Sustainable Environment Research 26 (2016) 142–148

Contents lists available at ScienceDirect

Sustainable Environment Research

journal homepage: www.journals.elsevier.com/sustainable-environment-research/

Original research article

Statistical behavior of ozone in urban environment

Sumit Sharma [a, b, *], Prateek Sharma [c], Mukesh Khare [b], Swati Kwatra [d]

[a] Centre for Environmental Studies, The Energy and Resources Institute, New Delhi 110003, India
[b] Civil Engineering Department, Indian Institute of Technology, New Delhi 110016, India
[c] Natural Resource and Management Department, Teri University, New Delhi 110070, India
[d] Department of Development Communication and Extension, Lady Irwin College, New Delhi 110001, India

A R T I C L E I N F O

Article history:
Received 27 July 2015
Received in revised form
2 October 2015
Accepted 3 December 2015
Available online 12 April 2016

Keywords:
Probability density function
Ground level ozone concentration
Statistical distribution models
Goodness-of-fit tests
Probability of exceedence
NAAQS

A B S T R A C T

This paper analyzes the statistical behavior of the ground level ozone concentrations (GLO) observed at a major traffic intersection in Delhi. Five sets of data, i.e. summer (May to July, high solar radiation data), winter (November to January, low solar radiation data), spring (March to April), autumn (September to October), and the entire year have been used to study the seasonal variation in the statistical behavior of GLO. Appropriate statistical distribution form has been identified from alternative candidate distribution models using the goodness-of-fit methods and parameters have been estimated using the method of maximum likelihood. The yearly, winters, spring, and summer datasets were found to follow the log-normal distribution model, while autumn dataset followed Weibull distribution. Analysis shows that ozone concentrations also show similar statistical behavior like other air pollutants and fit mainly to the log-normal distribution as reported for other pollutants in different studies. The seasonality of the datasets shows higher skewness during summers due to longish tail of the distribution mainly on account of higher photo−chemical activity. The probability density functions corresponding to the five datasets were used to compute the probability of exceedence of the National Ambient Air Quality Standards and return period of violation of standards. The distributions have also been used to classify the study region under various air quality descriptor categories. The region is found to violate the air quality compliance criteria 17% of the recorded times in the year. Alternative measures have been discussed to reduce the precursor emissions in order to achieve the air quality goals.

1. Introduction

Air pollution has emerged as a major problem in the developing countries. It affects not only the human health but also ecology [1], buildings [2] and agricultural productivity [3]. Maintaining the pollution levels below the respective standards for different pollutants is one of the primary goals of an air quality management plan. The regulatory authorities like Central Pollution Control Board of India (CPCB) lay down the compliance criterion of ambient air quality as National Ambient Air Quality Standards (NAAQS). In order to assess the status of air quality, regular monitoring of air pollutants is carried out and large amount of data is collected that is related to these standards. The usual practice followed is the computation of descriptive statistics and construction of time−series plots of these pollutants. The probability density function (pdf) of a pollutant is a useful tool of summarizing the information contained in the entire data set in a concise manner. It depicts the entire range, mean, extremes, probability of occurrences and typical graphical distribution pattern in a typical setting. The probability distribution function also helps in directly relating to the extent of meeting the requirement of NAAQS [4]; it provides a means to compute probability of exceedence of a standard and return period of violation of standard, if any. Several studies have been conducted in the past to understand the statistical behavior of primary pollutants like particulate matter (PM), carbon monoxide (CO) [5–7]. Jia et al. [8] studied the distribution of volatile organic compounds (VOC) exposures. However, limited work has been done in assessing the statistical behavior of ground level ozone (GLO) concentrations in an urban context.

* Corresponding author. Centre for Environmental Studies, The Energy and Resources Institute, IHC Complex, Lodi Road, New Delhi 110003, India.
E-mail address: sumits@teri.res.in (S. Sharma).
Peer review under responsibility of Chinese Institute of Environmental Engineering.

http://dx.doi.org/10.1016/j.serj.2016.04.006

Ozone can be easily differentiated from the criteria pollutants like PM, NOx, (oxides of nitrogen) and SO_2 (oxides of sulphur). While stratospheric influx is the primary source of ozone, it is also formed through reactions precursor species as a secondary pollutant. Understanding its behavior becomes even more difficult considering the photo-chemistry involved in the reactions of NOx and VOCs. VOCs help in oxidizing primary NO released from various sources to form NO_2. VOCs also help in retaining the existing ozone by competing with it to react with NO. NO_2 is then photolysed to generate atomic oxygen, which combines with oxygen to form O_3 in the troposphere [9]. Ozone is not only linked with effects on human health but also significant impact on crop productivities. There are few studies conducted in past on assessment of GLO in Indian context. Guardani et al. [10] studied the behavior of GLO concentration in urban area. Gilleland et al. [11] described statistical models for monitoring and regulating GLO. Recently, Chelani [12] assessed the statistical persistence of ozone concentrations in an urban setting. Unlike criteria pollutants, the GLO concentration is affected by the photochemistry, which in turn is dependent on the amount of solar radiation. It therefore becomes an interesting research proposition to study the statistical behavior of GLO and change in its behavior, if any, with change in solar radiation and hence with season. As discussed above the GLO concentration is expected to be high during daytime due to the photochemical reactions and thus due to longer sunshine hours during summers, the average concentration is expected to be higher in comparison to the winter's concentration. The annual average concentration, by the same argument is expected to be in between the summer and winter average. The other statistical characteristics are also likely to be affected due to the photochemistry. However, statistical characteristics such as mean and variance provide loose information about the distribution of GLO concentration. The pdf on the other hand provides a more complete and higher order description of pollutant concentration data [13]. The present work thus aims to model the pdf of GLO for five sets of data: (a) summer, (b) winter, (c) spring, (d) autumn, and (e) yearly data, and study the influence of season on the type of pdf, if any.

This paper attempts to analyze the statistical behavior of the GLO concentrations in an urban setting. For this purpose 8-h average data for the year 2010 has been taken. The data have been further divided in four data sets – summer (May to July, high solar radiation data), spring (March to April), autumn (September to October), and winter (November to January, low solar radiation data) to study seasonal variation in the statistical behavior of GLO. Appropriate pdfs identified for the entire year and the four seasons have been used to compute the exceedence probability of NAAQS violations.

2. Materials and methods

2.1. Study location and data

CPCB runs an automated air quality monitoring station near a major traffic intersection in the central region of the capital city of Delhi. The traffic intersection has a very high vehicular density. 8-h ambient ozone concentration data have been collected from the CPCB for the year 2010. The ozone concentrations are measured using online ozone analyzer (model O342M, Environment SA, France) which works on UV absorption technology. The lower detection limit is 0.4 ppb. The instantaneous ozone concentration data are averaged for 1-h/8-h reporting. The entire data set has been analyzed to model the statistical form of GLO. In addition, summer (May to July), spring (March to April), autumn (September to October), and winter (November to January) months data have been extracted to study variation in the distribution form with change in solar radiation. Z-scores method [14] is used to scrutinize the outliers. The outliers with z-score values more than 3 were scrutinized and removed from the dataset. Further analysis is carried out based on the dataset without outliers.

2.2. Methodology

Exploratory analysis has been carried out for the five data sets. For this purpose, the descriptive statistics has been computed; box-and-whisker plot with frequency histogram constructed to preliminarily assess the distribution form. This is followed by the modeling of the distribution form of the GLO data using the MINITAB version 14.1 software. It may be noted that "there is no a priori reason to expect that atmospheric distribution should adhere to a specific distribution" [15]. However, most of the distributions that have been found to fit the air quality are special cases of the four-parameter generalized gamma distribution, which include the one- and two-parameter exponential distributions; the one-, two- and three-parameter Weibull and standard gamma distributions; and lognormal distribution [16]. Thus, the challenge is to determine the best distribution model that fits the data set under study amongst the alternative potential candidate models. This involves modeling the distributions of air pollutant concentration, GLO in the present study. The modeling broadly includes two steps: (a) identification of appropriate distribution form from alternative candidate models; and (b) estimation of parameters of the identified model. The model identification has been carried out by Chi-square, Kolmogorov–Smirnov and Anderson-Darling goodness-of-fit tests. However, the final model selection is based on the Anderson-Darling test as it gives more weight to the tails, which is more relevant in the air pollution context. Moreover, it is more sensitive test as it makes use of the specific distribution in calculating critical values [17]. In all, fourteen distribution forms most commonly used in air pollution studies have been fitted to identify the best-fit distribution. The parameter estimation is done by method of moments, method of least squares (MLS) and method of maximum likelihood (MLE). Mage and Ott [18] based on several Monte Carlo simulations found that MLE provides the best estimates. Thus, MLE was preferred as the method for estimating model parameters in the present study. The details of the above mentioned model identification and parameter estimation techniques have been provided in detail in Ref. [17]. Using the best-fit distribution models, the compliance of the ozone standards at the major traffic intersection location is assessed for different seasons of 2010. The observed concentrations of ozone are also compared with the NOx concentrations to analyze the effect of titration chemistry in which primary NO emissions from sources like vehicles react with ozone to destruct it. This is a primary cause for lower ozone concentrations in the city centres in comparison to the surrounding regions.

3. Results and discussion

The 8-h ozone concentrations at the major traffic intersection monitoring station during the year 2010 are shown in Fig. 1. It shows clear violation of 8-h standard (100 $\mu g\ m^{-3}$) many times during the year. The annual average ozone concentration at major traffic intersection is 61 $\mu g\ m^{-3}$ with a standard deviation of 50. Seasonal variation in the ozone is clearly evident from the graph. The GLO concentration is observed to be high during the summer months due to higher photochemical activity. The winter months show lower concentration values. Ali et al. [19] shows the regression analyses of surface ozone with maximum temperature in Delhi, which also stated that ozone was mainly produced by photochemistry. The average diurnal variation across different days

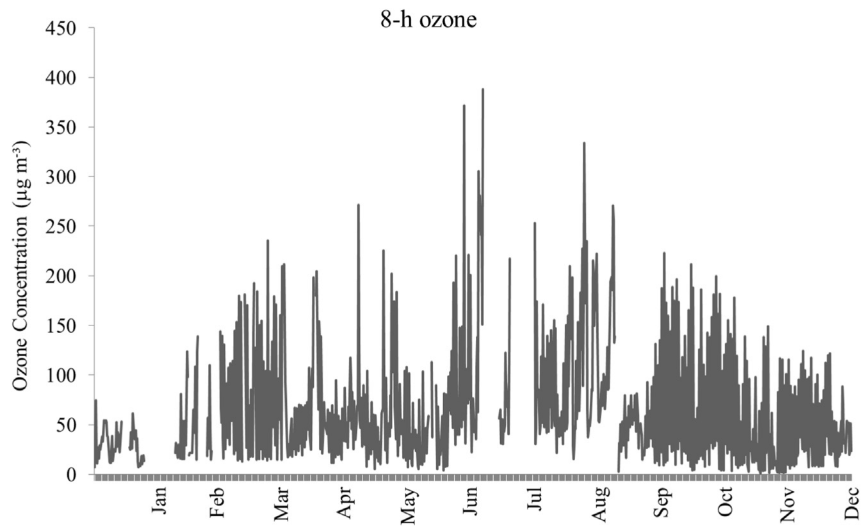

Fig. 1. 8-h variation of ozone at major traffic intersection in Delhi during 2010.

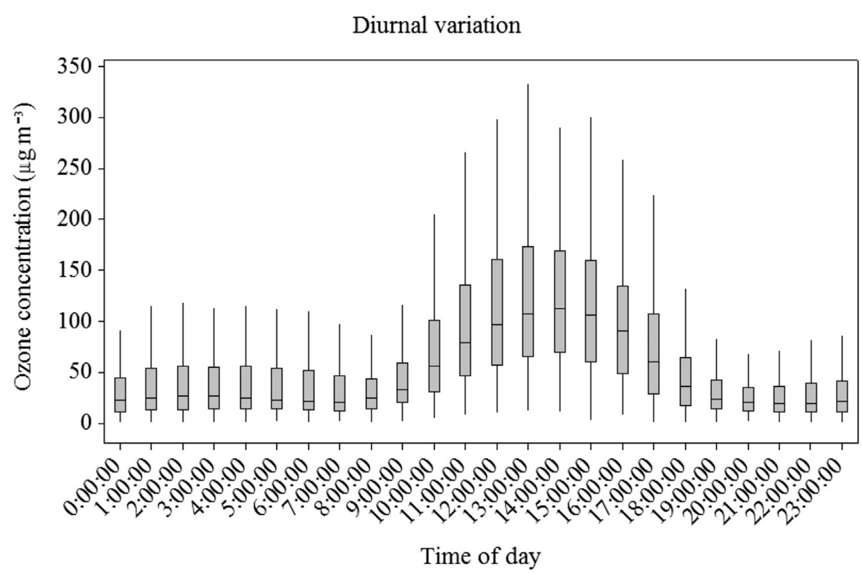

Fig. 2. Average diurnal variation across different days of the year 2010 at major traffic intersection.

of the year is shown in Fig. 2, which shows peak ozone formation due to higher photo–chemical activity during daytime (12 noon to 4 pm). The ozone formation is generally found to be higher during high sunshine hours; during night, ozone depletes due to its reactivity with NO (nitric oxide). Table 1 shows the summary statistics for the five data sets.

Table 1
Summary statistics for the yearly, seasonal 8-h average GLO data (μg m^{-3}).

Parameter	Winter	Summer	Spring	Autumn	Yearly
Number of data	244	205	210	177	925
Mean	40	63	63	70	61
Standard Deviation	36	48	48	54	50
Median	26	52	49	55	47
Minimum	1	4	11	3	1
Maximum	179	256	210	223	227
Skewness	1.5	1.8	1.2	1.0	1.3
Kurtosis	1.5	3.6	0.7	0.1	1.1

The summary statistics reveals variation in the summer and winter data. As expected, the mean and median values are found to be higher during the summer months compared to the winter months. The maximum value recorded during the winter months is 179 μg m^{-3}, which is 77 μg m^{-3} lesser than the maximum recorded for the summer months. The maximum values for spring and autumn months are in between 210 and 223 μg m^{-3}. This indicates shifting of the right tail towards the left side and reduction in the skewness for the winter data. A higher value of coefficient of skewness for the summer data further substantiates this fact. It also indicates variation in the statistical behavior of the ozone concentrations across different seasons. This is also revealed in the frequency histograms (Fig. 3). This can be attributed to higher observed GLO concentrations due to the photochemical activity and longer sunshine hours.

Fig. 4 shows the diurnal variation (averaged for all days in 2010) of NOx and ozone concentrations at the study site. NOx concentrations are found to be increasing during night time, mainly due to movement of heavy duty trucks which are only allowed to enter the

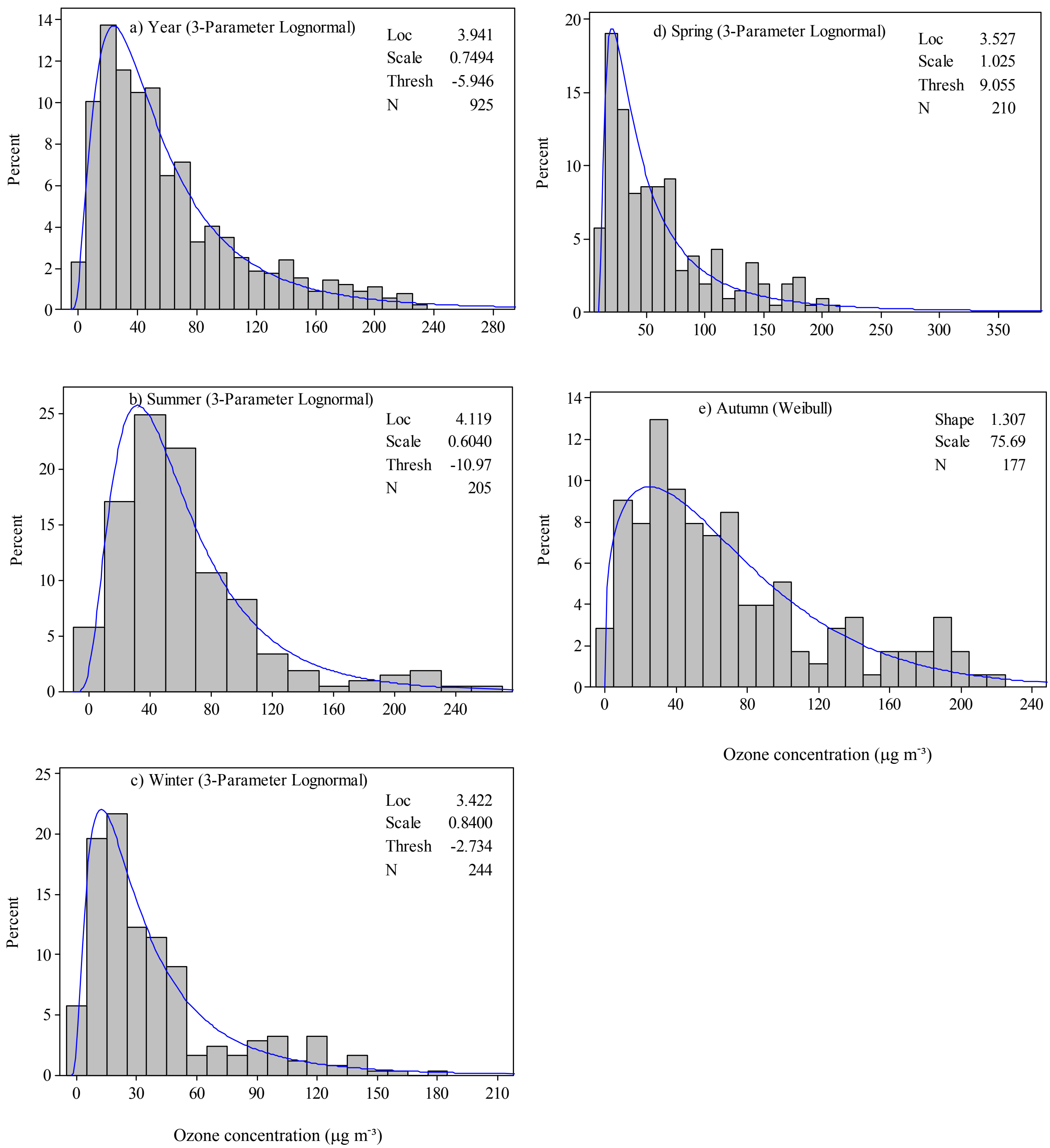

Fig. 3. Histogram of 8-h ozone concentrations (a) entire year, (b) summer, (c) winter, (d) spring and (e) autumn.

city during night. Ozone concentrations show inverse relationship with the NOx concentrations due to titration chemistry [20]. While, nitrogen dioxide (NO_2) helps in ozone formation during daytime, primary emissions of nitrogen monoxide (NO) released from diesel driven vehicles destruct ozone and reduce its ambient concentrations. Moreover, the ozone formation process stops during nighttime due to absence of sunlight leading to lower concentrations in night. VOCs and meteorological parameters also play important roles in ozone formation and transport; however, relevant data for the study site could not be collected.

Table 2 shows the summary of Anderson-Darling (AD) goodness-of-fit test along with the ML estimates of the model parameters of the various candidate models for the entire year. Similar tables for summer and winter, spring and autumn dataset are shown in Tables SI-1 to SI-4 in Supplementary Information (SI). The selection of best-fit distribution has been done on the basis of

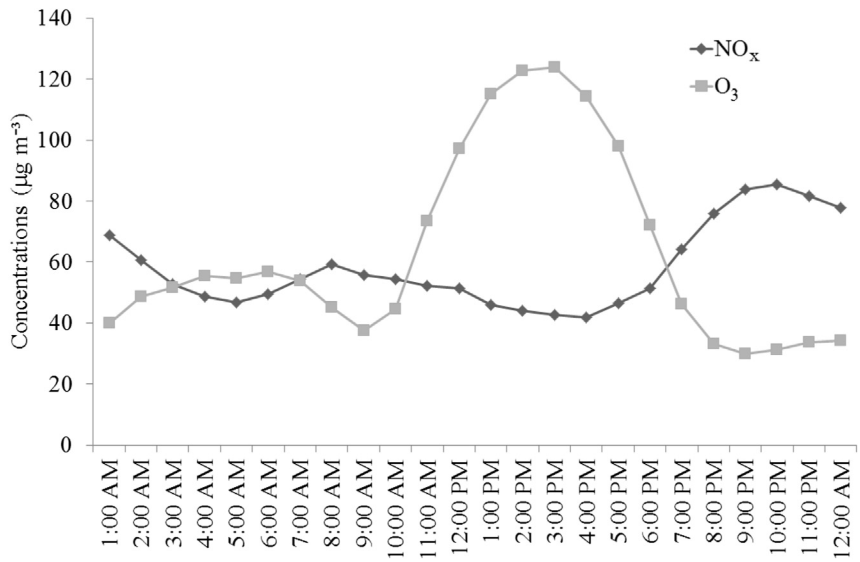

Fig. 4. Diurnal variation (averaged for all days in 2010) of NO_x and ozone concentrations at the study site.

Table 2
Summary of Anderson-Darling (AD) goodness-of-fit test along with the ML estimates of the model parameters for various candidate distribution models for the yearly dataset.

Distribution	AD	P	LRT[a] P	Location	Shape	Scale	Threshold
Normal	37	< 0.005		61		50	
Lognormal	3.2	< 0.005		3.8		0.9	
3-Parameter Lognormal	1.2	[a]	0.00	3.9		0.8	−5.95
Exponential	17	< 0.003				61	
2-Parameter Exponential	13	< 0.010	0.00			60	1.05
Weibull	3.0	< 0.010			1.28	66	
3-Parameter Weibull	2.2	< 0.005	0.00		1.24	65	1.02
Smallest Extreme Value	69	< 0.010		89		61	
Largest Extreme Value	13	< 0.010		40		33	
Gamma	1.7	< 0.005			1.55	39	
3-Parameter Gamma	1.28	[a]	0.03		1.46	41.27	0.91
Logistic	26.46	< 0.005		53.87		26.54	
Log(logistic)	2.50	< 0.005		3.80		0.52	
3-Parameter Log(logistic)	2.19	[a]	0.06	3.85		0.49	−1.94

[a] LRT: Likelihood ratio test.

AD statistics and p-value. A distribution with the largest p-value and lowest AD statistics is selected as the best-fit distribution. The table also shows likelihood ratio test (LRT) values, which is used to determine whether there is significant improvement in fit with larger distribution (i.e., 3-parameter over 2-parameter distribution); if a larger distribution significantly improves the fit, then the p-value for LRT statistics will be very small. Based on the above criteria the best-fit distributions for the five data sets have been identified. Thus, the best distribution for four of the five datasets (yearly, winter, spring, and summer) is three parameter log-normal. It may be noted that the log-normal is the special case of the four-parameter generalized gamma distribution which have been found to be useful for fitting air quality data for different pollutants in past studies [21–24]. Table SI-1 suggests that log normal and log-logistics do appear to be really close to the best-fit distributions for summers, however, log-normal is chosen based on its past reliability for fitting the air quality datasets. Only for the autumn data the AD statistic was the least for the Weibull distribution (Table SI-4) and hence was chosen as the best-fit distribution. However, Weibull distribution is also a special case of the four-parameter generalized gamma distribution.

The pdf for the best-fit distributions for the five data sets is shown in Fig. 5. The pdf also shows the line corresponding to the NAAQS for 8-h average ozone concentration; this clearly shows that the ozone standard is violated during different seasons across the year. However, the tail of the distribution is longish in case of summers.

The identified distributions for the five data sets have been used to compute the probability of exceedence of NAAQS and return period of violation. Table 3 provides the exceedence probability along with the return period of violations. The table shows non-compliance of the air quality for the entire year and for different seasons. On an average, every fourth reading in autumn, sixth reading during the summer and every 14th reading in winter months (return period 4–14) exceed the air quality criteria. It may be noted from pdf of different seasons that the number of violation and the magnitude of these violations are much higher in summer and autumn than in the winter season.

The data were further analyzed to describe the air quality of the study region in terms of GLO for the five data sets. The NAAQS standard for 8-h average ozone concentration is 100 µg m^{-3}. The CPCB uses exceedence indicator to compare the pollutant concentration with respect to prescribed standards for classifying different areas into descriptor categories (less than 0.5 times: low, between 0.5 and 1 times: moderate, between 1 and 1.5 times: high, and more than 1.5 times the standard: critical). Table 4 data provide percentage times the air quality remains in the CPCB prescribed descriptive categories for the five data sets.

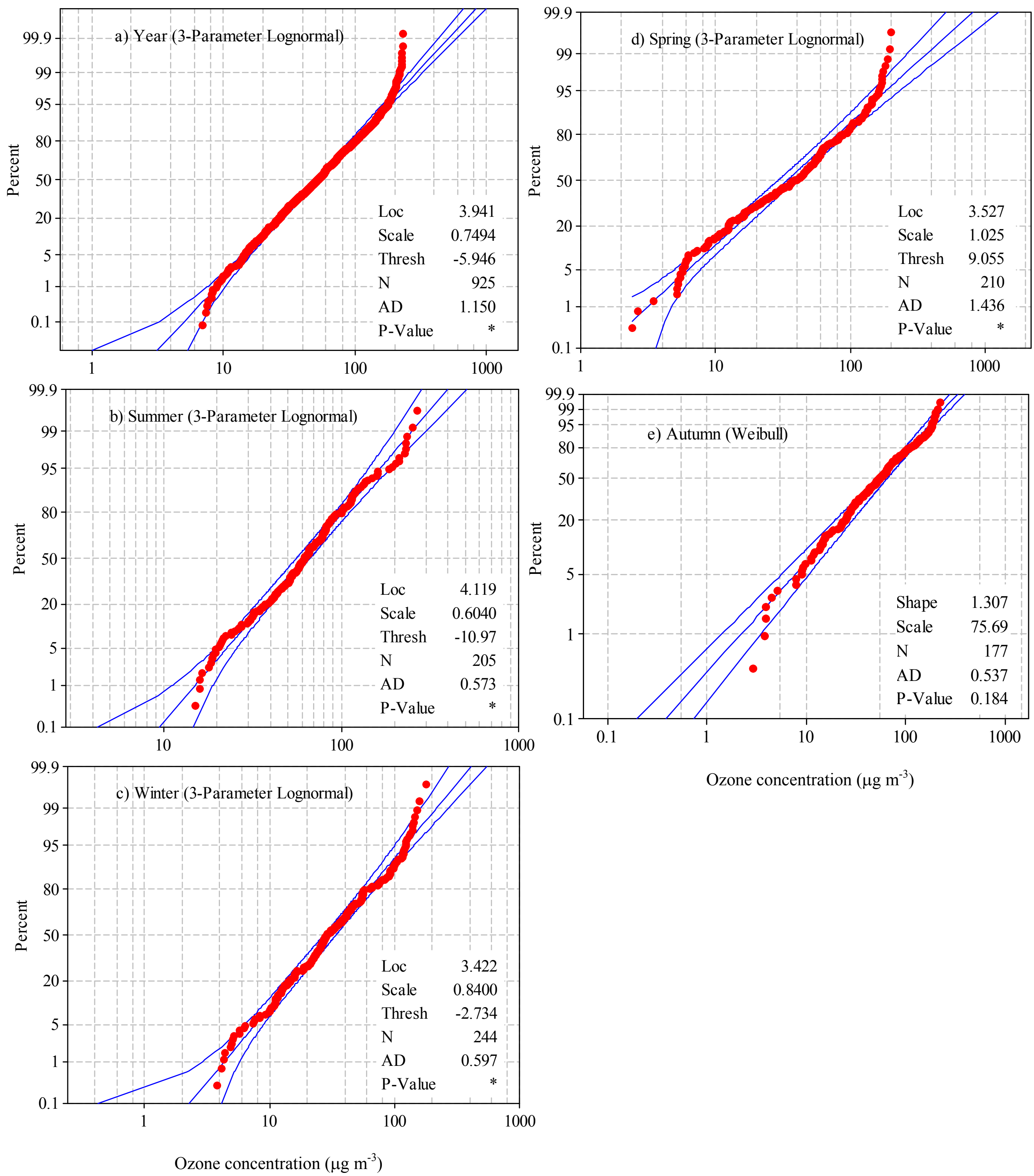

Fig. 5. Probability density function (pdf) for the best-fit distributions for (a) entire year, (b) summer, (c) winter, (d) spring and (e) autumn.

4. Conclusions

The statistical behavior of 8-h average GLO concentration monitored at a major traffic intersection was studied. Seasonal data were extracted from the original data set in order to study the seasonal variation in the GLO concentration. Summer and autumn seasons show higher GLO concentration in comparison to the winters due to higher photochemical activity. Diurnal variation of ozone shows the peak ozone formation during daytime (12 noon- 4 pm). This could be attributed to the ozone forming reactions of NO_x

Table 3
Exceedence probability along with the return period of violations for the five datasets.

Parameter	Summer	Winter	Spring	Autumn	Yearly
Pr ($X \leq x_{NAAQS}$) (in %)	83	93	83	76	83
Pr ($X \geq x_{NAAQS}$) (in %)	17	7	17	24	17
Return period (number)	6	14	6	4	6
NAAQS criteria being met or not (Yes/No)	No	No	No	No	No

Table 4
Percentage times the air quality remains in the CPCB prescribed descriptive categories for the five datasets.

Air quality	Summer	Winter	Spring	Autumn	Yearly
Low	49	74	57	44	54
Moderate	34	19	26	32	29
High	12	4	9	15	10
Critical	5	3	8	9	7

and VOCs in the presence of the sunlight. However, longer nights in winter ensure enhanced reduction of ozone by titrating reaction of NO with O_3 during night.

Statistical distribution models were used to assess the statistical behavior of GLO in the four seasons and across the year. For this purpose, the best-fit distribution models were identified using the goodness-of-fit tests for the five data sets. It is concluded that yearly, spring, winter, and summer datasets follow the 3-paramter log-normal distribution model, while autumn season followed the Weibull distribution. The results seem to be in accordance with the past studies which show that air quality data for different pollutants fit either of the special cases of the four-parameter generalized gamma distribution. Moreover, in the present study an attempt was made to study the variation, if any, in the distribution form of the pdf due to seasonal effects. Although, there is no change observed in the pdf in different seasons, but due to higher concentration during summer months, the right tail of the distribution is expected to be longer in comparison to the winter month data. This was also reflected in the data, as evident from the coefficient of skewness values. The values for coefficient of skewness for summers are higher than those of winter.

The pdf shows that the 8-h average ozone standard is violated during all the seasons with varying magnitude during the entire year. The probability of exceedence is found to be 7% for winters, 17% for spring, 24% for autumn, and 17% for summer and yearly datasets. The probability of exceedence and its magnitude are found to be less in winter and higher in autumn and summer. The maximum recorded value is 1.7 times the standard in winter, while it goes up to 2.6 times in summer. Moreover, higher number of readings was found to be in low pollution category in winter than in summer and autumn seasons, due to higher solar insolation. This longish tail of extreme values in summer is explained by the positively skewed 3-paramter log-normal distribution. The study region being a major traffic intersection has a large population of vehicles, which are the primary contributors of the precursor pollutants for ozone - NO_2 and VOCs. In order to roll back the source emissions to attain compliance criteria for acceptable air quality, several steps can be taken. These include plying restrictions on high emitting older vehicles, synchronization of traffic lights, traffic management during peak hours, in addition to the conventional measures such as regular inspection and maintenance of older vehicles, further advancement of vehicular emission norms, and improvement in fuel quality. The study provides useful insights and motivation to conduct more detailed analysis of short- and long-range sources contribution to ozone formation through dispersion models.

Acknowledgement

We acknowledge the secretarial assistance provided by Ms Valsa Charles (Centre for Environmental Studies, TERI) for formatting of this article.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.serj.2016.04.006.

References

[1] Bignal KL, Ashmore MR, Headley AD, Stewart K, Weigert K. Ecological impacts of air pollution from road transport on local vegetation. Appl Geochem 2007;22:1265–71.
[2] Kim ST, Maeda Y, Tsujino Y. Assessment of the effect of air pollution on material damages in Northeast Asia. Atmos Environ 2004;38:37–48.
[3] Agrawal M, Singh B, Rajput M, Marshall F, Bell JNB. Effect of air pollution on peri-urban agriculture: a case study. Environ Pollut 2003;126:323–9.
[4] CPCB. National Ambient Air Quality Standards. New Delhi, India: Central Pollution Control Board; 2009.
[5] Rumburg B, Alldredge R, Claiborn C. Statistical distributions of particulate matter and the error associated with sampling frequency. Atmos Environ 2001;35:2907–20.
[6] Lu HC. The statistical characters of PM_{10} concentration in Taiwan area. Atmos Environ 2002;36:491–502.
[7] Gokhale S, Khare M. Statistical behavior of carbon monoxide from vehicular exhausts in urban environments. Environ Modell Softw 2007;22:526–35.
[8] Jia CR, D'Souza J, Batterman S. Distributions of personal VOC exposures: a population-based analysis. Environ Int 2008;34:922–31.
[9] Jacob DJ. Heterogeneous chemistry and tropospheric ozone. Atmos Environ 2000;34:2131–59.
[10] Guardani R, Aguiar JL, Nascimento CAO, Lacava CIV, Yanagi Y. Ground-level ozone mapping in large urban areas using multivariate statistical analysis: application to the São Paulo Metropolitan Area. J Air Waste Manage 2003;53: 553–9.
[11] Gilleland E, Nychka D. Statistical models for monitoring and regulating ground-level ozone. Environmetrics 2005;16:535–46.
[12] Chelani AB. Statistical persistence analysis of hourly ground level ozone concentrations in Delhi. Atmos Res 2009;92:244–50.
[13] Hadley A, Toumi R. Assessing changes to the probability distribution of sulphur dioxide in the UK using a lognormal model. Atmos Environ 2003;37: 1461–74.
[14] Shiffler RE. Maximum Z scores and outliers. Am Stat 1988;42:79–80.
[15] Seinfeld J, Pandis SN. Atmospheric Chemistry and Physics. New York: Wiley; 1998.
[16] Jakeman AJ, Simpson RW, Taylor JA. Modeling distributions of air pollutant concentrations. III. The hybrid deterministic-statistical distribution approach. Atmos Environ 1988;22:163–74.
[17] Kottegoda NT, Rosso R. Applied Statistics for Civil and Environmental Engineers. 2nd ed. Oxford, UK: Blackwell; 2008.
[18] Mage DT, Ott WR. An evaluation of the methods of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmos Environ 1984;18: 163–71.
[19] Ali K, Inamdar SR, Beig G, Ghude S, Peshin S. Surface ozone scenario at Pune and Delhi during the decade of 1990s. J Earth Syst Sci 2012;121:373–83.
[20] Sillman S. Tropospheric ozone and photochemical smog. In: Sherwood Lollar B, editor. Treatise on Geochemistry. Environmental Geochemistry, Vol. 9. Amsterdam, Netherlands: Elsevier; 2003.
[21] Kao AS, Friedlander SK. Frequency distributions of PM_{10} chemical components and their sources. Environ Sci Technol 1995;29:19–28.
[22] Georgopoulos PG, Seinfeld JH. Statistical distributions of air pollutant concentrations. Environ Sci Technol 1982;16:401–16.
[23] Sharma S, Sharma P, Khare M. Hybrid modelling approach for effective simulation of reactive pollutants like ozone. Atmos Environ 2013;80:408–14.
[24] Sharma P, Sharma P, Jain S, Kumar P. An integrated statistical approach for evaluating the exceedence of criteria pollutants in the ambient air of megacity Delhi. Atmos Environ 2013;70:7–17.