A Methodology for Data-Driven Decision-Making in the Monitoring of Particulate Matter Environmental Contamination in Santiago of Chile

María Fernanda Cavieres, Víctor Leiva, Carolina Marchant, and Fernando Rojas

Contents

M. F. Cavieres (✉) · F. Rojas
Faculty of Pharmacy, Universidad de Valparaíso, Valparaíso, Chile
e-mail: fernanda.cavieres@uv.cl; fernando.rojas@uv.cl

V. Leiva
School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
e-mail: victorleivasanchez@gmail.com

C. Marchant
Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile
e-mail: carolina.marchant.fuentes@gmail.com

Reviews of Environmental Contamination and Toxicology,
DOI 10.1007/398_2020_41

1 Introduction

The main components of air are nitrogen (approximately 78%), oxygen (21%) and argon, carbon dioxide, helium, hydrogen, neon, and water vapor (together ~1%). Atmospheric pollution is the lasting presence in the air of these and other chemicals at concentrations above their natural levels, which could potentially lead to adverse health effects. It derives mainly from anthropogenic activities that use combustion. Effects of air contamination are important if climatological and geographical factors reduce its dissipation, especially in areas with huge anthropogenic activity. Due to this, many people breathe contaminated air, and the World Health Organization (WHO) has estimated that 4.2 million deaths every year are a result of poor air quality, with 91% of the world's population being exposed to air pollutants by living in places where air quality exceeds WHO safety guidelines (https://www.who.int/airpollution/en/).

It is a fact that meteorological and climatic variables play an important role in the determination of air pollution patterns and the global climate change is foreseen to cause an increase in the concentration of some pollutants (Kinney 2008). Santiago, the capital city of Chile, is among the cities with higher air pollution levels in the world. Its location and weather, when combined with high anthropological emissions, create critical air pollution conditions. A recent model explained elevated particulate matter (PM) concentrations during high pollution events in Santiago, as a function of weather conditions in central Chile and in Argentina, which at the local level generate a depression at the base of the inversion layer, an increase in the vertical thermal stability, lower humidity and low-wind conditions. Pollutant dispersion is thus decreased leading to poor ventilation of contaminated air (Toro et al. 2019).

Due to its geography, people in the city of Santiago experience dry and hot summers and damp winters, which contribute to the air pollution episodes that put younger and older people at risk of respiratory and cardiovascular diseases. Henríquez and Urrea (2017) showed a rise in daily emergency visits in winter compared to summer (odds ratio of 2.2646), which were associated with higher daily concentrations of PM, carbon monoxide, and sulfur and nitrogen oxide during winter. Noteworthy, the press has recently covered the news that Chile has nine of the ten more air-contaminated cities in South America (https://bit.ly/2ovYLiu).

The effects on health of air pollution vary according to type of pollutants, their concentration, and duration of exposure. It is generally accepted that air contamination causes cardiovascular and pulmonary morbidity in addition to increased mortality after exposure, but other epidemiological associations have also been described, including cancer as well as reproductive and immunological toxicity (https://www.who.int/airpollution/en/). Due to the multicomponent aspect of contaminated air, it is hard to establish what component of the contaminant mixture produces a specific health problem. However, there is a growing amount of evidence that indicates that PM plays a key role in the induction of cardiovascular and respiratory diseases (Kim et al. 2015).

Data analytics may be employed to generate information of air quality within the context of data-driven decision-making (DDDM), a process associated with big data and data science (Baesen 2014; Dietrich 2015; Aykroyd et al. 2019). DDDM allows us to study the impact of atmospheric contaminants on human health and the urban environment. In epidemiological studies, average air contaminant concentrations are employed as indicators of air pollutant levels. However, since concentrations of air pollutants vary with geographical and meteorological conditions, they are treated as random variables taking values greater than zero. Then, these random variables are described by a statistical distribution, which is frequently asymmetrical with a positive skewness (Marchant et al. 2013). Note that pollutant concentrations are expressed as number of units of mass of a certain substance (or agent) per a defined unit of mass in the set, so it can never take negative values (Ott 1990). Thus, the popular normal or Gaussian distribution is not applicable and authors resort to data transformation which has limitations. Alternatively, one can avoid data transformation by modeling with a suitable distribution (Leiva et al. 2015; Marchant et al. 2019) such as Birnbaum-Saunders.

The Birnbaum-Saunders distribution has positive skewness (asymmetry). It has been defined over a range of continuous values greater than zero, allowing its use to describe random variables with positive support such as atmospheric pollutant concentrations. This distribution was derived from physical considerations of material failure due to fatigue (Leiva 2016) and has been successfully applied to describe air pollutant concentrations (Leiva et al. 2008, 2015; Vilca et al. 2010; Ferreira et al. 2012; Marchant et al. 2018, 2019). Leiva et al. (2015) provided a mathematical formulation based on the proportionate-effect law (also known as the Gibrat law) to justify the use of the Birnbaum-Saunders model as environmental contaminant statistical distribution, justification which was previously associated also with the lognormal distribution (Ott 1990). According to Leiva et al. (2015), a contaminant concentration follows the proportionate-effect law if the growth in the concentration at any step of the contamination process is a random proportion of the previous value of the concentration. The Birnbaum-Saunders distribution has properties that are similar to those of the lognormal distribution (Aitchison and Brown 1973; Leiva 2016), including the relationship with the proportionate-effect law and contamination processes (Aitchison and Brown 1973, p. 22; Ott 1990; Leiva et al. 2015). In addition, in both Birnbaum-Saunders and lognormal distributions, their parameter estimation is sensitive to atypical (extreme or outliers) data; a situation frequently found when one analyzes air contaminant concentrations. Díaz-García and Leiva (2005) derived an extension of the Birnbaum-Saunders distribution which is known as the generalized Birnbaum-Saunders distribution. These distributions are a flexible general family based on the Birnbaum-Saunders distribution, which contains several particular cases. One of these cases is the Birnbaum-Saunders-Student-t distribution, from now on Birnbaum-Saunders-t distribution (Athayde et al. 2019), whose estimation of its parameters is not sensitive (robust) to the presence of atypical observations, which are common in environmental contamination data. It is important to mention this issue of robustness to be kept in mind when using the Birnbaum-

Saunders-t distribution, due to the relevant role that the Birnbaum-Saunders distribution and its extensions are taking in environmental modeling.

Air chemical pollutants can be broadly grouped into four classes: gaseous compounds; heavy metals; persistent organic pollutants; and suspended particles or PM. Given the importance of PM in air pollution toxicity and that the Birnbaum-Saunders and Birnbaum-Saunders-t models are adequate to statistically describe pollutant distributions, the objectives of this article are (1) to provide a notion of the serious threat of PM10 and PM2.5 for human health; (2) to describe the air contamination problem in Santiago, Chile; and (3) to propose a data science methodology that can be applied for modeling air quality. We exemplify this methodology using air contamination real data from the city of Santiago, Chile.

2 Particulate Matter and Contamination in Santiago of Chile

2.1 Adverse Effects of Particulate Matter

PM is a complex mixture of particles and liquid droplets that get into the air (Adams et al. 2015; Hime et al. 2018). It is classified according to its diameter, which is important for risk evaluation, as particle size determines site of deposition within the respiratory tract. Thus, particles with a diameter over 10 μm do not penetrate into airways. Hence, these particles are usually considered to be of low risk, as they are deposited in the upper respiratory tract (on the nose and throat epithelium, above the larynx) and are cleared by mucociliary function. On the contrary, particles with a diameter smaller than 10 μm (PM10) are considered to be inhalable, that is, they get past the larynx, and, according to their size, they are deposited either on lower airways (particles between 2.5 and 10 μm) or on the alveoli of the lungs (particles smaller than 2.5 μm, PM2.5). Particles smaller than 0.1 μm are called ultrafine particles (UFPs) and easily reach the lung, where they are absorbed into the blood (WHO 2000, 2013).

PM is very complex as it varies greatly in source and composition. Coarser particles (those between 2.5 and 10 μm) are formed by the breakup of larger particles and usually contain minerals as well as carbon. Finer particles ($<$2.5 μm and including UFPs) derive mainly from combustion and may be either a carbon core with adsorbed hydrocarbons, such as polycyclic aromatic hydrocarbons and metals or secondary particles formed from sulfur and nitrogen oxides. Coarser particles may also include biological material such as mold, pollen, endotoxins, and bacteria (WHO 2000; Adams et al. 2015; Falcon-Rodriguez et al. 2016; Thompson 2018).

Anthropogenic activities, such as public transportation and industrial combustion, are the main contributors to the pollution of air in urban environments. The health effects of inhalable PM, after both acute and chronic exposure, have been described in the scientific literature. For instance, a 1953 article describes how the London Fog

Incident in December 1952 led to the death of at least 4000 people mainly from respiratory and cardiovascular conditions (Logan 1953). Nowadays, it is greatly accepted that toxicity from exposure to air pollution results in great part from the action of airborne PM. In addition, these health effects generally include not only respiratory and cardiovascular diseases but also cancer (WHO 2013). Furthermore, there is some evidence that PM may also cause or contribute to neurotoxicity and developmental toxicity (Thompson 2018). Animal and human studies have reported that PM causes systemic inflammation increasing respiratory and cardiovascular morbidity, as well as mortality from respiratory and cardiovascular diseases and cancer (WHO 2013; Wu et al. 2018). In fact, outdoor air pollution and PM in outdoor air pollution are both classified by the International Agency for Research on Cancer (IARC) as carcinogenic to humans (Group 1). The IARC concluded that PM not only is associated with an increase in genetic damage predictive of cancer but that it also may promote cancer progression by inducing oxidative stress and sustained inflammation (IARC 2016).

2.2 *Geography, Topography, and Location of Santiago and Its Air Monitoring Stations*

Santiago, the capital of Chile, is the largest city in the country, with an area of 867.75 km^2 and a population of about 7.1 million people, which is approximately 40.5% of the Chilean population (according to the information obtained from the Population and Housing Census 2017 conducted by the Chilean government). Santiago city is located in subtropical South America (33°27′S, 70°40′W), between a coastal mountain range to the West (with an altitude close to 1000 m above sea level) and the Andes mountain range to the East (with an altitude of around 3000 m above sea level). Santiago has been facing air pollution problems for more than three decades, becoming one of the cities with the highest levels of air pollution in the world (Ostro 2003). Its poor air quality is believed to be the result of the growing industrial sector, fast-growing population, and increased number of motor vehicles, worsened by geophysical constraints for pollutant dispersion in Santiago's basin (Préndez et al. 2011; Mendoza et al. 2019). In fall and winter seasons, subsidence conditions induce thermal inversion layers that increase levels of pollutant concentrations, creating a characteristic seasonality of air quality in the city (Villalobos et al. 2015).

Santiago has eleven (11) monitoring stations (which make the air quality assessment network for the Metropolitan Region of Santiago, denominated MACAM), located at different zones in the Metropolitan region of Chile. Figure 1 shows these stations which are named as (MS1) Independencia; (MS2) La Florida; (MS3) Las Condes; (MS4) Santiago city; (MS5) Pudahuel; (MS6) Cerrillos; (MS7) El Bosque; (MS8) Cerro Navia; (MS9) Puente Alto; (MS10) Talagante; and (MS11) Quilicura. The monitoring stations are geographically located with their respective numbers on the map of Fig. 1.

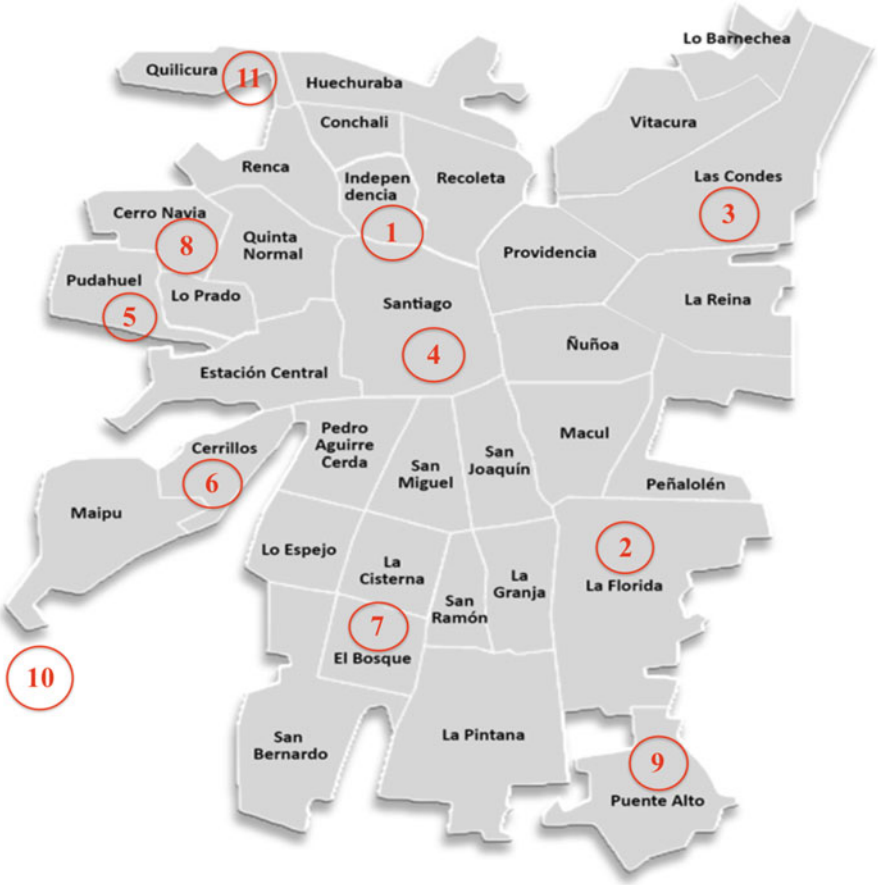

Fig. 1 Map of MACAM network of Santiago (Source: Metropolitan Regional Secretariat of the Chilean Ministry of Health)

2.3 Local Air Quality Guidelines

The current official methodology used by the Chilean authority in Santiago to predict PM10 concentrations is based on a multiple regression model (Morales et al. 2012). It helps to forecast the maximum value of the 24 h average concentration of PM10 in μg/normalized cubic meters (Nm^3) for the period from 00:00 to 24:00 h of the next day. In 2015, through Supreme Decree number 15/2015 and resolution number 9664/2015, it was instructed by the Chilean Ministry of Health to declare sanitary alert based on PM2.5 concentrations. Chilean guidelines for PM2.5 and PM10 concentrations are established at maximum values of 50 and 150 (in $\mu g/Nm^3$), during 24 h, respectively (CONAMA 1998; MMA 2011).

2.4 Health Effects of PM in Santiago

Several studies indicate that people living in Santiago are at risk due to the poor air quality of the city (Préndez et al. 2011; Romieu et al. 2012; Requia et al. 2018; Perez-Padilla and Menezes 2019). Ilabaca et al. (1999) evaluated the impact of daily variation of PM2.5 and other pollutants on the number of daily respiratory emergency visits to an important pediatric hospital of Santiago. The authors concluded that atmospheric air pollutant mixtures, especially fine PM, adversely affected the respiratory health of children residing in Santiago, evidenced by an increase in the number of daily respiratory emergency visits. Cifuentes et al. (2000) studied the effect of concentrations of inhalable PM, as well as of other gaseous pollutants, finding an association with increased daily mortality. Traffic combustion-related particles were found to be associated with emergency visits in Santiago (Cakmak et al. 2009). Franck et al. (2014, 2015) provided evidence for increased hospital admissions related to respiratory and cardiovascular diseases, after critical air pollution events in Santiago, showing the influence of combined exposure to airborne pollutants. Recently, Matus and Oyarzún (2019) indicated that an increase of 10 $\mu g/m^3$ of PM2.5 with 1 and 2 days of lag was associated with an increase of near 2% in children's hospitalizations due to respiratory diseases. This percentage increased to 5% when the exposure was with 8 days of lag, reflecting synergism between PM and respiratory viruses.

3 Data Science Methodology for Monitoring Urban Environmental Contamination

3.1 Birnbaum-Saunders np Control Charts

An np-chart is an adaptation of the control chart for nonconforming fraction when samples of equal size (n) are taken from the process (Leiva et al. 2015; Aykroyd et al. 2019). The np-chart is based on the binomial distribution as detailed below. In quality monitoring processes, one could be concerned about a random variable corresponding to the number (D) of times that the quality variable (X) exceeds a fixed value (x) established for the process, given an exceedance probability (p). Here, p can be computed by means of a continuous statistical distribution of the quality variable X as $p = \mathrm{P}(X > x) = 1 - F_X(x)$, where F_X is the cumulative distribution function of X. Thus, D follows a binomial distribution with parameters n and p. Based on this distribution, an np-chart is proposed with lower control limit (LCL), central line (CL), and upper control limit (UCL) given by

$$\text{LCL} = \max\left\{0, np_0 - k(np_0(1-p_0))^{1/2}\right\}, \quad \text{CL} = np_0, \quad \text{UCL} = np_0 + k(np_0(1-p_0))^{1/2},$$

where k is a control coefficient such that $k = 2$ indicates a warning level and $k = 3$ a dangerous level; p_0 is the nonconforming fraction corresponding to a target mean $\mu_X{}^{(0)}$ of the quality variable X, when the process is in control; and n is the size of each subgroup to be monitored. Note that the nonconforming fraction is the probability that the random variable X exceeds a dangerous concentration (x_0), and therefore, this probability is $P(X > x_0) = 1 - F_X(x_0)$. The Birnbaum-Saunders distribution has as one of its parameters the median (Leiva 2016). One can reparameterize the Birnbaum-Saunders distribution switching its median to its mean μ_X (Santos-Neto et al. 2014), with μ_X being the mean of the quality variable X previously defined. Note that this mean-based reparameterization of the Birnbaum-Saunders distribution allows us to have a similar setting as the Gaussian or normal distribution. Therefore, considering x_0 as proportional to $\mu_X{}^{(0)}$, that is, $x_0 = a\mu_X{}^{(0)}$, this permit us to establish a monitoring criterion, where $a > 0$ is a proportionality constant. Note that the target mean $\mu_X{}^{(0)}$ and the dangerous level x_0 can be taken from process specifications. Thus, when a monitoring process is in control ($\mu_X = \mu_X{}^{(0)}$) for a quality variable X following a Birnbaum-Saunders distribution (Leiva et al. 2015), the nonconforming fraction is given by $p_0 = 1 - F_X(x_0)$. Note that the specification of the point x_0 is equivalent to specifying the inspection point $a > 0$, because $x_0 = a\mu_X{}^{(0)}$, in which $\mu_X{}^{(0)}$ is the target mean, which is assumed to be known. Algorithm 1 provides a criterion for monitoring processes using an np-chart for a quality variable X following a Birnbaum-Saunders distribution.

Algorithm 1 np Control Chart Based on the Birnbaum-Saunders Distribution

1. Consider N subgroups of size n.
2. Collect n data $x_1, \ldots, x_n$ of the random variable of interest X for each subgroup.
3. Set the target mean $\mu_X{}^{(0)}$, the inspection constant a, and the control coefficient k.
4. Count in each subgroup of n data the number d of times that x_i exceeds $x_0 = a\mu_X{}^{(0)}$, for $i = 1,\ldots,n$.
5. Compute $\text{LCL} = \max\{0, np_0 - k(n\widehat{p}_0(1-\widehat{p}_0))^{1/2}\}$, $\text{CL} = n\widehat{p}_0$, and $\text{UCL} = n\widehat{p}_0 + k(n\widehat{p}_0(1-\widehat{p}_0))^{1/2}$, where $\widehat{p}_0$ is the maximum likelihood estimate of $p_0 = 1 - F_X(x_0)$, where F_X is the cumulative distribution function of the random variable X defined in step 2.
6. Declare the process as out-of-control if $d >$ UCL or $d <$ LCL or as in-control if $\text{LCL} \leq d \leq \text{UCL}$.

3.2 Standard Bivariate Control Charts

The standard Hotelling T^2 chart (Jackson 1985) is a useful tool for bivariate process control under a normal distribution. Specifically, it assumes that the vector summarizing the quality characteristics, (X_1, X_2), namely, follows a bivariate normal

distribution. To monitor a bivariate process, the following statistical hypotheses are considered:

$$\mathrm{H}_0 : (\mu_1, \mu_2) = \left(\mu_1{}^{(0)}, \mu_2{}^{(0)}\right) \text{ versus } \mathrm{H}_1 : (\mu_1, \mu_2) \neq \left(\mu_1{}^{(0)}, \mu_2{}^{(0)}\right), \tag{1}$$

where $(\mu_1{}^{(0)}, \mu_2{}^{(0)})$ is the target mean vector of an in-control process. Then, the standard T^2 statistic for testing the hypotheses above under a normal distribution is used. In general, the construction of a bivariate control chart considers two phases. In Phase I, a data set of size $N = m \times n$ is obtained from an in-control status of the underlying process, where m is the number of subgroups and n is the size of each subgroup. This data set is used (a) to estimate the parameters of interest; (b) to verify the distributional assumption with goodness-of-fit techniques; (c) to calculate LCL and UCL; and (d) to detect bivariate outliers. Note that a bivariate outlier is considered to be atypical by considering the whole bivariate data and not the value of one given random variable (Marchant et al. 2019). In Phase II, LCL and UCL obtained in Phase I are used to assess whether a data sample for a new subgroup from the process is in control or not. Hence, in Phase II, LCL and UCL are used to detect deviations of the new data set for a target mean value, $(\mu_1{}^{(0)}, \mu_2{}^{(0)})$, namely, or another target parameter of interest. In particular, for standard bivariate control charts, in Phase I, considering a number $m \geq 20$ of subgroups and a size of subgroups greater than one ($n > 1$), the distribution of the standard T^2 test statistic has a closed mathematically form. Then, the corresponding LCL and UCL obtained from T^2 are used (Lowry and Montgomery 1995; Montgomery 2009). Next, in Algorithm 2, we detail how to compute the LCL and UCL of a standard bivariate control chart. The average run length (ARL) is the mean number of points that must be plotted before one of them to indicate an out-of-control status. ARL can be used to evaluate the performance of a control chart. The probability that an observation is considered as out of control, if the process is actually in control, indicates a false alarm rate (FAR) η, which often is in the range 0.01–0.05 (1–5%).

Algorithm 2 Computation of Control Limits in Phase I for Standard Hotelling T^2 Charts

1. Consider two quality characteristics $(X_{1i}, X_{2i}) \sim \mathrm{N}_2((\mu_1, \mu_2), \sum)$ and the data vector (x_{1hi}, x_{2hi}), containing the observations of these two quality characteristics from an in-control process, where $h = 1,\ldots, m$ and $i = 1,\ldots, n$.
2. Obtain the maximum likelihood estimates of (μ_1, μ_2) and $\sum$ using the data of the pooled sample of size $N = m \times n$.
3. Verify the distributional assumption, as well as the presence of bivariate outliers. If the distributional assumption is verified and no bivariate outliers are detected, go to Step 4; otherwise, non-normal and/or robust control charts must be considered.
4. Calculate T^2 assuming a target $(\mu_1{}^{(0)}, \mu_2{}^{(0)})$.
5. Obtain the LCL and UCL for the bivariate control chart of FAR η.

Now, the LCL and UCL obtained in Phase I are used to monitor the process in Phase II, that is, to observe whether the process remains in control as the data of new subgroups are obtained. In Phase II, the T^2 statistic is now denoted by T^2_{new}. Thus, we plot the sequence of values for T^2_{new} in the bivariate control chart corresponding to r subgroups generated in this phase. Next, in Algorithm 3, we indicate how the control chart based on the bivariate normal distribution is utilized to monitor the underlying process.

Algorithm 3 Process Monitoring Using the Standard Hotelling T^2 Chart in Phase II

1. Obtain the new data vector (x_{1hi}, x_{2hi}), for $h = 1,\ldots, r$ and $i = 1,\ldots, n$, but, in this case, it is not necessary that the new data are collected from an in-control process.
2. Obtain the maximum likelihood estimates of (μ_1, μ_2) and $\sum$ with the data of Step 1.
3. Compute T^2_{new} for each sample of new data generated in the hth subgroup, with $h = 1,\ldots,r$, for regular time intervals, obtaining the sequence $t^2_{\text{new}1},\ldots, t^2_{\text{new}r}$.
4. Plot the points $t^2_{\text{new}1},\ldots, t^2_{\text{new}r}$ in the bivariate control chart with LCL and UCL obtained in Phase I.
5. Establish that the process is in control if all points $t^2_{\text{new}1},\ldots,t^2_{\text{new}r}$ fall between LCL and UCL; otherwise, if any of the points $t^2_{\text{new}1},\ldots, t^2_{\text{new}r}$ falls below the LCL or above the UCL, the process is in an out-of-control condition.

Bivariate control charts under a normal distribution use mean vector and variance-covariance matrix estimates, which are sensitive to outliers in Phase I. Bivariate outliers can influence parameter estimates and cause out-of-control conditions not be detected. The identification of outliers is usually based on the Mahalanobis distance (MD. Note that the MD is useful to test goodness of fit in regression models (Marchant et al. 2016b). However, sometimes outliers do not have a large MD, which is known as masking effect (Ben-Gal 2005). This effect is because the maximum likelihood estimators of the model parameters employed to generate the MD are statistically non-robust. Masking effects occur when a group of outliers distorts the estimates of the mean vector and/or variance-covariance matrix, resulting in a small difference from the outlier to the mean. For details on the masking effect, see Marchant et al. (2018) and references therein.

3.3 Bivariate Birnbaum-Saunders Control Charts (Phase I)

There are many practical applications where the normality assumption is not fulfilled, because the data exhibit asymmetrical behavior or heavy tails, as in the case of environmental pollution data. In this perspective, the bivariate Birnbaum-Saunders and Birnbaum-Saunders-t distributions are a good alternative to the bivariate normal distribution. In addition, as mentioned, these distributions have attractive properties, including robustness and a theoretical justification for modeling environmental data (Leiva et al. 2015). Thus, bivariate control charts based on Birnbaum-Saunders and

Birnbaum-Saunders-t distributions can provide a useful methodology for monitoring urban environmental pollution and particularly PM2.5 and PM10. In order to present the methodology to be used, the type Hotelling chart for bivariate process control is considered under Birnbaum-Saunders and Birnbaum-Saunders-t distributions. Specifically, we assume that the vector summarizing the quality characteristics (X_{1i}, X_{2i}) follows a bivariate Birnbaum-Saunders or Birnbaum-Saunders-t distribution, for $i = 1,\ldots, n$. The observed values (data) of these characteristics are (x_{1hi}, x_{2hi}), corresponding to the ith case in the hth subgroup, for $h = 1,\ldots,m$ and $i = 1,\ldots,n$, with $m \geq 20$ and $n > 1$, as mentioned. In addition, suppose that the underlying process is in control and the vectors (X_{1i}, X_{2i}) are independent over time. Now, in practice, we work with the observations (y_{1hi}, y_{2hi}), where $y_{jhi} = \log(x_{jhi})$, for the hth subgroup of the random vector (Y_{1i}, Y_{2i}), which follows a bivariate log-Birnbaum-Saunders and log-Birnbaum-Saunders-t with $h = 1,\ldots,m$, $i = 1,\ldots, n$ and $j = 1, 2$. For details on the bivariate log-Birnbaum-Saunders and log-Birnbaum-Saunders-t distributions, see Marchant et al. (2016b).

Such as in (1), to monitor a bivariate process control, the following statistical hypotheses are considered:

$$H_0 : (\mu_{Y1}, \mu_{Y2}) = \left(\mu_{Y1}^{(0)}, \mu_{Y2}^{(0)}\right) \quad \text{versus} \quad H_1 : (\mu_{Y1}, \mu_{Y2}) \neq \left(\mu_{Y1}^{(0)}, \mu_{Y2}^{(0)}\right) \tag{2}$$

where $\left(\mu_{Y1}^{(0)}, \mu_{Y2}^{(0)}\right) = \left(\log\left(\beta_1^{(0)}\right), \log\left(\beta_2^{(0)}\right)\right)$ is the target mean vector of an in-control process in logarithmic scale (see details in Marchant et al. 2018), with $\beta_j^{(0)}$ being the jth element of the target median vector $\left(\beta_1^{(0)}, \beta_2^{(0)}\right)$. Note that, differently from the normal case, the distribution of the T^2 statistic adapted (T^2_{a}) to the case of bivariate Birnbaum-Saunders distributions does not have a closed mathematically form. Then, we use the parametric bootstrap technique to determine its distribution (Hall 2013). After this distribution is obtained, we compute its quantiles, and then the LCL and UCL of the bivariate quality control chart used in this work are obtained in Phase I. Next, in Algorithms 4 and 5, we detail how to compute the LCL and UCL of bivariate Birnbaum-Saunders (with no outliers) and bivariate Birnbaum-Saunders-t (with outliers) control charts, respectively, with the parametric bootstrap distribution of the T^2_{a} statistic.

Algorithm 4 Computation of Bivariate Birnbaum-Saunders Control Chart Limits in Phase I

1. Consider two positive quality characteristics (X_{1i}, X_{2i}) which follow a bivariate Birnbaum-Saunders distribution and their data vector (t_{1hi}, t_{2hi}) contains the observations of these quality characteristics, where $h = 1,\ldots,m$ and $i = 1,\ldots,n$, with $m \geq 20$ subgroups of size $n > 1$ from an in-control process, as mentioned.
2. Generate the data vector (y_{1hi}, y_{2hi}) containing the logarithms of the data collected in Step 1, where (y_{1hi}, y_{2hi}) can be considered as an observation of (Y_{1i}, Y_{2i}), which follows a bivariate log-Birnbaum-Saunders distribution, with $h = 1,\ldots, m$, $i = 1,\ldots, n$ and $j = 1, 2$.

3. Obtain the maximum likelihood estimates of the corresponding parameters using the data of Step 2 with the pooled sample of size $N = m \times n$, and verify the distributional assumption. If the bivariate Birnbaum-Saunders distributional assumption is verified and no bivariate outliers are detected, go to Step 4; otherwise, non-normal and/or robust control charts must be considered, for example, as that proposed in Algorithm 5.
4. Produce a parametric bootstrap sample $((y^*_{1h1}, y^*_{2h1}),\ldots, (y^*_{1hn}, y^*_{2hn}))$ of size n from a bivariate log-Birnbaum-Saunders distribution using the maximum likelihood estimates computed in Step 3.
5. Calculate T^2_{a} with the bootstrap sample generated in Step 4, denoted by T^{2*}_{a}, assuming a target mean $\left(\mu^{(0)}_{Y1}, \mu^{(0)}_{Y2}\right)$ from the process.
6. Repeat Steps 4–5 a number B of times (for example B = 10,000) and compute B values of the bootstrap statistic of T^2_{a}, denoted by $t^{2*}_{\mathrm{a}1},\ldots, t^{2*}_{\mathrm{aB}}$.
7. Set the desired FAR η of the control chart.
8. Use the B values of the bootstrap statistic obtained in Step 6 to find out the $100 \times (\eta/2)$th and $100 \times (1 - \eta/2)$th quantiles of the distribution of T^2_{a}, which represent the LCL and UCL for the bivariate Birnbaum-Saunders control chart of FAR η, respectively.

Algorithm 5 Computation of Bivariate Birnbaum-Saunders-t Control Chart Limits in Phase I

1. Consider two positive quality characteristics $(X_{1\mathrm{i}}, X_{2\mathrm{i}})$ which follow a bivariate Birnbaum-Saunders-t distribution and their data vector (t_{1hi}, t_{2hi}) which contains the observations of these quality characteristics, where $h = 1,\ldots,m$ and $i = 1,\ldots, n$, with $m \geq 20$ subgroups of size $n > 1$ from an in-control process, as mentioned.
2. Generate the data vector (y_{1hi}, y_{2hi}) containing the logarithms of the data collected in Step 1, where (y_{1hi}, y_{2hi}) can be considered as an observation of (Y_{1i}, Y_{2i}), which follows a bivariate log-Birnbaum-Saunders-t distribution, with $h = 1,\ldots, m$, $i = 1,\ldots,n$ and $j = 1, 2$.
3. Obtain the maximum likelihood estimates of the corresponding parameters using the data of Step 2 with the pooled sample of size $N = m \times n$, and verify the distributional assumption. If the bivariate Birnbaum-Saunders-t distributional assumption is verified, go to Step 4; otherwise, another robust non-normal control charts must be considered.
4. Produce a parametric bootstrap sample $((y^*_{1h1}, y^*_{2h1}),\ldots, (y^*_{1hn}, y^*_{2hn}))$ of size n from a bivariate log-Birnbaum-Saunders-t distribution using the maximum likelihood estimates computed in Step 3.
5. Calculate T^2_{a} with the bootstrap sample generated in Step 4, denoted by T^{2*}_{a}, assuming a target mean $\left(\mu^{(0)}_{Y1}, \mu^{(0)}_{Y2}\right)$ from the process.
6. Repeat Steps 4–5 a number B of times (for example B = 10,000) and compute B values of the bootstrap statistic of T^2_{a}, denoted by $t^{2*}_{\mathrm{a}1},\ldots, t^{2*}_{\mathrm{aB}}$.
7. Set the desired FAR η of the control chart.

8. Use the B values of the bootstrap statistic obtained in Step 6 to find out the $100 \times (\eta/2)$th and $100 \times (1 - \eta/2)$th quantiles of the distribution of T^2_a, which represent the LCL and UCL for the bivariate Birnbaum-Saunders-t control chart of FAR η, respectively.

3.4 Bivariate Birnbaum-Saunders Control Charts (Phase II)

Note that, in Phase I, it is also necessary to verify the distribution assumption by using goodness-of-fit techniques and evaluate bivariate outliers with appropriate methods. Subsequently, the LCL and UCL of the bivariate Birnbaum-Saunders or Birnbaum-Saunders-t control charts obtained in Phase I, and summarized in Algorithms 4 and 5, are used to monitor the process in Phase II, that is, to assess whether the underlying process remains in control as the data of new subgroups are obtained. Therefore, the monitoring process using the bivariate Birnbaum-Saunders control charts is carried out in Phase II and summarized in Algorithm 6, with the adapted Hotelling statistic being denoted by T^2_{anew}.

Algorithm 6 Process Monitoring Using the Bivariate Birnbaum-Saunders and Birnbaum-Saunders-t Chart in Phase II

1. Repeat Steps 1–2 of Algorithms 4 and 5, obtaining the data vector (y_{1hi}, y_{2hi}), for $h = 1,\ldots,r$ and $i = 1,\ldots,n$, but, as mentioned, in this case, it is not necessary that the new data are collected from an in-control process.
2. Calculate the T^2_{anew} statistic for each sample of the new data obtained in Step 1, generated in the hth subgroup, with $h = 1,\ldots, r$, for regular time intervals, generating the sequence $t^2_{anew1},\ldots, t^2_{anewr}$.
3. Plot the points $t^2_{anew1},\ldots, t^2_{anewr}$ in the bivariate Birnbaum-Saunders and Birnbaum-Saunders–t control charts, with LCL and UCL obtained in Phase I.
4. Establish that the process is in control if all points $t^2_{anew1},\ldots, t^2_{anewr}$ fall between LCL and UCL; otherwise, if any of the points $t^2_{anew1},\ldots, t^2_{anewr}$ falls below the LCL or above the UCL, the process is in an out-of-control condition.

4 Case Study in Santiago of Chile

4.1 Data and Air Monitoring Stations in Santiago

We use data collected by the Chilean Metropolitan Environmental Health Service corresponding to the random variables: PM2.5 (X_1) and PM10 (X_2) concentrations, both of them measured in μg/Nm^3. These data are available at https://sinca.mma.gob.cl/ and were collected in 2015 as 1 h (hourly) average values. PM2.5 and

PM10 concentrations were observed in the monitoring stations (MS1)-(MS10) of Fig. 1. For our data analytics, we selected the MS5 station because it had high levels of pollution in the period of critical events of air quality (April 1 to August 31) in Santiago (Marchant et al. 2013). We use data of PM2.5 and PM10 concentrations from the MS5 station and denote them as "PM2015MS5." We employ the Chilean guidelines values as targets in this case study. First, we carry out a correlation analysis to detect if PM2.5 and PM10 concentrations of PM2015MS5 data are statistically associated.

4.2 Data Exploratory Analysis for PM2.5 and PM10

Figure 2 shows the scatterplot for PM2.5 and PM10 concentrations. From this figure, we detect that there is a high positive association between these concentrations being corroborated by a Pearson coefficient of correlation equal to 0.85 in MS5 station. Therefore, in order to monitor urban environmental pollution in Santiago, following the guideline of the Chilean Ministry of Health, which indicates that both PM2.5 and PM10 must be considered, we propose to use a methodology based on np and bivariate statistical control charts. This methodology allows us, on the one hand, to determine the number of times that the Chilean guidelines for PM2.5 and PM10 are exceeded each hour of the day and, on the other hand, to monitor PM10 and

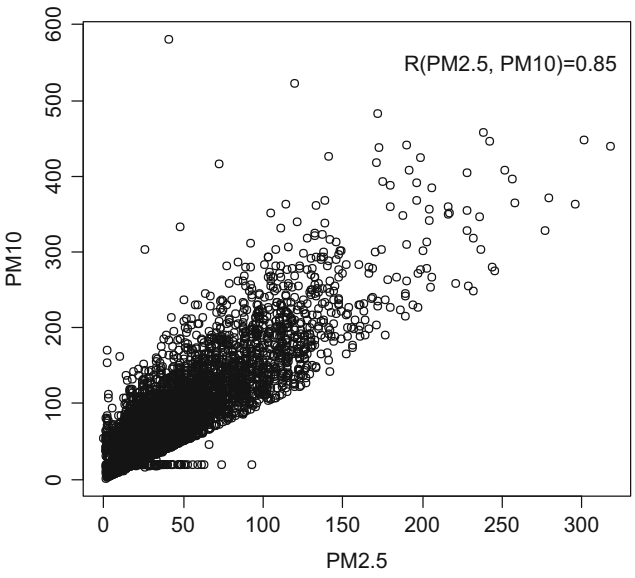

Fig. 2 Scatterplot and correlation between PM2.5 and PM10 concentrations with PM2015MS5 data

Table 1 Summary statistics for data sets in study

	Phase I		Phase II	
Statistic	PM2.5	PM10	PM2.5	PM10
n	1416	1416	744	744
Minimun	3	10	1	3
Mean	19.08	54.75	34.77	60.62
Median	18	51.5	29	47
Maximun	76	214	216	350
SD	8.68	24.13	26.33	50.15
CV	45.49%	44.06%	75.73%	82.73%
CS	1.78	1.304	2.04	1.9
CK	5.55	3.61	7.83	4.62

PM2.5 concentrations simultaneously, predicting critical periods of contamination adequately.

We use the methodology proposed in Sect. 3 to monitor PM pollution in Santiago, Chile. This methodology was implemented by the authors in a computational routine in R, a noncommercial and open-source software for statistical computing and graphs, which may be secured at no cost from http://www.r-project.org. The R software is currently very popular in the international scientific community and is one of the more used around of world. We carry out an exploratory data analysis for PM2015MS5. Table 1 provides descriptive statistics for each variable, including minimum and maximum concentrations, central tendency statistics, standard deviation (SD), and coefficients of variation (CV), skewness (CS), and kurtosis (CK). This table shows empirical distributions with positive skewness, different degrees of kurtosis, and a considerable amount of concentrations that exceed the Chilean guidelines for PM2.5 and PM10, that is, 50 μg/Nm^3 and 150 μg/Nm^3, respectively. Note that the exploratory analysis for each variable indicates marginal Birnbaum-Saunders or Birnbaum-Saunders-t distributions that seem to be good candidates for describing PM2015MS5 data.

4.3 Univariate and Bivariate Control Charts

To calculate the control limits in Phase I, we utilize data for the months of January and February of 2019 with $k = 59$, $n = 24$, $N = 1416$, B = 10,000 (bootstrap replications) and FAR $\eta = 0.0027$. We use these months since their air quality is stable (i.e., assumed as an in-control process), because the meteorological and topographical conditions favor no saturation of PM concentrations. We consider the degrees of freedom parameter $v = 4$ for the Birnbaum-Saunders-t distribution according to the robustness aspects mentioned in Marchant et al. (2016a). Note that this parameter v allows us accommodate outliers suitably.

Figure 3a, b displays the theoretical probability versus empirical probability (PP) plots with acceptance bands for a significance level of 1% in MS5 station

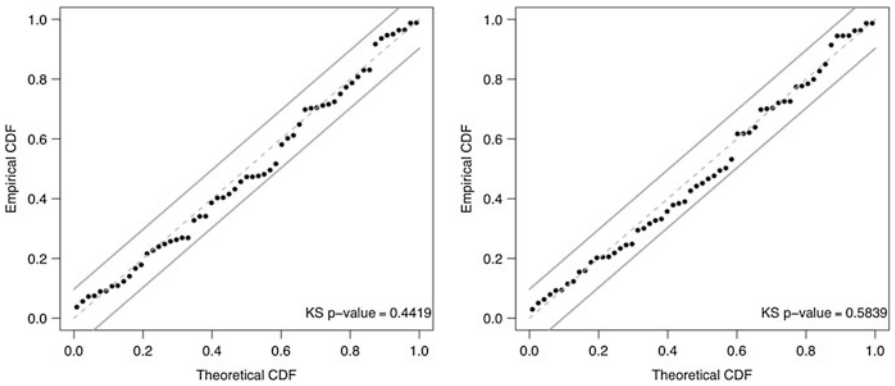

Fig. 3 PP plots in Phase I for bivariate Birnbaum-Saunders (left) and Birnbaum-Saunders-t (right) distributions with PM2015MS5 data

based on bivariate Birnbaum-Saunders and bivariate Birnbaum-Saunders-t distributions, respectively. From this figure, we confirm the good fit of the bivariate Birnbaum-Saunders and bivariate Birnbaum-Saunders-t distributions to the data in Phase I, which is supported by the p-values 0.4419 and 0.5839, respectively, of the Kolmogorov-Smirnov (KS) test associated with these PP plots (Marchant et al. 2016a). To monitor air quality of August 2015 in Phase II, we employ bivariate Birnbaum-Saunders and np univariate Birnbaum-Saunders charts. For bivariate Birnbaum-Saunders-t charts, we use the LCL and UCL obtained in Phase I summarized in Algorithm 5. For the control chart of this month, the number of subgroups and the subgroup size are $r = 31$ days and $n = 24$ h, respectively, giving a total of 744 observations. Furthermore, we use the transformed MD with the Wilson-Hilferty approximation to obtain a normal distribution and then to assess the fit of the most appropriate distribution to these data. For details about the Wilson-Hilferty approximation, see Marchant et al. (2016a, b).

Figure 4a, b displays the PP plots with acceptance bands for a significance level of 1% in MS5 station based on bivariate Birnbaum-Saunders and Birnbaum-Saunders-t distributions, respectively. From this figure, the Birnbaum-Saunders-t distribution has a better fit than the Birnbaum-Saunders distribution to the data in Phase II, which is supported by the p-values 0.609 and 0.3091, respectively, of the KS test associated with these PP plots (Marchant et al. 2016a). Due to such a situation, we continue this study only with the Birnbaum-Saunders-t distribution because of its robust estimation of parameters and the good fit to these data.

Figure 5 shows the bivariate Birnbaum-Saunders-t control chart for PM2015MS5 data. From this figure, it is possible to observe that there are no out-of-control episodes during this month, when considering the joint behavior of PM2.5 and PM10 concentrations. This type of chart is appropriate when there is a high correlation between the concentrations of PM2.5 and PM10 such as in our case; see Fig. 2. In addition, to monitor the number of times that the Chilean guidelines for PM2.5 and PM10 are exceeded each hour of the day of August-2015 in Phase II, we employ

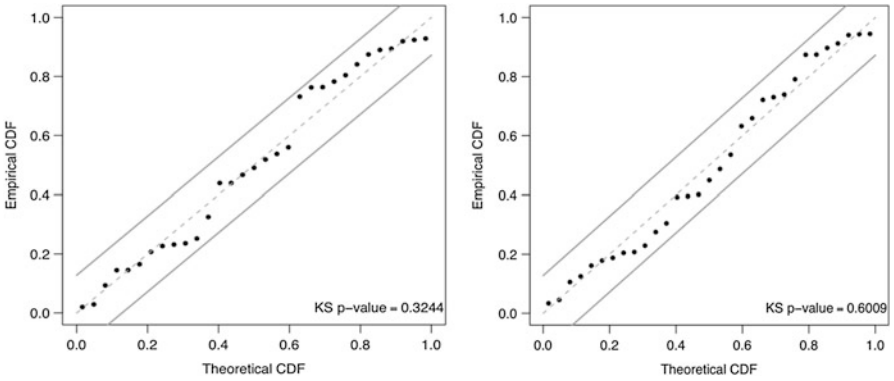

Fig. 4 PP plots in Phase II for bivariate Birnbaum-Saunders (left) and Birnbaum-Saunders-t (right) distributions with PM2015MS5 data

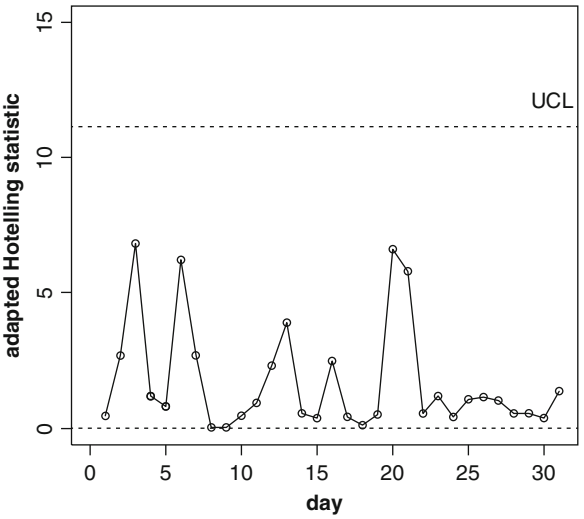

Fig. 5 Bivariate Birnbaum-Saunders-t chart for August 2015 with PM2015MS5 data

the univariate Birnbaum-Saunders np-chart. For the univariate chart of this month, the number of subgroups and the subgroup size are the same as for the bivariate chart.

Figure 6a, b shows the np Birnbaum-Saunders control charts for PM2.5 and PM10 concentrations, respectively. From this figure, it is possible to observe that the concentrations of PM2.5 exceed the Chilean guidelines much more than those of PM10. This occurs specifically in the first days of the month of August, where the Chilean guidelines are exceeded more than 15 times during a day, considering a maximum of 24 observations each day. Such a situation is highly detrimental to health as consequence of breathing air with high concentrations of PM, especially PM2.5, as mentioned in Sects. 1 and 2.

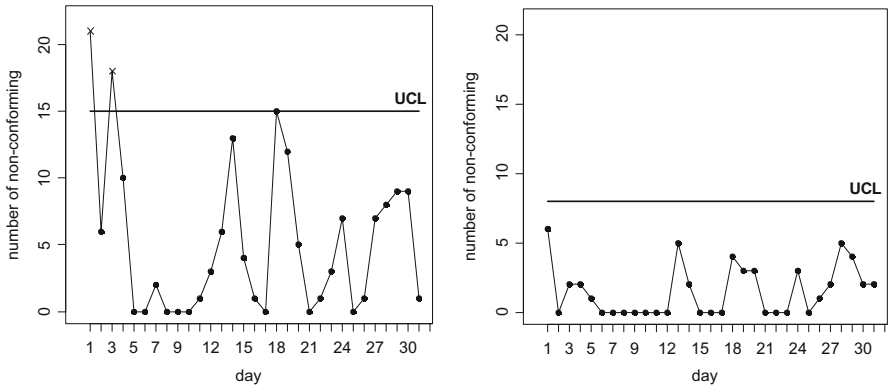

Fig. 6 np Birnbaum-Saunders chart for August 2015 for PM2.5 (left) and PM10 (right) concentrations with PM2015MS5 data

4.4 Big Data, Analytics Results, and Its Connection to Data-Driven Decision-Making

Current technologies, such as computer-based transactions, digital instruments, and sensors, allow us to generate large-scale data from different processes. These data may be collected efficiently, rapidly, and automatically and are frequently available online for decision-makers and analysts access. This is known as big data, a term often employed to describe large, diverse, and complex (structured and non-structured) data sets, which have high volume, variety, and velocity in their generation, otherwise known as the 3Vs (Baesen 2014; Dietrich 2015). Such a situation results in enormous opportunities for data-based knowledge discovery, and it is expected that the importance of data science will continue to increase in the future, becoming relevant for researchers in diverse areas who will be ready to exploit new opportunities for data-driven decision-making. In this new big data era, many methodologies need to be updated, as, for example, control charts (Aykroyd et al. 2019), where big data sources are providing new avenues for such charts because of continuous monitoring in diverse fields.

Control charts have primarily been used to monitor industry processes (Montgomery 2009), but recently, these methodologies are also being used to monitor service processes, such as banking and finance, distribution of electrical energy, public transportation, and retail (Aykroyd et al. 2019). Furthermore, control charts have also been used in education, government policies, healthcare, and marketing. The use of control charts in environmental monitoring processes is currently limited, but we expect it may become a popular alternative in the big data era.

The data-driven methodology proposed in this paper, based on tools of control charts for environmental monitoring, shows a good performance when assessing air quality, particularly when two correlated statistically variables are considered. The

analytics results obtained with our methodology are consistent with the official information of the Chilean Ministry of the Environment for PM10 (https://bit.ly/2W2jb1V). Specifically, there is an agreement between the critical episodes empirically detected with our methodology based on the robust bivariate Birnbaum-Saunders control chart and those verified by the Chilean Health Authority, that is, if our methodology were used, the same environmental decision made by the authority would be established. Note that joint analysis of PM10 and PM2.5 concentrations permits us to monitor air quality using one model, instead of employing two models as currently applied to perform this monitoring. With the current model, the interaction and/or dependence of the PM10 and PM2.5 is not considered. In addition, this data science tool helps to prevent and/or adequately alert the population about possible critical episodes of air contamination, providing support to regulatory decision-making when appropriate mitigation measures are needed, such as the prohibition of outdoor physical activities or domestic coal or firewood burning or restrictions on the use of internal combustion vehicles.

5 Summary

Airborne particulate matter pollution is a serious environmental problem. We propose that monitoring of air quality may be achieved by employing data analytics to generate information within the context of data-driven decision making to prevent and/or adequately alert the population about possible critical episodes of air contamination. In this paper, we propose a methodology for monitoring particulate matter pollution in Santiago of Chile, based on bivariate quality control charts and an asymmetric distribution. A case study with real particulate matter pollution from Santiago is provided, which shows that the methodology is suitable to alert early episodes of extreme air pollution. The results are in agreement with the critical episodes reported with the current model used by the Chilean health authority.

6 Conclusions

In this work, we have proposed and implemented a methodology based on bivariate control charts with heavy-tailed asymmetric distributions. These distributions have a theoretical support and can be applied to atmospheric environmental data. This methodology is useful for monitoring environmental risk when the particulate matter concentrations follows bivariate Birnbaum-Saunders or Birnbaum-Saunders-Student-t distributions. We have illustrated the proposed methodology with a case study of real-world data of air quality in Santiago, Chile. This case study has shown that the new methodology is useful for alerting episodes of extreme urban environmental pollution, allowing us to prevent adverse effects on human health for the population of Santiago. We have empirically demonstrated an agreement

between our methodology and real-world situations, as when the Chilean health authority detected environmental critical episodes and dictated environmental alert, pre-emergency, and emergency in Santiago, Chile.

The random variables to be modeled, related to PM2.5 and PM10 concentrations, correspond to an aggregation of a great amount of compounds, which are adsorbed on a solid or liquid surface in the atmosphere. Depending on the diverse composition of these particles, they can show different reactivity and balance between their degradation and production processes. Sometimes these particles show seasonal time dependence, or if the sources vary, their concentration may also vary, such as when drastic measures are taken to reduce their emissions. Therefore, a limitation of this study is not considering time series components in the modeling, which is an open problem to be conducted in future research (Decanini and Volta 2003; Querol et al. 2004; Kessler et al. 2010). Regarding this, we must mention that PM2.5 and PM10 levels are considered simple dilutions and concentrations in the air masses, without taking into account factors such as composition, chemical reactivity, production-degradation equilibrium, and evolution over time of these quantities. In particular, the chemical reactivity of these compounds obeys to different kinetic processes of formation and degradation, described in detail by Sander et al. (2006). Thus, this last aspect is also an open issue to be considered in a future study.

Acknowledgments The research of C. Marchant was partially supported by UCM 434220 grant from the Dirección de Investigación of the Universidad Católica del Maule, Chile. The research of V. Leiva was partially supported by grant "Fondecyt 1160868" from the National Commission for Scientific and Technological Research of the Chilean government. The research of F. Rojas was partially supported by grant "Fondecyt Iniciación 11190004" from the National Commission for Scientific and Technological Research of the Chilean government.

Statement of Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Adams K, Greenbaum DS, Shaikh R, van Erp AM, Russell AG (2015) Particulate matter components, sources, and health: systematic approaches to testing effects. J Air Waste Manag Assoc 65:544–558

Aitchison J, Brown JAC (1973) The lognormal distribution. Cambridge University Press, London

Athayde E, Azevedo A, Barros M, Leiva V (2019) Failure rate of Birnbaum-Saunders distributions: shape, change-point, estimation and robustness. Braz J Probab Stat 33:301–328

Aykroyd RG, Leiva V, Ruggeri F (2019) Recent developments of control charts, identification of big data sources and future trends of current research. Technol Forecast Soc Change 144:221–232

Baesen B (2014) Analytics in a big data world: the essential guide to data science and its applications. Wiley, New York

Ben-Gal I (2005) Outlier detection. In: Maimon O, Rockach L (eds) Data mining and knowledge discovery handbook. Springer, Boston, pp 321–352

Cakmak S, Dales RE, Gultekin T, Vidal CB, Farnendaz M, Rubio MA, Oyola P (2009) Components of particulate air pollution and emergency department visits in Chile. Arch Environ Occup Health 64:148–155

Cifuentes LA, Vega J, Köpfer K, Lave LB (2000) Effect of the fine fraction of particulate matter versus the coarse mass and other pollutants on daily mortality in Santiago, Chile. J Air Waste Manag Assoc 50:1287–1298

CONAMA (1998) Establishment of primary quality guideline for PM10 that regulates environmental alerts. Technical Report Decree 59, Ministry of Environment (CONAMA) of the Chilean Government, Santiago, Chile

Decanini E, Volta M (2003) Application to Northern Italy of a new modelling system for air quality planning: a comparison between different chemical mechanisms. Int J Environ Pollut 20:85–95

Díaz-García JA, Leiva V (2005) A new family of life distributions based on elliptically contoured distributions. J Stat Plan Inference 128:445–457

Dietrich D (2015) Data science and big data analytics: discovering, analyzing, visualizing and presenting data. Wiley, New York

Falcon-Rodriguez CI, Osornio-Vargas AR, Sada-Ovalle I, Segura-Medina P (2016) Aeroparticles, composition, and lung diseases. Front Immunol 7:1–9

Ferreira M, Gomes MI, Leiva V (2012) On an extreme value version of the Birnbaum-Saunders distribution. Rev Stat J 10:181–210

Franck U, Leitte AM, Suppan P (2014) Multiple exposures to airborne pollutants and hospital admissions due to diseases of the circulatory system in Santiago de Chile. Sci Total Environ 15:746–756

Franck U, Leitte AM, Suppan P (2015) Multifactorial airborne exposures and respiratory hospital admissions – the example of Santiago de Chile. Sci Total Environ 502:114–121

Hall P (2013) The bootstrap and edge worth expansion. Springer, New York

Henríquez G, Urrea C (2017) Association between air pollution and emergency consultations for respiratory diseases. Rev Med Chil 145:1371–1377

Hime NJ, Marks GB, Cowie CT (2018) A comparison of the health effects of ambient particulate matter air pollution from five emission sources. Int J Environ Res Public Health 15:1206

Ilabaca M, Olaeta I, Campos E, Villaire J, Tellez-Rojo M, Romieu I (1999) Association between levels of fine particulate and emergency visits for pneumonia and other respiratory illnesses among children in Santiago, Chile. J Air Waste Manag Assoc 49:154–163

International Agency on Cancer Research (2016) Outdoor air pollution. IARC monographs on the evaluation of carcinogenic risks to humans, vol 109, Lyon, France

Jackson J (1985) Multivariate quality control. Comm Stat Theor Meth 14:2657–2688

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. Environ Sci Technol 44:7005–7010

Kim KH, Kabir E, Kabir S (2015) A review on the human health impact of airborne particulate matter. Environ Int 74:136–143

Kinney PL (2008) Climate change, air quality, and human health. Am J Prev Med 35:459–467

Leiva V (2016) The Birnbaum-Saunders distribution. Academic Press, New York

Leiva V, Barros M, Paula GA, Sanhueza A (2008) Generalized Birnbaum-Saunders distribution applied to air pollutant concentration. Environmetrics 19:235–249

Leiva V, Marchant C, Ruggeri F, Saulo H (2015) A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. Environmetrics 26:463–476

Logan WP (1953) Mortality in the London fog incident, 1952. Lancet 1:336–338

Lowry CA, Montgomery DC (1995) A review of multivariate control charts. IEE Trans 27:800–806

Marchant C, Leiva V, Cavieres M, Sanhueza A (2013) Air contaminant statistical distributions with application to PM10 in Santiago, Chile. Rev Environ Contam Toxicol 223:1–31

Marchant C, Leiva V, Cysneiros FJA (2016a) A multivariate log-linear model for Birnbaum-Saunders distributions. IEEE Trans Reliab 65:816–827

Marchant C, Leiva V, Cysneiros FJA, Vivanco JF (2016b) Diagnostics in multivariate generalized Birnbaum-Saunders regression models. J Appl Stat 43:2829–2849
Marchant C, Leiva V, Cysneiros FJA, Liu S (2018) Robust multivariate control charts based on Birnbaum-Saunders distributions. J Stat Comput Sim 88:182–202
Marchant C, Leiva V, Christakos G, Cavieres MF (2019) Monitoring urban environmental pollution by bivariate control charts: New methodology and case study in Santiago, Chile. Environmetrics 30:e2551
Matus CP, Oyarzún GM (2019) Impact of particulate matter (PM 2.5) and children's hospitalizations for respiratory diseases. A case cross-over study. Rev Chil Pediatr 90:166–174
Mendoza Y, Loyola R, Aguilar A, Escalante R (2019) Valuation of air quality in Chile: the life satisfaction approach. Soc Indic Res:1–21
MMA (2011) Establishment of primary quality guideline for inhalable fine particulate matter PM2.5. Technical Report Decree 12, Ministry of Environment (MMA) of the Chilean Government, Santiago, Chile
Montgomery DC (2009) Introduction to statistical quality control. Wiley, New York
Morales RGE, Llanos A, Merino M, Gonzalez-Rojas CH (2012) A semi-empirical method of PM10 atmospheric pollution forecast at Santiago de Chile city. NEPT 11:181–186
Ostro BD (2003) Air pollution and its impacts on health in Santiago, Chile. In: McGranahan G, Murray F (eds) Air pollution and health in rapidly developing countries, vol 189. Earthscan, London
Ott WR (1990) A physical explanation of the lognormality of pollution concentrations. J Air Waste Manag Assoc 40:1378–1383
Perez-Padilla R, Menezes AMB (2019) Chronic obstructive pulmonary disease in Latin America. Ann Glob Health 85:7
Préndez M, Alvarado G, Serey I (2011) Some guidelines to improve the air quality management of Santiago, Chile. In: Air quality monitoring, assessment and management. IntechOpen, London
Querol X, Alastuey A, Viana MM, Rodriguez S, Artiñano B, Salvador P, Garcia do Santos S, Fernandez Patier R, Ruiz CR, de la Rosae J, Sanchez de la Campa A, Menendez M, Gil JI (2004) Speciation and origin of PM10 and PM2. 5 in Spain. J Aerosol Sci 35:1151–1172
Requia WJ, Adams MD, Arain A, Papatheodorou S, Koutrakis P, Mahmoud M (2018) Global association of air pollution and cardiorespiratory diseases: a systematic review, meta-analysis, and investigation of modifier variables. Am J Public Health 108:S123–S130
Romieu I, Gouveia N, Cifuentes LA, de Leon AP, Junger W, Vera J, Strappa V, Hurtado-Díaz M, Miranda-Soberanis V, Rojas-Bracho L, Carbajal-Arroyo L, Tzintzun-Cervantes G, HEI Health Review Committee (2012) Multicity study of air pollution and mortality in Latin America (the ESCALA study). Res Rep Health Eff 171:5–86
Sander SP, Golden DM, Kurylo MJ et al (2006) Chemical kinetics and photochemical data for use in atmospheric studies evaluation number 15. Technical Report JPL Publication 06-2, Jet Propulsion Laboratory, NASA, Pasadena, CA
Santos-Neto M, Cysneiros FJA, Leiva V, Barros M (2014) On a reparameterized Birnbaum-Saunders distribution and its moments, estimation and applications. REVSTAT Stat J 12:247–272
Thompson JE (2018) Airborne particulate matter: human exposure and health effects. Int J Occup Environ Med 60:392–423
Toro R, Kvakić M, Klaić ZV, Koračin D, Morales RG, Leiva MA (2019) Exploring atmospheric stagnation during a severe particulate matter air pollution episode over complex terrain in Santiago, Chile. Environ Pollut 244:705–714
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. Stoch Environ Res Risk Assess 24:771–782
Villalobos AM, Barraza F, Jorquera H, Schauer JJ (2015) Chemical speciation and source apportionment of fine particulate matter in Santiago, Chile, 2013. Sci Total Environ 512:133–142

WHO (2000) Air quality guidelines for Europe. Chapter 7.3 Particulate matter. World Health Organization, Regional Office for Europe, Copenhagen

WHO (2013) Health effects of particulate matter. Technical report, World Health Organization, Regional Office for Europe, Copenhagen

Wu W, Jin Y, Carlsten C (2018) Inflammatory health effects of indoor and outdoor particulate matter. J Allergy Clin Immunol 14:833–844