration X_0 and an air mass V_0 at an initial instant time $t = 0$. Then, $q_0 = X_0\, V_0$ is the total quantity of pollutant in the environment. Assuming that no chemical reactions exist, conservation of matter needs that the total pollutant amount present after dilution must be equivalent to the total pollutant quantity before dilution. Thus, if V_1 is the volume of the mixture after first dilution and X_1 the pollutant concentration in that diluted mixture, then

$$q_0 = X_0\, V_0 = X_1\, V_1 \tag{10}$$

Solving the equation given in (10) for X_1, the concentration X_1 is inversely proportional to the ratio of the final volume to the initial volume expressed as $X_1 = [V_0/V_1]\, X_0$. Now, assume V_0/V_1 is represented by a dimensionless dilution factor D_1 that is treated as an RV, and therefore, $X_1 = D_1\, X_0$ is also a RV, with X_0 being non-random. This random setting is closer to reality because contaminants vary randomly through the time due to climatological, geographical, and human activity and topographical elements, among others. If D_j stands for the corresponding factor in the jth dilution, then the concentration X_2 in the second dilution is given by $X_2 = D_2\, X_1 = D_2\, D_1\, X_0$. In general, for m denoting the number of successive dilutions in a random dilution process, the final concentration is

$$X_m = X_0 \prod_{j=1}^{m} D_j \tag{11}$$

By taking logarithm in (11), we have

$$\sum_{j=1}^{m} \log(D_j) = \log(X_n) - \log(n_0) = \int_{X_0}^{X_m} \frac{\mathrm{d}x}{x} \tag{12}$$

Alternatively, (12) can be represented by a differential equation of first order as usual in the mass balance model, after permitting the change rate of the concentration to be infinitely small. Note the similarity between equations given in (5) and (12) when $g(x) = x$, which, as mentioned, corresponds to an LN model. However, one could assume a more general expression for (12) as

$$\sum_{j=1}^{m} Y_j \approx \int_{X_0}^{X_m} \frac{\mathrm{d}x}{g(x)} \tag{13}$$

which should conduct to a BS model under certain conditions, with (13) being analogous to (5). This allows a statistical distribution to be considered for modeling pollutant concentration data, which will hardly be symmetrical, but usually positively skewed; see Ott (1990).

3. A CRITERION FOR ENVIRONMENTAL ASSESSMENT USING BS NP CHARTS

An np chart is an adaptation of the control chart for non-conforming fraction (p) when samples of equal size (n) are taken from the process. The np chart is based on the binomial distribution as detailed in the succeeding text. In environmental assessment, one could be concerned about the RV corresponding to the number (D) of times that the concentration of an atmospheric contaminant (T) exceeds a determinate value (t) established by some official regulation, given an exceedance probability (p). Here, p can be computed by means of the continuous distribution of the RV T associated with the mentioned concentration as $p = \mathrm{P}(T > t) = 1 - F_T(t)$, where $F_T(\cdot)$ is the CDF of T. Thus, D follows a binomial distribution with parameters n and p and then

$$\mathrm{P}(D = d) = \binom{n}{d} p^d [1-p]^{n-d}, \quad d = 0, 1, \ldots, n \tag{14}$$

$\mathrm{E}[D] = np$, and $\mathrm{Var}[D] = np[1-p]$. Based on (14), we propose an np chart with lower control limit (LCL), central line (CL) and upper control limit (UCL) given by

$$\mathrm{LCL} = \max\left\{0, np_0 - k\sqrt{np_0[1-p_0]}\right\}, \quad \mathrm{CL} = np_0, \quad \mathrm{UCL} = np_0 + k\sqrt{np_0[1-p_0]} \tag{15}$$

where k is a control coefficient, with $k = 2$ indicating a warning level and $k = 3$ a dangerous level, p_0 is the non-conforming fraction corresponding to a target mean μ_0 of the RV T associated with the atmospheric pollutant, when the contamination is in control, and n is the size of the subgroup. Note that the non-conforming fraction is the probability that the atmospheric contaminant exceeds a dangerous concentration (t_0), and therefore, this probability is $\mathrm{P}(T > t_0) = 1 - F_T(t_0)$. We reparameterize the BS distribution from (b, σ) to (b, μ), that is, switching from the median σ in the original BS parameterization to its mean given by $\mu = \sigma[2 + b^2]/2$. We consider t_0 as proportional to μ_0, that is, $t_0 = a\mu_0$, relating them to establish our monitoring criterion, where $a > 0$ is a proportionality constant. Note that the target mean μ_0 and dangerous concentration t_0 can be taken from an official air quality guideline. Then, the BS CDF (see equation (A.2) in Appendix) can be reparameterized in terms of its mean and expressed in function of t_0 and a as

$$F_T(t_0; b, \mu) = \Phi\left(\frac{1}{b}\xi\left(\frac{a[1 + b^2/2]}{\mu/\mu_0}\right)\right) \tag{16}$$

Thus, when a monitoring process is in control ($\mu = \mu_0$) for a pollutant concentration following a BS distribution, from (16), the non-conforming fraction is given by

$$p_0 = 1 - F_T(t_0; b, a) = \Phi\left(-\frac{1}{b}\xi\left(a[1 + b^2/2]\right)\right) \tag{17}$$

Note that the specification of the point t_0 is equivalent to specify the inspection point constant $a > 0$, because $t_0 = a\mu_0$, with μ_0 being the target mean, which is assumed to be known. A criterion for environmental assessment using an np chart when the atmospheric contaminant $T \sim \mathrm{BS}(b, \mu)$ is given by Algorithm 1.

Remark Note that the monitoring criterion using the np control chart based on the BS distribution, detailed in Algorithm 1, is valid in a general setting. For example, it can also be used for monitoring production processes associated with items subject to failure. In the case of environmental monitoring, for which our criterion was proposed, the following nomenclature and adaptations must be used in Algorithm 1: for Step 1, N subgroups could be, for example, N days, and the sample size n may be formed by data representing, for example, each hour of the day; for Step 2, data might correspond to PM10 concentrations (in $\mu\mathrm{g/m^3N}$), which must be measured for each subgroup at one monitoring station; for Step 4, d must be the number of times that PM10 concentrations (t_i) exceed the critical value t_0, which must

Algorithm 1 np control chart based on the BS distribution

1: Take N subgroups of size n.
2: Collect n data $t_1, \ldots, t_n$ of the RV of interest T for each subgroup.
3: Carry out an autocorrelation study for data collected in Step 2 to detect possible seasonal and/or serial dependence. If any dependence is detected, it must be removed using suitable techniques and then to continue with Step 4.
4: Fix the target mean μ_0, the inspection constant a, and the control coefficient k.
5: Count in each subgroup of n data the number d of times that t_i exceeds $t_0 = a\mu_0$, for $i = 1, \ldots, n$.
6: Compute LCL $= \max\{0, n\widehat{p}_0 - k\sqrt{n\widehat{p}_0[1-\widehat{p}_0]}\}$ and UCL $= n\widehat{p}_0 + k\sqrt{n\widehat{p}_0[1-\widehat{p}_0]}$, which are obtained from (15), where $\widehat{p}_0 = \Phi(-[1/\widehat{b}]\xi(a[1+\widehat{b}^2/2]))$ is given as in (17), with $\widehat{b} = \sqrt{2[\sqrt{s/r}-1]}$ being the modified moment (MM) estimate of b given in (A.5) (see Appendix), for $s = [1/n]\sum_{i=1}^{n} t_i$ and $r = \left[\{1/n\}\sum_{i=1}^{n}\{1/t_i\}\right]^{-1}$.
7: Declare the process as out of control if $d \geqslant$ UCL or $d \leqslant$ LCL, or as in control if LCL $\leqslant d \leqslant$ UCL.

be proportional to the target mean μ_0, both of them, as mentioned, taken from an official air quality guideline; for Step 5, only the UCL must be considered, with the corresponding LCL being equal to zero, declaring the contamination level as dangerous (out of control) in one monitoring station if $d \geqslant$ UCL or, otherwise, as non-dangerous (in control).

Let us consider a shift in the process mean and assume that the new process (shifted) mean becomes $\mu_1 = l\mu_0$, for a shift constant $l > 0$. We consider that the value of the shift constant l is greater than one, because we are concerned with the case where the mean of the characteristic to be monitored becomes greater than the target mean. Note that μ_0 and μ_1 are different means corresponding to in-control and out-of-control processes, respectively, but both of them are means of the BS distribution. Hence, the non-conforming fraction corresponding to the new mean of the RV T is obtained from (16) as

$$p_1 = 1 - F_T(t_0; b, l) = \Phi\left(-\frac{1}{b}\xi\left(\frac{a}{l}\left[1 + b^2/2\right]\right)\right) \tag{18}$$

As mentioned, in general, we say a process is in control if LCL $\leqslant D \leqslant$ UCL. Thus, when the process is actually in control, the probability to be in control is given by

$$P_{\text{in}}^0 = \text{P}(\text{LCL} \leqslant D \leqslant \text{UCL}|p_0) = \sum_{d=\lfloor \text{LCL} \rfloor + 1}^{\lfloor \text{UCL} \rfloor} \binom{n}{d} p_0^d [1-p_0]^{n-d} \tag{19}$$

whereas if the process mean has shifted to the new mean μ_1, probability given in (19) is expressed as

$$P_{\text{in}}^1 = \text{P}(\text{LCL} \leqslant D \leqslant \text{UCL}|p_1) = \sum_{d=\lfloor \text{LCL} \rfloor + 1}^{\lfloor \text{UCL} \rfloor} \binom{n}{d} p_1^d [1-p_1]^{n-d} \tag{20}$$

where $\lfloor x \rfloor$ denotes the integer part of the number x. Efficiency of the proposed criterion can be evaluated by using ARL0 and ARL1. As mentioned earlier, ARL0 is defined as the expected number of observations taken from an in-control state until the chart falsely adverts an out-of-control. ARL0 is regarded as acceptable if it is large enough to keep the level of false alarms at a reasonable value. ARL1 is defined as the expected number of observations taken from an out-of-control state until the chart correctly adverts an out-of-control. Ideally, the ARL1 value should be as small as possible. In-control and out-of-control ARLs are, respectively, given by

$$\text{ARL0} = \frac{1}{1 - P_{\text{in}}^0} \quad \text{and} \quad \text{ARL1} = \frac{1}{1 - P_{\text{in}}^1}$$

where P_{in}^0 and P_{in}^1 are given in (19) and (20), respectively. We select the control coefficient k such that ARL0 is close to a specified ARL denoted by r_0. Then, with the selected value of k, we can obtain the value of ARL1 for the shift constant l given in (18).

4. SIMULATION STUDY

In order to implement the numerical applications of this work, we use the non-commercial and open-source software `R`, which can be obtained at no cost from CRAN.r-project.org; see R Team (2014). The statistical software `R` is currently very popular in the international scientific community. A computational framework for analyzing data using the criterion proposed in this article is being developed by the authors in an `R` package, whose "in progress" version is available from the authors of this paper under request.

Tables 1–4 display the control coefficients (k) and the inspection constants (a) for the BS distribution with shape parameter $b \in \{0.5, 1.0\}$, when $r_0 \in \{200, 300, 370\}$ and $n \in \{20, 30, 40, 50\}$, and the shift constant $l \in \{0.1, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 2.0\}$. From Tables 1 and 2, note that as l decreases, the ARL1 first increases and then decreases, it being more slow when $b = 1.0$. From Tables 3 and 4, when comparing the results for $b = 1.0$ with those for $b = 0.5$, the ARL1 for the former one decreases more slowly than for the last one, as l increases. The inspection point a is determined as shorter for $b = 1.0$ than for the case $b = 0.5$. Also, from Tables 1–4, we observe that most of the k values are between 2 and 3. Notice that the ARL1 becomes smaller as n increases, as expected. All of these results show the good performance of our criterion.

Table 1. Control parameters and out-of-control average run length for the indicated r_0 when $b = 0.5$ based on the Birnbaum–Saunders distribution

	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$r_0 = 200$	$k = 2.60, a = 0.801$	$k = 2.79, a = 0.801$	$k = 2.84, a = 0.760$	$k = 2.91, a = 0.677$
l	LCL = 2; UCL = 14	LCL = 5; UCL = 20	LCL = 6; UCL = 23	LCL = 5; UCL = 23
1.0	10.42	7.65	1.47	1.00
0.95	20.86	17.07	2.07	1.00
0.9	48.87	47.61	3.59	1.01
0.8	199.06	172.52	26.70	1.20
0.7	31.95	14.98	161.22	3.36
0.6	5.34	2.54	7.81	85.45
0.5	1.71	1.15	1.46	19.60
0.3	1.00	1.00	1.00	1.00
0.1	1.00	1.00	1.00	1.00
$r_0 = 300$	$k = 2.901, a = 0.858$	$k = 2.931, a = 0.851$	$k = 2.876, a = 0.704$	$k = 2.907, a = 0.773$
l	LCL = 2; UCL = 15	LCL = 5; UCL = 21	LCL = 4; UCL = 21	LCL = 9; UCL = 29
1.0	85.48	47.06	1.03	1.62
0.95	218.57	144.83	1.09	2.56
0.9	419.00	373.19	1.26	5.33
0.8	88.80	5.25	2.82	74.88
0.7	12.95	6.43	23.84	73.98
0.6	3.12	1.72	225.09	3.72
0.5	1.37	1.06	7.08	1.13
0.3	1.00	1.00	1.00	1.00
0.1	1.00	1.00	1.00	1.00
$r_0 = 370$	$k = 3.013, a = 0.876$	$k = 2.986, a = 0.837$	$k = 2.998, a = 0.680$	$k = 3.012, a = 0.570$
l	LCL = 3; UCL = 16	LCL = 5; UCL = 21	LCL = 3; UCL = 20	LCL = 0; UCL = 17
1.0	369.28	33.74	1.01	1.00
0.95	291.90	99.58	1.02	1.00
0.9	116.65	310.59	1.07	1.00
0.8	18.02	77.68	1.62	1.00
0.7	4.20	7.94	6.91	1.00
0.6	1.64	1.89	216.55	1.09
0.5	1.09	1.08	30.50	3.83
0.3	1.00	1.00	1.02	21.88
0.1	1.00	1.00	1.00	1.00

LCL, lower control limit; UCL, upper control limit.

5. PRACTICAL ILLUSTRATION

The capital and largest city of Chile, Santiago, is located between 32–56′ and 34–17′ south (latitude) and between 69–47′ and 71–43′ west (longitude), at an elevation of 520 m above mean sea level. Santiago is surrounded by both the Andes and coastal mountains, forming the Santiago Basin. A problem arising due to this particular geographical characteristic is the limitation of air renovation in the basin. The hills strongly restrict wind movement and air flow causing the pollutants to become trapped in the basin, and then producing air contamination in the city of Santiago. In addition to these geographical conditions, meteorological factors also engender contamination problems. Because of this mixture of meteorological and geographical factors, Santiago has serious atmospheric contamination problems during both winter and summer periods. In these periods, Santiago is exposed to episodes of sudden and critical rises in its air contamination levels, which are known as critical episodes, although they are short-lived. In winter, there is an accumulation of PM and gaseous contaminants, whereas increased solar radiation during the summer favors ozone-producing photochemical reactions. Carbon monoxide (CO), nitrogen dioxide (NO_2), PM10, sulfur dioxide (SO_2), and tropospheric ozone (O_3) are the main contributors to air quality problems in Santiago. However, NO_2, SO_2, and O_3 are all precursors of PM10; see Marchant *et al.* (2013). For this reason, the official guideline of the Metropolitan Environmental Health Authority of the Chilean government (www.mma.gob.cl) regards the prediction of extreme air pollution episodes by forecasting the maximum value of the average concentration of 24 h PM10 for the period of 00–24 h of the next day. Then, we consider PM10 concentration data collected during 2008 by this Chilean health authority. PM10 concentrations are observed in 10 monitoring stations located at different zones and named as follows (locations to the right):

(S1) Independencia;
(S2) La Florida;
(S3) Las Condes;
(S4) Santiago city;
(S5) Pudahuel;
(S6) Cerrillos;
(S7) El Bosque;
(S8) Cerro Navia;
(S9) Puente Alto;
(S10) Talagante.

Table 2. Control parameters and out-of-control average run length for the indicated r_0 when $b = 1.0$ based on the Birnbaum–Saunders distribution

	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$r_0 = 200$	$k = 2.689, a = 0.496$	$k = 2.605, a = 0.413$	$k = 2.774, a = 0.463$	$k = 2.759, a = 0.401$
l	LCL = 6; UCL = 18	LCL = 2; UCL = 16	LCL = 5; UCL = 22	LCL = 6; UCL = 24
1.0	201.53	1.06	1.17	1.00
0.95	133.17	1.10	1.27	1.00
0.9	82.09	1.16	1.43	1.01
0.8	30.01	1.39	2.12	1.04
0.7	11.44	2.01	4.41	1.19
0.6	4.77	4.07	15.96	1.82
0.5	2.29	14.76	135.99	5.49
0.3	1.05	50.92	5.37	34.36
0.1	1.00	1.02	1.00	1.00
$r_0 = 300$	$k = 2.928, a = 0.548$	$k = 2.923, a = 0.611$	$k = 2.931, a = 0.535$	$k = 2.912, a = 0.504$
l	LCL = 1; UCL = 14	LCL = 5; UCL = 21	LCL = 7; UCL = 25	LCL = 9; UCL = 29
1.0	11.23	47.02	3.83	1.61
0.95	15.80	81.33	5.41	1.97
0.9	23.33	149.31	8.25	2.59
0.8	61.09	382.05	25.63	5.88
0.7	214.72	141.59	130.23	22.90
0.6	554.34	29.21	236.01	187.07
0.5	156.78	6.98	28.63	94.52
0.3	5.37	1.17	1.35	1.50
0.1	1.01	1.00	1.00	1.00
$r_0 = 370$	$k = 2.79, a = 0.602$	$k = 3.020, a = 0.59$	$k = 3.0401, a = 0.392$	$k = 2.910, a = 0.390$
l	LCL = 2; UCL = 15	LCL = 5; UCL = 21	LCL = 3; UCL = 20	LCL = 5; UCL = 24
1.0	64.78	33.11	1.01	1.00
0.95	102.47	55.74	1.01	1.00
0.9	168.76	100.60	1.02	1.00
0.8	401.16	334.91	1.09	1.03
0.7	250.50	206.34	1.30	1.14
0.6	66.16	40.64	2.09	1.63
0.5	17.08	8.86	6.13	4.41
0.3	1.84	1.22	181.02	129.52
0.1	1.00	1.00	1.01	1.00

LCL, lower control limit; UCL, upper control limit.

Table 3. Control parameters and out-of-control average run length for the indicated r_0 when $b = 0.5$ based on the Birnbaum–Saunders distribution

	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$r_0 = 200$	$k = 2.60, a = 0.801$	$k = 2.79, a = 0.801$	$k = 2.84, a = 0.760$	$k = 2.91, a = 0.677$
l	LCL = 2; UCL = 14	LCL = 5; UCL = 20	LCL = 6; UCL = 23	LCL = 5; UCL = 23
1.1	3.83	2.63	1.08	1.00
1.2	2.06	1.50	1.01	1.00
1.3	1.44	1.15	1.00	1.00
1.4	1.18	1.04	1.00	1.00
1.5	1.08	1.01	1.00	1.00
1.6	1.03	1.00	1.00	1.00
1.7	1.01	1.00	1.00	1.00
1.8	1.00	1.00	1.00	1.00
2.0	1.00	1.00	1.00	1.00
$r_0 = 300$	$k = 2.901, a = 0.858$	$k = 2.931, a = 0.851$	$k = 2.876, a = 0.704$	$k = 2.907, a = 0.773$
l	LCL = 2; UCL = 15	LCL = 5; UCL = 21	LCL = 4; UCL = 21	LCL = 9; UCL = 29
1.1	18.53	8.86	1.00	1.09
1.2	6.41	3.14	1.00	1.01
1.3	3.13	1.73	1.00	1.00
1.4	1.96	1.26	1.00	1.00
1.5	1.46	1.09	1.00	1.00
1.6	1.22	1.03	1.00	1.00
1.7	1.11	1.01	1.00	1.00
1.8	1.05	1.00	1.00	1.00
2.0	1.01	1.00	1.00	1.00
$r_0 = 370$	$k = 3.013, a = 0.876$	$k = 2.986, a = 0.837$	$k = 2.998, a = 0.680$	$k = 3.012, a = 0.570$
l	LCL = 3; UCL = 16	LCL = 5; UCL = 21	LCL = 3; UCL = 20	LCL = 0; UCL = 17
1.1	83.13	7.03	1.00	1.00
1.2	21.13	2.70	1.00	1.00
1.3	7.86	1.58	1.00	1.00
1.4	3.90	1.20	1.00	1.00
1.5	2.40	1.07	1.00	1.00
1.6	1.72	1.02	1.00	1.00
1.7	1.39	1.01	1.00	1.00
1.8	1.21	1.00	1.00	1.00
2.0	1.06	1.00	1.00	1.00

LCL, lower control limit; UCL, upper control limit.

Maximum permitted PM10 concentrations according to Chilean, American, and European guidelines are 150, 150, and 100 $\mu g/m^3N$, respectively. Chilean air quality guideline indicates the maximum level at which contaminant concentrations become harmful to human health; see Table 5. At such levels, one can calculate exceedance probabilities, that is, the probability that a concentration level exceeds a value established by the official guideline for a given period of time.

We use the criterion proposed in this paper for assessing environmental risk in Santiago, Chile. We provide information that can be useful for preventing adverse effects on human health of the population of Santiago. We show the coherence between our criterion and what happened in real-world situations, in which Chilean health authority ruled environmental alerts, such as our criterion adverts.

The used PM10 concentration data are obtained as 1 h (hourly) average values and are based on the month of May 2008 and eight monitoring stations (S1–S8). We select data from stations S1 to S8 mainly because they are better conformed to situations of low and high stability, which allow us to analyze for different pollution patterns. Thus, S9 and S10 stations have been omitted from this analysis.

Table 6 provides descriptive statistics of the data, including central tendency statistics, as well as the standard deviation and coefficients of variation, skewness, and kurtosis, among others. Note that the BS distribution can be reasonably assumed for modeling these data, due to their asymmetric nature and their level of kurtosis, in addition to its theoretical arguments for modeling this type of data; see Section 2. Figure 1 depicts graphical plots of the partial autocorrelation function for the indicated sets (a similar behavior is observed for the other data sets). From this figure, note that, practically, the presence of serial correlation is not detected, in agreement with results reported by Morel *et al.* (1999), Gokhale and Khare (2007), Vilca *et al.* (2010), and Marchant *et al.* (2013). Seasonality over time is not detected either. However, studies during wider periods could detect seasonal dependence, which may be treated according to Step 3 of Algorithm 1.

Table 4. Control parameters and out-of-control average run length for the indicated r_0 when $b = 1.0$ based on the Birnbaum–Saunders distribution

	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$r_0 = 200$	$k = 2.689, a = 0.496$	$k = 2.605, a = 0.413$	$k = 2.774, a = 0.463$	$k = 2.759, a = 0.401$
l	LCL = 6; UCL = 18	LCL = 2; UCL = 16	LCL = 5; UCL = 22	LCL = 6; UCL = 24
1.1	270.80	1.02	1.06	1.00
1.2	173.26	1.01	1.02	1.00
1.3	92.14	1.00	1.01	1.00
1.4	51.47	1.00	1.00	1.00
1.5	31.06	1.00	1.00	1.00
1.6	20.10	1.00	1.00	1.00
1.7	13.80	1.00	1.00	1.00
1.8	9.96	1.00	1.00	1.00
2.0	5.86	1.00	1.00	1.00
$r_0 = 300$	$k = 2.928, a = 0.548$	$k = 2.923, a = 0.611$	$k = 2.931, a = 0.535$	$k = 2.912, a = 0.504$
l	LCL = 1; UCL = 14	LCL = 5; UCL = 21	LCL = 7; UCL = 25	LCL = 9; UCL = 29
1.1	6.41	18.93	2.29	1.25
1.2	4.16	9.37	1.64	1.10
1.3	2.98	5.43	1.33	1.04
1.4	2.30	3.57	1.17	1.01
1.5	1.88	2.58	1.09	1.00
1.6	1.61	2.01	1.04	1.00
1.7	1.43	1.66	1.02	1.00
1.8	1.30	1.44	1.01	1.00
2.0	1.15	1.20	1.00	1.00
$r_0 = 370$	$k = 2.79, a = 0.602$	$k = 3.020, a = 0.59$	$k = 3.0401, a = 0.392$	$k = 2.910, a = 0.390$
l	LCL = 2; UCL = 15	LCL = 5; UCL = 21	LCL = 3; UCL = 20	LCL = 5; UCL = 24
1.1	29.59	14.07	1.00	1.00
1.2	15.74	7.30	1.00	1.00
1.3	9.44	4.41	1.00	1.00
1.4	6.22	3.00	1.00	1.00
1.5	4.43	2.24	1.00	1.00
1.6	3.35	1.79	1.00	1.00
1.7	2.67	1.52	1.00	1.00
1.8	2.21	1.34	1.00	1.00
2.0	1.67	1.15	1.00	1.00

LCL, lower control limit; UCL, upper control limit.

Table 5. Levels of environmental risk corresponding to 24-h PM10 concentrations (in $\mu g/m^3N$)

Level	PM10 concentration
Warning	195–239
Pre-emergency	240–329
Emergency	⩾330

PM, particulate matter.

Histograms for the data, displayed in Figures 1(a)–(d), provide evidence for positive skewness and heavy tails of their distributions. Usual and adjusted boxplots are also showed in Figures 1(c)–(f). There, a significant number of observations classified as outliers by the usual boxplot are not outliers when boxplot adjusted for asymmetrical data is considered. A similar behavior is detected for the other stations (omitted here).

Table 6. Summary statistics for PM10 concentrations (in $\mu g/m^3N$) in the indicated data set

Data set	n	Min.	Median	Mean	Max.	SD	CV	CS	CK
S1	744	1	78.0	83.30	254	47.75	0.57	0.67	3.33
S2	744	1	79.0	95.64	548	78.18	0.82	1.69	7.37
S3	744	1	51.0	59.65	506	49.09	0.82	2.14	14.88
S4	744	1	78.5	91.97	379	64.90	0.71	1.44	5.70
S5	744	1	88.0	110.70	600	95.00	0.86	1.79	7.49
S6	744	1	77.7	91.71	367	62.40	0.68	1.38	5.72
S7	744	1	84.0	102.59	434	79.20	0.77	1.19	4.26
S8	744	1	90.5	112.56	598	95.84	0.85	1.95	8.37

PM, particulate matter; SD, standard deviation; CV, coefficient of variation; CS, coefficient of skewness; CK, coefficient of kurtosis.

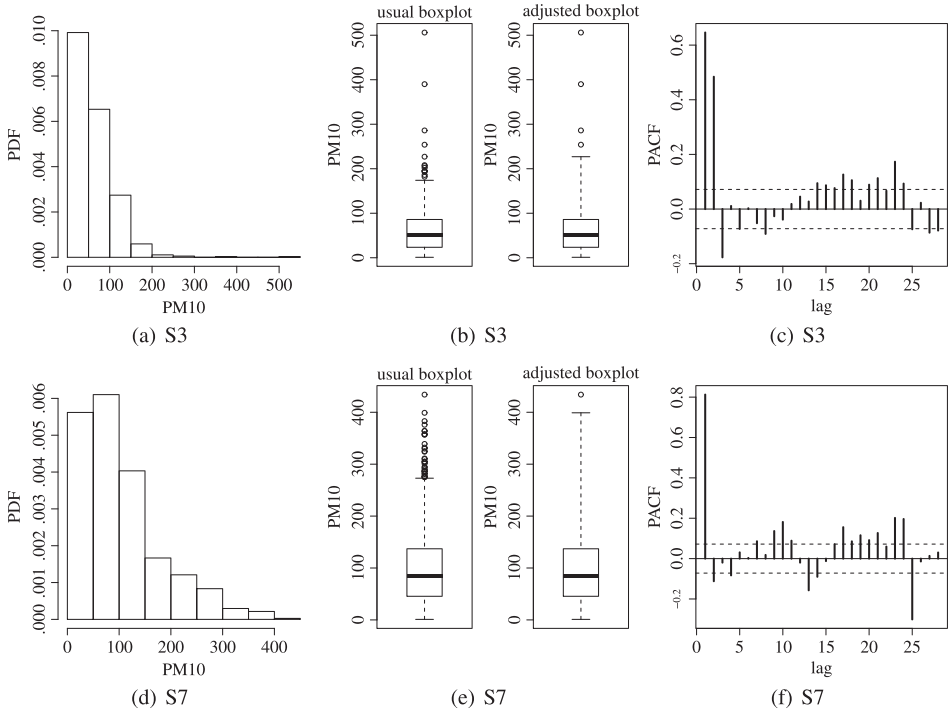
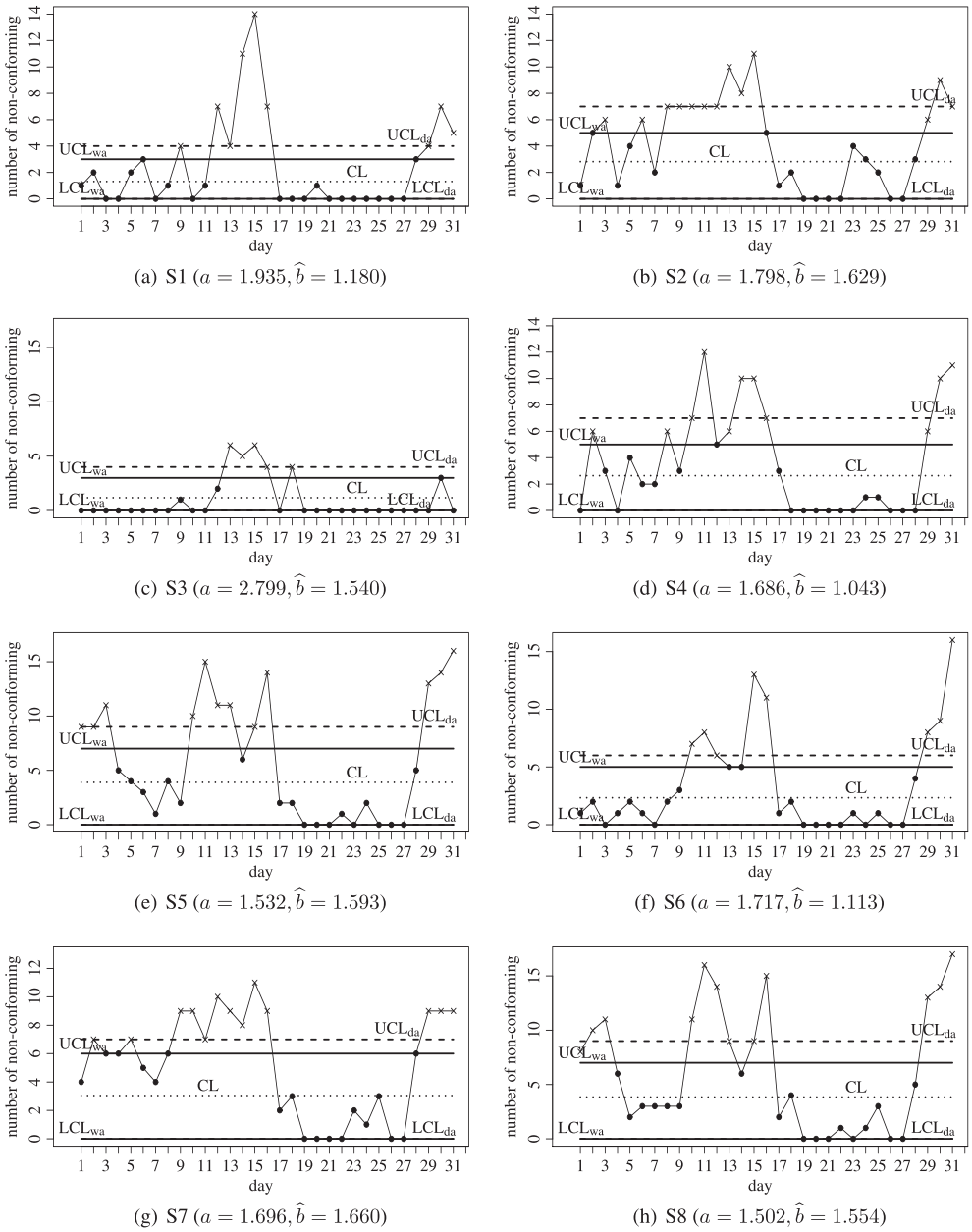


Figure 1. Histogram, boxplot, and partial autocorrelation function plot for PM10 concentrations (in $\mu g/m^3N$) in the indicated data set: PM, particulate matter

Next, we apply our criterion based on the BS np chart proposed in Section 3. Specifically, this chart is used to monitor the number of non-conforming PM10 concentration levels in N subgroups of size n, where N denotes the number of days in the month of May 2008 and n is the number of hours of each subgroup, that is, $N = 31$ and $n = 24$. The control chart is constructed following Algorithm 1. Figure 2 shows BS np charts for the data under analysis. In the figure, limits with $k = 2$ and $k = 3$ are plotted, indicating warning and dangerous levels, respectively. Notice that several points are below the UCL, indicating that the contamination levels are out of control. Then, an environmental alert must be declared for those out-of-control days. Note that out-of-control detection is based on the fact that at least one of the air quality monitoring stations presents a dangerous PM10 level for human health; see Comisión Nacional del Medio Ambiente (1998) for details of the official decree of the Ministry of Environment (CONAMA in Spanish) of the Chilean government that indicates this regulation, which is valid for the analyzed 2008 data. Station S3 is less contaminated than other stations, due to it is located at a higher altitude, and therefore, it has better ventilation. For this reason, Station S3 has less points outside of the limits; see Marchant *et al.* (2013). Note the coherence between our criterion and the official information provided by the Chilean Ministry of Health, which established environmental alerts during days 12–16 May 2008 and an environmental pre-emergence for 30 May 2008; see www.seremisaludrm.cl/sitio/pag/aire/indexjs3airee001.asp



(a) S1 ($a = 1.935, \widehat{b} = 1.180$) (b) S2 ($a = 1.798, \widehat{b} = 1.629$)

(c) S3 ($a = 2.799, \widehat{b} = 1.540$) (d) S4 ($a = 1.686, \widehat{b} = 1.043$)

(e) S5 ($a = 1.532, \widehat{b} = 1.593$) (f) S6 ($a = 1.717, \widehat{b} = 1.113$)

(g) S7 ($a = 1.696, \widehat{b} = 1.660$) (h) S8 ($a = 1.502, \widehat{b} = 1.554$)

Figure 2. Birnbaum–Saunders np chart with $t_0 = 150$, where index "wa" adverts warning ($k = 2$, solid line) and index "da" dangerous ($k = 3$, dashed line) situations for PM10 concentrations (in $\mu g/m^3N$) in the indicated station: (a)–(c) S3 and (d)–(f) S7. PM, particulate matter

6. CONCLUSIONS AND FUTURE STUDIES

We have proposed a criterion based on an attribute control chart for assessing environmental risk when the air pollutant concentration follows a Birnbaum–Saunders distribution. We have provided a novel justification for the usage of this distribution in environmental sciences. We have conducted a simulation study to evaluate the proposed criterion, which has reported its good performance to provide an earlier alert of an out-of-control process. An application with air pollution concentration data, collected by the Chilean official environmental authority, has been considered to validate our criterion. The application has shown the coherence between our criterion and what is happening in real-world situations in which Chilean health authority ruled environmental alerts, such as our criterion adverted. Thereby, the proposed criterion has shown to be useful for alerting episodes of extreme air pollution, which must be monitored, alerted, and corrected for protection of human health.

Some possible issues to be considered in future studies are the following. First, seasonal and serial dependence can be considered in our criterion by using time series models; see Box *et al.* (2008). Second, because monitoring particulate matter pollution is not an easy task due to its complex composition, one could consider more than one random variable to model pollutant concentrations by means of

multivariate structures for the Birnbaum–Saunders distribution, for example, carbon monoxide, particulate matter, and tropospheric ozone; see Marchant *et al.* (2015). Third, particulate matter can also be degraded. However, the Birnbaum–Saunders distribution only allows us to model accumulation (aggregation) due to how it was formulated. Some recent studies allow us to count with new cumulative damage models that consider aggregation and degradation in a same framework; see Leiva *et al.* (2015). Work on these three issues is currently under progress, and we hope to report some findings in a future paper.

Acknowledgements

The authors thank the editors and two anonymous referees for their valuable comments on an earlier version of this manuscript. This research was partially supported by FONDECYT 1120879 and CONICYT "Becas-Chile" grants from the Chilean government, and by CAPES, CNPq, and FACEPE grants from the Brazilian government.

REFERENCES

Aitchison J, Brown JAC. 1973. *The Lognormal Distribution*. Cambridge University Press: London.
Alencar AP, Santos BR. 2014. Association of pollution with quantiles and expectations of the hospitalization rate of elderly people by respiratory diseases in the city of Sao Paulo, Brazil. *Environmetrics* **25**:165–171.
Almaskut A, Farrel P, Krewski D. 2012. Statistical methods for estimating the environmental burden of disease in Canada, with applications to mortality from fine particulate matter. *Environmetrics* **23**:329–344.
Aslam M, Jun C-H, Ahmad M. 2011. New acceptance sampling plans based on life tests for Birnbaum–Saunders distributions. *Journal of Statistical Computation and Simulation* **81**:461–470.
Balakrishnan N, Leiva V, Lopez J. 2007. Acceptance sampling plans from truncated life tests from generalized Birnbaum–Saunders distribution. *Communications in Statistics: Simulation and Computation* **36**:643–656.
Birnbaum ZW, Saunders SC. 1969. A new family of life distributions. *Journal of Applied Probability* **6**:319–327.
Box GEP, Jenkins GM, Reinsel GC. 2008. *Time Series Analysis: Forecasting and Control*. Wiley: New York.
Clyde M. 2000. Model uncertainty and health effect studies for particulate matter. *Environmetrics* **11**:745–763.
Comisión Nacional del Medio Ambiente. 1998. Establishment of primary quality guideline for PM10 that regulates environmental alerts. *Ministry of Environment (CONAMA) of the Chilean Government Decree 59,* Santiago, Chile.
Cox L. 2000. Statistical issues in the study of air pollution involving airborne particulate matter. *Environmetrics* **11**:611–626.
Cramér H. 1946. *Mathematical Methods of Statistics*. Princeton University Press: Princeton.
Desmond A. 1985. Stochastic models of failure in random environments. *Canadian Journal of Statistics* **13**:171–183.
Epprecht EK, Costa AFB, Mendes FCT. 2003. Adaptive control charts for attributes. *IIE Transactions* **35**:567–582.
Ferreira M, Gomes MI, Leiva V. 2012. On an extreme value version of the Birnbaum–Saunders distribution. *Revstat Statistical Journal* **10**:181–210.
Ferreira-Baptista L, De Miguel E. 2005. Geochemistry and risk assessment of street dust in Luanda, Angola: a tropical urban environment. *Atmospheric Environment* **39**:4501–4512.
Gokhale S, Khare M. 2007. Statistical behavior of carbon monoxide from vehicular exhausts in urban environments. *Environmental Modelling and Software* **22**:526–535.
Grigg OA, Farewell VT. 2004. A risk-adjusted sets method for monitoring adverse medical outcomes. *Statistics in Medicine* **23**:1593–1602.
Hsu LF. 2004. Note on 'design of double-and triple-sampling control charts using genetic algorithms'. *International Journal of Production Research* **42**:1043–1047.
Joekes S, Barbosa EP. 2013. An improved attribute control chart for monitoring non-conforming proportion in high quality processes. *Control Engineering Practice* **21**:407–412.
Kessler SH, Smith JD, Che DL, Worsnop DR, Wilson KR, Kroll JH. 2010. Chemical sinks of organic aerosol: kinetics and products of the heterogeneous oxidation of erythritol and levoglucosan. *Environmental Science & Technology* **44**:7005–7010.
Leiva V, Barros M, Paula GA, Sanhueza A. 2008. Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. *Environmetrics* **19**:235–249.
Leiva V, Ferreira M, Gomes MI, Lillo C. 2015. Extreme value Birnbaum–Saunders regression models applied to environmental data. *Stochastic Environmental Research and Risk Assessment,* (in press), available at DOI:10.1007/s00477-015-1069-6.
Leiva V, Marchant C, Saulo H, Aslam M, Rojas F. 2014. Capability indices for Birnbaum–Saunders processes applied to electronic and food industries. *Journal of Applied Statistics* **41**:1881–1902.
Leiva V, Ponce MG, Marchant C, Bustos O. 2012. Fatigue statistical distributions useful for modeling diameter and mortality of trees. *Colombian Journal of Statistics* **35**:349–367.
Leiva V, Sanhueza A, Angulo JM. 2009. A length-biased version of the Birnbaum–Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment* **23**:299–307.
Leiva V, Tejo M, Guiraud P, Schmachtenberg O, Orio P, Marmolejo F. 2015. Modeling neural activity with cumulative damage distributions. *Biological Cybernetics* (in press). DOI: 10.1007/s00422-015-0651-9.
Lio YL, Tsai T-R, Wu S-J. 2010. Acceptance sampling plans from truncated life tests based on the Birnbaum–Saunders distribution for percentiles. *Communications in Statistics: Simulation and Computation* **39**:1–18.
Lund R, Seymour L. 1999. Assessing temperature anomalies for a geographical region: a control chart approach. *Environmetrics* **10**:163–177.
Manly B, Mackenzie D. 2000. A cumulative sum type of method for environmental monitoring. *Environmetrics* **11**:151–166.
Marchant C, Leiva V, Cavieres MF, Sanhueza A. 2013. Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Reviews of Environmental Contamination and Toxicology* **223**:1–31.
Marchant C, Leiva V, Cysneiros FJA. 2015. Multivariate Birnbaum–Saunders regression models for metal fatigue. Under 2dn review in *IEEE Transactions on Reliability*.
Montgomery D. 2005. *Introduction to Statistical Quality Control*. Wiley: New York.
Morel B, Yeh S, Cifuentes L. 1999. Statistical distributions for air pollution applied to the study of the particulate problem in Santiago. *Atmospheric Environment* **33**:2575–2585.
Morrison LW. 2008. The use of control charts to interpret environmental monitoring data. *Natural Areas Journal* **28**:66–73.

Ng HKT, Kundu D, Balakrishnan N. 2003. Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics and Data Analysis* **43**:283–298.
Ott WR. 1990. A physical explanation of the lognormality of pollution concentrations. *Journal of the Air and Waste Management Association* **40**:1378–1383.
Querol X, Alastuey A, Viana M, Rodriguez S, Artiñano B, Salvador P, Garcia S, Patier R, Ruiz C, de la Rosa J, Sanchez A, Menendez M, Gil J. 2004. Speciation and origin of PM10 and PM2.5 in Spain. *Journal of Aerosol Science* **35**:1151–1172.
R Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. Available at www.R-project.org).
Reich BJ, Fuentes M, Burke J. 2008. Analysis of the effects of ultrafine particulate matter while accounting for human exposure. *Environmetrics* **20**: 131–146.
Rodrigues AD, Epprecht EK, de Magalhães MS. 2011. Double-sampling control charts for attributes. *Journal of Applied Statistics* **38**:87–112.
Saulo H, Leiva V, Ruggeri F, Kitsos C, Oliveira TA, Rigas A, Gulati S (eds.) 2015. Monitoring environmental risk by a methodology based on control charts In *Theory and Practice of Risk Assessment*, Springer Proceedings in Mathematics & Statistics, vol. 136, Kitsos C, Oliveira TA, Rigas A, Gulati S (eds). Springer: Switzerland, *177–197*, DOI:10.1007/978-3-319-18029-8_14.
Saulo H, Leiva V, Ziegelmann FA, Marchant C. 2013. A nonparametric method for estimating asymmetric densities based on skewed Birnbaum–Saunders distributions applied to environmental data. *Stochastic Environmental Research and Risk Assessment* **27**:1479–149.
Vilca F, Sanhueza A, Leiva V, Christakos G. 2010. An extended Birnbaum–Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment* **24**:771–782.
Vione D, Maurino V, Minero C, Pelizzetti E, Harrison MA, Olariu RI, Arsene C. 2006. Photochemical reactions in the tropospheric aqueous phase and on particulate matter. *Chemical Society Reviews* **35**:441–453.
WHO –World Health Organization–. 2003. *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide*. WHO Working Group: Bonn, Germany.
WHO –World Health Organization–. 2006. *Air Quality Guidelines - Global Update 2005: Particulate Matter, Ozone, Nitrogen Oxide and Sulfur Dioxide*. WHO Regional Office for Europe: Copenhagen, Denmark.
Woodall WH. 2006. The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology* **38**:89–104.
Wu Z, Wang Q. 2007. An np control chart using double inspections. *Journal of Applied Statistics* **34**:843–855.

APPENDIX: THE BIRNBAUM–SAUNDERS DISTRIBUTION

The BS distribution is unimodal and has shape ($b > 0$) and scale ($\sigma > 0$) parameters, asymmetry to the right, and positive support. However, σ is also a position parameter because it corresponds to the median of the distribution. The RV

$$T = \sigma\left[\frac{bZ}{2} + \sqrt{\left\{\frac{bZ}{2}\right\}^2 + 1}\right]^2 \tag{A.1}$$

is said to follow a BS distribution with parameters b and σ if $Z = [1/b]\xi(T/\sigma) \sim \mathrm{N}(0,1)$, with $\xi(y) = \sqrt{y} - 1/\sqrt{y} = 2\sinh(\log(\sqrt{y}))$, which is denoted by $T \sim \mathrm{BS}(b,\sigma)$. The corresponding CDF is given by

$$F_T(t;b,\sigma) = \mathrm{P}(T \leqslant t) = \Phi\left(\frac{1}{b}\xi(t/\sigma)\right), \quad t > 0 \tag{A.2}$$

The qth quantile of T is $t(q;b,\sigma) = \sigma\left[bz(q)/2 + \sqrt{\{bz(q)/2\}^2 + 1}\right]^2$, for $0 < q < 1$, where $z(q)$ is the qth $\mathrm{N}(0,1)$ quantile. Thus, if $q = 0.5$, then $t(0.5) = \sigma$, and as mentioned, σ is the median of the BS distribution. A random number generator from $T \sim \mathrm{BS}(b,\sigma)$ obtained from (A.1) is given by Algorithm 2.

Algorithm 2 Random number generator for the BS distribution

1: Obtain a random number z from $Z \sim \mathrm{N}(0,1)$.
2: Set values for b and σ of $T \sim \mathrm{BS}(b,\sigma)$.
3: Compute a BS random number with $t = \sigma\left[bz/2 + \sqrt{\{bz/2\}^2 + 1}\right]^2$.
4: Repeat Steps 1 to 3 until the required amount of random numbers is completed.

Some properties of $T \sim \mathrm{BS}(b,\sigma)$ are as follows: (P1) $W = [1/b^2]\xi^2(T/\sigma) \sim \chi^2(1)$, (P2) $r\,T \sim \mathrm{BS}(b, r\sigma)$, for $r > 0$, and (P3) $1/T \sim \mathrm{BS}(b, 1/\sigma)$, that is, the BS distribution is related to the normal and chi-squared distributions and belongs to the scale and closed under reciprocation families. The mean of $T \sim \mathrm{BS}(b,\sigma)$ is

$$\mathrm{E}[T] = \mu = \frac{\sigma}{2}[2 + b^2] \tag{A.3}$$

whereas its variance and coefficients of variation, skewness, and kurtosis are, respectively, defined as

$$\mathrm{Var}[T] = \frac{\sigma^2}{4}[4b^2 + 5b^4]$$
$$\mathrm{CV}[T] = \frac{\sqrt{4b^2 + 5b^4}}{2 + b^2}$$
$$\mathrm{CS}[T] = \frac{24b + 44b^3}{\sqrt{[4 + 5b^2]^3}}$$
$$\mathrm{CK}[T] = 3 + \frac{240b^2 + 558b^4}{[4 + 5b^2]^2}$$

In addition, if T has a BS distribution with parameters b and σ, and because $1/T$ has a BS distribution with parameters b and $1/\sigma$, we have that

$$\mathrm{E}\left[\frac{1}{T}\right] = \frac{1}{2\sigma}\left[2 + b^2\right] \tag{A.4}$$

Several methods have been proposed for estimating the parameters of the BS distribution. However, in all these methods, it is not possible to find explicit expressions for the estimators, and then numerical procedures must be used for calculating their estimates. Ng *et al.* (2003) introduced an MM method based on (A.4) for estimating the BS parameters, which provide easy analytical expressions to compute them. Specifically, let $T_1, \ldots, T_n$ be a random sample of size n from $T \sim \mathrm{BS}(b, \sigma)$ and $t_1, \ldots, t_n$ be their observations (data). Then, estimators of b and σ are obtained using the MM method equating $\mathrm{E}[T]$ and $\mathrm{E}[1/T]$ given in (A.3) and (A.4) to its corresponding sample moments, obtaining the MM estimators of b and σ as

$$\widehat{b} = \sqrt{2\left[\sqrt{\frac{S}{R}} - 1\right]} \quad \text{and} \quad \widehat{\sigma} = \sqrt{SR} \tag{A.5}$$

where

$$S = \frac{1}{n}\sum_{i=1}^{n} T_i \quad \text{and} \quad R = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\right]^{-1}$$

are the arithmetic and harmonic means of $T_1, \ldots, T_n$, respectively. Some studies have shown similar statistical properties between the MM and maximum likelihood estimators. Therefore, because of their simplicity, we use the MM estimators; see Leiva *et al.* (2008).

