Journal of Engineering and Science Research 7 (2): 07-11, 2023
e-ISSN: 2289-7127
© RMP Publications, 2023
DOI: 10.26666/rmp.jesr.2023.2.2


# Fitting Statistical Distribution Functions of Air Pollutant Concentration in Different Urban Locations in Malaysia

**Muhammad Ismail Jaffar[1], Hazrul Abdul Hamid[2,*], Riduan Yunus[1], Ahmad Fauzi Raffee[1]**

[1]Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn, 86400 Batu Pahat, Malaysia
[2]School of Distance Education, Universiti Sains Malaysia, 18000 Gelugor, Penang, Malaysia
*Corresponding Author: hazrul@usm.my



*Received: 15 February 2023; Revised: 25 February 2023; Accepted: 20 March 2023; Published: 30 April 2023*

**Abstract:** Ground level ozone is an unstable substance in the atmosphere which reacts with nitrogen dioxide (NOx) in the presence of sunlight normally found near ground level. Ground level ozone is known as one of the major air quality issues worldwide. The presence of the NOx volatile organic compounds (VOCs) can influence the concentration levels of ground level ozone. This pollutant can affect crops and human health. The main aim of this research is to find the best fit distribution for urban monitoring stations in Kuala Terengganu, Kota Bharu and Alor Setar. Secondary data from 2011 to 2015 used in this study was obtained from the Department of Environment (DOE). This research used the central fitting probability distribution which is between gamma, Nakagami, lognormal and log logistic distributions. Meanwhile, the method of moments was used to estimate the parameters. The best distribution represented the monitoring station and predicted the return period of the concentration. The results show that the Nakagami distribution represented the Kuala Terengganu station from 2011 to 2015. In Kota Bharu however, the gamma distribution fit better compared to other distributions in 2013. The gamma distribution seems to fit the data from 2011 to 2014 whereas in 2015, the Nakagami distribution fit better than other distributions for the Alor Setar station. In addition, no return period was predicted for concentrations above 0.10 ppm found at the Kuala Terengganu monitoring station.

**Keywords:** *Long-term prediction, lognormal distribution, Nakagami distribution.*

## 1. Introduction

The main issues in environmental pollution nowadays have shifted to air pollution due to the increasing trend of polluted ambient air worldwide. Ambient air pollution affects developed countries in America and Europe more due to urbanization and industrialization compared to developing countries in Asia and Africa [1]. Moreover, air pollution has been shown to affect human health, agriculture, material and the environment significantly. Air pollution has been found to cause numerous human diseases and even death in some cases [2].

The $O_3$ pollutant is formed by inducing the emission of nitrogen dioxide ($NO_2$) and volatile organic compounds (VOCs) in the presence of solar radiation (sunlight) [3]. According to Kim et al. [4], the air pollutant which is most dangerous to human health is particulate matter followed by $O_3$. Particulate matter and $O_3$ have resulted in deaths of 2.1 million and 0.47 million people, respectively. In the past decade, $PM_{10}$ was reported as a significant pollutant in air quality studies. For example, studies on $PM_{10}$ profiles

**Corresponding Author:** Hazrul Abdul Hamid, School of Distance Education, Universiti Sains Malaysia, 18000 Gelugor, Penang, Malaysia., Email: hazrul@usm.my

associated with human health were conducted in Iran [5]. The forecasting of $PM_{10}$ concentrations was also done in Malaysia using several statistical methods in order to predict future $PM_{10}$ concentrations [6-7]. Compared to $PM_{10}$, $O_3$ receives substantial concern worldwide due to significant deleterious effects to human health.

The modelling of air pollution is an important tool for providing early information related to ambient air quality status in future. This study used the statistical distribution method in order to predict future $O_3$ concentrations. In previous studies, the $PM_{10}$ pollutant was investigated by numerous researchers to predict the exceedance level using statistical distribution. The prediction of exceedances and return period at four different locations in Malaysia, namely Kuantan, Kota Kinabalu, Nilai and Johor Bharu was done using gamma distribution [7]. Similar studies comparing generalized extreme value distribution (GEV) and generalized Pareto distribution (GPD) to develop a model for extreme $PM_{10}$ data in Johor Bharu to estimate the return period were conducted [8]. These studies mostly focused on $PM_{10}$ concentrations. This indicated that there are limited statistical distribution studies on $O_3$ concentrations. For instance, gamma and lognormal distributions were applied in Pasir Gudang to examine the return period of $O_3$ [9]. The studies on statistical distributions to predict the exceedance probability and return period of $O_3$ focused more on common, widely used statistical distributions.

The objective of this research is to fit models using two types of statistical distribution in order to predict the exceedances and return period of $O_3$ concentrations in three different urban areas in Malaysia. New knowledge is expected to be contributed by this research in terms of statistical distribution modelling of air quality data as statistical distribution methods such as the Nakagami distribution and gamma distribution applied to $O_3$ data have never been applied in previous studies.

## 2. Monitoring Stations and Air Quality Data

Three urban monitoring locations were selected in this study. The first monitoring station is located in Kota Bharu which is the capital city of Kelantan with total estimated area is 115.6 $km^2$. The second monitoring station involved in this study is Kuala Terengganu which is the capital city of Terengganu with a total area of 605 $km^2$. The third monitoring station is located in the capital city of Kedah, Alor Setar. The total area of Alor Star is 666 $km^2$. All three monitoring stations are situated in the city center where the $O_3$ concentration is influenced by the number of motor vehicles as the main source of its precursor [10]. The total number of motor vehicles in these three states is 626,603, 894,265 and 1,345,100 for Terengganu, Kelantan and Kedah, respectively. Additionally, these three monitoring stations are located near the coastal area with an estimated distance of 8km for Kuala Terengganu and Alor Setar while Kota Bahru has an estimated distance of 7km from the coastal area. The areas situated near the coastal area are probably influenced by sea breeze which transports sea salt aerosol which contains the $O_3$ precursor, $NO_x$. Thus, this area may have similar $O_3$ characteristics due to the location and number of motor vehicles which influence the dispersion of $O_3$ concentrations. Hence, the selection of these areas is due to the conditions mentioned above.

The hourly average of $O_3$ concentration data from 2011 to 2015 was used and analyzed in this study. This data from the monitoring stations in Kuala Terengganu, Kota Bharu dan Alor Setar was provided by the air division, Department of Environment, Malaysia.

## 3. Research Methodology

This research used two statistical distributions to fit the $O_3$ concentration data, thus predicting the return period of the pollutant concentration. Probability density function (PDF) and cumulative distribution function (CDF) were used to estimate exceedance probability of $O_3$ concentration. In this research, two-parameter gamma and Nakagami, distributions were used to analyze $O_3$ concentration. According to Hamid et al. [9], method of moments (MoM) can be used as an estimator for all parameter distributions in this research.

### 3.1 Gamma distribution

Forbes et al. [11] defined gamma distribution as follows:

$$f(x)=\left[\frac{1}{\mu\Gamma(x)}\right]\left(\frac{x}{\mu}\right)^{\lambda-1}\exp\left(-\frac{x}{\mu}\right) \qquad (1)$$

$$x>0, \mu>0, \lambda>0$$

where λ and μ is the shape parameter and scale parameter respectively. Method of Moment is used to estimate the parameter is given by Forbes et al. [11] as below:

$$\mu=\frac{s^2}{\bar{x}} \qquad (2)$$

$$\lambda=\left[\frac{\bar{x}}{s}\right]^2 \qquad (3)$$

### 3.2 Nakagami distribution

The Nakagami distribution defined by Binoti et al. [12] is shown below:

$$f(x) = \frac{2\lambda^{\lambda}}{\Gamma(\lambda)\mu^{\lambda}} x^{2\lambda-1} \exp\left(-\frac{\lambda}{\mu}x^2\right) \quad (4)$$

$$x > 0, \mu > 0, \lambda > 0$$

where λ and μ is the shape parameter and scale parameter respectively. The parameter can be estimated by the method of moments which is defined by Noga and Studanski [13].

$$\lambda = \frac{\bar{x}}{\left(s^2 - \bar{x}^2\right)} \quad (5)$$

$$\mu = \bar{x} \quad (6)$$

### 3.3 Performance Indicator

Five performance indicators will be used to describe how well and fit the set of observation to the distribution as shown in Table 1.

Table 1 Performance indicator.

| No | Performance Indicator | Equation |
|---|---|---|
| 1 | Root Mean Square Error (RMSE) | $\sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{1}\left(P_i - O_i\right)^2}$ Smallest value of RMSE indicate the best estimator. |
| 2 | Normalized Absolute Error (NAE) | $\frac{\sum_{i=1}^{n} ABS\left(P_i - O_i\right)}{\sum_{i=1}^{n} O_i}$ Smallest value of NAE indicate the best estimator. |
| 3 | Coefficient of Determination ($R^2$) | $\left[\frac{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)\left(O_i - \bar{O}\right)}{N.S_{pred}.S_{obs}}\right]^2$ Value of $R^2$ near to 1 indicate best estimator. |
| 4 | Index of Agreement (IA) | $1 - \left[\frac{\sum_{i=1}^{N}\left(P - O_i\right)^2}{\sum_{i=1}^{N} O_i}\right]$ Value IA near to 1 indicate best estimator. |
| 5 | Prediction Accuracy (PA) | $\frac{\sum_{i=1}^{N}\left(P_i - \bar{O}\right)^2}{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2}$ Value of PA near to 1 indicates as best estimator. |

## 4. Selection of the best fit distribution for $O_3$ concentration

The best statistical distribution was determined by using performance indicators which represent the ground level ozone concentration for the urban monitoring stations. The results of these stations are tabulated in Table 2 according to the year and the monitoring station, respectively. Based on the estimated value of scale and shape parameters, the probability density function (PDF) and cumulative density functions (CDF) can be plotted as shown in Figure 1, Figure 2 and Figure 3 (PDF) and Figure 4, Figure 5 and Figure 6 (CDF) below, respectively.

Table 2 Parameter estimation for urban monitoring stations.

| Year | Distribution | Scale parameter | Shape parameter |
|---|---|---|---|
| **Kuala Terengganu** | | | |
| 2011 | Nakagami | 0.00040 | 1.4960 |
| 2012 | Nakagami | 0.00037 | 1.2402 |
| 2013 | Nakagami | 0.00046 | 1.3945 |
| 2014 | Nakagami | 0.00070 | 1.7159 |
| 2015 | Nakagami | 0.00060 | 1.4350 |
| **Kota Bharu** | | | |
| 2011 | Nakagami | 0.00039 | 2.0081 |
| 2012 | Nakagami | 0.00022 | 6.4892 |
| 2013 | Gamma | 0.00036 | 5.9959 |
| 2014 | Nakagami | 0.00012 | 2.2301 |
| 2015 | Nakagami | 0.00049 | 2.0992 |
| **Alor Setar** | | | |
| 2011 | Gamma | 0.00120 | 1.6318 |
| 2012 | Gamma | 0.00088 | 0.0169 |
| 2013 | Gamma | 0.00052 | 3.5403 |
| 2014 | Gamma | 0.00088 | 2.2397 |
| 2015 | Nakagami | 0.00025 | 7.8223 |

## 5. Probability of exceedances

The probability density function plots are shown in Figure 1, Figure 2 and Figure 3 whereas the cumulative distribution function plots are shown in Figure 4, Figure 5 and Figure 6 respectively. The result show that the predicted hourly average for ground level ozone will not exceed the limits set by the Malaysian Ambient Air Quality Guidelines (MAAQG) which is 0.10 ppm. According to the results based on the performance indicator, the Nakagami distribution has a better fit compared to other distributions from 2011 to 2015. The probability of the ground level ozone equal to or less than 0.10 ppm is 1.0 and the probability greater than 0.10 ppm is 0. In addition, this indicated that for five conservative years, the ground level ozone stays below 0.10 ppm and that there is no return period predicted for concentrations above 0.10 ppm. For the Kota Bharu station, the results show that the Nakagami distribution fits better in 2011, 2012, 2014 and 2015 but in 2013, the gamma distribution fit better than other distributions. In 2013, the probability of ground level ozone equal to or less than 0.10 ppm is 0. and the probability greater than 0.10 ppm is 0. In addition, there is no return period for other years from 2011 to 2015.

The results obtained by the Alor Setar station show that the gamma distribution has the best fit even though the Nakagami distribution showed a better fit in 2015. In 2011, the probability of the ground level ozone equal to or less than 0.10 ppm is 0.99912. Moreover, there is no return period from 2012 to 2015. The probability of exceedances and the return period derived from the CDF plot are shown in Table 4 for all urban monitoring stations.

Table 3 Probability of exceedances for ground level ozone in urban areas.

| Year | Probability | Actual return period (days) | Predicted return period (days) |
|---|---|---|---|
| **Kuala Terengganu** | | | |
| 2011 | 1.0 | 0 | 0 |
| 2012 | 1.0 | 0 | 0 |
| 2013 | 1.0 | 0 | 0 |
| 2014 | 1.0 | 0 | 0 |
| 2015 | 1.0 | 0 | 0 |
| **Kota Bharu** | | | |
| 2011 | 1.0 | 0 | 0 |
| 2012 | 1.0 | 0 | 0 |
| 2013 | 0.99986 | 8902 | 0 |
| 2014 | 1.0 | 0 | 0 |
| 2015 | 1.0 | 0 | 0 |
| **Alor Setar** | | | |
| 2011 | 0.99912 | 1250 | 0 |
| 2012 | 1.0 | 0 | 0 |
| 2013 | 1.0 | 0 | 0 |
| 2014 | 1.0 | 0 | 0 |
| 2015 | 1.0 | 0 | 0 |

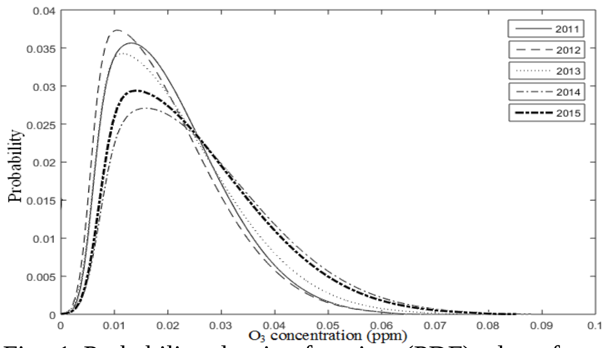

Fig. 1 Probability density function (PDF) plot of ground level ozone for Kuala Terengganu.

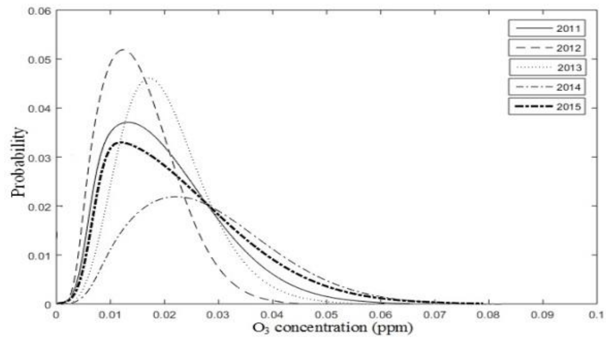

Fig. 2 Probability density function (PDF) plot of ground level ozone for Kota Bharu

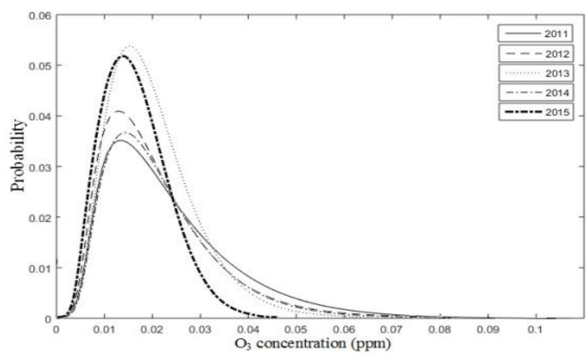

Fig. 3 Probability density function (PDF) plot of ground level ozone for Alor Setar.

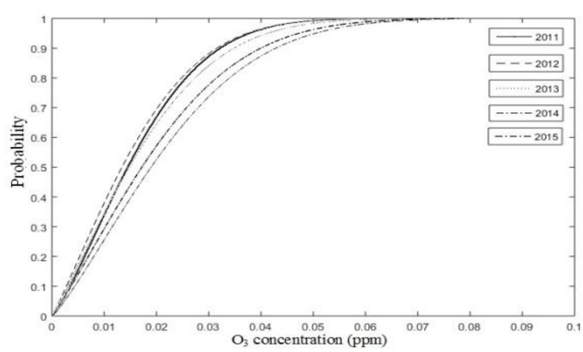

Fig. 4 Cumulative distribution function (CDF) of ground level ozone for Kuala Terengganu.

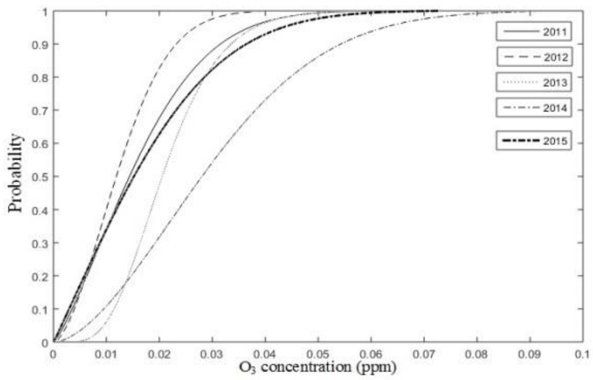


Fig. 5 Cumulative distribution function (CDF) of ground level ozone for Kota Bharu.

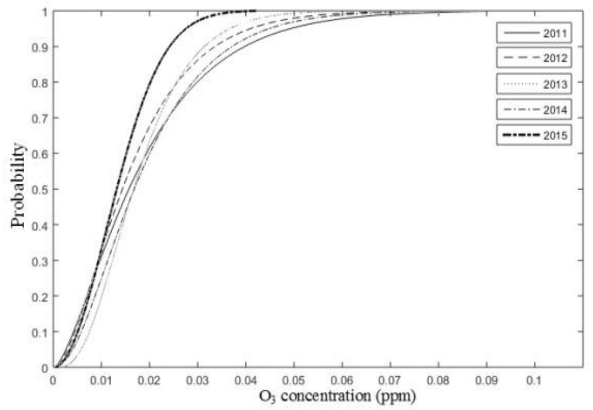


Fig. 6 Cumulative distribution function (CDF) of ground level ozone for Alor Setar.

## 6. Conclusion

Based on the results obtained from the analysis at urban monitoring stations for ground level ozone from 2011 to 2015, the concentrations recorded by the urban monitoring stations have a lower annual average of 0.10 ppm. The probability exceedances did not show that $O_3$ concentrations will exceed the limit for future predictions. The comparison between gamma and Nakagami distributions show that the Nakagami distribution is a much more appropriate fit for the actual monitoring data recorded by the Kuala Terengganu station from 2011 to 2015 as well as the monitoring data recorded by the Kota Bharu station in 2011, 2012, 2014 and 2015. However, the gamma distribution was found to be a better fit for the monitoring data recorded by the Alor Setar station except in 2015 where the Nakagami distribution was found to be more appropriate.

## Acknowledgements

The author would like to thank Ministry of Education and Universiti Sains Malaysia for the funding (FRGS - PJJAUH/6711750). The authors also wishes to thanks Universiti Tun Hussein Onn Malaysia (UTHM) for Post Graduate Research Grant Vot. U700.

## References

[1] World Health Organization, 2016, "Ambient Air Pollution: A global assessment of exposure and burden of disease", World Health Organization.

[2] Goudarzi G. Cardiovascular and respiratory mortality attributed to ground-level ozone in Ahvaz, Iran, Environmental Monitoring and Assessment, Vol. 187, No. 8, 1–9. 2015.

[3] Ghazali, N. A. Ramli, A. S. Yahaya, N. F. F. M. Yusof, N. Sansuddin, and W. A. Al Madhoun. Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques, Environmental Monitoring and Assessment, Vol. 165, No. 1–4, 475–489, 2010.

[4] Kim, E. Kabir, and S. Kabir. A review on the human health impact of airborne particulate matter, Environment International, Vol. 74, 136–143, 2015.

[5] Maleki, A. Sorooshian, G. Goudarzi, and A. Nikfal. Temporal profile of $PM_{10}$ and associated health effects in one of the most polluted cities of the world (Ahvaz, Iran ) between 2009 and 2014, Aeolian Research, Vol. 22, 135–140, 2016.

[6] Ul-Saufie, A. S. Yahaya, N. A. Ramli, and H. A. Hamid. Performance of Multiple Linear Regression Model for Long-term PM10 Concentration Prediction Based on Gaseous and Meteorological Parameters, Journal of Applied Sciences, Vol. 12, No. 14. 1488–1494, 2012.

[7] Sansuddin, N. A. Ramli, A. S. Yahaya, N. F. F. M. Yusof, N. A. Ghazali, and W. A. Al Madhoun. Statistical analysis of PM10 concentrations at different locations in Malaysia, Environmental Monitoring and Assessment, Vol. 180, No. 1–4, 573–588, 2011.

[8] Amin, M. B. Adam, and A. Z. Aris. Extreme Value Analysis for Modeling High PM10 Level in Johor Bahru, Jurnal Teknologi, Vol. 76, No. 1, 171–179, 2015.

[9] Hamid, H.A., Jaffar I. and A. F. Raffee. Two-parameter central fitting distribution to predict the concentration of ground level ozone: Case study in industrial area, in International Conference on Mathematics, Engineering and Industrial Applications 2018.

[10] Department of Environment Malaysia, Malaysia Environmental Quality Report 2016.

[11] Forbes C, Evans M, Hastings N, Peacock B (2011) Statistical distributions 4$^{th}$ edition. A John Wiley & Sons, Inc., Publication. Canada.

[12] Binoti D H B, Binoti M L M S, Leite H G, Fardin L, Oliveira J C. Probability density functions for description of diameter distribution in thinned stands of Tectona grandis. Cerne, Lavras, Vol.18,185-196, 2012.

[13] Noga K M, Studanski R. Estimation of Nakagami distribution parameters in describing a fading radio-communication channel. Scientific Journal of Polish Naval Academy Vol.204, No.1,69–81, 2016.