Research Article

Modeling the Amount of Pollutants Ozone Using Moments Method and Generalized Extreme Value Distribution

Mohammed Mohammed El Genidy and Aya Kamal Ali

Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt

Abstract

Background: Air pollution is a major reason of the depletion of the ozone which allows harmful rays to reach the earth's surface. The problem is how to create a mathematical model by determining the appropriate distribution function for the quantities of the pollutants ozone in the Tenth of Ramadan city of Zagazig province at Egypt. Methodology: In this study, the moment's method, the properties of the cumulative distribution function and Anderson-Darling test were applied to determine. Results: The best values of the parameters of the generalized extreme value distribution. Conclusion: Moreover, the extreme values and statistical measurements were obtained with high accuracy.

Key words: Pollutants ozone, atmospheric pollutants, generalized extreme value distribution, moments method, Anderson-Darling test

Received: July 13, 2016 Accepted: August 20, 2016 Published: September 15, 2016


Corresponding Author: Mohammed Mohammed El Genidy, Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt

Copyright: © 2016 Mohammed Mohammed El Genidy and Aya Kamal Ali. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.
INTRODUCTION

Ozone pollution is a major problem in some regions of the world, which is characterized by high concentrations of ozone at ground level. Exposure to ozone can cause serious health problems in plants and people who lead to several respiratory diseases. Ozone is molecule of three oxygen atoms bound together (O₃) and it is highly unstable and poisonous, while there is also good ozone in the upper earth’s atmosphere that protects us from harmful UV radiation. Our cars, industries and numerous other human sources emit harmful organic compounds (VOCs) and nitrogen oxide gases (NOx) that combined with high temperatures and enough sunlight result in ozone pollution. Ozone pollution may start in urban areas with a high concentration of human activity, but it can spread across vast distances. Albert et al. considered the first quantitative assessment of the impact of physical processes in the snow on air-snow chemical exchange of ozone. Measurements of snow properties, interstitial ozone concentrations and an ozone kinetic depletion experiment results are presented along with two-dimensional model results of the diffusion and ventilation processes affecting gas exchange at Alert, Nunavut, Canada.

Arryndip and Joseph presented extreme temperature forecast in Mbong, Cameroon through return level analysis of the generalized extreme value distribution. Bali has done study on determining the type of asymptotic distribution for the extreme changes in stock prices, foreign exchange rates and interest rates and used regression method to determine the correct specification of the limit distribution for maximum and minimum. Barakat et al. performed a study of the air pollution by extreme value models in the Tenth of Ramadan city of Zagazig province in Egypt. Modelling non-stationary extremes with application to surface level ozone have done study by Eastoe and Tawn. El Damsisy et al. deal with maximum likelihood function to estimate reliability and failure rate of the electronic system by using mixture Lindley distribution. The rule of Br and BrCl in surface ozone destruction at polar sunrise have conducted study by Foster et al. Hammitt carried out study on subjective probability based scenarios for uncertain input parameters: stratospheric ozone depletion. Hasan et al. presented modeling of extreme temperature using generalized extreme value distribution: A case study of Penang. Estimation of the generalized extreme value distribution by the method of probability weighted moments was performed by Hosking et al. Multivariate extreme value distribution with applications to environmental data was presented by Joe. Smith has done study in extreme value analysis of environmental time series: An application to trend detection in ground level ozone. Gong and Ordieres-Mere performed study on prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. Porter has done study in modelling of pollutants in complex environmental systems. Coman et al. presented hourly ozone prediction for a 24 h horizon using networks. Sousa et al. considered multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Arvanitis and Moussiopoulos carried out estimating long term urban exposure to particulate matter and ozone in Europe.

The moments’ method and the properties of the cumulative distribution function were applied in this study to find the best values for the shape, scale and location parameters of the generalized extreme value distribution relating to the quantities of pollutants ozone in the Tenth of Ramadan city. To ensure accuracy and appropriate of the results in this study Anderson-Darling test was performed on the given actual data with the generalized extreme value distribution.

MATERIALS AND METHODS

Atmospheric pollutants and their sources: Air pollution is the presence of harmful substances in the air and harm to human health, the environment and also on the ozone in the upper atmosphere. Air pollutants come from of natural sources and not from human action, such as gas and dust from volcanic eruptions, forest fires, dust storms and the resulting emissions from the intensity of the sun’s rays. In addition to emissions from natural gas leaks usually it is limited in certain areas governed by geographical and geological factors. The types of air pollutants are carbon monoxide CO, carbon dioxide CO₂, nitrogen oxides NOₓ and particulate matter, where the human is the main reason of the production of industrial gases, dust and fumes.

Effects of air pollution on the ozone layer: Pollutants emitted from the earth lead to the presence of ozone in the lower layers of the atmosphere and in this case the ozone is dangerous ingredients on human health, if inhaled human little of it happening to him rampage in the respiratory tract and may cause death. While ozone in the upper atmosphere acts as a shield or a protective filter that protects the earth from harmful ultraviolet radiation, it means that without the presence of the ozone layer leads to the end of life on earth’s surface.
**Generalized extreme value distribution:** Let $X$ be a random variable represented the quantities of pollutants ozone in the Tenth of Ramadan city, which is measured by micro-grams per cubic meter ($\mu g \text{ m}^{-3}$) and has generalized extreme value distribution where the probability density function is defined as:

$$ f(x) = \frac{1}{\theta} g(x)^{-\frac{1}{\alpha}} e^{-\frac{1}{\alpha} g(x)} \quad ; \quad x \in \begin{cases} (\infty, \frac{\theta}{\alpha}) \quad \text{if} \quad \alpha < 0 \\ (\frac{\theta}{\alpha}, \infty) \quad \text{if} \quad \alpha = 0 \\ (\frac{\theta}{\alpha}, \infty) \quad \text{if} \quad \alpha > 0 \end{cases} \quad (1) $$

$$ g(x) = \begin{cases} 1 + \frac{\alpha(x-\lambda)}{\theta} \quad , \quad \text{if} \quad \alpha \neq 0 \\ e^{-\frac{x}{\theta}} \quad , \quad \text{if} \quad \alpha = 0 \end{cases} \quad (2) $$

where, the cumulative distribution function:

$$ F(x) = e^{-\frac{g(x)}{\alpha}} \quad (3) $$

such that, $\alpha$ is a shape parameter, $\theta$ is a scale parameter and $\lambda$ is a location parameter.

**Moment methods:** The moment generating function is defined as:

$$ M(t) = E(e^tX) = \int e^{tx} f(x) \, dx \quad ; \quad t = \frac{\theta}{\alpha} \text{ and } \alpha < 0 $$

Let:

$$ y = \left(1 + \frac{\alpha(x-\lambda)}{\theta}\right)^{-\frac{1}{\alpha}} $$

Then:

$$ x = \frac{\lambda - \frac{\theta}{\alpha}}{\alpha} y^{-\alpha} $$

And:

$$ dx = -\frac{\theta}{\alpha^2} y^{-(\alpha+1)} dy $$

Also, if $x-a$ then $y-0$ and if $x-\infty$ then $y-\infty$, thus:

$$ M(t) = \int e^{tx} f(x) \, dx \quad ; \quad x = \frac{\theta}{\alpha} \text{ and } \alpha < 0 $$

Then:

$$ M(t) = e^{-\frac{\theta}{\alpha} t} \int e^{-\frac{1}{\alpha} \left(1 + \frac{\alpha(x-\lambda)}{\theta}\right)^{-\frac{1}{\alpha}}} \, dx $$

$$ M(t) = e^{-\frac{\theta}{\alpha} t} \int e^{-\frac{1}{\alpha} \left(1 + \frac{\alpha(x-\lambda)}{\theta}\right)^{-\frac{1}{\alpha}}} \, dx $$

**Mean:**

$$ M'(0) = \text{Mean} = \frac{\theta}{\alpha} + \lambda + \frac{\alpha \Gamma(1-\alpha)}{\theta} $$

**Variance:**

$$ \text{Variance} = M''(0) - (M'(0))^2 $$

On the other hand, the median was obtained from Eq. 3 as follows:
\[
F(x) = e^{-\lambda x} = \frac{1}{2}
\]
\[
\ln\left(\frac{1}{2}\right) = -g(x)
\]
\[
\left[1 + \frac{\alpha x - \lambda}{\theta}\right]^{\frac{1}{\alpha}} = \ln(2)
\]
\[
1 + \frac{\alpha x - \lambda}{\theta} = \left(\ln(2)\right)^{-\alpha}
\]

Then:
\[
\text{Median} = x = \lambda + \theta\left[\left(\ln(2)^{-\alpha}\right)^{-1}\right] = \frac{\theta\Gamma(1-\alpha)-1}{\alpha}
\] (8)

**Data:** A set of ozone pollutants was measured during the year 2009 to assess the quality of the air in the 10th of Ramadan city of Zagazig province in Egypt. This study was performed according to the supported project by Zagazig university during 2008-2009, jointly with the Egyptian national center of nuclear safety and radiation control. The observations for this pollutant are recorded every hour on the 24 h through year 2009. The resulted data was used by Barakat et al.\textsuperscript{10-20} in the extreme value modeling.

**Software:** The programs, which have been implemented on the set of the data of ozone pollutants in this study were as follows:


**RESULTS**

Consider the data set of pollutants ozone were selected during one year (365 days) from the Tenth of Ramadan city of Zagazig province at Egypt (Table 1). The measurement unite of the pollutants ozone is micro-grams per cubic meter (\(\mu g \cdot m^{-3}\)).

Hence, from Eq. 5, 7 and 8, the mean, median, variance and standard deviation were obtained as follows:

**Mean:**
\[
\text{Mean} = \lambda + \theta\left(\frac{\Gamma(1-\alpha)-1}{\alpha}\right) = 59.48356
\] (9)

**Median:**
\[
\text{Median} = 1 + \theta\left[\left(\ln(2)^{-\alpha}\right)^{-1}\right] = 58.07
\] (10)

**Variance:**
\[
\text{Variance} = \frac{\theta^2\left(\Gamma(1-2\alpha)-\Gamma(1-\alpha)^2\right)}{\alpha^2} = 116.7622
\] (11)

**Standarddeviation:**
\[
\text{Standarddeviation} = \frac{\theta}{\alpha}\sqrt{\frac{\left(\Gamma(1-2\alpha)-\Gamma(1-\alpha)^2\right)}{\alpha^2}} = 10.8056593
\] (12)

**Subtracting Eq. 9 and 10, then:**
\[
\theta = \frac{1.41356a}{\Gamma(1-\alpha)-((\ln 2)^{-\alpha}-1)}
\]

**Substitution** \(\theta^2\) from Eq. 13 in Eq. 11, then:
\[
\frac{\Gamma(1-2\alpha)-(\Gamma(1-\alpha)^2)}{\Gamma(1-\alpha) - (\ln 2)^{\alpha} \cdot \Gamma(1-\alpha) - (\ln 2)^{-\alpha}} \approx 58.44
\] (13)

Solving Eq. 14 by running mathematica software on the computer using the command:

\[
\text{FindRoot} \left[\frac{\Gamma(1-2\alpha)-(\Gamma(1-\alpha)^2)}{\Gamma(1-\alpha) - (\ln 2)^{\alpha} \cdot \Gamma(1-\alpha) - (\ln 2)^{-\alpha}} \approx 58.44\{x, -1\}\right]
\]

\[
\text{FindRoot} \left[\frac{\Gamma(1-2\alpha)-(\Gamma(1-\alpha)^2)}{\Gamma(1-\alpha) - (\ln 2)^{\alpha} \cdot \Gamma(1-\alpha) - (\ln 2)^{-\alpha}} \approx 58.44\{x, -1.5\}\right]
\]

\[
\text{FindRoot} \left[\frac{\Gamma(1-2\alpha)-(\Gamma(1-\alpha)^2)}{\Gamma(1-\alpha) - (\ln 2)^{\alpha} \cdot \Gamma(1-\alpha) - (\ln 2)^{-\alpha}} \approx 58.44\{x, -0.7\}\right]
\]
Thus, the values of shape parameter $\alpha$ were found as follows:

$$a_1 = 0.471199, a_2 = -0.0635913, a_3 = -0.529371, a_4 = -3.85113$$

Then, the corresponding values of the location and scale parameters ($\lambda$ and $\theta$) were obtained as follows:

$$a_1 = 0.471199, a_2 = -0.0635913, a_3 = -0.529371, a_4 = -3.85113$$

if $a_i = 0.471199$, then $\lambda_i = 57.53$ and $\theta_i = 1.361$

if $a_2 = -0.0635913$, then $\lambda_2 = 54.7763$ and $\theta_2 = 9.09171$

if $a_3 = -0.529371$, then $\lambda_3 = 56.9953$ and $\theta_3 = -11.712$

if $a_4 = -3.85113$, then $\lambda_4 = 58.1263$ and $\theta_4 = -0.286769$

On the other hand, mathematica software have been implemented on the computer to find the graphical representation of $F(x)$ in Eq. 3 as follows:

$$F(x) = \alpha(x-\lambda)^{\alpha-1} e^{-(x-\lambda)^{\alpha}}$$

Plot[$\{F_1(x)\}, \{x, 55, 79.8545\}$,
PlotStyle $\rightarrow \{\text{Thickness}[0.01]\}$,
AxesLabel $\rightarrow \{"x value", "F(x)"\}]

$$F(x) = \alpha(x-\lambda)^{\alpha-1} e^{-(x-\lambda)^{\alpha}}$$

Plot[$\{F_2(x)\}, \{x, 23, 141.954\}$,
PlotStyle $\rightarrow \{\text{Thickness}[0.01]\}$,
AxesLabel $\rightarrow \{"x value", "F(x)"\}]

$$F(x) = \alpha(x-\lambda)^{\alpha-1} e^{-(x-\lambda)^{\alpha}}$$

Plot[$\{F_3(x)\}, \{x, 37, 123.747\}$,
PlotStyle $\rightarrow \{\text{Thickness}[0.01]\}$,
AxesLabel $\rightarrow \{"x value", "F(x)"\}]

The graphical representations of $F(x)$ are shown in Fig. 1-4 at the different values of parameters $\alpha$, $\lambda$ and $\theta$ for the generalized extreme value distribution.

**DISCUSSION**

The main objective of this study is to clarify how modeling the amount of pollutants ozone in the tenth of Ramadan city.
in Egypt using generalized extreme value function with the best values of the shape, location and scale parameters and obtainment the extreme values from \( F(x) \) and comparing the results obtained with real data.

Determine the values of the parameters for the generalized extreme value function: Fig. 3 and 4 show that \( F(x) \) is not a cumulative distribution function where \( F(x) \) is decreasing when \( x \to -\infty \) therefore, the following values of the parameters were rejected:

\[
\alpha = \alpha_1 = -0.529371, \quad \lambda = \lambda_1 = 56.9953, \quad \theta = \theta_1 = -11.712 \quad \text{and} \quad \alpha = \alpha_2 = -3.85113, \quad \lambda = \lambda_2 = 58.1263, \quad \theta = \theta_2 = -0.286769
\]
Formula of the cumulative distribution function $F(x)$ and the probability density function $f(x)$: The cumulative distribution function $F(x)$ had been taken the formula:

$$F(x) = e^{\left(\frac{-(x-54.7763)}{9.09171}\right)}\left(0.0635913\right)^{-1}, \ x \in [23, 141.954]$$  \hspace{1cm} (15)

Then, the probability density function of the generalized extreme value distribution was defined as follows:

$$f(x) = \frac{1}{9.09171}\left(g(x)\right)^{-0.0635913}e^{g(x)}, \ \text{where}$$

$$g(x) = \left(1 + \frac{0.0635913(x-54.7763)}{9.09171}\right)^{-1}$$  \hspace{1cm} (16)

The graphical representation of $f(x)$ had shown in Fig. 5, moreover $f(x)$ satisfies the condition:

$$\int_{23}^{141.954} f(x) \, dx = 1$$

Computing each of the mean, variance and standard deviation of generalized extreme value distribution: The mean and variance are defined as follows:

$$E(x) = \int_{23}^{141.954} x \cdot f(x) \, dx = 59.4835 = \text{The actual mean}$$

$$E(x^2) - (E(x))^2 = \int_{23}^{141.954} x^2 f(x) \, dx - (59.4835)^2 = 116.7671 = \text{The actual variance}$$

where, the actual variance is equal to 116.7622. While the standard deviation is equal to 10.80606311 which it was approached to actual standard deviation 10.8056593.

Extreme values and median: Let $u = F(x)$ where by solving the equation, extreme values and the corresponding values of $f(x)$ were obtained in Table 2. While the median is the solving of the equation $F(x) = 0.5$ and equal to 58.07. On the other hand, in Table 2, at $F(x) = 0.5$ the median has the same value 58.07.

Anderson-Darling test: The Anderson-Darling test is used to test if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails. The K-S test is distribution free in the sense that the critical values do not

On the other hand, the actual data of the amount of pollutants ozone belong to the interval [32.51, 123] (Table 1), while the data in Fig. 1, belong to the interval [55, 79.8545] and in Fig. 2, belong to [23, 141.954].

Thus [32.51, 123] ∈ [55, 79.8545] and (32.51, 123) ∈ [23, 1413954], for this reason the values of parameters $\alpha = \alpha_1 = 0.471199$, $\lambda = \lambda_1 = 57.53$, $\theta = \theta_1 = 1.36$ were excluded and the values $\alpha = \alpha_2 = -0.0635913$, $\lambda = \lambda_2 = 54.7763$, $\theta = \theta_2 = 9.09171$ are acceptable and more accurate.

While, in Fig. 1 and 2 show that $F(x)$ is a cumulative distribution function where $F(x)$ is increasing and $F(x) - 1$ when $x \rightarrow \infty$, thus the following values of the parameters were accepted:

$$\alpha = \alpha_2 = -0.0635913, \ \lambda = \lambda_2 = 54.7763, \ \theta = \theta_2 = 9.09171$$

Fig. 3: $F(x)$ is not a cumulative distribution function because it is decreasing function when $x \rightarrow \infty$, thus the values of the parameters $\alpha = -0.529371$, $\lambda = 56.9953$ and $\theta = -11.712$ were rejected.

Fig. 4: $F(x)$ is not a cumulative distribution function because it is decreasing function when $x \rightarrow \infty$, thus the values of the parameters $\alpha = -3.85113$, $\lambda = 58.1263$ and $\theta = -0.286769$ were rejected.

$$\alpha = \alpha_1 = 0.471199, \ \lambda = \lambda_1 = 57.53, \ \theta = \theta_1 = 1.36 \ \text{and} \ \alpha = \alpha_2 = -0.0635913, \ \lambda = \lambda_2 = 54.7763, \ \theta = \theta_2 = 9.09171$$

$$\alpha = \alpha_2 = -0.0635913, \ \lambda = \lambda_2 = 54.7763, \ \theta = \theta_2 = 9.09171$$
Table 2: Extreme values with corresponding values of cumulative distribution function and probability density function

<table>
<thead>
<tr>
<th>x</th>
<th>F(x)</th>
<th>f(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7E-07</td>
<td>0.00001</td>
<td>0.0003</td>
</tr>
<tr>
<td>40.1951</td>
<td>43.7428</td>
<td>44.4448</td>
</tr>
<tr>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>0.0046</td>
<td>0.00789</td>
<td>0.01315</td>
</tr>
<tr>
<td>60.7548</td>
<td>63.84</td>
<td>67.7831</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>0.035183</td>
<td>0.029339</td>
<td>0.0216</td>
</tr>
<tr>
<td>97.0969</td>
<td>98.924</td>
<td>101.443</td>
</tr>
<tr>
<td>0.996</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td>0.000624</td>
<td>0.000477</td>
<td>0.000326</td>
</tr>
<tr>
<td>122.656</td>
<td>124.012</td>
<td>125.855</td>
</tr>
<tr>
<td>0.99996</td>
<td>0.99997</td>
<td>0.99998</td>
</tr>
<tr>
<td>8.38E-06</td>
<td>6.41E-06</td>
<td>4.41E-06</td>
</tr>
</tbody>
</table>

Table 3: Comparing the results of software (EasyFit) and the method used in this study with the actual values for mean, median, variance and standard deviation

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Median</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual value</td>
<td>59.48356</td>
<td>58.07</td>
<td>116.7622</td>
<td>10.8056539</td>
</tr>
<tr>
<td>Software (EasyFit)</td>
<td>59.484</td>
<td>58.445</td>
<td>109.19</td>
<td>10.45</td>
</tr>
<tr>
<td>Method used in this study</td>
<td>59.48356</td>
<td>58.07</td>
<td>116.771</td>
<td>10.8060311</td>
</tr>
</tbody>
</table>

depend on the specific distribution being tested (note that this is true only for a fully specified distribution, i.e., the parameters are known). The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

The Anderson-Darling test is defined as:

- **H₀**: The data, follows the generalized extreme value distribution f(x) in Eq. 16
- **H₁**: The data, does not follow the generalized extreme value distribution f(x) in Eq. 16

**Test statistic:** The Anderson-Darling test statistic is defined as:

\[
AD = -N \sum_{i=1}^{N} \left[ \ln(F(x_i)) + \ln(1-F(x_{N-i+1})) \right]
\]

then, \(AD = 1.066342\). Such that, \(F(x)\) in Eq. 15, is the cumulative distribution function of the generalized extreme value distribution, \(N = 365\) is the population size of the pollutants ozone in the tenth of Ramadan city in Egypt during one year and \(x\) are the ordered data.

The critical values for the Anderson-Darling test are dependent on the specific distribution that is being tested which it is the generalized extreme value distribution where the test is a one-sided test and the hypothesis that the distribution is of a specific form is rejected if the test statistic \(AD\) is greater than the critical value 2.5018 at significant level 0.05, but \(AD\) is less than the critical value, then \(H₀\) was accepted. Therefore, the actual data, follows the generalized extreme value distribution f(x) in Eq. 16.

Comparing the statistical measurements in each of the software EasyFit program and the method used in this study: The software (EasyFit program) has been run on the actual values listed in Table 1, thus we got the values of mean, median, variance and standard deviation where \(\alpha = -0.0635913\), \(\lambda = 54.7763\) and \(\theta = 9.09171\) in Table 3, also the actual values and the results of the method used in this study had shown in the same table. As a result, it is clear that the method used in this study better and more accurate than the method used by the software (EasyFit program).

Fig. 5: f(x) is the probability density function of the generalized extreme value distribution at the best estimating parameters \(\alpha = -0.0635913\), \(\lambda = 54.7763\) and \(\theta = 9.09171\)
CONCLUSION

The statistical procedures in this study give us more accurate results than the other methods which used maximum likelihood function in estimation of parameters for the distribution function or software such as EasyFit program. Furthermore, the extreme values, the arithmetic mean, variance, standard deviation and median of pollutants ozone can be obtained by using generalized extreme value distribution. The method used in this study enable the researchers to apply it in relevant fields. In addition, it gives very accurate results of statistics and distribution function for any other phenomena data.

REFERENCES