Asian Journal of Scientific Research

ISSN 1992-1454

OPEN ACCESS Asian Journal of Scientific Research

ISSN 1992-1454
DOI: 10.3923/ajsr.2016.143.151

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

Citation: Mohammed Mohammed El Genidy and Aya Kamal Ali, 2016. Modeling the amount of pollutants ozone using moments method and generalized extreme value distribution. Asian J. Sci. Res., 9: 143-151.

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

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 (O_3) 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.*[1] 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.

Arreyndip and Joseph[2] presented extreme temperature forecast in Mbong, Cameroon through return level analysis of the generalized extreme value distribution. Bali[3] 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.*[4] performed a study of the air pollution by extreme value models in the Tenth of Ramadan city of Zagazig province at Egypt. Modelling non-stationary extremes with application to surface level ozone have done study by Eastoe and Tawn[5]. El Damsesey *et al.*[6] deal with maximum likelihood function to estimate reliability and failure rate of the electronic system by using mixture Lindley distribution. The rule of Br_2 and BrCl in surface ozone destruction at polar sunrise have conducted study by Foster *et al.*[7]. Hammitt[8] carried out study on subjective probability based scenarios for uncertain input parameters: stratospheric ozone depletion. Hasan *et al.*[9] 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.*[10]. Multivariate extreme value distribution with applications to environmental data was presented by Joe[11]. Smith[12] has done study in extreme value analysis of environmental time series: An application to trend detection in ground level ozone. Gong and Ordieres-Mere[13] performed study on prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. Porter[14] has done study in modelling of pollutants in complex environmental systems. Coman *et al.*[15] presented hourly ozone prediction for a 24 h horizon using networks. Sousa *et al.*[16] considered multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Arvanitis and Moussiopoulos[17] 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 CO_2, 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 (µg m^{-3}) and has generalized extreme value distribution where the probability density function is defined as:

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

θ>0, λ, α∈R; R is the set o f real numbers:

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

where, the cumulative distribution function:

$$F(x) = e^{-g(x)} \tag{3}$$

such that, α is a shape parameter, θ is a scale parameter and λ is a location parameter.

Moment methods: The moment generating function is defined as:

$$M(t) = E(e^{tx}) = \int_{-\infty}^{a} e^{tx} f(x)\, dx, \text{ where } a = \lambda - \frac{\theta}{\alpha} \text{ and } \alpha < 0$$

Let:

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

Then:

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

And:

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

Also, if x→a then y→0 and if x→-∞ then y→-∞, thus:

$$M(t) = e^{\left(\lambda - \frac{\theta}{\alpha}\right)t} \int_{-\infty}^{0} e^{\frac{\theta t}{\alpha} y^{-\alpha}} \frac{1}{\theta} y^{(\alpha+1)} e^{-y} (-\theta y^{-(\alpha+1)}) dy$$

Then:

$$M(t) = e^{(\lambda - \frac{\theta}{\alpha})t} \int_{0}^{\infty} e^{\frac{\theta t}{\alpha} y^{-\alpha}} e^{-y} dy$$

$$M(t) = e^{(\lambda - \frac{\theta}{\alpha})t} \left[\int_0^\infty e^{-y} dy + \int_0^\infty \left(\frac{\theta t}{\alpha} y^{-\alpha}\right) e^{-y} dy + \int_0^\infty \frac{\left(\left(\frac{\theta t}{\alpha}\right)^2 y^{-2\alpha}\right)}{2!} e^{-y} dy + \dots \right]$$

$$M(t) = e^{(\lambda - \frac{\theta}{\alpha})t} \int_0^\infty \left[\sum_{n=0}^{\infty} \frac{\left(\frac{\theta t}{\alpha} y^{-\alpha}\right)^n}{n!} \right] e^{-y} dy$$

$$M(t) = e^{\left(\lambda - \frac{\theta}{\alpha}\right)t} \left[1 + \frac{\theta t}{1!\alpha} \Gamma(1-\alpha) + \frac{(\theta t)^2}{2!\, \alpha^2} (1\text{-}2\alpha) + \dots \right]$$

$$M(t) = e^{\left(\lambda - \frac{\theta}{\alpha}\right)t} \left[1 + \sum_{n=1}^{\infty} \frac{(\theta t)^n}{n!\, \alpha^n} \Gamma(1\text{-}n\alpha) \right] \tag{4}$$

Then, the first moment around zero will be as follows:

$$M'(0) = \text{Mean} = -\frac{\theta}{\alpha} + \lambda + \frac{\theta \Gamma(1-\alpha)}{\alpha} \tag{5}$$

And the second moment around zero takes the following formula:

$$M''(0) = \left(-\frac{\theta}{\alpha} + \lambda\right)^2 + \frac{\theta^2 \Gamma(1-2\alpha)}{\alpha^2} + \frac{2\theta\left(-\frac{\theta}{\alpha} + \lambda\right)\Gamma(1-\alpha)}{\alpha} \tag{6}$$

Thus:

$$\text{Variance} = M''(0) - (M'(0))^2 = \frac{\theta^2\left(\Gamma(1-2\alpha) - \Gamma(1-\alpha)^2\right)}{\alpha^2} \tag{7}$$

On the other hand, the median was obtained from Eq. 3 as follows:

$$F(x) = e^{-g(x)} = \frac{1}{2}$$

$$Ln\left(\frac{1}{2}\right) = -g(x)$$

$$\left[1+\frac{\alpha(x-\lambda)}{\theta}\right]^{-\frac{1}{\alpha}} = Ln(2)$$

$$1+\frac{\alpha(x-\lambda)}{\theta} = \left(Ln(2)\right)^{-\alpha}$$

Then:

$$Median = x = \lambda + \frac{\theta\left[(Ln(2))^{-\alpha} - 1\right]}{\alpha} \quad (8)$$

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 at Egypt. This study was performed according a supported project by Zagazig university during 2008-2009, jointly with the Egyptian national center of nuclear safety and radiation control. The observations for this poolutant are recorded every hour on the 24 h through year 2009. The resulted data was used by Barakat *et al.*[18-20] 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:

- EasyFit professional, version 5.5 (Released: 2010–2-05), 2004-2010 MathWave
  Technologies, http://www.mathwave.com
- Mathematica4, version number 4.0.1.0, copyrights 1988-1999 Wolfram research, http://www.wolfram.com

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 (μg m^{-3}).

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

$$Mean = \lambda + \theta\left(\frac{\Gamma(1-\alpha)-1}{\alpha}\right) = 59.48356 \quad (9)$$

$$Median = l + \frac{\theta\left[(Ln(2))^{-\alpha} - 1\right]}{\alpha} = 58.07 \quad (10)$$

$$Variance = \frac{\theta^2\left(\Gamma(1-2\alpha) - \Gamma(1-\alpha)^2\right)}{\alpha^2} = 116.7622 \quad (11)$$

$$Standard devaition = \frac{\theta}{\alpha}\sqrt{\Gamma(1\text{-}2\alpha) - (\Gamma(1\text{-}\alpha))^2} = 10.80565593 \quad (12)$$

Subtracting Eq. 9 and 10, then:

$$\theta\left[\frac{\Gamma(1-\alpha)-1}{\alpha} - \frac{(\ln 2)^{-\alpha}-1}{\alpha}\right] = 1.41356$$

$$\theta = \frac{1.41356\alpha}{(\Gamma(1\alpha)-1) - ((\ln 2)^{-\alpha}-1)}$$

$$\theta = \frac{1.41356\alpha}{\Gamma(1-\alpha) - (\ln 2)^{-\alpha}}$$

$$\theta^2 \cong \frac{1.998\alpha^2}{\left[\Gamma(1-\alpha) - (\ln 2)^{-\alpha}\right]^2} \quad (13)$$

Substitution θ^2 from Eq. 13 in Eq. 11, then:

$$\frac{\Gamma(1\text{-}2\alpha) - (\Gamma(1\text{-}\alpha))^2}{\left[\Gamma(1\text{-}\alpha) - (\ln 2)^{-\alpha}\right]^2} \cong 58.44 \quad (14)$$

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

$$FindRoot\left[\frac{Gamma[1-2\alpha] - (Gamma[1-\alpha])^2}{\left(Gamma[1-\alpha] - (Log[2])^{-\alpha}\right)^2} == 58.44\{x,-1\}\right]$$

$$FindRoot\left[\frac{Gamma[1-2\alpha] - (Gamma[1-\alpha])^2}{\left(Gamma[1-\alpha] - (Log[2])^{-\alpha}\right)^2} == 58.44\{x,-1.5\}\right]$$

$$FindRoot\left[\frac{Gamma[1-2\alpha] - (Gamma[1-\alpha])^2}{\left(Gamma[1-\alpha] - (Log[2])^{-\alpha}\right)^2} == 58.44\{x,-0.7\}\right]$$

$$\text{FindRoot}\left[\frac{\text{Gamma}[1-2\alpha]-(\text{Gamma}[1-\alpha])^2}{\left(\text{Gamma}[1-\alpha]-(\text{Log}[2])^{-\alpha}\right)^2}==58.44\{x,-2\}\right]$$

Thus, the values of shape parameter α were found as follows:

$$\alpha_1 = 0.471199, \alpha_2 = -0.0635913, \alpha_3 = -0.529371, \alpha_4 = -3.85113$$

Then, the corresponding values of the location and scale parameters (λ and θ) were obtained as follows:

$$\alpha_1 = 0.471199, \alpha_2 = -0.0635913, \alpha_3 = -0.529371, \alpha_4 = -3.85113$$
$$\alpha_1 = 0.471199, \alpha_2 = -0.0635913, \alpha_3 = -0.529371, \alpha_4 = -3.85113$$
$$\text{if } \alpha_1 = 0.471199, \text{ then } \lambda_1 = 57.53 \text{ and } \theta_1 = 1.361$$
$$\text{if } \alpha_2 = -0.0635913, \text{ then } \lambda_2 = 54.7763 \text{ and } \theta_2 = 9.09171$$
$$\text{if } \alpha_3 = -0.529371, \text{ then } \lambda_3 = 56.9953 \text{ and } \theta_3 = -11.712$$
$$\text{if } \alpha_4 = -3.85113, \text{ then } \lambda_4 = 58.1263 \text{ and } \theta_4 = -0.286769$$

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:

$$\alpha = \alpha_1$$
$$\lambda = \lambda_1$$
$$\theta = \theta_1$$
$$F(x) = F_1(x) = E^{\left(1+\frac{\alpha_1(x-\lambda_1)}{\theta_1}\right)^{\frac{-1}{\alpha_1}}}$$

Plot[{F_1(x)},{x, 55, 79.8545},
PlotStyle → {{Thickness[0.01]}},
AxesLable → {"x value", "F(x)"}]

$$\alpha = \alpha_2$$
$$\lambda = \lambda_2$$
$$\theta = \theta_2$$
$$F(x) = F_2(x) = EE^{-\left(1+\frac{\alpha_2(x-\lambda_2)}{\theta_2}\right)^{\frac{-1}{\alpha_2}}}$$

Plot[{F_2(x)},{x, 23, 141.954},
PlotStyle → {{Thickness[0.01]}},
AxesLable → {"x value" , "F(x)"}]

$$\alpha = \alpha_3$$
$$\lambda = \lambda_3$$
$$\theta = \theta_3$$
$$F(x) = F_3(x) = E^{-\left(1+\frac{\alpha_3(x-\lambda_3)}{\theta_3}\right)^{\frac{-1}{\alpha_3}}}$$

Plot[{F_3(x)},{x, 37, 123.747},
PlotStyle → {{Thickness[0.01]}},
AxesLable → {"x value" , "F(x)"}]

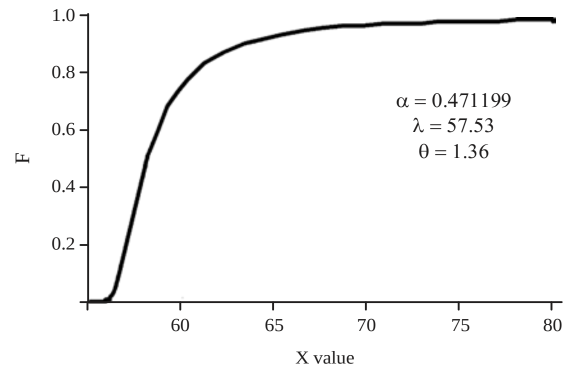

Fig. 1: F(x) is a cumulative distribution function because it is increasing function and F(x)→1 when x→∞, thus the values of the parameters α: 0.471199, λ: 57.53 and θ: 1.36 were accepted

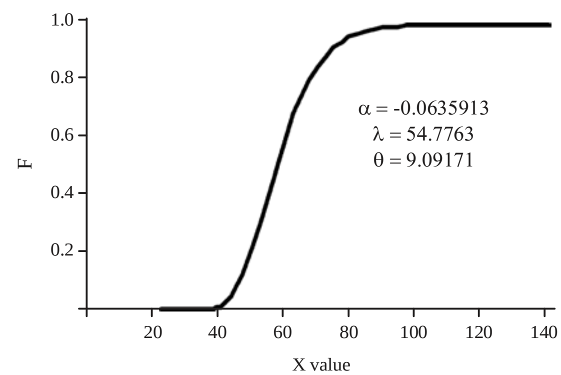

Fig. 2: F(x) is a cumulative distribution function because it is increasing function and F(x)→1 when x→∞, thus the values of the parameters α: -0.0635913, λ: 54.7763 and θ: 9.09171 were accepted

$$\alpha = \alpha_4$$
$$\lambda = \lambda_4$$
$$\theta = \theta_4$$
$$F(x) = F_4(x) = E^{-\left(1+\frac{\alpha_4(x-\lambda_4)}{\theta_4}\right)^{\frac{-1}{\alpha_4}}}$$

Plot[{F_4(x)},{x, 60, 80},
PlotStyle → {{Thickness[0.01]}},
AxesLable → {"x value", "F(x)"}]

The graphical representations of F(x) are shown in Fig. 1-4 at the different values of parameters α, λ and θ 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

Table 1: Data set of the pollutants ozone (µg m^{-3}) during one year (365 days) in the Tenth of Ramadan city of El-Sharkia governorate in Egypt

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Pollutant ozone	32.51	32.61	33.29	36.77	36.97	38.67	38.84	39.64	39.81	40.22	42.42	42.7	42.87	42.97
Day	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Pollutant ozone	43.21	43.8	43.93	44.14	44.3	44.63	44.77	44.94	45.12	45.37	45.54	45.59	45.88	45.94
Day	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Pollutant ozone	46.26	46.32	46.68	46.7	46.71	46.72	46.93	47.18	47.34	47.48	47.64	47.69	47.85	47.97
Day	43	44	45	46	47	48	49	50	51	52	53	54	55	56
Pollutant ozone	48.25	48.4	48.41	48.46	48.52	48.93	49.17	49.39	49.4	49.71	49.96	50.18	50.18	50.29
Day	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Pollutant ozone	50.31	50.54	50.58	50.72	50.78	50.82	50.88	50.91	50.93	50.95	50.96	50.97	51.1	51.12
Day	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Pollutant ozone	51.19	51.33	51.37	51.43	51.45	51.47	51.51	51.59	51.67	51.81	51.82	51.93	51.96	52.01
Day	85	86	87	88	89	90	91	92	93	94	95	96	97	98
Pollutant ozone	52.02	52.12	52.13	52.14	52.15	52.16	52.22	52.26	52.31	52.32	52.33	52.34	52.37	52.4
Day	99	100	101	102	103	104	105	106	107	108	109	110	111	112
Pollutant ozone	52.46	52.57	52.61	52.63	52.7	52.7	52.71	52.87	52.96	52.98	53.04	53.15	53.35	53.38
Day	113	114	115	116	117	118	119	120	121	122	123	124	125	126
Pollutant ozone	53.52	53.69	53.72	53.73	53.98	54.08	54.18	54.19	54.31	54.32	54.33	54.41	54.41	54.43
Day	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Pollutant ozone	54.46	54.56	54.62	54.69	54.83	54.94	54.95	55.09	55.14	55.23	55.33	55.35	55.4	55.45
Day	141	142	143	144	145	146	147	148	149	150	151	152	153	154
Pollutant ozone	55.49	55.71	55.72	55.73	55.74	55.78	55.79	55.82	55.83	55.83	55.89	56.05	56.09	56.1
Day	155	156	157	158	159	160	161	162	163	164	165	166	167	168
Pollutant ozone	56.18	56.26	56.38	56.4	56.43	56.57	56.58	56.6	56.81	56.87	56.87	56.87	56.9	56.93
Day	169	170	171	172	173	174	175	176	177	178	179	180	181	182
Pollutant ozone	56.97	56.98	57.06	57.19	57.37	57.49	57.68	57.72	57.72	57.84	57.85	57.91	58.05	58.05
Day	183	184	185	186	187	188	189	190	191	192	193	194	195	196
Pollutant ozone	58.07	58.25	58.48	58.49	58.53	58.53	58.58	58.87	58.87	58.91	59.07	59.12	59.19	59.2
Day	197	198	199	200	201	202	203	204	205	206	207	208	209	210
Pollutant ozone	59.31	59.39	59.55	59.68	59.83	59.85	59.85	59.93	59.97	60.01	60.04	60.14	60.26	60.28
Day	211	212	213	214	215	216	217	218	219	220	221	222	223	224
Pollutant ozone	60.29	60.36	60.56	60.64	60.84	60.98	61.04	61.08	61.17	61.18	61.21	61.24	61.34	61.82
Day	225	226	227	228	229	230	231	232	233	234	235	236	237	238
Pollutant ozone	61.95	62.08	62.12	62.18	62.33	62.4	62.43	62.47	62.47	62.51	62.56	62.56	62.56	62.62
Day	239	240	241	242	243	244	245	246	247	248	249	250	251	252
Pollutant ozone	62.62	62.76	62.76	62.82	62.84	63.02	63.02	63.18	63.27	63.47	63.48	63.96	63.98	64.06
Day	253	254	255	256	257	258	259	260	261	262	263	264	265	266
Pollutant ozone	64.33	64.34	64.43	64.49	64.68	64.69	64.76	65.04	65.27	65.28	65.28	65.35	65.37	65.37
Day	267	268	269	270	271	272	273	274	275	276	277	278	279	280
Pollutant ozone	65.5	65.56	65.69	65.73	65.78	65.9	65.98	66.14	66.18	66.23	66.4	66.57	66.63	66.64
Day	281	282	283	284	285	286	287	288	289	290	291	292	293	294
Pollutant ozone	66.73	66.75	66.81	66.9	66.99	67.1	67.37	67.51	67.56	67.59	67.6	67.67	67.83	67.86
Day	295	296	297	298	299	300	301	302	303	304	305	306	307	308
Pollutant ozone	68.1	68.15	68.36	68.4	68.47	68.55	68.78	68.88	68.98	69.29	69.38	69.39	69.58	69.66
Day	309	310	311	312	313	314	315	316	317	318	319	320	321	322
Pollutant ozone	69.73	69.74	69.81	69.88	69.98	70.2	70.21	70.32	70.46	70.74	70.74	70.77	70.89	70.99
Day	323	324	325	326	327	328	329	330	331	332	333	334	335	336
Pollutant ozone	71.04	71.1	71.14	71.31	71.6	71.6	71.69	71.7	71.7	71.71	71.73	71.75	71.98	72.1
Day	337	338	339	340	341	342	343	344	345	346	347	348	349	350
Pollutant ozone	72.12	72.23	72.82	74	74.03	74.31	75.7	75.8	76.4	76.6	78.52	79.04	79.11	80.76
Day	351	352	353	354	355	356	357	358	359	360	361	362	363	364
Pollutant ozone	80.87	81.2	81.37	81.7	82.2	82.9	82.94	84.05	84.48	84.94	86.69	87.5	87.78	110.4
Day	365													
Pollutant ozone	123													

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:

$$\alpha = \alpha_3 = -0.529371,\ \lambda = \lambda_3 = 56.9953,\ \theta = \theta_3 = -11.712 \text{ and}$$
$$\alpha = \alpha_4 = -3.85113,\ \lambda = \lambda_4 = 58.1263,\ \theta = \theta_4 = -0.286769$$

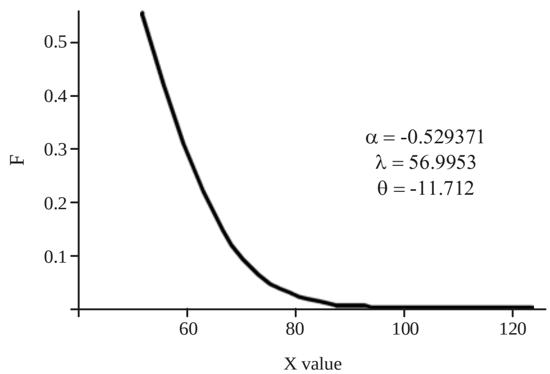

Fig. 3: F(x) is not a cumulative distribution function because it is decreasing function when x→∞, thus the values of the parameters α: -0.529371, λ: 56.9953 and θ: -11.712 were rejected

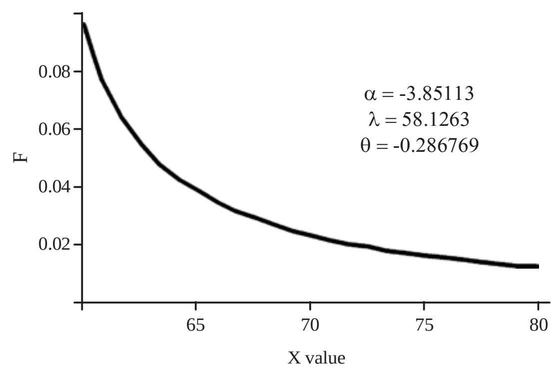

Fig. 4: F(x) is not a cumulative distribution function because it is decreasing function when x→∞, thus the values of the parameters α: -3.85113, λ: 58.1263 and θ: -0.286769 were rejected

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→∞, thus the following values of the parameters were accepted:

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

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.

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:

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

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

$$f(x) = \frac{1}{9.09171}[g(x)]^{-0.0635913} e^{-[g(x)]},\ \text{where}$$
$$g(x) = \left(1 + \frac{-0.0635913(x-54.7763)}{9.09171}\right)^{\frac{-1}{(-0.0635913)}} \quad (16)$$

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:

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

$$E(x^2)-\left(E(x)\right)^2 = \int_{23}^{141.954} x^2 f(x)\, dx - (59.4835)^2 = 116.771 \cong \text{The actual variance}$$

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 10.80565593.

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

Table 2: Extreme values with corresponding values of cumulative distribution function and probability density function

x	28.3347	30.8037	31.8192	32.9778	33.9124	34.7943	35.9248	36.0799	37.8509	38.784
F(x)	0	0.00001	0.00003	0.00009	0.0002	0.0004	0.0009	0.001	0.003	0.005
f(x)	7E-07	0.00001	0.00003	0.00008	0.00016	0.0003	0.00061	0.00067	0.0017	0.0026
x	40.1951	41.821	43.7428	44.4448	45.6018	46.9888	50.3836	53.0786	55.5689	58.07
F(x)	0.01	0.02	0.03	0.04	0.07	0.1	0.2	0.3	0.4	0.5
f(x)	0.0046	0.00789	0.01315	0.015365	0.01924	0.024018	0.034349	0.039261	0.040538	0.039019
x	60.7548	63.84	67.7831	73.8395	79.3835	81.0897	83.2419	86.1929	91.0374	95.6553
F(x)	0.6	0.7	0.8	0.9	0.95	0.96	0.97	0.98	0.99	0.995
f(x)	0.035183	0.029339	0.0216	0.012034	0.006474	0.005283	0.004058	0.002791	0.001466	0.000768
x	97.0969	98.924	101.443	105.598	109.574	110.816	112.39	114.547	118.149	121.519
F(x)	0.996	0.997	0.998	0.999	0.9995	0.9996	0.9997	0.9998	0.9999	0.99995
f(x)	0.000624	0.000477	0.000326	0.000171	8.91E-05	7.23E-05	5.53E-05	3.79E-05	1.98E-05	1.05E-05
x	122.656	124.012	125.855	128.69	131.839	132.67	133.651	134.834	136.299	141.954
F(x)	0.99996	0.99997	0.99998	0.99999	0.999995	0.999996	0.999997	0.999998	0.999999	1
f(x)	8.38E-06	6.41E-06	4.41E-06	2.44E-06	1.23E-06	1.02E-06	8.1E-07	6.19E-07	4.38E-07	1.06E-07

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

	Statistics			
Method	Mean	Median	Variance	Standard deviation
Actual value	59.48356	58.07	116.7622	10.80565593
Software (EasyFit)	59.484	58.445	109.19	10.45
Method used in this study	59.4835	58.07	116.771	10.8060311

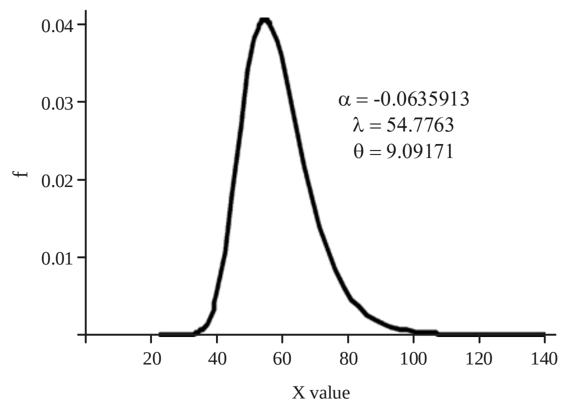

Fig. 5: f(x) is the probability density function of the generalized extreme value distribution at the best estimating parameters α: -0.0635913, λ: 54.7763 and θ: 9.09171

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:

- **H_0:** The data, follows the generalized extreme value distribution f(x) in Eq. 16
- **H_a:** 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:

$$AD = -N-\frac{1}{N}\sum_{i=1}^{N}(2i-1)[Ln(F(x_i)) + Ln(1-F(x_{N-i+1}))]$$

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_i 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[2] 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 α = -0.11985, λ = 55.109 and θ = 9.3043 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).

CONCULOSION

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

1. Albert, M.R., A.M. Grannas, J. Bottenheim, P.B. Shepson and F.E. Perron, 2002. Processes and properties of snow-air transfer in the high Arctic with application to interstitial ozone at Alert, Canada. Atmospheric Environ., 36: 2779-2787.
2. Arreyndip, N.A. and E. Joseph, 2015. Extreme temperature forecast in mbonge, cameroon through return level analysis of the Generalized Extreme Value (GEV) distribution. Int. J. Math. Comput. Phys. Electr. Comput. Eng., 9: 347-352.
3. Bali, T.G., 2003. The generalized extreme value distribution. Econ. Lett., 79: 423-427.
4. Barakat, H.M., E.M. Nigm, A.A. Ramadan and O.M. Khaled, 2010. A study of the air pollutants by extreme value models. J. Applied Stat. Sci., 18: 199-209.
5. Eastoe, E.F. and J.A. Tawn, 2009. Modelling non stationary extremes with application to surface level ozone. J. Royal Stat. Soc.: Series C (Applied Stat.), 58: 25-45.
6. El Damsesy, M.A., M.M. El Genidy and A.M. El Gazar, 2014. Estimation of a system performance in Pareto distribution with two independent random variables. J. Applied Sci., 14: 2854-2856.
7. Foster, K.L., R.A. Plastridge, J.W. Bottenheim, P.B. Shepson, B.J. Finlayson-Pitts and C.W. Spicer, 2001. The role of Br_2 and BrCl in surface ozone destruction at polar sunrise. Science, 291: 471-474.
8. Hammitt, J.K., 1990. Subjective-probability-based scenarios for uncertain input parameters: Stratospheric ozone depletion. Risk Analysis J., 10: 93-102.
9. Hasan, H.B., N.F. Ahmad Radi and S. Kassim, 2012. Modeling of extreme temperature using Generalized Extreme Value (GEV) distribution: A case study of Penang. Proceedings of the World Congress on Engineering, Volume 1, July 4-6, 2012, London, UK., pp: 181-186.
10. Hosking, J.R.M., J.R. Wallis and E.F. Wood, 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27: 251-261.
11. Joe, H., 1994. Multivariate extreme value distributions with applications to environmental data. Can. J. Stat., 22: 47-64.
12. Smith, R.L., 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. Stat. Sci., 4: 367-377.
13. Gong, B. and J. Ordieres-Mere, 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. Environ. Modell. Software, 84: 290-303.
14. Porter, N., 2010. Modelling of pollutants in complex environmental systems. Environ. Modell. Software, 25: 817-817.
15. Coman, A., A. Ionescu and Y. Candau, 2008. Hourly ozone prediction for a 24-h horizon using neural networks. Environ. Modell. Software, 23: 1407-1421.
16. Sousa, S.I.V., F.G. Martins, M.C.M. Alvim-Ferraz and M.C. Pereira, 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Exp. Model. Software, 22: 97-103.
17. Arvanitis, A. and N. Moussiopoulos, 2006. Estimating long term urban exposure to particulate matter and ozone in Europe. Environ. Modell. Software, 21: 447-453.
18. Barakat, H.M., E.M. Nigm and O.M. Khaled, 2014. Statistical modeling of extremes under linear and power normalizations with applications to air pollution. Kuwait J. Sci., 41: 1-19.
19. Barakat, H.M., E.M. Nigm, O.M. Khaled and F.A. Momenkhan, 2015. Bootstrap method for order statistics and modeling study of the air pollution. Commun. Stat. Simulation Comput., 44: 1477-1491.
20. Barakat, H.M., E.M. Nigm and O.M. Khaled, 2012. Statistical modeling of extreme values with applications to air pollution. Life Sci. J., 9: 124-132.