Atmospheric Research 241 (2020) 104956

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

The statistical behavior of *PM10* events over guadeloupean archipelago: Stationarity, modelling and extreme events

Thomas Plocoste[a,b,*], Rudy Calif[b], Lovely Euphrasie-Clotilde[a], France-Nor Brute[b]

[a] Department of Research in Geoscience, KaruSphère SASU, Abymes 97139, Guadeloupe (F.W.I.), France
[b] Univ Antilles, LaRGE Laboratoire de Recherche en Géosciences et Energies (EA 4935), Pointe-à-Pitre F-97100, France

ARTICLE INFO

Keywords:
*PM*10 Statistical analysis
Stationarity
Mixture models
Extreme events
Caribbean area

ABSTRACT

Environmental pollution management is one of the most important features in pollution risk assessment. Several studies have shown that exposure to particulate matter with an aerodynamic diameter of 10 μm or less, i.e. *PM*10, were associated to adverse health effects. To our knowledge, no study has yet investigated the modelling of *PM*10 frequency distribution and extreme events in the Caribbean basin. Here, the descriptive statistics and four theoretical distributions (lognormal, Weibull, Burr and stable) were used to fit the parent distribution of *PM*10 daily average concentrations in Guadeloupe archipelago with a database of 11 years. In order to determine the best distribution, the Kolmogorov–Smirnov statistic test (*KS* test) was computed as performance indicator value. With an annual average of 26.4 ± 16.1 $\mu g/m^3$, the descriptive statistics highlighted that *PM*10 concentrations in Guadeloupe are lower than those measured in cities of Europe, Asia or Africa. Contrary to other megacities, we found that high *PM*10 levels in Guadeloupe are mainly due to natural large-scale sources, i.e. African dust. From May to September, i.e. high dust season, *PM*10 concentrations are 1.5 times larger since dust outbreaks are more frequent. A statistical stationarity threshold of 66 months is estimated using the distribution analysis. This underlines the cycle stability of African dust over this last decade. Concerning the statistical modelling, our results showed that Burr & Weibull mixture model is the best distribution to represent *PM*10 daily average concentrations with a first statistical behavior corresponding to the low dust season and an another to the high dust season. By analysing the extreme events statistic with the classical power-law distribution, we observed that Burr & Weibull mixture model could also improve the modelling of these events. In summary, the Burr & Weibull mixture model is suitable to model both classical and extreme events.

1. Introduction

Air pollution is one of the challenging environmental problems in Caribbean area for last decade. The concentrations of air pollutants are usually random variables influenced by the emission level, meteorological conditions and topography (Lu, 2002). Each area is specific and the required emission reduction to meet air quality standard is different (Mijić et al., 2009). Information about the frequency distribution of pollutants is crucial for elaborating air pollution strategies. When the specific probability function of an air pollutant is identified, it is easier to predict statistically the required emission reduction, the mean concentration, the frequency of exceedance of the air quality standard, as well as the return period (Lu and Fang, 2003; Mijić et al., 2009). It is also efficient in order to investigate the similarities and differences among the types of air pollution of different areas.

In literature, different types of probability distribution have been used to fit the pollutant concentrations, including lognormal distribution (Gavriil et al., 2006; Dong et al., 2017), Weibull distribution (Georgopoulos and Seinfeld, 1982), gamma distribution (Lu, 2004)] log-logistic distribution (Karaca et al., 2005) and type V Pearson distribution (Morel et al., 1999), to cite a few. The lognormal distribution was widely used to represent the type of air pollutant concentration distribution (Lu, 2002; Lu and Fang, 2003; Xi et al., 2013). Indeed, the fact that air pollutants concentration tends to be a lognormal distribution has been explained by the theory of successive random dilutions (Ott, 1990). After the pollutants are emitted by the source, in the transport process before they reach the receptor, they undergo successive mixing and diluting, resulting in a lognormal frequency distribution (Ott, 1990; Lu, 2002). According to Kao and Friedlander (1995); Kan and Chen (2004), this distribution is the most suitable to represent

* Corresponding author at: Department of Research in Geoscience, KaruSphère SASU, Abymes 97139, Guadeloupe
E-mail addresses: karusphere@gmail.com (T. Plocoste), rudy.calif@univ-antilles.fr (R. Calif), lovelyeuphrasie@hotmail.fr (L. Euphrasie-Clotilde), francenor.brute@univ-antilles.fr (F.-N. Brute).

https://doi.org/10.1016/j.atmosres.2020.104956
Received 12 October 2019; Received in revised form 9 February 2020; Accepted 17 March 2020
Available online 19 March 2020
0169-8095/

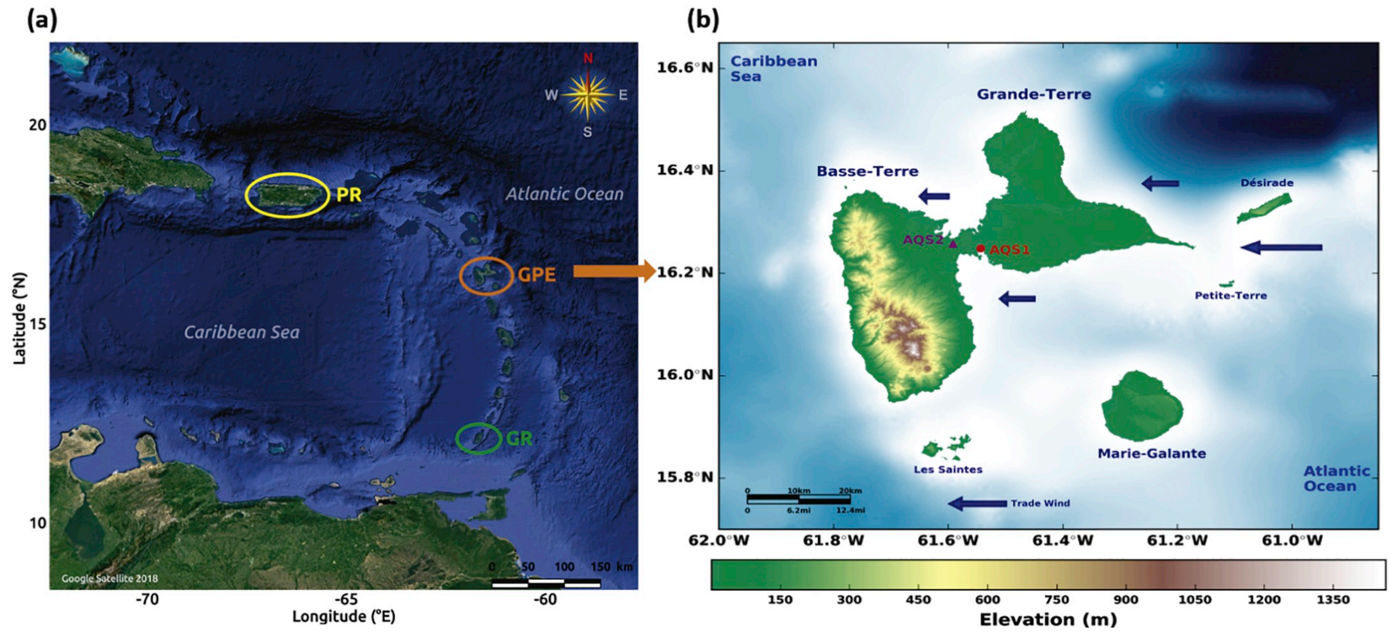

Fig. 1. (a) shows an overview of the West Indies arc with Puerto-Rico at the top (18.23°N, −66.50°W; PR in yellow), Guadeloupe archipelago in the middle (16.25°N, −61.58°W; GPE in orange), and Grenada at the bottom (12.10°N, −61.68°W; GR in green). (b) is a zoom of Guadeloupe archipelago (Topography of IGN 25 m and Bathymetry GEBCO 900 m) with the locations of the Air Quality stations at Pointe-à-Pitre (AQS1, in a red circle) and Baie-Mahault (AQS2, in a purple triangle). The blue arrows indicate Trade winds direction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

particulate matters concentrations with an aerodynamic diameter of less than 10 μm, i.e. $PM10$. Particulate matter are coming from several sources such as vehicle emissions, industrial and domestic emissions, forest fires, cigarette smoke, natural trees, marine aerosols and mineral dust. To our knowledge, no study has yet investigated the modelling of $PM10$ frequency distribution and extreme events in the Caribbean area.

Several epidemiological studies have already shown that exposure to $PM10$ concentrations induce adverse health effects as respiratory (Pope III et al., 1991; Wong et al., 1999; Hoek et al., 2012) and cardiovascular diseases (Schwartz and Morris, 1995; Le Tertre et al., 2002; Polichetti et al., 2009). According to a study by Medina et al. (2004) in 19 European cities, a mere reduction in $PM10$ concentrations by only 5 $\mu g/m^3$ would prevent between 3300 and 7700 deaths per year. Additional research has confirmed that an increase of 10 $\mu g/m^3$ in $PM10$ results in an increased risk of hospitalization for myocardial infarction (Maheswaran et al., 2005; Zanobetti and Schwartz, 2006). In other studies, a relationship between $PM10$ levels and preterm births was highlighted (Ritz et al., 2000; Hansen et al., 2006; Suh et al., 2009; Stieb et al., 2012; Zhao et al., 2015). All these studies show the importance of a good knowledge of $PM10$ concentrations behavior in the Atmospheric Boundary Layer (ABL) in order to elaborate strategies to reduce these impacts.

Here, the descriptive statistics and distribution characteristics of daily average $PM10$ concentration events in Guadeloupe archipelago were analyzed for 11 years between 2005 and 2017. The best probability density function, i.e. goodness of fit, were validated using the Kolmogorov–Smirnov test. This approach has been applied previously in numerous studies (Karaca et al., 2005; Bigi and Harrison, 2010; Lonati et al., 2011; Kwon et al., 2015; Sajjadi et al., 2017), to mention a few.

In order to carry out this work, that paper is organized as follows. Firstly, Section 2 describes the study area and the data sets used. Thereafter, Section 3 recalls the statistical framework and the statistical models. Then, the achieved results are presented and discussed in section 4. To conclude, an outlook for future studies are given in Section 5.

2. Study area and data collection

Guadeloupe archipelago is a French West Indies island with an area of ~1800 km^2 located at the North of the Lesser Antilles (16.25°N −61.58°W, 440,000 inhabitants, Fig. 1(a) (Plocoste et al., 2019). Hourly $PM10$ data were provided by Guadeloupe air quality network which is managed by Gwad'Air agency (http://www.gwadair.fr/). The Air Quality Stations (AQS) are located at Pointe-à-Pitre (AQS1, 16.2422°N 61.5414°W, urban area) and Baie-Mahault (AQS2, 16.2561°N 61.5903°W, suburban area). Between 2005 and 2017, there was only one $PM10$ sensor available on this air quality network which was successively placed in Pointe-à-Pitre from 2005 to 2012 and Baie-Mahault from 2015 to nowadays. Between AQS1 and AQS2 the distance is approximately 8.1 km. With wind speed measurements made by Météo France on the international airport of Pôle Caraïbes at Abymes (16.2630°N 61.5147°W), an average of 3.2 m/s (11.5 km/h) were computed between 2005 and 2017. This means that an air mass can travel a distance of 11.5 km in 1 h. Thus, in the study area, the distance travelled by an air mass in 1 h is greater than the distance from AQS1 to AQS2. Considering the frozen turbulence assumption (Tennekes et al., 1972), we assume in first approximation that data from AQS1 and AQS2 can be concatenate. To conduct this study, hourly data is converted into daily average values. Between 2005 and 2017, 3849 points are available; i.e. 96% of potential data. Fig. 2 displays $PM10$ daily measurements available for this study. $PM10$ signals show huge fluctuations at AQS1 and AQS2 indicating a strong variability of $PM10$ data.

3. Methods

Here, the statistical framework used to analyze $PM10$ data is presented.

3.1. Descriptive statistics

In this section, we briefly recall the first four central moments

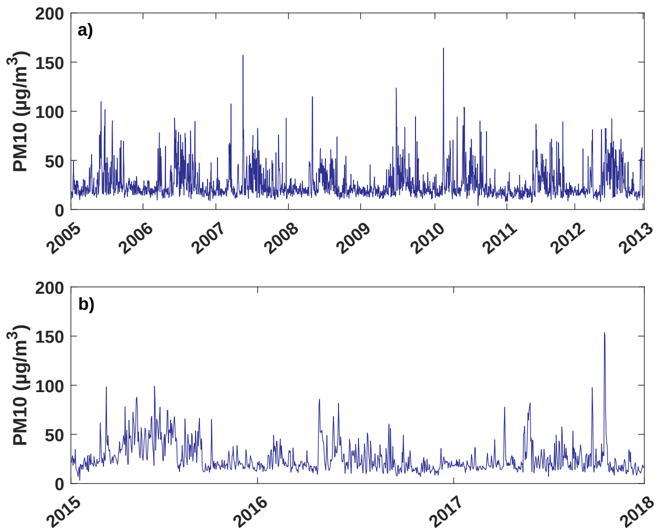

Fig. 2. *PM*10 times series analyzed in this work for (a) AQS1 and (b) AQS2.

frequently used to analyze a discrete process, i.e. the mean (Eq. 1), standard deviation (Eq. 2), skewness (Eq. 3) and kurtosis (Eq. 4) (Papoulis and Pillai, 2002):

$$\overline{M} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{2}$$

$$S = \frac{\overline{(x-\overline{x})^3}}{(\overline{(x-\overline{x})^2})^{\frac{3}{2}}} \tag{3}$$

$$K = \frac{\overline{(x-\overline{x})^4}}{(\overline{(x-\overline{x})^2})^2} \tag{4}$$

Because of the high order, kurtosis is particularly sensitive to extreme or intermittent fluctuations (Windsor and Toumi, 2001). Indeed, kurtosis can be a useful indicator of intermittency in pollution studies (Windsor and Toumi, 2001; Plocoste et al., 2018). Highly intermittent time series will have a higher kurtosis.

Classically, the four first moments are used to describe statistically data behavior. For a profound statistic description, the data distribution can be sharply estimated.

3.2. Probability density function

The Probability Density Function (PDF) of a random variable X is a description of the distribution of the values of the random variable. We denote the density function of a signal x as $f(x)$. According to Papoulis and Pillai (2002), the PDF is defined as:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{5}$$

3.3. Kernel density estimation

To estimate the empirical distribution of *PM*10 concentrations, we use a kernel density estimation. In statistics, a kernel distribution is a non-parametric representation of the probability density function of a random variable (Taskin and Zaim, 2000). Also known as Parzen–Rosenblatt window method (Koutsoukas et al., 2013), this distribution provides a simple way of finding structure in data sets without the imposition of a parametric model (Wand and Jones, 1994). Unlike a histogram, that ranks values into discrete bins, a kernel distribution sums the component smoothing functions for each data value to produce a smooth, continuous probability curve. For all x real values, kernel density estimator's formula is given by Silverman (2018) with:

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left[\frac{x - x_i}{h}\right] \tag{6}$$

where $x_1, x_2, \ldots, x_n$ are random samples from an unknown distribution, n is the sample size, $K(\cdot)$ is the kernel smoothing function and h is the bandwidth commonly called the smoothing parameter. As it can be noticed in Eq. 6, our estimate $\widehat{f}_h$ is bin-independent regardless of our choice of K. The aim of K is to spread out the contribution of each data point in our estimate of the parent distribution (Cranmer, 2001). There are different types of kernel smoother: normal, box, triangle, epanechnikov (Bowman et al., 1998). In the literature, normal kernel is usually recommended (Silverman, 2018). The following kernel is used for estimating the distribution of *PM*10 concentrations (Calif, 2012):

$$K\left[\frac{x - x_t}{h}\right] = \left[\frac{1}{\sigma\sqrt{\pi}}\right]\exp\left[-\frac{(x - x_i)^2}{2h^2}\right] \tag{7}$$

3.4. Statistical unimodal models

3.4.1. Lognormal distribution

The lognormal distribution, sometimes called Galton's distribution, is the probability distribution of a random variable X if $log(X)$ has a normal distribution (Crow and Shimizu, 1987; Limpert et al., 2001). In order to use this distribution, the quantity of interest must be positive because $log(X)$ exists only when X is positive. Lognormal distribution has been applied as a model for several topics as water resources (Kosugi, 1996), geophysics (Campbell, 1995), optical engineering (Al-Habash et al., 2001) or statistics (Reed and Jorgensen, 2004). The lognormal distribution is given by (Forbes et al., 2011):

$$F_L(x \mid \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{\frac{-(log(x) - \mu)^2}{2\sigma^2}\right\},\ x > 0. \tag{8}$$

where μ and σ are respectively the mean and the standard deviation of the logarithmic values.

3.4.2. Weibull distribution

The Weibull distribution, named for its inventor Waloddi Weibull, is a continuous probability distribution valid when the quantity of interest is positive. In literature, this distribution has been successfully used as a statistical model for many types of fields as ferrography (Roylance and Pocock, 1983), environmental science (Seguro and Lambert, 2000), finance (Chen and Gerlach, 2013), medicine (Pujades-Rodrguez et al., 2011) or agricultural engineering (Bai et al., 2013). The Weibull distribution is given by (Lu, 2004):

$$F_W(x \mid a, b) = \frac{b}{a}\left(\frac{x}{a}\right)^{(b-1)}\exp\left[-\left(\frac{x}{a}\right)^b\right],\ x > 0. \tag{9}$$

where a and b are respectively the scale parameter and the shape parameter.

3.4.3. Burr distribution

The Burr Type XII distribution, more commonly called Burr distribution in probability theory, is a continuous probability distribution for a non-negative random variable (Burr, 1942). Also known as the Singh–Maddala distribution (Singhi and Maddala, 1976), it can express a wide range of distribution as gamma, log-normal, log-logistic, bell-shaped, and J-shaped beta distributions. Burr distribution can fit a wide range of empirical data (Papalexiou and Koutsoyiannis, 2012). Indeed, different values of its parameters cover a wide range of skewness and kurtosis. Burr distributions have been proposed as a model for many types of fields as finance (Pacurar, 2008), insurance (Burnecki et al., 2000), hydrology (Ganora and Laio, 2015), image processing (Sumaiya and Kumari, 2017) or environmental science (Saulo et al., 2013), to mention a few. The burr distribution is given by (Kleiber and Kotz, 2003):

$$F_{\mathrm{B}}(x \mid \alpha, c, k) = \frac{\frac{kc}{\alpha}\left(\frac{x}{\alpha}\right)^{c-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{c}\right)^{k+1}},\ x>0,\ \alpha>0,\ c>0,\ k>0. \tag{10}$$

with α the scale parameter, c and k are shape parameters.

3.4.4. Stable distribution

In probability theory, a distribution is called stable if a linear combination of two independent random variables with this distribution has the same distribution, up to location and scale parameters. A random variable is termed stable if its distribution is stable. Stable distributions are suitable for modelling heavy tails and skewness. In literature, stable distributions have been proposed as a model for many types of fields as finance (Nolan, 2014), radar and image processing (Belkacemi and Marcos, 2007; Boccignone and Ferraro, 2013), acoustics (Pereyra and Batatia, 2012) or geology and geophysics (Lavallée and Archuleta, 2003; Zaliapin et al., 2005), to cite a few. Sometimes, stable distribution family is referred to the Lévy alpha-stable distribution (Ditlevsen, 1999; Liang and Chen, 2013).

The stable is an application of the generalized central limit theorem which mentions that limit of normalized sums of independent identically distributed variables is stable. There are different parameterizations for this distribution. The parameterization depicted in Nolan (2003) was used for this study. In this instance, a random variable X has the stable distribution $S(\alpha,\beta,\gamma,\delta_0;0)$ if its characteristic function is given by (Nolan, 2003):

$$F_S(e^{(itX)}) = \begin{cases} \exp\left(-\gamma^{\alpha}\,|t|^{\alpha}\left[1+i\beta sign(t)\tan\frac{\pi\alpha}{2}((\gamma|t|)^{1-\alpha}-1)\right]+i\delta_0 t\right) & \text{for } \alpha \neq 1,\\ \exp\left(-\gamma\,|\,t\,|\left[1+i\beta sign(t)\frac{2}{\pi}\ln(\gamma|t|)\right]+i\delta_0 t\right) & \text{for } \alpha = 1 \end{cases} \tag{11}$$

with α the first shape parameter which describes the tails of the distribution ($0 < \alpha \leq 2$), β the second shape parameter which indicates the skewness of the distribution ($-1 \leq \beta \leq 1$), γ the scale parameter ($0 < \gamma < \infty$) and δ the location parameter ($-\infty < \delta < \infty$).

3.5. Mixture models

In this study, we also considered a mixture of statistical model to fit *PM*10 time series. Consequently, several mixture models are used in order to provide an analytical approach to estimate *PM*10 frequency distribution. In literature, mixture model has been used for many types of fields as aquaculture (Thorpe, 1977), pharmacology (Haehner et al., 1996), epidemiology (Court-Brown and Caesar, 2006), astrophysics (Schwab et al., 2010) or environmental science (Jaramillo and Borja, 2004; Calif, 2012), to mention a few.

3.5.1. Lognormal and lognormal PDF

The lognormal & lognormal PDF is defined by:

$$F_{\mathrm{L_1L_2}}(x \mid \mu_1,\sigma_1,\mu_2,\sigma_2) = p[F_{\mathrm{L_1}}(x \mid \mu_1,\sigma_1)] + (1-p)[F_{\mathrm{L_2}}(x \mid \mu_2,\sigma_2)] \tag{12}$$

where $F_{\mathrm{L_1L_2}}(x|\mu_1,\sigma_1,\mu_2,\sigma_2)$ is the mixture lognormal & lognormal PDF, μ_1/σ_1 and μ_2/σ_2 are respectively the parameters values of lognormal distribution for the first mode and the second mode; and p is the weight parameter the mixed proportion of a component distribution ($0 < p < 1$). In other words, p represents the proportion of each distribution in the mixture model (Lu, 2003). p can be obtained by using the following equations (Jaramillo and Borja, 2004):

$$\overline{M} = p\overline{M}_1 + (1-p)\overline{M}_2 \tag{13}$$

and

$$\sigma^2 = p(\sigma_1^2 - (p-1)(\overline{M}_1 - \overline{M}_2)^2) - (p-1)\sigma_2^2 \tag{14}$$

where $\overline{M}$ is the average and σ is the standard deviation of the times series studied; $\overline{M}_1$ and $\overline{M}_2$ are the average of the empirical data for the first mode and the second mode; ${\sigma_1}^2$ and ${\sigma_2}^2$ are the variance for the first mode and the second mode.

3.5.2. Weibull and Weibull PDF

The Weibull & Weibull PDF is expressed as:

$$F_{\mathrm{W_1W_2}}(x \mid a_1,b_1,a_2,b_2) = p[F_{\mathrm{W_1}}(x \mid a_1,b_1)] + (1-p)[F_{\mathrm{W_2}}(x \mid a_2,b_2)] \tag{15}$$

where $F_{\mathrm{W_1W_2}}(x|a_1,b_1,a_2,b_2)$ is the mixture Weibull & Weibull PDF, a_1/b_1 and a_2/b_2 are respectively the parameters values of Weibull distribution for the first mode and the second mode; and p is the weight parameter.

3.5.3. Burr and burr PDF

The burr & burr PDF can be written as:

$$\begin{aligned} &F_{\mathrm{B_1B_2}}(x \mid \alpha_1,c_1,k_1,\alpha_2,c_2,k_2)\\ &= p[F_{\mathrm{B_1}}(x \mid \alpha_1,c_1,k_1)] + (1-p)[F_{\mathrm{B_2}}(x \mid \alpha_2,c_2,k_2)] \end{aligned} \tag{16}$$

where $F_{\mathrm{B_1B_2}}(x|\alpha_1,c_1,k_1,\alpha_2,c_2,k_2)$ is the mixture burr & burr PDF, $\alpha_1/c_1/k_1$ and $\alpha_2/c_2/k_2$ are respectively the parameters values of burr distribution for the first mode and the second mode; and p is the weight parameter.

3.5.4. Lognormal and Weibull PDF

The lognormal & Weibull PDF is defined by:

$$F_{\mathrm{L_1W_2}}(x \mid \mu_1,\sigma_1,a_2,b_2) = p[F_{\mathrm{L_1}}(x \mid \mu_1,\sigma_1)] + ((1-p)[F_{\mathrm{W_2}}(x \mid a_2,b_2)] \tag{17}$$

where $F_{\mathrm{L_1W_2}}(x|\mu_1,\sigma_1,a_2,b_2)$ is the mixture lognormal & Weibull PDF, μ_1/σ_1 and a_2/b_2 are respectively the parameters values of lognormal distribution for the first mode and Weibull distribution for the second mode; and p is the weight parameter.

3.5.5. Burr and lognormal PDF

The burr & lognormal PDF can be written as:

$$F_{\mathrm{B_1L_2}}(x \mid \alpha_1,c_1,k_1,\mu_2,\sigma_2) = p[F_{\mathrm{B_1}}(x \mid \alpha_1,c_1,k_1)] + (1-p)[F_{\mathrm{L_2}}(x \mid \mu_2,\sigma_2)] \tag{18}$$

where $F_{\mathrm{B_1L_2}}(x|\alpha_1,c_1,k_1,\mu_2,\sigma_2)$ is the mixture burr & lognormal PDF, $\alpha_1/c_1/k_1$ and μ_2/σ_2 are respectively the parameters values of burr distribution for the first mode and lognormal distribution for the second mode; and p is the weight parameter.

3.5.6. Burr and Weibull PDF

The burr & Weibull PDF can be written as:

$$F_{\mathrm{B_1W_2}}(x \mid \alpha_1,c_1,k_1,a_2,b_2) = p[F_{\mathrm{B_1}}(x \mid \alpha_1,c_1,k_{1,})] + (1-p)[F_{\mathrm{W_2}}(x \mid a_2,b_2)] \tag{19}$$

where $F_{\mathrm{B_1W_2}}(x|\alpha_1,c_1,k_1,a_2,b_2)$ is the mixture burr & Weibull PDF, $\alpha_1/c_1/k_1$ and a_2/b_2 are respectively the parameters values of burr distribution for the first mode and Weibull distribution for the second mode; and p is the weight parameter.

3.6. Kolmogorv–Smirnov test

For a sample, the Kolmogorov–Smirnov statistic (*KS*) quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution (Kolmogorov, 1933). *KS* can be written as:

$$KS = \sup_x |F_n(x) - F(x)| \tag{20}$$

where 'sup' means supremum, i.e. the largest set of distances. $F_n(x)$ is the hypothesized distribution function whereas $F(x)$ is the empirical distribution function estimated based on the random sample.

Here, in order to test the hypothesis that the distribution of experimental data is modeled with a particular probability density function, KS is computed by comparing their Cumulative Distribution Function (CDF). Then, KS is compared to the critical value $CV_\circ$ (Massey Jr, 1951):

$$CV_\circ = c_\delta \sqrt{\frac{m+n}{mn}} \tag{21}$$

with m and n respectively the number of samples for the empirical CDF and the theoretical CDF and the treshold $c_\delta = 1.36$ at the significance level $\delta = 0.05$ (Smirnov, 1948). The empirical CDF and the theoretical CDF must have the same number of sample. When $KS < CV_\circ$, the null hypothesis H_0 is true, i.e. both time series x_1 and x_2 are from the same distribution. Otherwise, the null hypothesis is rejected, x_1 and x_2 are from a different distribution.

3.7. Extreme events modelling

An extreme event is defined as the event exceeding a certain threshold of normalized logarithmic return (Zhao et al., 2010). In other words, it is an event occurring rarely and suddenly over a short time period compared to the characteristics times scales of their posterior evolution (Sornette, 2002). Extreme events take place frequently in both nature and society (Santhanam and Kantz, 2008). Traditionally, study of extreme events is a specific field of statistics (Ghil et al., 2011). In literature, power-law distributions have often been used to study the recurrence of extreme events as earthquakes (Bak et al., 2002), floods (Malamud and Turcotte, 2006), droughts (Ault et al., 2014) or economic recession (Ormerod and Mounfield, 2001) as examples. Mathematically, a continuous random variable X exhibits a power-law if it drawn from a probability distribution with a density of the form (Clauset et al., 2009):

$$p(x) \propto x^{-\alpha} \tag{22}$$

where $\alpha > 1$ is the exponent or scaling parameter and $x > x_{min} > 0$. The power law pattern holds only above some value x_{min}. Consequently, it is considered that the tail of the distribution follows a power-law. Some researchers represent this by apply a slowly varying functions often denoted by $L(x)$ such that the tail of the probability density follows a power-law (Virkar et al., 2014):

$$p(x) \propto L(x)x^{-\alpha} \tag{23}$$

where in the limit of large x, $L(cx)/L(x) \to 1$ for any $c > 0$.

3.8. Seasonal air mass back trajectory classification

In order to investigate the seasonal atmospheric circulation of air masses in Guadeloupe, the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) is used (Stein et al., 2015; Rolph et al., 2017). In literature, HYSPLIT back trajectories are frequently applied to determine dusty air masses origin in the Caribbean area (Prospero et al., 2005; Dunion, 2011; Prospero et al., 2014). Meteorological database used as input for HYSPLIT is the National Center for Atmospheric Research/National Centers for Environmental Prediction (NCAR/NCEP) re-analysis data (Kalnay et al., 1996). Guadeloupe back trajectories have been generated day-to-day according to the following parameters from 2006 to 2016:

- Saharan Air Layer (SAL) properties has been considered. This is a hot and dry dust-laden layer which extends generally from 1500 to 5000 m and located above the top of the marine boundary layer (Carlson and Prospero, 1972; Ben-Ami et al., 2012). Consequently, in literature, all studies on back trajectories generally consider altitudes between 1500 m and 3000 m referring to the SAL (Carlson and Prospero, 1972; Dunion, 2011). To simplify our statistic evaluation, a single altitude level was chosen with 1500 m;
- Starting location is 16.24°N, −61.53°E;
- At 12 UTC (8 a.m. local time);
- Duration is 10 days (240 h).

Thereafter, back trajectories data as latitude/longitude and positions are generated by HYSPLIT and imported in QGIS geographic information system for representation. This procedure was recently validated by Euphrasie-Clotilde et al., 2020 study which presents the methodology for trajectory clustering used in this work. Here, back trajectories were studied according to dust seasons of Caribbean region.

Next sections describe the results achieved by applying the methods described previously. These results are then discussed.

4. Results and discussions

4.1. Descriptive statistics

4.1.1. Comparison between AQS1 and AQS2 data

To confirm the frozen turbulence mentioned in section 2, it is essential to ensure that AQS1 (2005–2012) and AQS2 (2015–2017) are subject to the same level of pollution. In insular context, the background atmosphere is mainly consisted of marine aerosols and anthropogenic pollution (Clergue et al., 2015; Rastelli et al., 2017). In Table 1, one can note that achieved results for $PM10$ are very close for all statistical parameters in both periods. This highlights the stationarity of $PM10$ data between 2005 and 2017. It is important to emphasize that $PM10$ statistics is the same at AQS1 (urban area) and AQS2 (suburban area) showing that marine aerosols and anthropogenic pollution in Guadeloupe archipelago generate low $PM10$ concentrations, i.e. ~20 $\mu g/m^3$ (Euphrasie-Clotilde et al., 2017), contrary to megacity. Indeed, due to strong anthropogenic pollution, $PM10$ annual mean concentrations can reach 53 $\mu g/m^3$ in Europe (Querol et al., 2004), 133 $\mu g/m^3$ in South America (Silva et al., 2017), 150 $\mu g/m^3$ in United States (Baldasano et al., 2003) and 160 $\mu g/m^3$ in China (Matus et al., 2012).

4.1.2. Overall analysis

Table 2 shows the descriptive statistics for all $PM10$ data, i.e. 11 years. With an average of 26.4 $\mu g/m^3$, $PM10$ concentrations in Guadeloupe are lower than those measured in Nilai Malaysia 59.1 $\mu g/m^3$(Sansuddin et al., 2011), Jawaharlal Nehru Port India 66.1 $\mu g/m^3$ (Gupta et al., 2004), Belgrade Serbia 68.3 $\mu g/m^3$ (Mijić et al., 2009), Anatolia Turkey 78.0 $\mu g/m^3$ (Ozel and Cakmakyapan, 2015), Beijing China 145.1 $\mu g/m^3$ (Xi et al., 2013) or Abadan Iran 186.1 $\mu g/m^3$ (Momtazan et al., 2018).

With σ equal to 16.1 $\mu g/m^3$, standard deviation value found in Guadeloupe is lower than those observed for megacity. Indeed, due to the wide heterogeneity of $PM10$ sources, standard deviation values in Anatolia Turkey (Ozel and Cakmakyapan, 2015), Abadan Iran (Momtazan et al., 2018), Nilai Malaysia (Sansuddin et al., 2011) and Beijing China (Xi et al., 2013) are higher with respectively 26.0, 26.4, 28.6 and 91.4 $\mu g/m^3$. Consequently, when standard deviation is high, the variability is likewise high, indicating huge concentrations. One can

Table 1
Statistical parameters (Mean ($\overline{M}$), Standard deviation (σ), Minimum (Min), Maximum (Max), Skewness (S) and Kurtosis (K)) of $PM10$ data at AQS1 and AQS2.

Period	Location	$\overline{M}$	σ	S	K	Min	Max
2005–2012 (N = 2814)	AQS1	26.6	16.1	2.4	11.6	4.0	164.4
2015–2017 (N = 1035)	AQS2	26.1	15.9	2.4	12.6	3.3	153.5

$\overline{M}$, σ, Min and Max are in $\mu g/m^3$ and N represents the data point number.

Table 2
Mean ($\overline{M}$), standard deviation (σ), Minimum (*Min*), Maximum (*Max*), Skewness (*S*) and Kurtosis (*K*) of *PM*10 data for all data, per year and per season with season 1 from October to April and season 2 from May to September.

Study period	$\overline{M}$	σ	*Min*	*Max*	*S*	*K*
Overall (N = 3849)	26.4	16.1	3.3	164.4	2.4	11.8
2005 (N = 354)	27.3	15.3	9.8	109.9	2.3	9.1
2006 (N = 358)	27.9	16.6	9.0	84.2	1.7	5.4
2007 (N = 357)	27.8	17.5	9.9	*157.2*	2.4	12.8
2008 (N = 355)	24.9	13.1	10.1	114.9	3.2	18.3
2009 (N = 365)	24.8	14.5	10.6	123.8	2.9	14.9
2010 (N = 354)	27.5	19.7	4.0	*164.4*	2.9	15.0
2011 (N = 334)	24.4	13.8	6.9	89.3	2.0	7.3
2012 (N = 337)	28.4	17.2	8.0	92.2	1.3	3.9
2015 (N = 336)	31.9	17.4	3.3	98.9	1.2	4.1
2016 (N = 354)	23.1	12.3	7.1	85.7	1.9	7.8
2017 (N = 345)	23.2	14.0	7.1	153.5	3.6	25.0
Season 1 (N = 2253)	22.0	12.4	3.3	164.4	4.2	33.0
Sesaon 2 (N = 1596)	32.9	18.5	4.0	157.2	1.5	5.9

$\overline{M}$, σ, *Min* and *Max* are in $\mu g/m^3$ and N represents the data point number.

notice that *PM*10 maximum concentrations can reach 445.0 $\mu g/m^3$ in Malaysia (Sansuddin et al., 2011) or 540.0 $\mu g/m^3$ in China (Dong et al., 2017) but do not exceed 164.4 $\mu g/m^3$ in Guadeloupe.

*PM*10 time series has positive skewness, i.e. $S = 2.4$, which means the frequency distribution moves away from a normal distribution on the right with a larger right tail. The positive value of kurtosis obtained, i.e. $K = 11.8$, is greater than the peak of a Gaussian distribution which is $K = 3$ (Dong et al., 2017). This indicates that *PM*10 datasets exhibit an evident peak tail distribution. Positive values of skewness and kurtosis was previously found by Yusof et al. (2010); Sansuddin et al. (2011); Dong et al. (2017) with respectively $S = 1.0$, $K = 3.0$; $S = 3.4$, $K = 34.8$ and $S = 1.9$, $K = 8.6$.

4.1.3. Yearly analysis

Yearly average results for *PM*10 concentrations are presented in Table 2. From 1 year to the next, there is a heterogeneity in *PM*10 average concentrations with $23.2 \leqslant \overline{M} \leqslant 31.9$. Due to African dust presence in air masses crossing over Guadeloupe, the 20 $\mu g/m^3$ limit recommended by World Health Organization (World Health Organization, 2013) for *PM*10 annual average concentration is always exceeded. Many factors may explain the change in annual *PM*10 concentrations from year to year. The main one is activation of dust sources in Africa. According to Knippertz et al. (2010), during summer, several sources as northern Chad, Mali, Mauritania and southern Algeria become more active. This activations are associated to the development and movement of African easterly waves in concert with extra-tropical disturbances. Furthermore, the modifications of terrain properties in these source regions, e.g. rainfall, vegetation cover or land use, coupled with the meteorological processes that affect them will act to modulate transport to receptor sites in the western Atlantic (Ginoux et al., 2012; Prospero et al., 2014). Previously, Smirnov et al. (2000) showed that high concentrations of dust at the surface of Barbados are correlated with high column optical depths measured by a collocated AERONET instrument. Velasco-Merino et al. (2018) investigated time of transport linked to dusty air mass from western Africa sites to Caribbean sites, using the combination of HYSPLIT back trajectories and the Aerosol Optical Depth (AOD) measuring in these both areas. They showed that dust from African coast, i.e. Dakar, arrives over Guadeloupe after a transit of 5–7 days across the Atlantic Ocean.

Table 3 presents the number of dusty days per year found by Euphrasie-Clotilde (2018) where *PM*10 daily average concentrations exceed the regulatory information and alert threshold defined by European Union (2008) with respectively 50 and 80 $\mu g/m^3$. As a French department, Guadeloupe is under European legislation. Dusty days were identified by Euphrasie-Clotilde (2018) using HYSPLIT day-to-day back trajectories. When $PM10 \geq 50\mu g/m^3$, the number of dusty days seems close between 2006 2012 and 2015. However, when $PM10 \geq 80\mu g/m^3$, the number of dusty days is clearly much higher for 2015. Despite its high annual average, 2015 does not present extreme events contrary to 2007, 2010 and 2017. For 2007 and 2017, African dust episodes are fewer but more intense (see maximum values in Table 2). In 2007, *PM*10 maximum daily concentration reached 157.2 $\mu g/m^3$ on May 15th. For 2017, October has experienced one of the most intense desert dust episode of this last 10 years with 153.5 $\mu g/m^3$ on the 18th and 150.9 $\mu g/m^3$ on the 19th. As regard 2010, these maximal concentrations are linked to the eruption of Soufrière on Montserrat (Plocoste and Calif, 2019) in February 11th with 150.0 and 164.4 $\mu g/m^3$ measured in Guadeloupe respectively February 12th and 13th. All these events and values are listed in Gwad'Air annual reports available on their website http://www.gwadair.fr/.

Table 3
Number of dusty days per year where daily average *PM*10 concentration are respectively superior or equal to 50 $\mu g/m^3$ ($PM10 \geq 50\mu g/m^3$) and 80 $\mu g/m^3$ ($PM10 \geq 80\mu g/m^3$) (Euphrasie-Clotilde, 2018).

Year	$PM10 \geq 50\mu g/m^3$	$PM10 \geq 80\mu g/m^3$
2005	25	6
2006	42	5
2007	38	5
2008	15	3
2009	25	3
2010	38	8
2011	21	3
2012	45	5
2015	51	39
2016	12	3
2017	20	4

All years have a positive skewness with $1.2 \leqslant S \leqslant 3.6$ indicating a clear right-slide feature of *PM*10 time series. One can notice that 2012 and 2015, which exhibit the higher annual concentrations, show the lowest values of skewness with respectively 1.3 and 1.2. Conversely, 2017 which exhibits the lowest concentration, shows the highest value of skewness with 3.6. As skewness, all years exhibit a positive kurtosis with $3.9 \leqslant K \leqslant 25.0$ and the same behavior is observed with the lowest values for 2012 $K = 3.9$, 2015 $K = 4.1$ and the highest value for 2017 $K = 25.0$. Indeed, as seen in Tables 3, 2015 is experiencing recurrent pollution of cross-border origin throughout the year, i.e. African dust. This explains why 2015 exhibit the highest *PM*10 annual concentration and the lowest values for skewness and kurtosis. While 2017 has one of the lowest annual concentrations, this year is experiencing the shortest and most intense African dust outbreaks episode of the last decade. The highest kurtosis value (in red in Table 2) clearly shows the strong intermittent feature of African dust in 2017. As regard 2010, one can notice that the eruption of Soufrière on Montserrat does not influence significantly skewness and kurtosis values due to the scarcity of this type of phenomenon during a year.

4.1.4. Seasonal analysis

According to literature (Prospero et al., 2014; Velasco-Merino et al., 2018; Plocoste and Pavón-Dominguez, 2020), two seasons can be observed for dust outbreaks, i.e. season 1 from October to April and season 2 from May to September. Descriptive statistics of both seasons for 11 years are shown in Table 2. *PM*10 concentrations are higher in season 2 (32.9 ± 18.5 $\mu g/m^3$) than season 1 (22.0 ± 12.4 $\mu g/m^3$). Season 1 average concentration is close to the 20 $\mu g/m^3$ found by Euphrasie-Clotilde et al. (2017) for local sources. In insular context, the background atmosphere is mostly composed of marine sea salt aerosols which significantly impact *PM*10 concentrations due to their large size (Sellegri et al., 2001; Reid et al., 2003; Prats et al., 2011). Optical characterization of particle size and satellite detections, i.e. Volume Particle Size Distribution and Ångström Exponent, showed other types

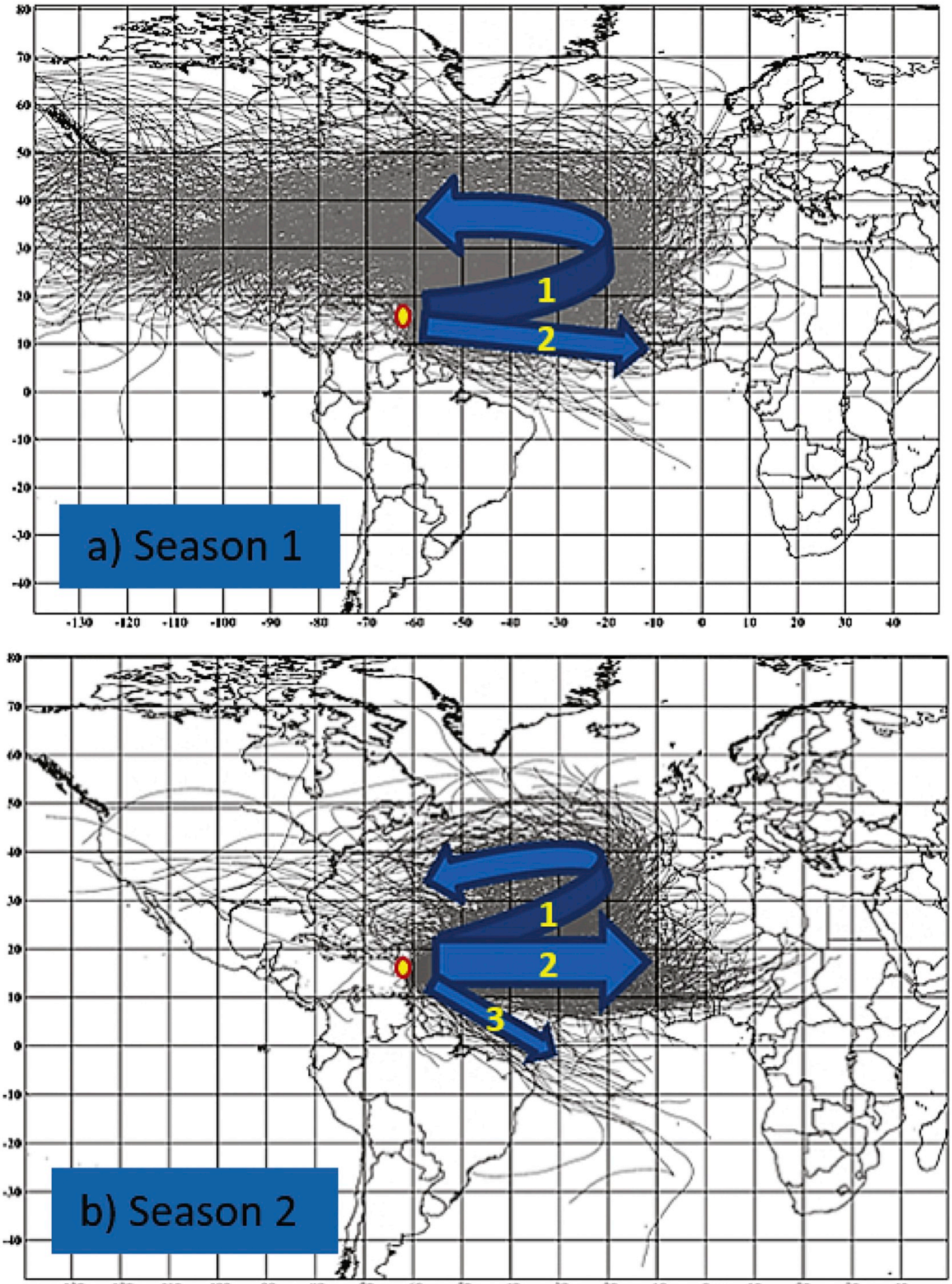

Fig. 3. Guadeloupe seasonal atmospheric circulation represented by day-to-day HYSPLIT back trajectories for (a) season 1 and (b) season 2 between 2006 and 2016. Back trajectories parameters are height = 1500 *m* and duration = 240 h (10 days). Yellow circle shows back trajectories starting location and arrows indicate the origin of air masses with (1) from the United States, (2) from West African coast and (3) from South America. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of aerosols are also present in the Caribbean area. From November to March, the seasonal transport of biomass and polluted dust aerosol (mixture of desert aerosols with biomass) to the Caribbean was modeled by Adams et al. (2012). In other words, there is few episode of African dust during season 1. As regards season 2, the average and the standard deviation values are 1.5 times that of season 1. In addition, kurtosis and skewness values obtained in season 2 are weaker than those of season 1. As noticed in Table 2, two statistical behaviors seem to be observed for *PM*10 data.

In order to better understand air masses paths behavior between seasons 1 and 2, all day-to-day HYSPLIT back trajectories for Guadeloupe between 2006 and 2016 are studied (see procedure in Section 3.8). Here, inconsistent traces due to trajectories turning on itself was removed. These cases are mainly caused by hurricanes over

Table 4
Number of available day-to-day back trajectories associated with *PM*10 statistical parameters (mean " $\overline{M}$ ", standard deviation "σ") for seasons 1 and 2 from 2006 to 2016.

Period	Air masses path	Number of cases (%)	$\overline{M}_{PM10} \pm \sigma_{PM10}$ ($\mu g/m^3$)
Season 1	Path 1	530 (63.8%)	19.19 ± 6.11
	Path 2	281 (33.8%)	21.99 ± 9.27
	Path 3	20 (2.4%)	25.60 ± 9.59
Season 2	Path 1	310 (33.4%)	27.52 ± 13.63
	Path 2	540 (58.3%)	40.17 ± 19.11
	Path 3	77 (8.3%)	29.88 ± 13.48

the North Atlantic. Thus, it is impossible to precisely determine the origin of air masses. This allowed us to identify 3 main types of trajectories previously found in literature with (see Fig. 3):

- Path 1: from the United States, looping over the Atlantic then return to the Caribbean (Dunion, 2011);
- Path 2: a direct corridor between western African coast and the Caribbean (Gläser et al., 2015);
- Path 3: from South America to the Caribbean (Gläser et al., 2015).

As shown in Fig. 3(a-b), frequency of these types of paths differs according to seasons. To quantitatively observe the impact of air masses origin on air pollution level, each back trajectory path is linked to *PM*10 measurements. Table 4 presents achieved results for this analysis.

During season 1, the main path is (1) with 63.8% of cases while season 2 is dominated by path (2) with 58.3% of cases. Between both main paths, *PM*10 average concentrations are 19.19 ± 6.11 $\mu g/m^3$ and 40.17 ± 19.11 $\mu g/m^3$. One can notice that path (2) brings twice as much *PM*10 in season 2. Path (3) which is caused by the ascent of the Intertropical Convergence Zone towards the north is not significant for season 1. All these results confirm that dust outbreaks are more frequent in season 2, i.e. from May to September.

4.2. PM10 data distribution: analysis, stationarity, and modelling

4.2.1. Analysis

The probability density function (PDF) is a useful tool for sharply describing the statistical information contained in a dataset. It depicts the entire range, mean and probability of occurrences of the studied variable (Sharma et al., 2016). Fig. 4 illustrates the distribution plot for all *PM*10 data, by year and by season. Maximum probability value provided by Fig. 4 are listed in Table 5. For the overall scale, the PDF plot in Fig. 4(a) shows the distribution is skewed to the right with a maximum probability value of 17.4 $\mu g/m^3$. This PDF seems to highlight a unimodal distribution with a large tail.

At yearly scale, in order to observe the visual clarity of distribution behavior, years should be clustered by similar distribution. For that, it seems to us that a first discrimination can be made from the fourth moment. Here, we group years by taking the overall kurtosis value presented in Table 2, i.e. K=11.8, as a selection criterion. Fig. 4(b) shows years with K > 11.8 while Fig. 4(c) depicts years with K < 11.8. By analysing these distributions plots, we may notice that distributions differ from one year to another with a maximum probability value range between 17.1 $\mu g/m^3$ in 2017 and 19.7 $\mu g/m^3$ in 2015. 2015 distribution exhibits the most extended wing to the right due to the large number of dust events listed in Table 3. Contrary to Fig. 4(a), Fig. 4(c) clearly shows a strong occurrence of dust events during the high dust season for 2005, 2006, 2011, 2012, 2015 and 2016. As regards the other years in Fig. 4(b), the distributions seem illustrate the same pattern with a second mode less significant than Fig. 4(c).

As regards the seasonal scale, we can observe the distribution difference between season 1 and season 2 for 11 years in Fig. 4(d). Nevertheless maximum probability values remain close with

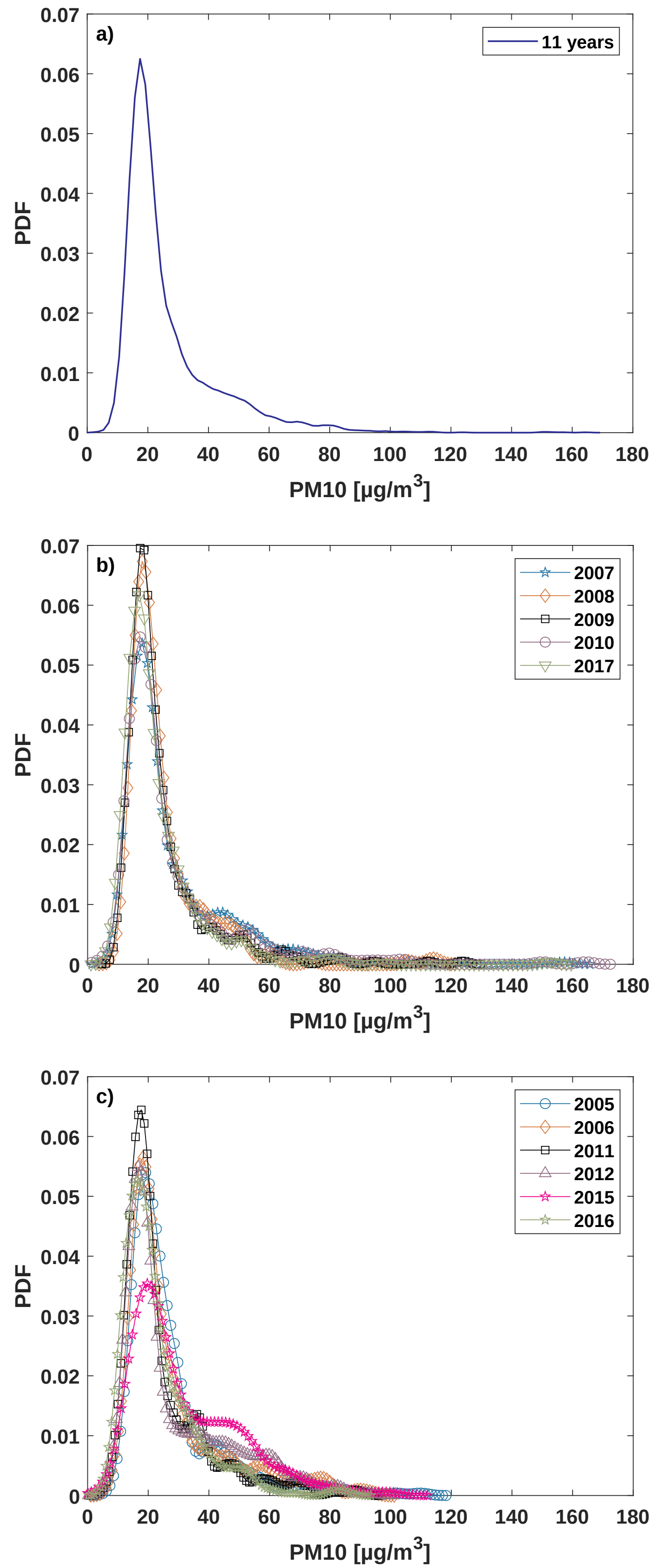

Fig. 4. Probability Density Function (PDF) for all data (a), by year (b-c) and by season (d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively 16.9 and 19.6 $\mu g/m^3$. Indeed, distribution curve of season 1 is more skewed to the right whereas distribution curve of season 2 is more flattened. All these results confirm the skewness and kurtosis

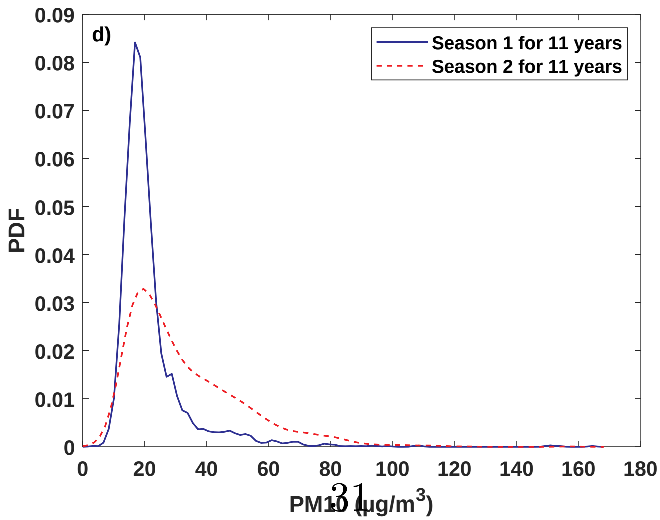

Fig. 4. (*continued*)

Table 5
Maximum probability value (*MaxProb*) provided by the distributions for *PM*10 data.

Year	*MaxProb* ($\mu g/m^3$)
Overall (N = 3849)	17.4
2005 (N = 354)	19.2
2006 (N = 358)	18.2
2007 (N = 357)	18.0
2008 (N = 355)	18.0
2009 (N = 365)	18.6
2010 (N = 354)	16.7
2011 (N = 334)	17.7
2012 (N = 337)	17.3
2015 (N = 336)	19.7
2016 (N = 354)	16.5
2017 (N = 345)	17.1
Season 1 (N = 2253)	16.9
Sesaon 2 (N = 1596)	19.6

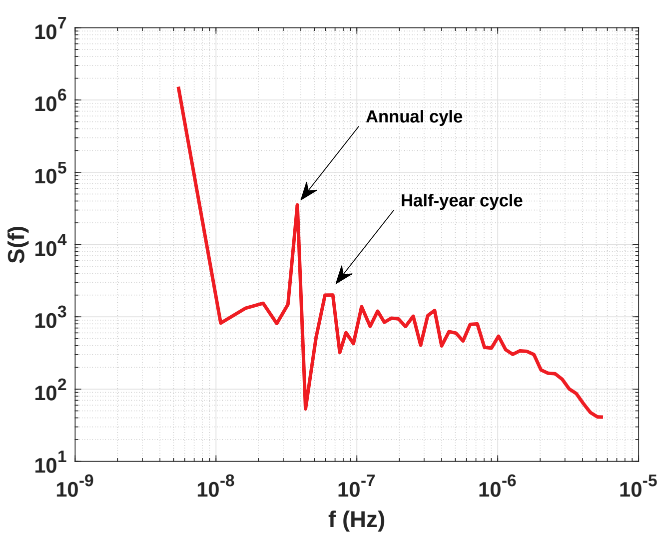

Fig. 5. Fourier spectrum *S*(*f*) of *PM*10 time series between 2005 and 2011 from Plocoste and Calif (2019) study.

values presented in Table 2.

4.2.2. Distribution stationarity threshold

A good way to check the depictiveness of a model is to use empirical distribution plots. The first step consists to determine the length of the increment time between each season. For that, a recent work made by Plocoste and Calif (2019) on spectral observation of *PM*10 fluctuations between 2005 and 2011 in the same study area shows two main peaks in Fourier spectrum as illustrated in Fig. 5. One at 3.48×10^{-8} Hz which corresponding to 11 months (approximately the annual cycle) and the other at 6.75×10^{-8} Hz which corresponding to 6 months (half-year cycle). In order to study the strict sense stationarity, i.e. its statistical properties are invariant to a shift of the origin (Papoulis and Pillai, 2002) of *PM*10 events, distributions with 6 months increments are plotted for the entire period, i.e. 11 years. This means for each distribution plot, 6 months of data are added from June 2005 ($t6 = 6$ months) to December 2017 ($t132 = 132$ months). Fig. 6 shows the 22 distributions obtained. Visually, from 66 months (*t*66), one can observe that distributions values seem close with weak variations. To quantify this stationarity threshold, the first four central moments versus 6 months increment length are estimated and showed in Fig. 7. For these moments, the distribution shape remains roughly stationary from *t*66. In others words, 66 months of *PM*10 data would be sufficient to be representative statistically and to model 11 years. This underlines the scarcity of extreme events and confirms the cycle stability of African dust.

4.2.3. Unimodal model

In order to fit the occurrences of the daily *PM*10 concentrations, several statistical models were performed. Here we consider four statistical models, including the classical lognormal model (Xi et al., 2013; Dong et al., 2017) and Weibull model (Yusof et al., 2010) which are frequently used for *PM*10 data statistic modelling; and for the first time in this field, Burr and stable models are applied. Table 6 shows the parameters obtained for these distributions in all studied cases. Visually, in Figs. 8–10, Burr and stable distribution appear to give the best fit. Lognormal distribution seems only significant for season 2 and year 2015 which are two cases with high *PM*10 concentrations due to African dust. As regards Weibull distribution, the results achieved with this model look the worst.

Fig. 11 illustrates *PM*10 data distribution with the statistical models for the statistical stationarity threshold, i.e. *t*66 months. In order to assess the impact of year 2015 on statistical models parameters, we performed the same analysis for all years without the special year 2015, i.e. 10 years. Distribution parameters values for these both cases are depicted in Table 7. Between the statistical stationarity threshold, all years without 2015 and the overall data (see Table 6), parameters values of each model are quite similar. This confirms the statistical stationarity threshold time scale found with PDF method. In addition, these results show that the influence of 2015 data on distribution parameters values for 11 years is not significant.

4.2.4. Mixture model

Traditionally, in statistics, a mixture model is the combination of at least two models (Reynolds, 2009). Here, despite the good results achieved with Burr model, we can notice that Burr PDF do not perfectly fit a second mode between 35 and 55 $\mu g/m^3$ (see Figs. 8 and 10). In order to improve these results, it seems natural to consider mixture models to take into account the seasonal variability. In these mixture models, we consider that the first mode represents the low dust season, i.e. parameters values of statistical distribution in season 1; and the second mode represents the high dust season, i.e. parameters values of statistical distribution in season 2. Parameters values used are presented in Table 6 for all data and Tables A1 and A2 in Appendix A for each year. Several combinations of mixture models have been applied: Lognormal & Lognormal, Weibull & Weibull, Burr & Burr, Lognormal &

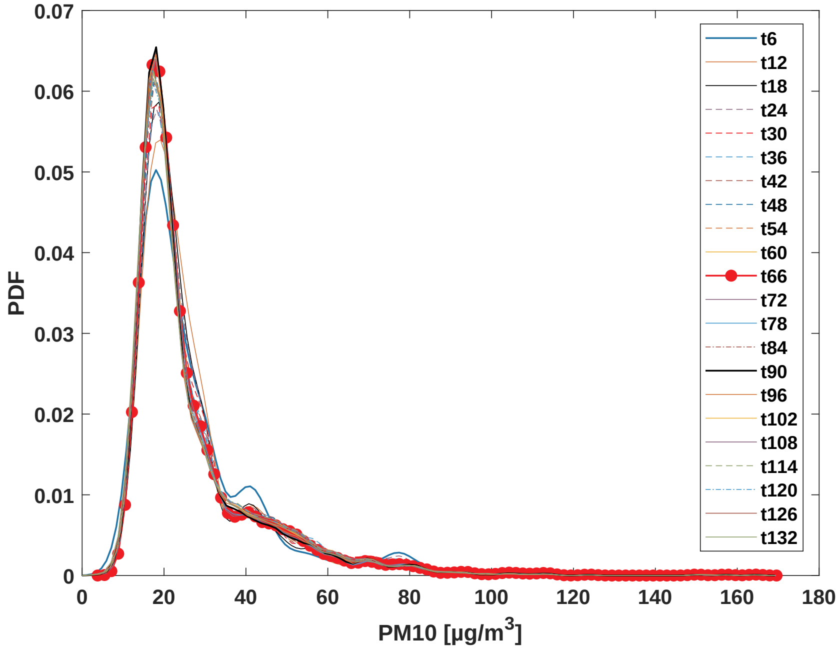

Fig. 6. Probability Density Function (PDF) for 11 years with 6 months increments, i.e. 22 distributions. Red circles show the stationarity threshold estimated at *t*66. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Weibull, Burr & Lognormal and Burr & Weibull. Fig. 12 shows the PDF of *PM*10 concentrations for the entire study period, i.e. 11 years, with all mixture models. Here, we only illustrate the results for all *PM*10 data because each year highlights the same fit pattern for each mixture model. In Fig. 12(a), the best fit seems to be the Burr & Burr model for the low dust season at 18 $\mu g/m^3$ however the high dust season at 45 $\mu g/m^3$ is still not fit. In Fig. 12(b), Burr & Weibull model seems to perfectly fit both seasons.

In order ton confirm the results visually observed in 4.2.3 and 4.2.4, the *KS* test is performed in the following section.

4.2.5. Kolmogorov–Smirnov statistic

Here, the achieved results for *KS* and CV_c are presented in Table 8 for unimodal models and mixture models.

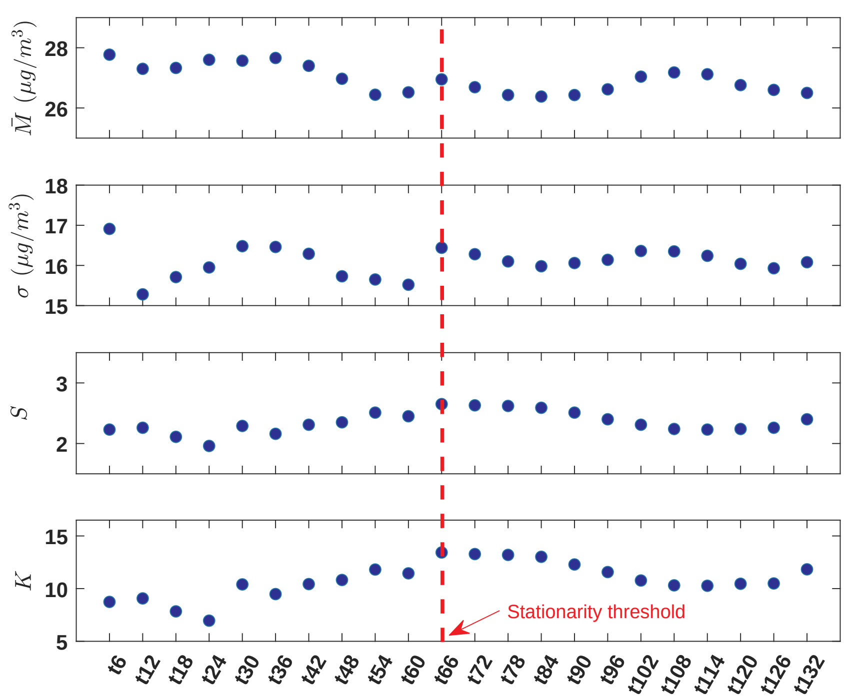

Fig. 7. First four central moments (Mean “ $\bar{M}$“, standard deviation “σ“, Skewness “*S*“ and Kurtosis “*K*“) versus 6 months increment length from June 2005 (*t*6) to December 2017 (*t*132). Red vertical dashed line indicates the stationarity threshold quantitatively estimated at *t*66. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Parameters values of lognormal (μ, σ), Weibull (a, b), Burr (α, c, k) and stable (α, β, γ, δ) distributions for all data, per year and per season with season 1 from October to April and season 2 from May to September.

Study period	Lognormal		Weibull		Burr			Stable			
	μ	σ	a	b	α	c	k	α	β	γ	δ
Overall (N = 3849)	3.15	0.48	30.05	1.81	15.32	7.88	0.27	1.12	0.99	4.76	18.86
2005 (N = 354)	3.20	0.44	30.99	1.94	16.31	9.94	0.23	1.14	1.00	4.53	20.17
2006 (N = 358)	3.20	0.49	31.70	1.84	15.49	8.98	0.23	1.08	1.00	4.88	19.41
2007 (N = 357)	3.18	0.50	31.50	1.76	14.71	9.65	0.20	1.00	1.00	4.70	18.75
2008 (N = 355)	3.12	0.40	28.15	2.01	16.03	10.42	0.25	1.18	1.00	3.87	19.17
2009 (N = 365)	3.10	0.43	28.13	1.87	15.36	9.72	0.25	1.14	1.00	3.86	18.39
2010 (N = 354)	3.15	0.53	31.05	1.61	14.96	7.66	0.26	1.06	0.96	4.67	18.51
2011 (N = 334)	3.08	0.46	27.73	1.92	14.92	7.58	0.31	1.19	1.00	4.48	18.00
2012 (N = 337)	3.19	0.53	32.20	1.80	13.94	10.11	0.17	0.97	1.00	5.03	18.36
2015 (N = 336)	3.33	0.52	36.17	1.98	19.76	4.80	0.45	1.14	0.98	7.05	22.80
2016 (N = 354)	3.03	0.46	26.16	2.01	15.40	5.61	0.46	1.25	1.00	4.74	17.54
2017 (N = 345)	3.04	0.46	26.30	1.82	14.66	7.82	0.31	1.22	1.00	4.09	17.51
Season 1 (N = 2253)	3.00	0.40	24.93	1.92	15.36	8.57	0.36	1.31	0.98	3.43	17.58
Season 2 (N = 1596)	3.35	0.52	37.27	1.91	23.42	3.84	0.67	1.21	1.00	7.81	23.89

N represents the data point number.

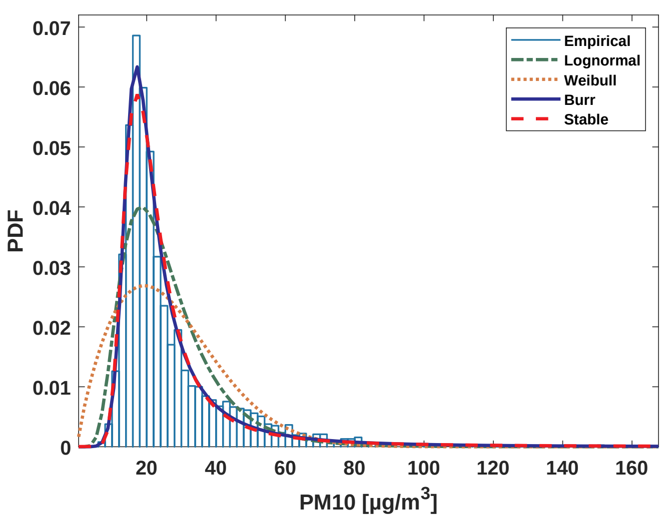

Fig. 8. Empirical distribution (in bar) superimposed to Lognormal, Weibull, Burr and stable PDFs of daily average $PM10$ concentrations for all data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For unimodal models, except for 2015 and season 2, KS_{Log} values exceed $CV_{\circ}$ values. As regards the other classical distribution, KS_{Wei} has the highest KS values and is greater than $CV_{\circ}$. For KS_{Bur} and KS_{Stab}, their values are lower than $CV_{\circ}$ values for all cases. Overall, one may notice that KS_{Bur} values are smaller than KS_{Stab} values. In other words, KS test confirms that Burr and stable models are from the same distribution as $PM10$ concentrations. However, due to its smallest KS values, Burr is the best unimodal model to fit $PM10$ concentrations in all cases. For dusty periods, i.e. 2015 and season 2, lognormal distribution may become an alternative to model $PM10$ concentrations. Using it as a unimodal model, Weibull distribution is never representative.

As regard mixture models, all KS_{LogLog} and KS_{WeiWei} values are greater than $CV_{\circ}$. For some cases, KS_{LogWei} is smaller than $CV_{\circ}$. However, the results obtained with this mixture model do not improve those obtained with the unimodal Burr model, i.e. $KS_{Bur} < KS_{LogWei}$ KS_{BurBur} presents good results but has close values to KS_{Bur}. Contrary to the others mixture models, KS_{BurLog} and KS_{BurWei} really improve KS_{Bur} results. Nevertheless, KS_{BurWei} exhibits the lowest values, i.e. it is the best

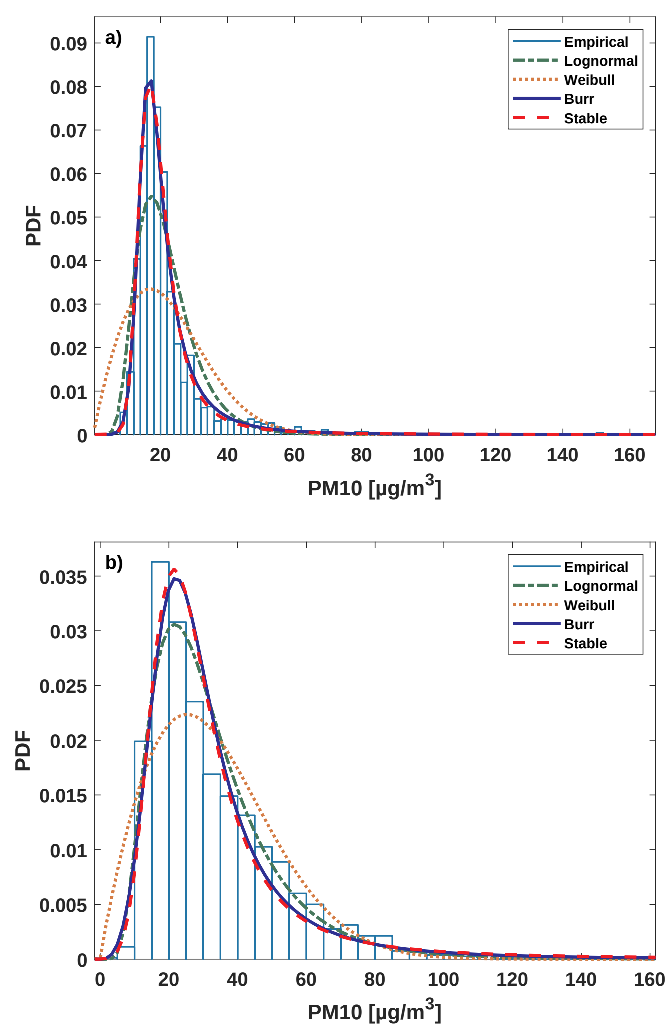

Fig. 9. Empirical distribution (in bar) superimposed to Lognormal, Weibull, Burr and stable PDFs of daily average $PM10$ concentrations for (a) season 1 (October–April) and (b) season 2 (May to September). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

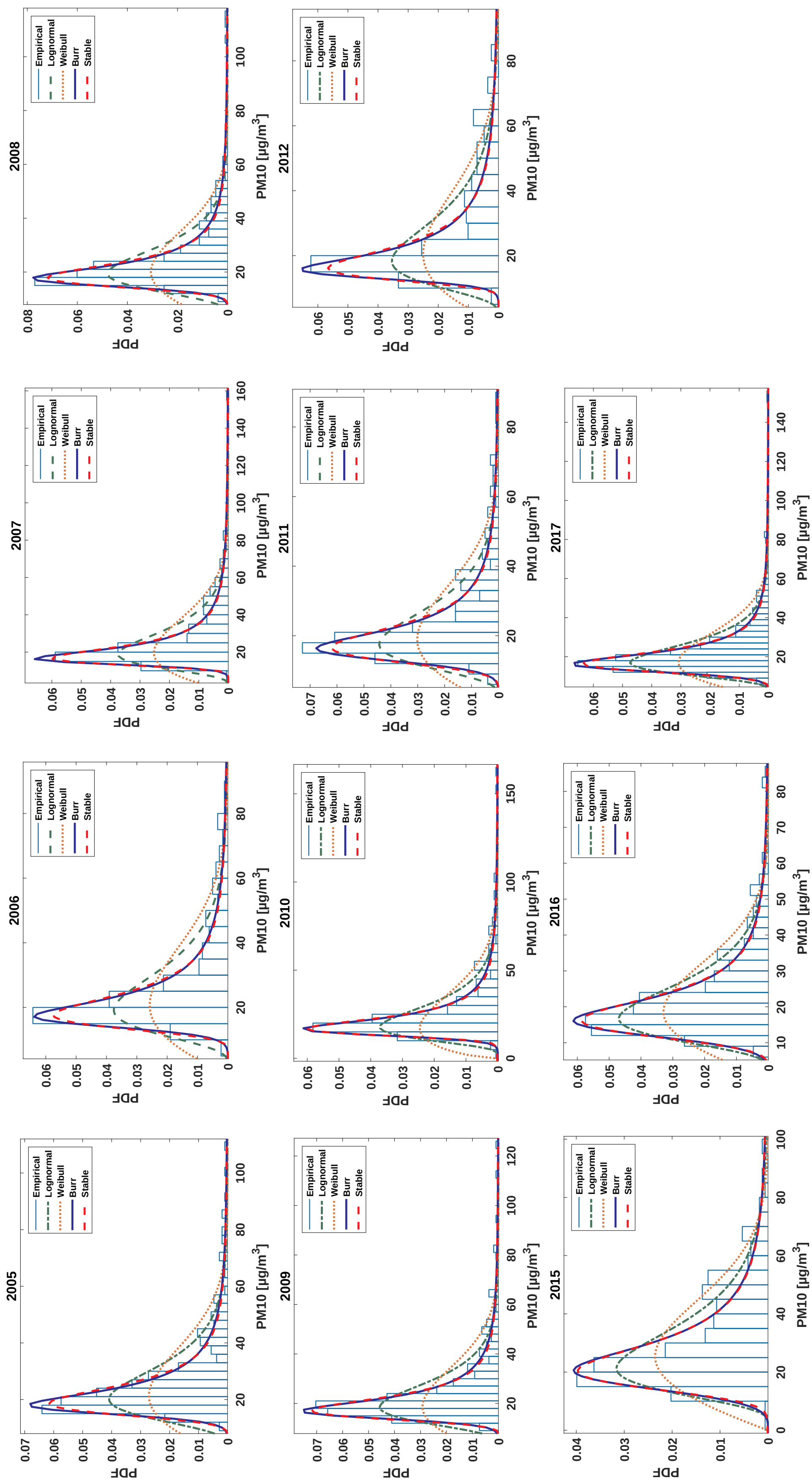

Fig. 10. Empirical distribution (in bar) superimposed to Lognormal, Weibull, Burr and stable PDFs of daily average *PM*10 concentrations by year from 2005 to 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

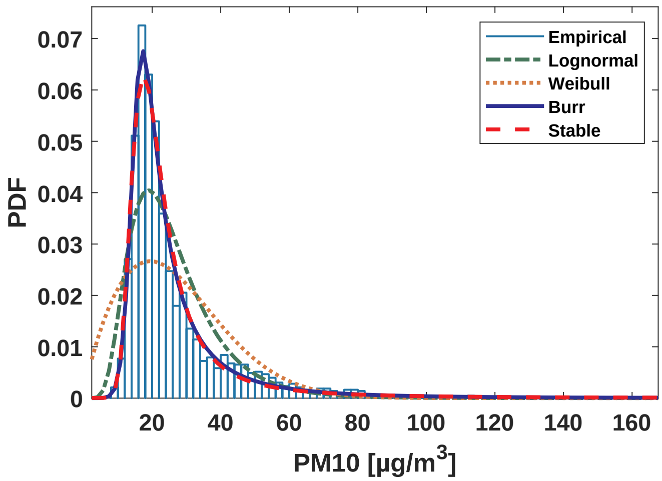

Fig. 11. Empirical distribution (in bar) superimposed to Lognormal, Weibull, Burr and stable PDFs at the statistical stationarity threshold, i.e. t = 66 months. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Parameters values of Lognormal (μ, σ), Weibull (a, b), Burr (α, c, k) and Stable (α, β, γ, δ) distributions for the statistical stationarity threshold (t66 months) and all the years without 2015 (t120 months).

Distribution	Parameters	t66 months (N = 1963)	t120 months (N = 3513)
Lognormal	μ	3.15	3.12
	σ	0.47	0.47
Weibull	a	30.26	29.46
	b	1.81	1.81
Burr	α	15.51	15.21
	c	9.01	8.06
	k	0.24	0.27
Stable	α	1.11	1.14
	β	0.99	0.99
	γ	4.45	4.60
	δ	19.07	18.63

N represents the data point number.

fit for $PM10$ concentrations. In order to observe the fit behavior between the best unimodal model (Burr) and the best mixture model (Burr & Weibull), both distributions are plotted in Fig. 13. The same pattern is found for each year (see Figs. B1 and B2 in Appendix B).

In literature, the most frequently statistical models used to represent the distribution of $PM10$ concentrations are lognormal (Lu, 2002; Lu and Fang, 2003; Gavriil et al., 2006; Yusof et al., 2010; Lonati et al., 2011; Sansuddin et al., 2011; Xi et al., 2013; Dong et al., 2017), Weibull (Yusof et al., 2010), gamma (Sansuddin et al., 2011; Ozel and Cakmakyapan, 2015), log-logistic (Karaca et al., 2005) or Pearson type V (Gavriil et al., 2006; Mijić et al., 2009). One may notice that lognormal distribution is always valid for cities with high $PM10$ concentrations where the background atmosphere is strongly impacted by anthropogenic pollution. Indeed, in Europe and China for example, exhaust from motor vehicles has been regarded as the major source of particulate matter in the ABL due to heavy traffic (Künzli et al., 2000; He et al., 2016). In addition, city configuration may also play a crucial feature in atmospheric pollution dispersion. Contrary to Guadeloupe where most of buildings are four stories high or less (Plocoste et al., 2014), megacities have many high-rise buildings which influence wind circulation and can promote an accumulation of $PM10$ in the surface layer. In Caribbean islands, road traffic is less important (Plocoste et al.,

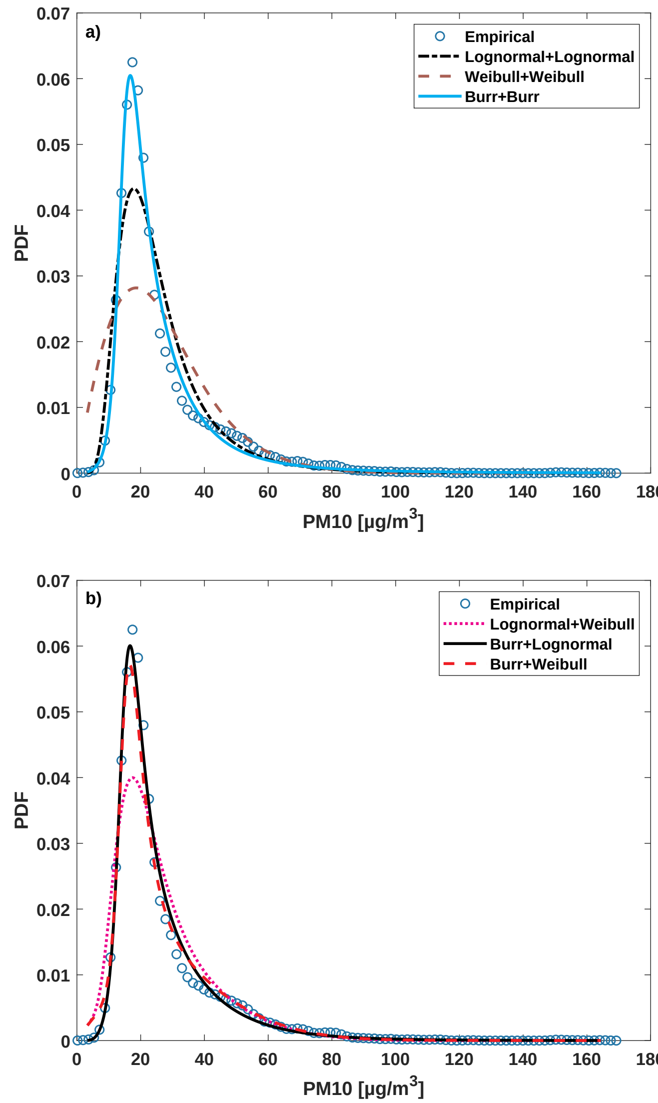

Fig. 12. Probability Density Function (PDF) for all data with different mixture models combinations: a) Lognormal & Lognormal, Weibull & Weibull and Burr & Burr; b) Lognormal & Weibull, Burr & Lognormal and Burr & Weibull. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2018) and the background atmosphere is mostly composed of marine sea salt aerosols (Sellegri et al., 2001; Reid et al., 2003; Prats et al., 2011). High levels of $PM10$ are mainly caused by African seasonal dust (Prospero et al., 2014; Plocoste et al., 2017) which generate two different statistical behaviors, i.e. one for the low dust season and the other for the high dust season. Consequently, in order to enhance unimodal model performance, a Burr & Weibull mixture model is required to fit $PM10$ data. This study is the first of its kind in the Caribbean Basin.

4.3. Extreme events

Classically, study of extreme events is a specific field of statistics (Ghil et al., 2011). It seeks to assess, from a given random variable, the probability of events that are more extreme than any previously observed. In literature, power-law distributions are frequently used to study this kind of events. There are several types of power-law

Table 8
Kolmogorov–Smirnov statistic for unimodal models (Lognormal (KS_{Log}), Weibull (KS_{Wei}), Burr (KS_{Bur}), Stable (KS_{Stab}) and mixture models (Lognormal & Lognormal (KS_{LogLog}), Weibull & Weibull (KS_{WeiWei}), Burr & Burr (KS_{BurBur}), Lognormal & Weibull (KS_{LogWei}), Burr & Lognormal (KS_{BurLog}), Burr & Weibull (KS_{BurWei}) with their critical value ($CV_{\circ}$) for the considered periods.

Study period	KS_{Log}	KS_{Wei}	KS_{Bur}	KS_{Stab}	KS_{LogLog}	KS_{WeiWei}	KS_{BurBur}	KS_{LogWei}	KS_{BurLog}	KS_{BurWei}	$CV_{\circ}$
Overall (N = 3849)	0.119	0.140	***0.025***	0.029	0.051	0.081	0.029	0.041	0.023	**0.017**	0.031
2005 (N = 354)	0.124	0.192	***0.059***	0.065	0.113	0.122	0.057	0.088	0.054	**0.051**	0.102
2006 (N = 358)	0.115	0.179	***0.050***	0.053	0.120	0.126	0.053	0.087	0.048	**0.042**	0.102
2007 (N = 357)	0.126	0.202	***0.053***	0.059	0.115	0.134	0.058	0.095	0.053	**0.048**	0.102
2008 (N = 355)	0.116	0.178	***0.051***	0.059	0.129	0.132	0.050	0.120	0.045	**0.040**	0.102
2009 (N = 365)	0.115	0.192	***0.052***	0.055	0.112	0.139	0.052	0.082	0.051	**0.049**	0.102
2010 (N = 354)	0.131	0.198	***0.057***	0.059	0.116	0.166	0.058	0.113	0.056	**0.054**	0.102
2011 (N = 334)	0.126	0.201	***0.054***	0.057	0.117	0.150	0.056	0.092	0.048	**0.033**	0.105
2012 (N = 337)	0.122	0.217	***0.056***	0.065	0.112	0.143	0.054	0.099	0.052	**0.043**	0.105
2015 (N = 336)	0.098	0.137	***0.074***	0.074	0.073	0.120	0.065	0.060	0.047	**0.040**	0.105
2016 (N = 354)	0.120	0.150	***0.057***	0.062	0.099	0.116	0.056	0.093	0.045	**0.039**	0.102
2017 (N = 345)	0.117	0.169	***0.052***	0.058	0.051	0.081	0.029	0.041	0.023	**0.049**	0.104
Season 1 (N = 2253)	0.124	0.179	***0.023***	0.024	–	–	–	–	–	–	0.041
Season 2 (N = 1596)	0.046	0.091	***0.042***	0.043	–	–	–	–	–	–	0.048

(−) indicates that *KS* value can not be computed.
Values in bold italics indicate the best unimodal distribution while values in bold highlight the best mixture model.

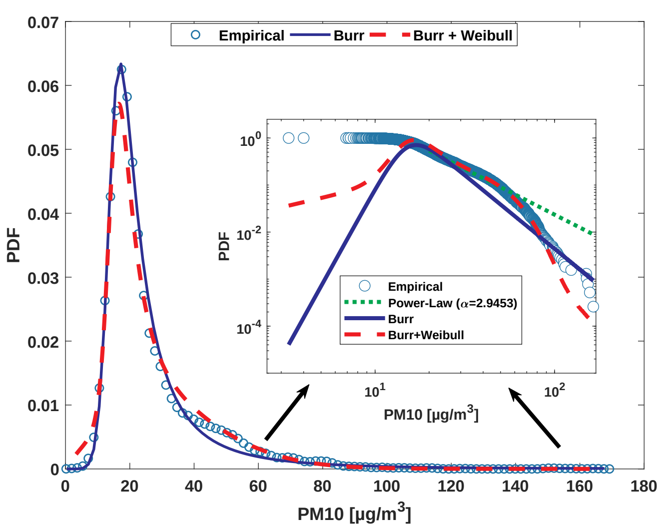

Fig. 13. Probability Density Function (PDF) and its correspondence in log-log plot (in the inset) with the best unimodal distribution (Burr, in blue solid line), the best bimodal distribution (Burr + Weibull, in red dash line) and the power-law fit (in green dot line) for all data. The arrows indicate extreme events location. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distributions as heavy-tail distributions, Pareto distributions and Zipfian distributions (Reed, 2001; Mitzenmacher, 2004), to cite a few. In this study, a robust power-law method developed by Clauset et al. (2009) is applied. This method is based on a statistical framework to discern and quantify the behavior of power-law in empirical data. Fig. 13 shows in the inset the PDF in log-log plot for all *PM*10 concentrations with the power-law fit. Indeed, extreme events are classically studied in log-log plot. An exponent parameter α equal to 2.9453 is found. This value is quite similar to that found by Hsu et al. (2011) in Taïwan with $\alpha = 3$ for heavy-tails of *PM*10 data. Overall, power-law distribution seems suitable. As an example, in Figs. B3 and B4 (in Appendix B), we can see that this distribution perfectly fit *PM*10 data for 2009 and 2015. Nevertheless, one can observe that power-law distribution does not properly fit the extreme events for all years. By using the Burr & Weibull mixture model, we found that this bimodal distribution improve the fit because he follow the same pattern as *PM*10 concentrations for heavy-tail. In Figs. B3 and B4, Burr & Weibull mixture model give a better fit than power-law distribution respectively for 2005–2006-2007 and 2011–2012–2016–2017. Consequently, the Burr & Weibull mixture model can also give a good representation of extreme events. To sum up, the Burr & Weibull mixture model is suitable for both classical events and extreme events.

5. Conclusion

Due to the adverse effects of air pollution on human health, it is essential to model particulate matter distribution. In literature, it is well known that significant problems (e.g., respiratory, preterm birth or cardiovascular) may be caused or worsened by exposure to *PM*10 concentration on a short and long-term basis. This work is the first to investigate *PM*10 frequency distribution and extreme events in the Caribbean area with 11 years of daily database.

With an annual average of 26.4 ± 16.1 $\mu g/m^3$, the descriptive statistics showed that *PM*10 concentrations in Guadeloupe are lower than those measured in cities of Europe, Asia or Africa. In Caribbean islands, high *PM*10 concentrations are mainly due to African dust contrary to megacities where anthropogenic pollution is strong. Two seasons are clearly observed, one from October to April (low dust season) and the other from May to September (high dust season) with respectively 22.0 ± 12.4 $\mu g/m^3$ and 32.9 ± 18.5 $\mu g/m^3$. From May to September, the Intertropical Convergence Zone allows the transport of African dust across the Atlantic Ocean in the Saharan Air Layer. *PM*10 time series exhibited positive values of skewness and kurtosis which indicating a clear right-side feature and an evident peak tail distribution.

The probability density function analysis enabled us to identify extreme events throughout the years but also allowed us to estimate a statistical stationarity threshold of 66 months for *PM*10 data. In others words, 66 months of *PM*10 data would be sufficient for a model to be representative of the 11 years of study. Overall, this highlights the cycle stability of African dust over this last decade.

Thereafter, four theoretical distributions, i.e. lognormal, Weibull, Burr and stable, were used to fit *PM*10 events. In order to identify the best fit, the Kolmogorov–Smirnov statistic test was performed. In this Caribbean context, the results highlighted that Burr & Weibull mixture model is the best distribution to represent *PM*10 daily average concentrations. In this mixture model, the first mode represents the low dust season and the second mode the high dust season.

Given the health risks, we were also interested by extreme events. A robust power-law distribution method developed by Clauset et al. (2009) was used. Achieved results show that the classical power-law distribution seems correct for modelling extreme events, nevertheless our study give a good representation of this type of events with Burr & Weibull mixture model. In other words, the Burr & Weibull mixture model is valid for modelling both classical and extreme events.

To conclude, it is hard to manage and reduce $PM10$ emissions in West Indian arc because one of the main emitters of particulate matter in this area is from natural large-scale sources. Here, only data from Guadeloupe were analyzed. In order to validate these distributions for the Caribbean basin, it would be necessary to extend this survey to other islands in the West Indian arc in future studies.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgements

The authors are very grateful to the anonymous reviewers for their valuable comments and constructive suggestions, which helped us to improve substantially the quality of the paper. The authors would like to thank Guadeloupe air quality network (Gwad'Air) for providing air quality data and Mr. Sylvio Laventure (geomatic specialist) for mapping assistance.

Funding

The authors declare that they have not received any fund for the present paper. The paper is the sole work of the authors and is not a part/product of any project.

Appendix A. Tables

Table A1
Parameters values of lognormal (μ_1, σ_1), Weibull (a_1, b_1) and Burr (α_1, c_1, k_1) distributions by year for season 1, i.e. from October to April.

Year	Lognormal		Weibull		Burr		
	μ_1	σ_1	a_1	b_1	α_1	c_1	k_1
2005 (N = 210)	3.0365	0.3024	24.3903	2.9968	17.1962	8.5940	0.4526
2006 (N = 208)	2.9996	0.3574	24.4498	2.1989	15.8217	9.4366	0.3645
2007 (N = 211)	3.0687	0.4470	27.4558	1.8474	14.5068	9.8303	0.2379
2008 (N = 203)	3.0121	0.2840	23.7420	2.8437	16.6206	11.2966	0.3634
2009 (N = 212)	2.9465	0.3079	22.5756	2.4312	16.0552	9.1809	0.4649
2010 (N = 203)	3.0454	0.4916	27.6370	1.5422	13.9596	10.3180	0.2198
2011 (N = 210)	2.9414	0.3660	23.1937	2.0663	15.2252	8.1388	0.4294
2012 (N = 190)	2.9915	0.4339	25.4266	1.7754	14.1911	11.6223	0.2336
2015 (N = 194)	3.0897	0.4043	27.0675	2.1488	18.0208	6.3543	0.5242
2016 (N = 208)	2.9420	0.4192	23.7107	2.0093	14.3341	6.8368	0.4108
2017 (N = 204)	2.9445	0.4089	24.0170	1.6126	14.8547	8.0342	0.3994

N represents the data point number.

Table A2
Parameters values of lognormal (μ_2, σ_2), Weibull (a_2, b_2) and Burr (α_2, c_2, k_2) distributions by year for season 2, i.e. from May to September.

Year	Lognormal		Weibull		Burr		
	μ_2	σ_2	a_2	b_2	α_2	c_2	k_2
2005 (N = 144)	3.4299	0.5023	40.0582	1.9309	18.1054	8.8988	0.1990
2006 (N = 150)	3.4655	0.5216	41.5980	2.0168	28.6209	3.5031	0.8112
2007 (N = 146)	3.3539	0.5224	37.4320	1.8083	17.0426	6.9047	0.2553
2008 (N = 152)	3.2715	0.4702	33.6866	1.8968	18.3638	6.0171	0.3778
2009 (N = 153)	3.3154	0.4829	35.5079	1.8534	18.6122	6.3573	0.3418
2010 (N = 151)	3.2959	0.5416	35.6927	1.7606	19.2040	4.9844	0.4422
2011 (N = 124)	3.3054	0.5157	35.1987	2.0642	34.4411	2.8845	1.5938
2012 (N = 150)	3.4677	0.5481	41.7728	2.1172	310.7700	2.1332	73.0695
2015 (N = 141)	3.6525	0.4832	48.2580	2.5584	209.4135	2.5930	45.7057
2016 (N = 146)	3.1386	0.4870	29.4682	2.0960	22.9322	3.5269	0.9997
2017 (N = 142)	3.1759	0.4927	30.9824	1.8740	16.6098	5.7531	0.3859

N represents the data point number.

Appendix B. Figures

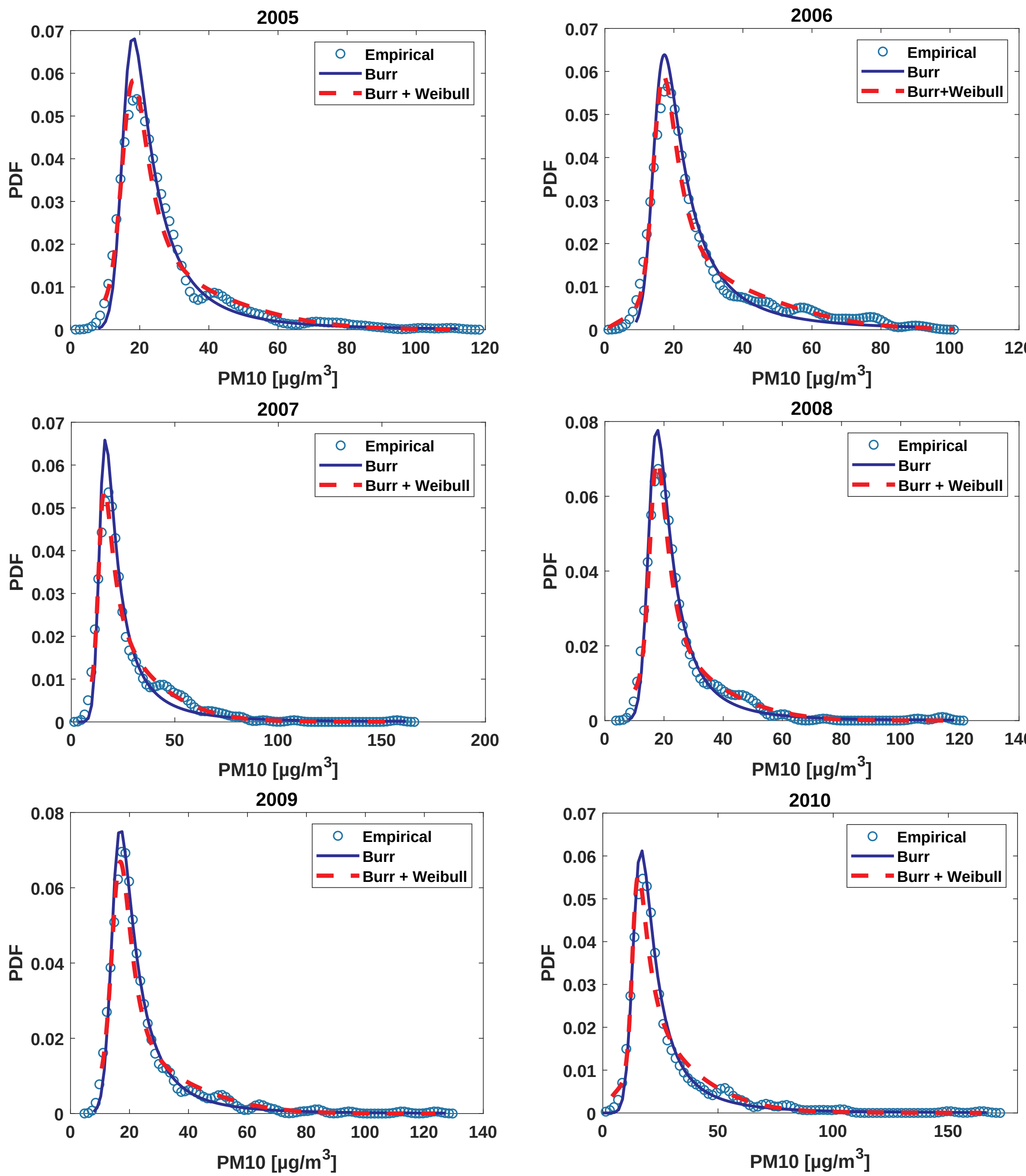

Fig. B1. Probability Density Function (PDF) of *PM*10 concentrations with the best unimodal distribution (Burr, in blue solid line) and the best distribution model (Burr + Weibull, in red dash line) by year from 2005 to 2010. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

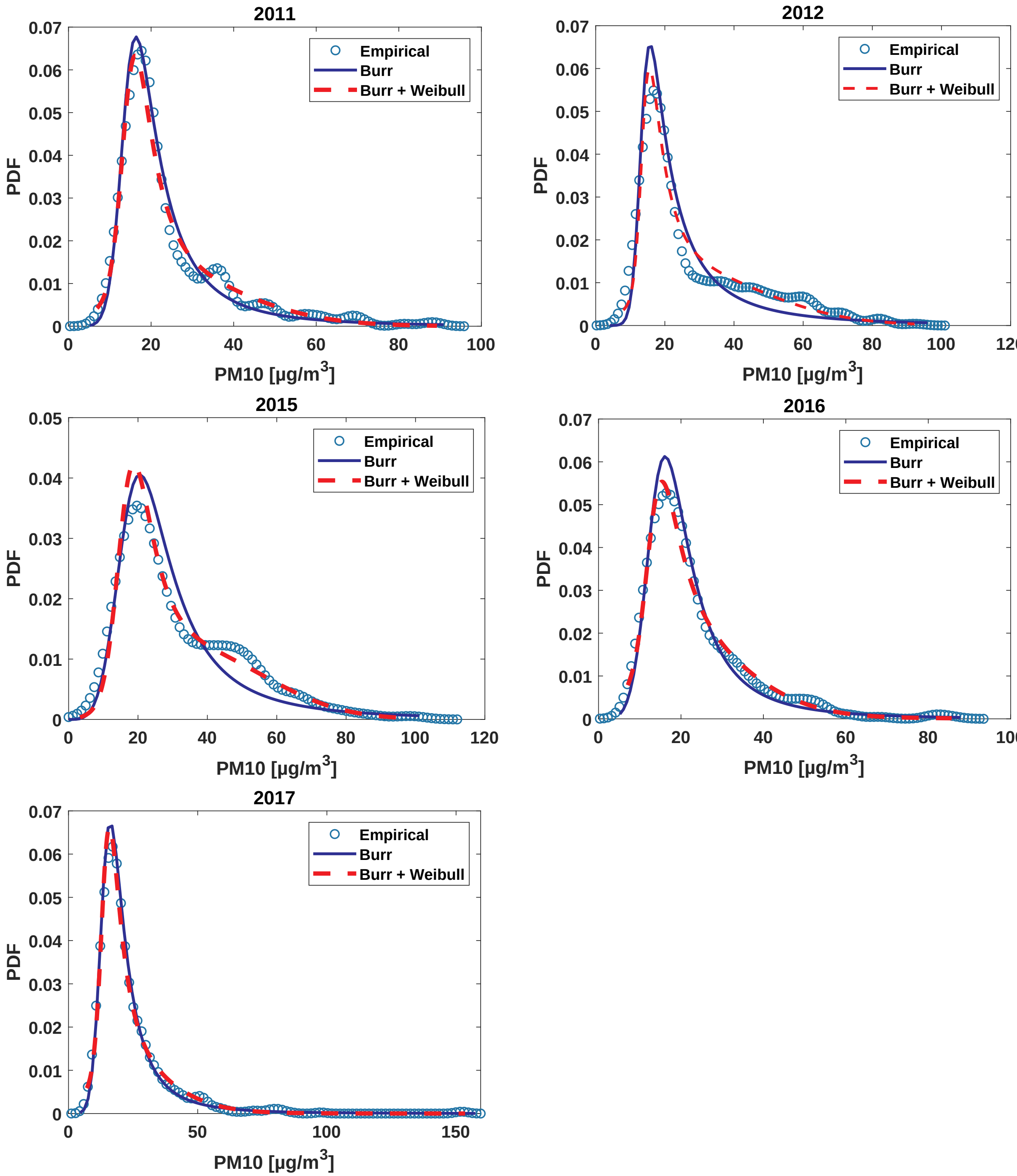

Fig. B2. Probability Density Function (PDF) of *PM*10 concentrations with the best unimodal distribution (Burr, in blue solid line) and the best distribution model (Burr + Weibull, in red dash line) by year from 2011 to 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

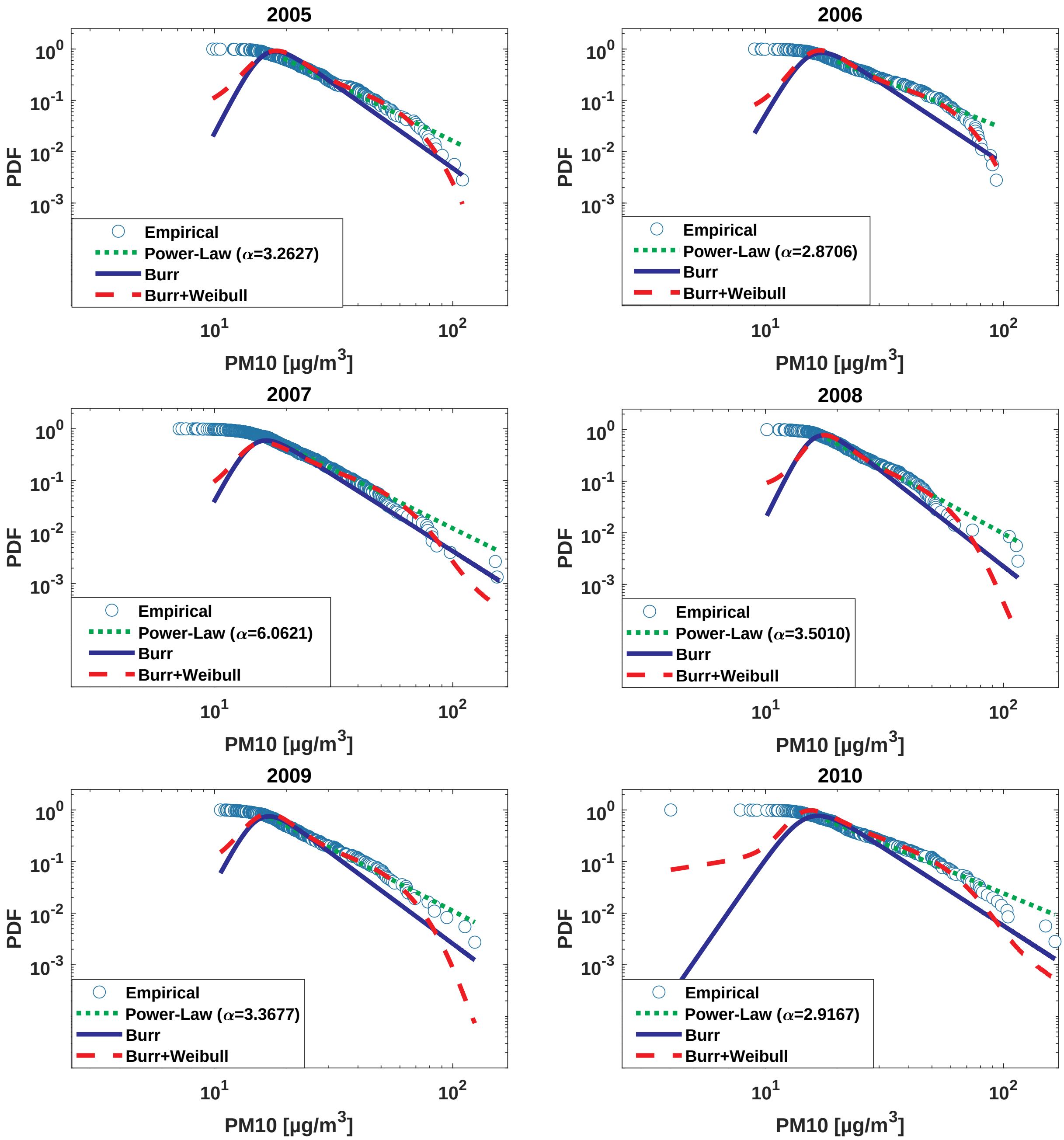

Fig. B3. Probability Density Function (PDF) in log-log plot with the best unimodal distribution (Burr, in blue solid line), the best bimodal distribution (Burr + Weibull, in red dash line) and the power-law fit (in green dot line) by year from 2005 to 2010. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

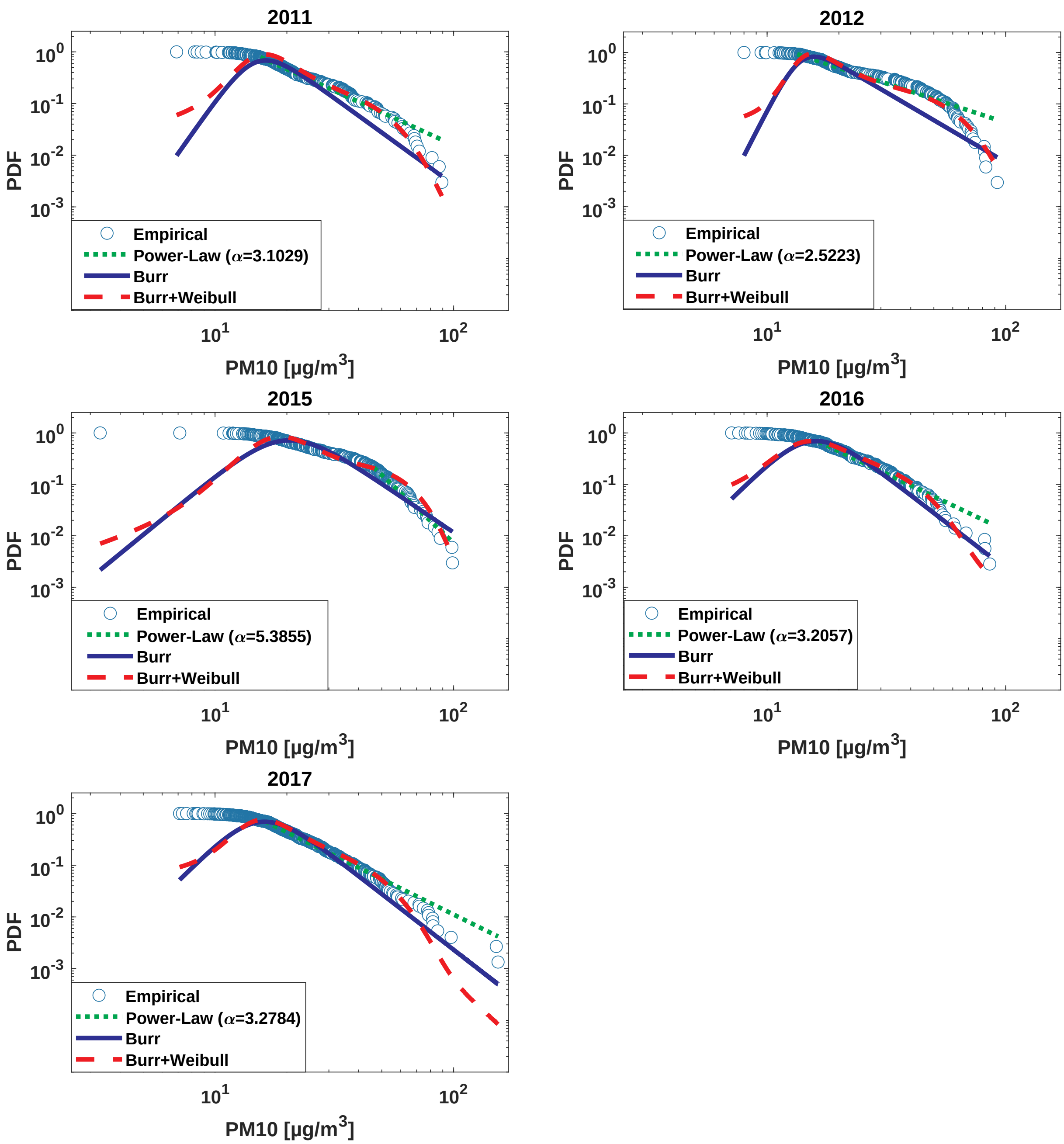

Fig. B4. Probability Density Function (PDF) in log-log plot with the best unimodal distribution (Burr, in blue solid line), the best bimodal distribution (Burr + Weibull, in red dash line) and the power-law fit (in green dot line) by year from 2011 to 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Adams, A.M., Prospero, J.M., Zhang, C., 2012. CALIPSO-derived three-dimensional structure of aerosol over the Atlantic Basin and adjacent continents. J. Clim. 25 (19), 6862–6879.

Al-Habash, A., Andrews, L.C., Phillips, R.L., 2001. Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media. Opt. Eng. 40 (8), 1554–1563.

Ault, T.R., Cole, J.E., Overpeck, J.T., Pederson, G.T., Meko, D.M., 2014. Assessing the risk of persistent drought using climate model simulations and paleoclimate data. J. Clim. 27 (20), 7529–7549.

Bai, J., Wang, J., Xiao, H., Ju, H., Liu, Y., Gao, Z., 2013. Weibull distribution for modeling drying of grapes and its application. Trans. Chin. Soc. Agric. Eng. 29 (16), 278–285.

Bak, P., Christensen, K., Danon, L., Scanlon, T., 2002. Unified scaling law for earthquakes. Phys. Rev. Lett. 88 (17), 178501.

Baldasano, J., Valera, E., Jimenez, P., 2003. Air quality data from large cities. Sci. Total Environ. 307 (1–3), 141–165.

Belkacemi, H., Marcos, S., 2007. Robust subspace-based algorithms for joint angle/Doppler estimation in non-Gaussian clutter. Signal Process. 87 (7), 1547–1558.

Ben-Ami, Y., Koren, I., Altaratz, O., Kostinski, A., Lehahn, Y., 2012. Discernible rhythm in the spatio/temporal distributions of transatlantic dust. Atmos. Chem. Phys. 12 (5), 2253–2262.

Bigi, A., Harrison, R.M., 2010. Analysis of the air pollution climate at a central urban background site. Atmos. Environ. 44 (16), 2004–2012.

Boccignone, G., Ferraro, M., 2013. Gaze shift behavior on video as composite information foraging. Signal Process. Image Commun. 28 (8), 949–966.

Bowman, A., Hall, P., Prvan, T., 1998. Bandwidth selection for the smoothing of distribution functions. Biometrika 85 (4), 799–808.

Burnecki, K., Kukla, G., Weron, R., 2000. Property insurance loss distributions. Phys. A Stat. Mech. Appl. 287 (1–2), 269–278.

Burr, I.W., 1942. Cumulative frequency functions. Ann. Math. Stat. 13 (2), 215–232.

Calif, R., 2012. Pdf models and synthetic model for the wind speed fluctuations based on the resolution of Langevin equation. Appl. Energy 99, 173–182.

Campbell, J.W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. Oceans 100 (C7), 13237–13254.
Carlson, T.N., Prospero, J.M., 1972. The large-scale movement of Saharan air outbreaks over the northern equatorial Atlantic. J. Appl. Meteorol. 11 (2), 283–297.
Chen, Q., Gerlach, R.H., 2013. The two-sided Weibull distribution and forecasting financial tail risk. Int. J. Forecast. 29 (4), 527–540.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.
Clergue, C., Dellinger, M., Buss, H., Gaillardet, J., Benedetti, M., Dessert, C., 2015. Influence of atmospheric deposits and secondary minerals on Li isotopes budget in a highly weathered catchment, Guadeloupe (Lesser Antilles). Chem. Geol. 414, 28–41.
Court-Brown, C.M., Caesar, B., 2006. Epidemiology of adult fractures: a review. Injury 37 (8), 691–697.
Cranmer, K., 2001. Kernel estimation in high-energy physics. Comput. Phys. Commun. 136 (3), 198–207.
Crow, E.L., Shimizu, K., 1987. Lognormal Distributions. Marcel Dekker, New York.
Ditlevsen, P.D., 1999. Observation of α-stable noise induced millennial climate changes from an ice-core record. Geophys. Res. Lett. 26 (10), 1441–1444.
Dong, Q., Wang, Y., Li, P., 2017. Multifractal behavior of an air pollutant time series and the relevance to the predictability. Environ. Pollut. 222, 444–457.
Dunion, J.P., 2011. Rewriting the climatology of the tropical North Atlantic and Caribbean Sea atmosphere. J. Clim. 24 (3), 893–908.
Euphrasie-Clotilde, L., 2018. Climatology of Desert Dust Transportation Over the Atlantic Toward Caribbean Area. Ph.D. thesis. Antilles University.
Euphrasie-Clotilde, L., Molinié, J., Feuillard, T., Brute, F., 2017. The relationship between coastal West African dust level and Caribbean island dust. WIT Trans. Ecol. Environ. 211, 121–127.
Euphrasie-Clotilde, L., Plocoste, T., Feuillard, T., Velasco-Merino, C., Mateos, D., Toledano, C., Brute, F.N., Bassette, C., Gobinddass, M., 2020. Assessment of a new detection threshold for $PM10$ concentrations linked to African dust events in the Caribbean Basin. Atmos. Environ. 224, 117354.
European Union, E.U., 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Off. J. Eur. Union 108, 1–44.
Forbes, C., Evans, M., Hastings, N., Peacock, B., 2011. Statistical Distributions. John Wiley & Sons.
Ganora, D., Laio, F., 2015. Hydrological applications of the Burr distribution: practical method for parameter estimation. J. Hydrol. Eng. 20 (11), 04015024.
Gavriil, I., Grivas, G., Kassomenos, P., Chaloulakou, A., Spyrellis, N., 2006. An application of theoretical probability distributions, to the study of PM10 and PM2.5 time series in Athens, Greece. Glob. NEST J. 8, 241–251.
Georgopoulos, P.G., Seinfeld, J.H., 1982. Statistical distributions of air pollutant concentrations. Environ. Sci. Technol. 16 (7), 401A–416A.
Ghil, M., Yiou, P., Hallegatte, S., Malamud, B., Naveau, P., Soloviev, A., Friederichs, P., Keilis-Borok, V., Kondrashov, D., Kossobokov, V., et al., 2011. Extreme events: dynamics, statistics and prediction. Nonlinear Process. Geophys. 18 (3), 295–350.
Ginoux, P., Prospero, J.M., Gill, T.E., Hsu, N.C., Zhao, M., 2012. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. Rev. Geophys. 50 (3).
Gläser, G., Wernli, H., Kerkweg, A., Teubler, F., 2015. The transatlantic dust transport from North Africa to the Americas—its characteristics and source regions. J. Geophys. Res.-Atmos. 120 (21), 11–231.
Gupta, A., Patil, R., Gupta, S., 2004. A statistical analysis of particulate data sets for Jawaharlal Nehru port and surrounding harbour region in India. Environ. Monit. Assess. 95 (1–3), 295–309.
Haehner, B.D., Gorski, J.C., Vandenbranden, M., Wrighton, S.A., Janardan, S.K., Watkins, P.B., Hall, S.D., 1996. Bimodal distribution of renal cytochrome P450 3A activity in humans. Mol. Pharmacol. 50 (1), 52–59.
Hansen, C., Neller, A., Williams, G., Simpson, R., 2006. Maternal exposure to low levels of ambient air pollution and preterm birth in Brisbane, Australia. BJOG Int. J. Obstet. Gynaecol. 113 (8), 935–941.
He, H.d., Pan, W., Lu, W. z, Xue, Y., Peng, G. h, 2016. Multifractal property and long-range cross-correlation behavior of particulate matters at urban traffic intersection in Shanghai. Stoch. Env. Res. Risk A. 30, 1515–1525.
Hoek, G., Pattenden, S., Willers, S., Antova, T., Fabianova, E., Braun-Fahrländer, C., Forastiere, F., Gehring, U., Luttmann-Gibson, H., Grize, L., et al., 2012. PM10, and children's respiratory symptoms and lung function in the PATY study. Eur. Respir. J. 40 (3), 538–547.
Hsu, H., Lin, C., Guenther, A., Tribbia, J., Liu, S., 2011. Air-chemistry "turbulence": power-law scaling and statistical regularity. Atmos. Chem. Phys. 11, 8395–8413.
Jaramillo, O., Borja, M., 2004. Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew. Energy 29 (10), 1613–1630.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Sana, S., White, G., Woollen, J., others, Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–472.
Kan, H., Chen, B.H., 2004. Statistical distributions of ambient air pollutants in Shanghai, China. Biomed. Environ. Sci. 17 (3), 366–372.
Kao, A.S., Friedlander, S.K., 1995. Frequency distributions of PM10 chemical components and their sources. Environ. Sci. Technol. 29 (1), 19–28.
Karaca, F., Alagha, O., Ertürk, F., 2005. Statistical characterization of atmospheric PM10 and PM2.5 concentrations at a non-impacted suburban site of Istanbul, Turkey. Chemosphere 59 (8), 1183–1190.
Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. Vol. 470 John Wiley & Sons.
Knippertz, P., Todd, M.C., et al., 2010. J. Geophys. Res.-Atmos. 115 (D12).
Kolmogorov, A., 1933. Sulla determinazione empirica di una legge di distribuzione. Inst. Ital. Attuari, Giorn 4, 83–91.
Kosugi, K., 1996. Lognormal distribution model for unsaturated soil hydraulic properties. Water Resour. Res. 32 (9), 2697–2703.
Koutsoukas, A., Lowe, R., KalantarMotamedi, Y., Mussa, H.Y., Klaffke, W., Mitchell, J.B., Glen, R.C., Bender, A., 2013. In silico target predictions: defining a benchmarking data set and comparison of performance of the multiclass Nave Bayes and Parzen-Rosenblatt window. J. Chem. Inf. Model. 53 (8), 1957–1966.
Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak, F., Puybonnieux-Texier, V., Quénel, P., et al., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. Lancet 356, 795–801.
Kwon, S.B., Jeong, W., Park, D., Kim, K.T., Cho, K.H., 2015. A multivariate study for characterizing particulate matter (PM10, PM2.5, and PM1) in Seoul metropolitan subway stations, Korea. J. Hazard. Mater. 297, 295–303.
Lavallée, D., Archuleta, R.J., 2003. Stochastic modeling of slip spatial complexities for the 1979 Imperial Valley, California, earthquake. Geophys. Res. Lett. 30 (5).
Le Tertre, A., Medina, S., Samoli, E., Forsberg, B., Michelozzi, P., Boumghar, A., Vonk, J., Bellini, A., Atkinson, R., Ayres, J., et al., 2002. Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. J. Epidemiol. Community Health 56 (10), 773–779.
Liang, Y., Chen, W., 2013. A survey on computing Lévy stable distributions and a new MATLAB toolbox. Signal Process. 93 (1), 242–251.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. AIBS Bull. 51 (5), 341–352.
Lonati, G., Cernuschi, S., Giugliano, M., 2011. The duration of PM10 concentration in a large metropolitan area. Atmos. Environ. 45 (1), 137–146.
Lu, H.C., 2002. The statistical characters of PM10 concentration in Taiwan area. Atmos. Environ. 36 (3), 491–502.
Lu, H.C., 2003. Comparisons of statistical characteristic of air pollutants in Taiwan by frequency distribution. J. Air Waste Manage. Assoc. 53 (5), 608–616.
Lu, H.C., 2004. Estimating the emission source reduction of PM10 in Central Taiwan. Chemosphere 54 (7), 805–814.
Lu, H.C., Fang, G.C., 2003. Predicting the exceedances of a critical PM10 concentration—a case study in Taiwan. Atmos. Environ. 37 (25), 3491–3499.
Maheswaran, R., Haining, R.P., Brindley, P., Law, J., Pearson, T., Fryers, P.R., Wise, S., Campbell, M.J., 2005. Outdoor air pollution, mortality, and hospital admissions from coronary heart disease in Sheffield, United Kingdom: a small-area level ecological study. Eur. Heart J. 26 (23), 2543–2549.
Malamud, B.D., Turcotte, D.L., 2006. The applicability of power-law frequency statistics to floods. J. Hydrol. 322 (1–4), 168–180.
Massey Jr., F.J., 1951. The Kolmogorov–Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46 (253), 68–78.
Matus, K., Nam, K.M., Selin, N.E., Lamsal, L.N., Reilly, J.M., Paltsev, S., 2012. Health damages from air pollution in China. Glob. Environ. Chang. 22 (1), 55–66.
Medina, S., Plasencia, A., Ballester, F., Mücke, H., Schwartz, J., 2004. Apheis: public health impact of PM10 in 19 European cities. J. Epidemiol. Community Health 58 (10), 831–836.
Mijić, Z., Tasić, M., Rajšić, S., Novaković, V., 2009. The statistical characters of PM10 in Belgrade area. Atmos. Res. 92 (4), 420–426.
Mitzenmacher, M., 2004. A brief history of generative models for power-law and lognormal distributions. Internet Math. 1 (2), 226–251.
Momtazan, M., Geravandi, S., Rastegarimehr, B., Valipour, A., Ranjbarzadeh, A., Yari, A.R., Dobaradaran, S., Bostan, H., Farhadi, M., Darabi, F., et al., 2018. An investigation of particulate matter and relevant cardiovascular risks in Abadan and Khorramshahr in 2014–2016. Toxin Rev. 1–8.
Morel, B., Yeh, S., Cifuentes, L., 1999. Statistical distributions for air pollution applied to the study of the particulate problem in Santiago. Atmos. Environ. 33 (16), 2575–2585.
Nolan, J., 2003. Stable Distributions: Models for Heavy-Tailed Data. Birkhauser, New York.
Nolan, J.P., 2014. Financial modeling with heavy-tailed stable distributions. Wiley Interdiscip. Rev. Comput. Stat. 6 (1), 45–55.
Ormerod, P., Mounfield, C., 2001. Power law distribution of the duration and magnitude of recessions in capitalist economies: breakdown of scaling. Phys. A Stat. Mech. Appl. 293 (3–4), 573–582.
Ott, W.R., 1990. A physical explanation of the lognormality of pollutant concentrations. J. Air Waste Manage. Assoc. 40 (10), 1378–1383.
Ozel, G., Cakmakyapan, S., 2015. A new approach to the prediction of PM10 concentrations in Central Anatolia Region, Turkey. Atmos. Pollut. Res. 6 (5), 735–741.
Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. J. Econ. Surv. 22 (4), 711–751.
Papalexiou, S.M., Koutsoyiannis, D., 2012. Entropy based derivation of probability distributions: a case study to daily rainfall. Adv. Water Resour. 45, 51–57.
Papoulis, A., Pillai, S.U., 2002. Probability, Random Variables, and Stochastic Processes. Tata McGraw-Hill Education.
Pereyra, M., Batatia, H., 2012. Modeling ultrasound echoes in skin tissues using symmetric α-stable processes. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 59 (1).
Plocoste, T., Calif, R., 2019. Spectral observations of PM10 fluctuations in the Hilbert space. In: Functional Calculus [Working Title]. IntechOpen, pp. 1–13.
Plocoste, T., Jacoby-Koaly, S., Molinié, J., Petit, R., 2014. Evidence of the effect of an urban heat island on air quality near a landfill. Urban Clim. 10, 745–757.
Plocoste, T., Calif, R., Jacoby-Koaly, S., 2017. Temporal multiscaling characteristics of particulate matter $PM10$ and ground-level ozone O_3 concentrations in Caribbean region. Atmos. Environ. 169, 22–35.
Plocoste, T., Dorville, J.F., Monjoly, S., Jacoby-Koaly, S., André, M., 2018. Assessment of

nitrogen oxides and ground-level ozone behavior in a dense air quality station network: case study in the lesser antilles arc. J. Air Waste Manage. Assoc. 68 (12), 1278–1300.
Plocoste, T., Calif, R., Jacoby-Koaly, S., 2019. Multi-scale time dependent correlation between synchronous measurements of ground-level ozone and meteorological parameters in the Caribbean Basin. Atmos. Environ. 211, 234–246.
Plocoste, T., Pavón-Dominguez, P., 2020. Temporal scaling study of particulate matter (PM10) and solar radiation influences on air temperature in the Caribbean basin using a 3D joint multifractal analysis. Atmos. Environ. 222, 117115.
Polichetti, G., Cocco, S., Spinali, A., Trimarco, V., Nunziata, A., 2009. Effects of particulate matter (PM10, PM2. 5 and PM1) on the cardiovascular system. Toxicology 261 (1–2), 1–8.
Pope III, C.A., Dockery, D.W., Spengler, J.D., Raizenne, M.E., 1991. Respiratory health and PM10 pollution: a daily time series analysis. Am. Rev. Respir. Dis. 144 (3_pt_1) nfopages668–674.
Prats, N., Cachorro, V., Berjón, A., Toledano, C., De Frutos, A., 2011. Column-integrated aerosol microphysical properties from AERONET Sun photometer over southwestern Spain. Atmos. Chem. Phys. 11 (24), 12535–12547.
Prospero, J.M., Blades, E., Mathison, G., Naidu, R., 2005. Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust. Aerobiologia 21 (1), 1–19.
Prospero, J.M., Collard, F.X., Molinié, J., Jeannot, A., 2014. Characterizing the annual cycle of African dust transport to the Caribbean Basin and South America and its impact on the environment and air quality. Glob. Biogeochem. Cycles 28, 757–773.
Pujades-Rodrguez, M., Dantony, E., Pinoges, L., Ecochard, R., Etard, J.F., Carrillo-Casas, E., Szumilin, E., AIDS Working Group of Médecins Sans Frontières, et al., 2011. Toxicity associated with stavudine dose reduction from 40 to 30 mg in first-line antiretroviral therapy. PLoS One 6 (11) e28112.
Querol, X., Alastuey, A., Ruiz, C., Artiñano, B., Hansson, H., Harrison, R., Buringh, E.T., Ten Brink, H., Lutz, M., Bruckmann, P., et al., 2004. Speciation and origin of PM10 and PM2.5 in selected European cities. Atmos. Environ. 38 (38), 6547–6555.
Rastelli, E., Corinaldesi, C., Dell'Anno, A., Martire, M.L., Greco, S., Facchini, M.C., Rinaldi, M., O'Dowd, C., Ceburnis, D., Danovaro, R., 2017. Transfer of labile organic matter and microbes from the ocean surface to the marine aerosol: an experimental approach. Sci. Rep. 7 (1) 11475.
Reed, W.J., 2001. The Pareto, Zipf and other power-laws. Econ. Lett. 74 (1), 15–19.
Reed, W.J., Jorgensen, M., 2004. The double Pareto-lognormal distribution — a new parametric model for size distributions. Commun. Stat. Theory Methods 33 (8), 1733–1753.
Reid, J.S., Jonsson, H.H., Maring, H.B., Smirnov, A., Savoie, D.L., Cliff, S.S., Reid, E.A., Livingston, J.M., Meier, M.M., Dubovik, O., Tsay, S., 2003. Comparison of size and morphological measurements of coarse mode dust particles from Africa. J. Geophys. Res.-Atmos. 108 (D19).
Reynolds, D.A., 2009. Gaussian mixture models. Encycl. Biom. 741, 1–15.
Ritz, B., Yu, F., Chapa, G., Fruin, S., 2000. Effect of air pollution on preterm birth among children born in Southern California between 1989 and 1993. Epidemiology 502–511.
Rolph, G., Stein, A., Stunder, B., 2017. Real-time environmental applications and display sYstem: READY. Environ. Model. Softw. 95, 210–228.
Roylance, B., Pocock, G., 1983. Wear studies through particle size distribution I: application of the Weibull distribution to ferrography. Wear 90 (1), 113–136.
Sajjadi, S.A., Zolfaghari, G., Adab, H., Allahabadi, A., Delsouz, M., 2017. Measurement and modeling of particulate matter concentrations: applying spatial analysis and regression techniques to assess air quality. MethodsX 4, 372–390.
Sansuddin, N., Ramli, N.A., Yahaya, A.S., Yusof, N.F.F.M., Ghazali, N.A., Al Madhoun, W.A., 2011. Statistical analysis of PM10 concentrations at different locations in Malaysia. Environ. Monit. Assess. 180 (1–4), 573–588.
Santhanam, M., Kantz, H., 2008. Return interval distribution of extreme events and long-term memory. Phys. Rev. E 78 (5), 051113.
Saulo, H., Leiva, V., Ziegelmann, F.A., Marchant, C., 2013. A nonparametric method for estimating asymmetric densities based on skewed Birnbaum–Saunders distributions applied to environmental data. Stoch. Env. Res. Risk A. 27 (6), 1479–1491.
Schwab, J., Podsiadlowski, P., Rappaport, S., 2010. Further evidence for the bimodal distribution of neutron-star masses. Astrophys. J. 719 (1), 722.
Schwartz, J., Morris, R., 1995. Air pollution and hospital admissions for cardiovascular disease in Detroit, Michigan. Am. J. Epidemiol. 142 (1), 23–35.
Seguro, J., Lambert, T., 2000. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J. Wind Eng. Ind. Aerodyn. 85 (1), 75–84.
Sellegri, K., Gourdeau, J., Putaud, J.P., Despiau, S., 2001. Chemical composition of marine aerosol in a Mediterranean coastal zone during the FETCH experiment. J. Geophys. Res.-Atmos. 106 (D11), 12023–12037.
Sharma, S., Sharma, P., Khare, M., Kwatra, S., 2016. Statistical behavior of ozone in urban environment. Sustain. Environ. Res. 26 (3), 142–148.
Silva, J., Rojas, J., Norabuena, M., Molina, C., Toro, R.A., Leiva-Guzmán, M.A., 2017. Particulate matter levels in a South American megacity: the metropolitan area of Lima-Callao, Peru. Environ. Monit. Assess. 189 (12) 635.
Silverman, B.W., 2018. Density Estimation for Statistics and Data Analysis. Routledge.
Singhi, S., Maddala, G., 1976. A function for the size distribution of incomes. Econometrica 44 (5), 963–970.
Smirnov, N., 1948. Table for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19 (2), 279–281.
Smirnov, A., Holben, B., Savoie, D., Prospero, J., Kaufman, Y., Tanre, D., Eck, T., Slutsker, I., 2000. Relationship between column aerosol optical thickness and in situ ground based dust concentrations over Barbados. Geophys. Res. Lett. 27 (11), 1643–1646.
Sornette, D., 2002. Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes, and human birth. Proc. Natl. Acad. Sci. 99 (Suppl. 1), 2522–2529.
Stein, A., Draxler, R.R., Rolph, G.D., Stunder, B.J., Cohen, M., Ngan, F., 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bull. Am. Meteorol. Soc. 96 (12), 2059–2077.
Stieb, D.M., Chen, L., Eshoul, M., Judek, S., 2012. Ambient air pollution, birth weight and preterm birth: a systematic review and meta-analysis. Environ. Res. 117, 100–111.
Suh, Y.J., Kim, H., Seo, J.H., Park, H., Kim, Y.J., Hong, Y.C., Ha, E.H., 2009. Different effects of PM10 exposure on preterm birth by gestational period estimated from time-dependent survival analyses. Int. Arch. Occup. Environ. Health 82 (5), 613–621.
Sumaiya, M., Kumari, R.S.S., 2017. Sar image despeckling using heavy-tailed Burr distribution. SIViP 11 (1), 49–55.
Taskin, F., Zaim, O., 2000. Searching for a Kuznets curve in environmental efficiency using kernel estimation. Econ. Lett. 68 (2), 217–223.
Tennekes, H., Lumley, J.L., Lumley, J., et al., 1972. A First Course in Turbulence. MIT press.
Thorpe, J., 1977. Bimodal distribution of length of juvenile Atlantic salmon (*Salmo salar* L.) under artificial rearing conditions. J. Fish Biol. 11 (2), 175–184.
Velasco-Merino, C., Mateos, D., Toledano, C., Prospero, J.M., Molinie, J., Euphrasie-Clotilde, L., González, R., Cachorro, V.E., Calle, A., Frutos, A. Md, 2018. Impact of long-range transport over the Atlantic Ocean on Saharan dust optical and micro-physical properties based on AERONET data. Atmos. Chem. Phys. 18 (13), 9411–9424.
Virkar, Y., Clauset, A., et al., 2014. Power-law distributions in binned empirical data. Ann. Appl. Stat. 8 (1), 89–119.
Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. Chapman and Hall/CRC.
Windsor, H., Toumi, R., 2001. Scaling and persistence of UK pollution. Atmos. Environ. 35 (27), 4545–4556.
Wong, T.W., Lau, T.S., Yu, T.S., Neller, A., Wong, S.L., Tam, W., Pang, S.W., 1999. Air pollution and hospital admissions for respiratory and cardiovascular diseases in Hong Kong. Occup. Environ. Med. 56 (10), 679–683.
World Health Organization, W.H.O., 2013. Review of Evidence on Health Aspects of Air Pollution – REVIHAAP Project. Technical Report. WHO Regional Office for Europe.
Xi, W., Chen, R.J., Chen, B.H., Kan, H.D., 2013. Application of statistical distribution of PM10 concentration in air quality management in 5 representative cities of China. Biomed. Environ. Sci. 26 (8), 638–646.
Yusof, N.F.F.M., Ramli, N.A., Yahaya, A.S., Sansuddin, N., Ghazali, N.A., al Madhoun, W., 2010. Monsoonal differences and probability distribution of PM10 concentration. Environ. Monit. Assess. 163 (1–4), 655–667.
Zaliapin, I., Kagan, Y.Y., Schoenberg, F.P., 2005. Approximating the distribution of Pareto sums. Pure Appl. Geophys. 162 (6–7), 1187–1228.
Zanobetti, A., Schwartz, J., 2006. Air pollution and emergency admissions in Boston, MA. J. Epidemiol. Community Health 60 (10), 890–895.
Zhao, X., Shang, P., Pang, Y., 2010. Power law and stretched exponential effects of extreme events in Chinese stock markets. Fluct. Noise Lett. 9 (2), 203–217.
Zhao, N., Qiu, J., Zhang, Y., He, X., Zhou, M., Li, M., Xu, X., Cui, H., Lv, L., Lin, X., et al., 2015. Ambient air pollutant PM10 and risk of preterm birth in Lanzhou, China. Environ. Int. 76, 71–77.