CHEMICAL ENGINEERING TRANSACTIONS

VOL. 74, 2019

A publication of AIDIC

The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Sauro Pierucci, Jiří Jaromír Klemeš, Laura Piazza

ISBN 978-88-95608-71-6; **ISSN** 2283-9216

DOI: 10.3303/CET1974059

Modelling Annual Data of PM_{10} Atmospheric in Campinas City from Probability Density Function

Marco A. Cremasco, Osvaldir P. Taranto, Guilherme J. Castilho, Edson Tomaz

Process Engineering Department – Chemical Engineering School – University of Campinas
Albert Einstein Avenue, 500 – ZIP CODE: 13083-852 – Campinas - SP – Brazil
cremasco@feq.unicamp.br

Inhalable particulates (PM_{10}) are those ones that present diameter less than 10µ in the atmosphere. The effects of atmospheric levels can cause since fatigue, burning eyes, nose and throat to serious risk of respiratory and cardiovascular disease manifestation, and an increase in premature deaths of sensitive groups (children, the elderly, and suffering from respiratory and cardiac diseases). Due to importance of this subject, the goal of this work is to verify the model, based on probability density function (PDF), that describes the annual concentrations of PM_{10}, obtained from time series in the period of 2015 to 2017, provided from monitoring of two stations of Campinas City, Sao Paulo State, Brazil. Known the time series, it is necessary to group the original data into classes. The strategy employed in this study is to choose a representative statically bin, which it is evaluated by saturation of coefficient of variation with increase of the classes. After that, several PDFs were evaluated. The Kolmogorov-Smirnov and sum of the quadratic errors criteria were applied on the determination of the best fit.

1. Introduction

Air pollution is a global epidemic, caused by chemical and biological molecules, and particulate matter (PM), which results in various environmental and human health impacts (Li et al., 2017), and is believed to kill more people worldwide than AIDS, malaria, breast cancer, or tuberculosis (Rhode and Muller, 2015). The World Health Organization estimates that more than 6 million premature deaths were caused by air pollution exposure in 2012 (WHO, 2017).

A critical component of air pollution is atmospheric PM, and it is a common proxy indicator for air pollution, and affects more people than any other pollutant. The major components of PM are sulfate, nitrates, ammonia, sodium chloride, black carbon, mineral dust and water (WHO, 2018). The PM includes fine particles with small diameters that remain suspended in air and do not settle (Gamble and Lewis, 1996). PM_{10} denotes particles with an aerodynamic diameter of 10 µm or less, and $PM_{2.5}$ denotes those with a diameter of 2.5 µm or less. These particles are respirable and 80 % or more will deposit somewhere in the respiratory system (Gamble and Lewis, 1996), while the thicker particles can aggravate preexisting respiratory problems, especially in children and elderly, increasing the risk of emergency hospitalization and premature death (Carneseca et al., 2012). High concentrations of air particulates can have environmental impacts, such as degraded atmospheric visibility (Li et al., 2017). Particulate sources include electric power plants, industrial facilities, automobiles, biomass burning (Rhode and Muller, 2015; Ljungman and Mittleman, 2014). Air quality measurements are typically reported in terms of daily or annual mean concentrations of PM_{10} particles per cubic meter of air volume (m^3). Routine air quality measurements typically describe such PM concentrations in terms of micrograms per cubic meter (µg m^{-3}), and the WHO recommends PM_{10} limit value of 20 µg m^{-3}/year and 50 µg m^{-3}/day. Then, decision-makers and researchers need accurate and reliable estimates of air pollution exposure and the related health impacts (WHO, 2016). In this aspect, the statistical distribution model can describe air pollutant concentrations in an organized and efficient manner including extreme and average concentrations (Giulia et al., 2017), as well as it is a tool of summarizing the information contained in the entire data set in a concise manner (Lu, 2002). Thus, information on the frequency distribution is necessary for developing air pollutant control strategies (Lu, 2003).

Paper Received: 10 March 2018; Revised: 6 July 2018; Accepted: 20 December 2018

Please cite this article as: Cremasco M., Taranto O., Castilho G.J., Tomaz E., 2019, Modelling Annual Data of Pm10 Atmospheric in Campinas City from Probability Density Function, Chemical Engineering Transactions, 74, 349-354 DOI:10.3303/CET1974059

Many types of distributions have been used to fit the air pollutant concentration data. Gravill et al. (2006) analyzed PM_{10} and $PM_{2.5}$ time series from June 1999 to May 2000, at central location in Athens (Aristotelous St.). Their results indicated Pearson type VI PDF as better fit for their data. Norazian et al. (2012) studied the data sets of PM_{10}, for 2006 and 2007, in an industrialized area with high population and traffic density, in Shah Alam, Selangor, founding the Log-normal distribution as the best fit. Giulia et al. (2017) founded for NO_2 concentrations that Log-normal and Log-logistic model for winter and summer seasons were the best fits, while $PM_{2.5}$ were fitted with the Log-normal model. Pietro and Cremasco (2017) founded, for a green area (Ibirapuera Park) in Sao Paulo City, the Gumbel model, in 2010 up to 2015, as indicates to express the atmospheric NO_x concentration. As can see the best fit of PDF depends on season, pollutant, and geography. The present work aims to study the atmospheric dispersion of PM_{10} from June 2015 (winter beginning in South Hemisphere) to June 2018 (end of fall in South Hemisphere) by time series from two monitoring stations at Campinas City, Brazil.

2. Method

Campinas City is located in the east-central portion of the State of São Paulo, 47°03'33" West Longitude and 22°48'57" South Latitude, and it has an average altitude of 680 meters above the sea level, with tropical climate (type Cwa as Köppen). Campinas is recognized as the Brazilian capital of Science, Technology and Innovation, with the major national centers of R & D & I, with 1.1 million inhabitants (in 2015), and its GDP is $ 18.8 billion, comprising 6 % of industrial activities, 47.5 % and 46.6 % of trade and services. Five major Brazilian highways intersect Campinas and connect it with major producers and consumers (Campinas, 2018). The time series of PM_{10} concentrations were directly exported from the electronic platform of the Environmental Sanitation Technology Company of the State of São Paulo (CETESB). These data are relative of an avenue with intense vehicle traffic in city downtown (Francisco Glicerio Av.), and from a green area (Taquaral Park), for start of winter of 2015 to end of fall of 2018 (June in Brazil). The sampling frequency is 1 h, and it was necessary to filter these data due to the presence of lack of information in some hours or periods. As done in previous work (Prieto and Cremasco, 2017) this work is divided in two parts. In the first one, the original data were divided into bins (K) to found the optimal bin, based on the saturation of coefficient of variation, CV(%)=100σ/E(x), with standard deviation, σ, and the expectation value, E(X), given from discrete approach from moment technique based on PDF for each K as

$$\sigma = [Var(X)]^{1/2} = \left[\sum_{i=1}^{K} x_i^2 p(x_i) - E(X)^2\right]^{1/2} \quad (1)$$

$$E(X) = \sum_{i=1}^{K} x_i p(x_i) \quad (2)$$

To adjust the probability density functions (PDF) is done from the distributions present in Table 1. The adjustment method used was least squares, and the quality of the adjustments was evaluated using the Kolmogorov-Smirnov (K–S) test, that it is defined as the maximum difference between the sample cumulative probability and the expected cumulative probability (Lu, 2003)

$$D_{max} = max|P(x_i) - G(x_i)| \quad (3)$$

with $P(x_i)$ and $G(x_i)$ as expected and observed cumulative frequency function. The D_{max} value is compared with a critical theoretic difference, D_C, for a certain significance level. Other criteria to observe the performance of the adjustment is the evaluation of the sum of the quadratic errors $S(e^2)$, defined as

$$S(e^2) = \frac{1}{K}[p(x_i) - g(x_i)]^2 \quad (4)$$

where $p(x_i)$ and $g(x_i)$ are experimental and expected values of the frequency distribution.

2. Results and Discussion

The time series of PM_{10} concentration for the beginning of the winter in 2015 to end of the fall of 2018 are presented in Figure 1 for Taquaral Park, and Figure 2 for Campinas downtown. For all time series, the mean value, μ, is calculated as

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i f(x_i) \quad (5)$$

Table 1: Probability density function (Pietro and Cremasco, 2017)

PDF	f(x) =
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$
Log-normal	$\frac{1}{x\sqrt{2\pi\sigma^2}}exp\left[-\frac{(\ell n(x)-\mu)^2}{2\sigma^2}\right]$
Logistic	$\frac{1}{k}exp\left(\frac{x-\alpha}{k}\right)\left[1+exp\left(\frac{x-\alpha}{k}\right)\right]^{-2}$
Gumbel	$\frac{1}{a}exp\left[-\frac{x-b}{a}-exp\left(\frac{x-b}{t}\right)\right]$
Weibull	$\frac{\eta}{b}\left(\frac{x}{b}\right)^{(\eta-1)}exp\left[-\left(\frac{x}{b}\right)^{\eta}\right]$
Rayleigh	$\frac{1}{a^2}exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^2\right]$
Maxwell	$\frac{1}{a^3}\left(\frac{2}{\pi}\right)^{1/2}x^2 exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^2\right]$
Gamma	$a\frac{(ax)^{(b-1)}}{\Gamma(b)}x^2 exp(-ax)$
Beta	$\frac{x^{(a-1)}(1-x)^{(b-1)}}{\Gamma(a)\Gamma(b)/\Gamma(a+b)}$

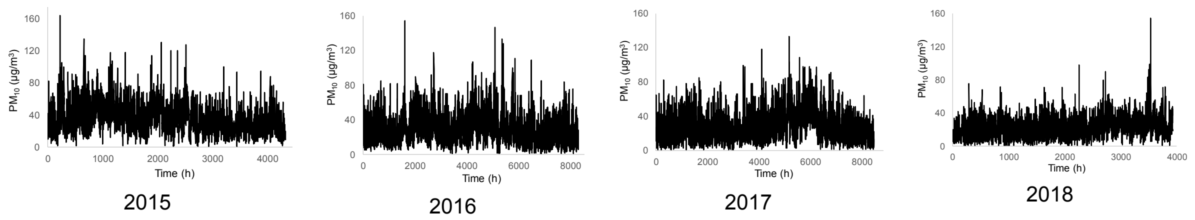

Figure 1: Times series of PM_{10} concentration for the year 2015 to 2018 – Francisco Glicerio Av. (Downtown)

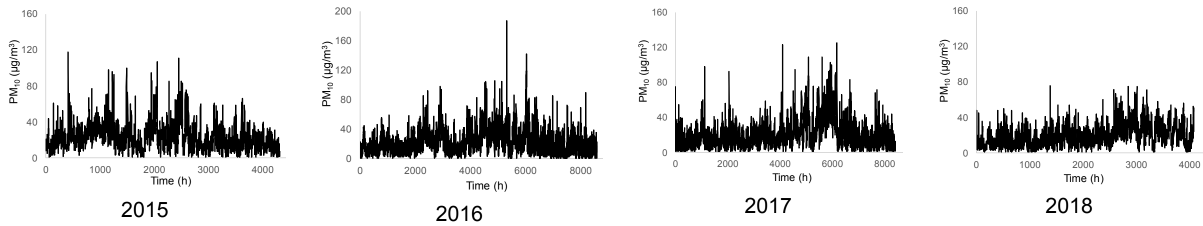

Figure 2: Times series of PM_{10} concentration for the year 2015 to 2018 – Taquaral Park

In Figures 1 and 2, for 2015, there is a description from winter to start of the summer; 2016 and 2017 all the seasons, and 2018 for summer to end of the fall. The majority picks of PM_{10} concentration are in winter, with increasing of this concentration in this season, which it is reduced in the summer and the fall, as showed for 2018, in Taquaral Park. In Taquaral Park, also, presents the less values for PM_{10} concentration as expected due to green area nature. These observations are confirmed by Figure 3, for annual average of PM_{10} concentration, that its done from Figures 1 and 2. Still from Figures 1 and 2 it was obtained experimental function density with different values of bins (starts with K = 2, minimum value, up to K = 30). Figure 4 shows the evolution of the coefficient of variation (Cv), calculated from Equations 1 and 2, with the increase of the bins. It is possible to check that Taquaral Park station presents, in all years, high dispersion of data points in a data series around the expected value than downtown station. This variability is associated with high and more constant traffic in Francisco Glicerio Street than that one around Taquaral Park. From Figure 4, ones can be notice that is no more significant variation in Cv for 15 < K < 18, and so on. Therefore, it was adopted the distribution in 18 bins, which it is close with K = 20 founded in Pietro and Cremasco (2017). After fixed K = 18, the functions present in Table 1 were adjusted.

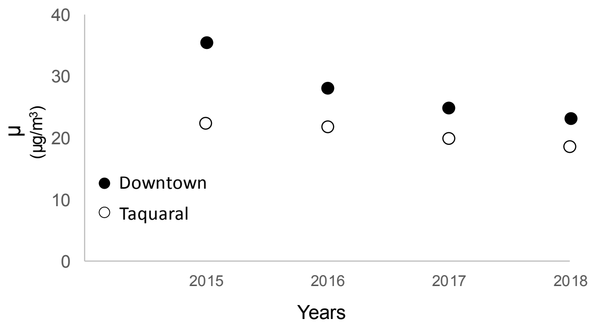

Figure 3: Evolution of PM_{10} average concentration for the year 2015 to 2018

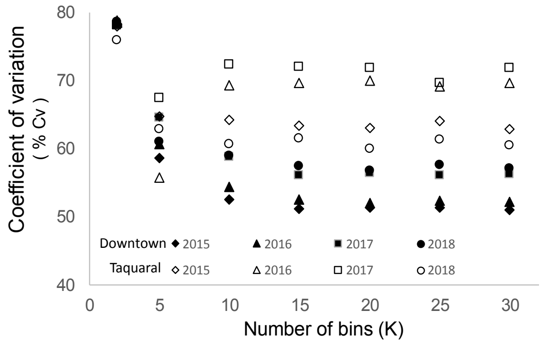

Figure 4: Evolution of the coefficient of variation with the increase of the bins

Table 2: Maximum deviation (D) of the Kolmogorov-Smirnov test

	2015		2016		2017		2018	
Station	Downtown	Taquaral	Downtown	Taquaral	Downtown	Taquaral	Downtown	Taquaral
Normal	0.0331	0.1008	0.0440	0.1366	0.0757	0.0987	0.0541	0.0970
Log-normal	0.0452	0.0619	0.0294	0.0441	0.0292	0.0101	0.0709	0.0718
Logistic	0.0500	0.1140	0.0478	0.1438	0.0496	0.1082	0.0706	0.1103
Gumbel	0.0119	0.0299	0.0083	0.0570	0.0355	0.0360	0.0333	0.0358
Weibull	0.0282	0.0163	0.0453	0.0078	0.0746	0.0602	0.0110	0.0111
Rayleigh	0.0411	0.0499	0.0485	0.0622	0.0655	0.0923	0.0102	0.0391
Maxwell	0.0455	0.1110	0.0628	0.1296	0.0894	0.1452	0.0603	0.1005
Gamma	0.0143	0.0363	0.0332	0.0781	0.0628	0.0405	0.0172	0.0296
Beta	0.0141	0.0174	0.0230	0.0229	0.0500	0.0526	0.0210	0.0122

Table 3: Sum of the quadratic errors [$S(e^2) \times 10^4$]

	2015		2016		2017		2018	
Station	Downtown	Taquaral	Downtown	Taquaral	Downtown	Taquaral	Downtown	Taquaral
Normal	23.31	43.71	40.67	33.51	63.72	58.64	17.93	23.35
Log-normal	13.17	28.64	8.42	20.40	21.68	2.01	47.13	30.56
Logistic	21.06	44.90	34.84	38.00	44.18	41.90	24.04	25.02
Gumbel	1.15	21.48	2.87	7.61	17.07	15.74	13.26	10.87
Weibull	6.35	19.31	17.25	1.50	48.39	34.25	5.50	8.27
Rayleigh	17.93	46.63	35.82	110.20	69.70	55.34	5.72	24.85
Maxwell	27.99	222.80	32.28	451.10	52.70	223.90	92.49	150.0
Gamma	3.34	23.32	12.31	9.48	34.00	31.61	8.02	9.78
Beta	1.32	14.13	3.53	7.67	31.19	17.88	9.85	7.95

Once the bin was defined, the adjustment of the nine probability density functions was performed using the least squares methodology. The Kolmogorov-Smirnov test was used, and the maximum deviation (D) among accumulative distributions from each model and experimental data were compared to the critical deviation value, D_C, with 5 % as significance level that is, for K = 18, D_C = 0.30936. In cases where $D < D_C$, the adjustment is approved. All distributions present in Table 1 attend these criteria, as shown in Table 2. From experimental PDF and models, it was applied the evaluation of the sum of the quadratic errors, $S(e^2)$, which results are in Table 3. If one adopts $S(e^2) < 5,0 \times 10^{-3}$ as criteria, the Normal, Rayleigh and Maxwell models are excluded. Except those models and Logistic, the Log-normal, Gumbel, Weibull, Gamma and Beta distribution presents $D_C < 0.1$. For Taquaral Park, the Weibull and Beta are the most indicates, while for downtown Gumbel model has the best fit, including a good performance for all situation analyzed in this work. Figures 5 to Figure 12 show the comparison of the adjustments of the Gumbel, Weibull and Beta models with the respective experimental PDF.

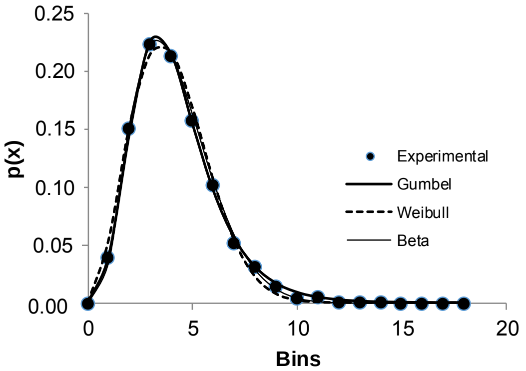

Figure 5: Downtown 2015

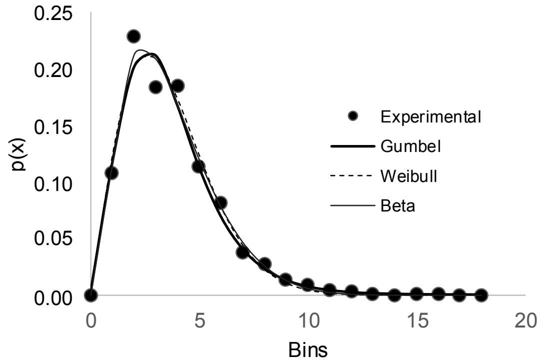

Figure 6: Taquaral 2015

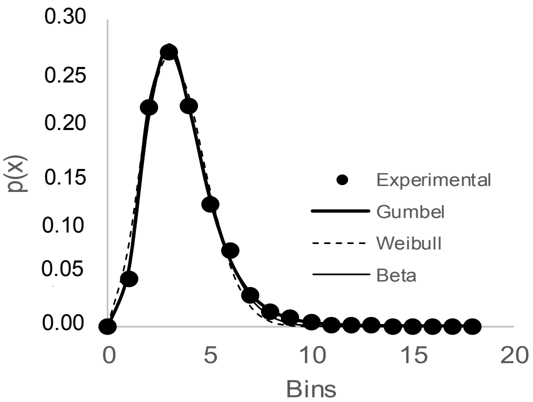

Figure 7: Downtown 2016

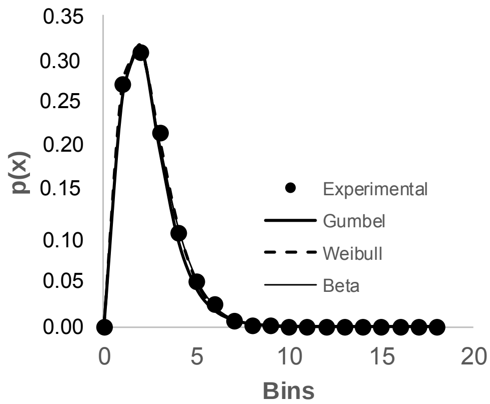

Figure 8: Taquaral 2016

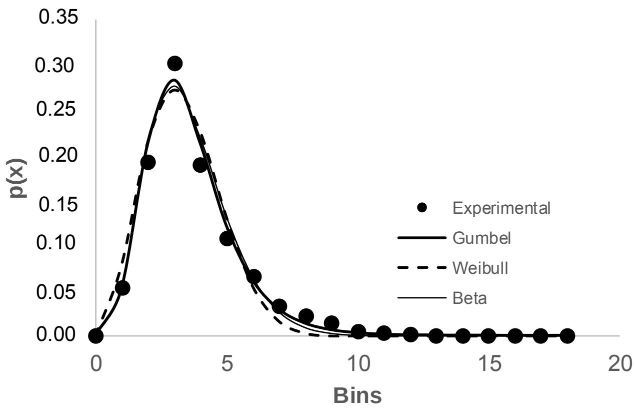

Figure 9: Downtown 2017

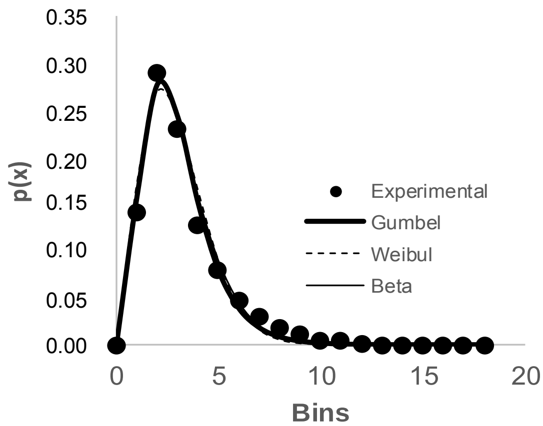

Figure 10: Taquaral 2017

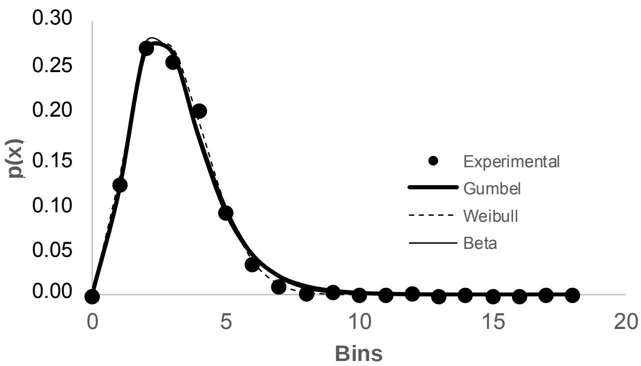

Figure 11: Downtown 2018

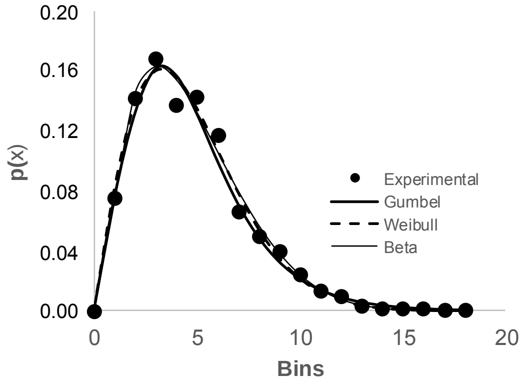

Figure 12: Taquaral 2018

4. Conclusion

This work shows that is possible to reconstruct the time series of PM_{10} concentration into an adequate PDF model, such as Log-normal, Gumbel, Weibull, Gamma and Beta distributions, particularly the Gumbel model that is adequate to different seasons and monitoring stations at Campinas City, Brazil, from June 2015 to June 2018. It is important to mention the necessity of the previous knowledge of the bins number to do the appropriate analysis of the frequency distribution from original time series for air pollutant control strategy.

Acknowledgements

The authors would like to thank DEPro/FEQ/Unicamp for the financial support.

References

Campinas. Welcome to Campinas, http://www.internationaloffice.unicamp.br/english/wp-content/uploads/2015/02/guia_tur_eventos_ing.pdf. accessed 14.11.2018.

Carneseca, E.C., Achcar, J.A.; Martinez, E.Z., 2012, Association between Particulate Matter Air Pollution and Monthly Inhalation and Nebulization Procedures in Ribeirão Preto, São Paulo State, Brazil. Cad. Saúde Pública, 28, 8, 1591-1598.

Gamble, J.F., Lewis, R.J., 1996, Health and Respirable Particulate (PM10) Air Pollution: A Causal or Statistical Association? Environmental Health Perspectives, 104, 8, 838 – 850.

Giulia, S., Nagendra, S.M., KHare, M., 2017, Extreme Events of Reactive Ambient Air Pollutants and Their Distribution Pattern at Urban Hotspots, Aerosol and Air Quality Research, 17, 394–405.

Gavrill, I., Grivas, G., Kassomenos, P., Chaloulakou, A., 2006, An Application of the Theoretical Probability Dstributions to the Study of PM10 and PM2.5 Time Series in Athens, Greece, Global Nest Journal, 8, 3, p. 241-251.

Li, X., Chen, X., Yuan, X., Zen, G., León, T., Liang, J., Chen, G., Yuan, X., 2017, Characteristics of Particulate Pollution (PM2.5 and PM10) and Their Spaces scale-dependent Relationships with Meteorological Elements in China, Sustainability, 9, 2330 – 2344.

Ljungman, P.L., Mittleman, M.A., 2014, Ambient Air Pollution and Stroke, Stroke, 45, 3734 – 3741.

Lu, H.-C., 2002, The Statistical Characters of PM10 Concentration in Taiwan Area. Atmospheric Environment, 36, 491– 502.

Lu, H.-C., 2003, Comparisons of statistical characteristic of air pollutants in Taiwan by frequency distribution, Journal of Air & Waste Management Association, 53, 5, 608-616.

Norazian, M.N., Abdulla, M.M.A., Tan, C.-Y., Ramlic, N.A., Yahayac, A.S.; Fitric, N.F.M.Y., 2011, Modelling of PM10 Concentration for Industrialized Area in Malaysia: A case Study in Shah Alam. Physics Procedia, 22, 318 – 324.

Prieto, W.H., Cremasco, M.A., 2017, Application of Probability Density functions in Modelling Annual Data of Atmospheric NOx Temporal concentration. Chemical Engineering Transactions, 57, 487-492.

Rhode R.A. Muller, R.A., 2015, Air Pollution in China: Mapping of Concentrations and Sources. PLOS ONE. August, 2015, 1 – 14.

WHO. World Health Organization, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease. Geneve: WHO Productions Services.

WHO. World Health Organization, 2018, Ambient (outdoor) Air Quality and Health. 2 May 2018, http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health accessed 20.11.2018.