Stationary and Non-Stationary Models of Extreme Ground-Level Ozone in Peninsular Malaysia

Siti Aisyah Zakaria^{1,*}, Nor Azrita Mohd Amin¹, Noor Fadhilah Ahmad Radi², Nasrul Hamidin³

¹Institute of Engineering Mathematics, Universiti Malaysia Perlis, KampusPauh Putra, 02600 Arau, Perlis, Malaysia

²Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

³School of Environmental Engineering, University Malaysia Perlis, 02600 Arau, Perlis, Malaysia

Received November 19, 2020; Revised March 5, 2021; Accepted March 16, 2021

Cite This Paper in the following Citation Styles

(a): [1] Siti Aisyah Zakaria, Nor Azrita Mohd Amin, Noor Fadhilah Ahmad Radi, Nasrul Hamidin, "Stationary and Non-Stationary Models of Extreme Ground-Level Ozone in Peninsular Malaysia," *Mathematics and Statistics*, Vol. 9, No. 3, pp. 357 - 370, 2021. DOI: 10.13189/ms.2021.090318.

(b): Siti Aisyah Zakaria, Nor Azrita Mohd Amin, Noor Fadhilah Ahmad Radi, Nasrul Hamidin (2021). Stationary and Non-Stationary Models of Extreme Ground-Level Ozone in Peninsular Malaysia. *Mathematics and Statistics*, 9(3), 357 - 370. DOI: 10.13189/ms.2021.090318.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract High ground-level ozone (GLO) concentrations will adversely affect human health, vegetations as well as the ecosystem. Therefore, continuous monitoring for GLO trends is a good practice to address issues related to air quality based on high concentrations of GLO. The purpose of this study is to introduce stationary and non-stationary model of extreme GLO. The method is applied to 25 selected stations in Peninsular Malaysia. The maximum daily GLO concentration data over 8 hours from year 2000 to 2016 are used. The factors of this distribution are anticipated using maximum likelihood estimation. A comparison between stationary (constant model) and non-stationary (linear and cyclic model) is performed using the likelihood ratio test (LRT). The LRT is based on the larger value of deviance statistics compared to a chi-square distribution providing the significance evidence to non-stationary model either there is linear trend or cyclic trend. The best fit model between selected models is tested by Akaike's Information Criterion. The results show that 25 stations conform to the non-stationary model either linear or cyclic model, with 14 stations showing significant improvement over the linear model in location parameter while 11 stations follow the cyclic model. This study is important to identify the trends of ozone phenomenon for better quality risk management.

Keywords Akaike's Information Criterion, Ground-Level Ozone, Likelihood Ratio Test, Generalized

Extreme Value Distribution, Maximum Likelihood Estimation

1. Introduction

Extreme ground-level ozone (GLO) in this study refers to the maximum value of GLO concentration in a specific period of time such as yearly, monthly, biweekly or weekly. Higher concentrations will adversely affect human health, indirectly affect the efficiency of photosynthesis in plants as well as other activities in the ecosystem as mentioned in several literatures [1]–[3]. Therefore, continuous monitoring for GLO trends is important to ensure that the level of ozone concentration does not exceed the standard recommendation stated by the National Ambient Air Quality Standard (NAAQS) and Malaysia Ambient Air Quality Standard (MAAQS). MAAQS has recommended that ozone concentration values should not exceed 0.06 ppm for an average of 8 hours.

Several studies have explored the trends of this GLO specifically in Peninsular Malaysia such as in [4]–[7]. Most of the studies found that there are seasonal variations in GLO according to locations which are influenced by the interchange of the two monsoons, namely the Northeast Monsoon and the Southeast Monsoon. However, most

literatures such as in [8]–[11] are focused on the centre of distribution rather than the tail distribution which may contain extreme values. Hence, we continue the analysis on GLO trends by focusing on extreme parts using extreme theory to classify the seasonal pattern based on stationary and non-stationary model. Stationary model will describe the characteristics of GLO data based on the constant value of all parameters. As for non-stationary model, linear and cyclic trends for the location parameter that is dependent on time were considered.

By taking into account the underlying processes driving the GLO generation process, the stationary and non-stationary characteristics of GLO data can be established. GLO is a secondary pollutant, which means that it is produced in the atmosphere, and the chemical reactions involved in ozone synthesis are depending on environmental factors or time. Seasonal trends that occur in the production of GLO are influenced by the increment rate of the reactions in the presence of the sun and human activities. A couple of good examples are, combustion of fossil fuels from the increasing number of vehicles in urban areas as well as emissions from major activities in industrial zones.

This study proposed the use of Generalized Extreme Value (GEV) model for both stationary and non-stationary models. Maximum likelihood estimation is used to evaluate the parameters. For stationary model, all parameters are assumed to be constant. Meanwhile, for non-stationary model, variations through time in the observed process are modelled as linear trends and yearly cyclic trends in the location parameter. The non-stationarity of the model is often apparent because of seasonal effects, or in the form of trends that occurs in environmental processes. This might help the analysis of non-stationary model to obtain the best-fitted model in selected GLO data as the novelty of this study.

The aims for this statistical study of GLO data are to predict the high levels of GLO concentration and give public health awareness, possibly in response to legislations regulating emissions of pollutants, and more importantly to identify trends in high ozone levels. These findings can hopefully increase the current scientific interest in ascertaining effects of human activities on the environment and to assess changes in ozone levels due to activities, either directly or through changing emission patterns. Therefore, the two major aims in this study are to introduce a stationary and non-stationary model of weekly maxima GLO, and to identify the best fitted model to describe extreme GLO data sets in the selected air monitoring stations.

2. Methodology

2.1. Data Description

The method is applied to GLO data sets in Peninsular

Malaysia, which consist of weekly maxima of GLO concentrations from year 2000 to 2016. In 25 stations operated by a private firm known as Alam Sekitar Malaysia Sdn Bhd (ASMA) employed by the Department of Environment (DOE) Malaysia, this datum was measured at the monitoring site. In compliance with Mokhtar et al. [5], Teledyne Ozone Analyzer Model 400E UV Absorption is used by Beer-Lambert Law to measure low-range ozone concentrations. This instrument was used to record ozone levels at all of the monitoring stations shown in Table 1.

Table 1. Location of air monitoring stations in Peninsular Malaysia

Negeri	Location	Code Stations	Category
Johor	Pasir Gudang	WK01	Industrial
Terengganu	Kemaman	WK02	Industrial
Penang	Perai	WK03	Industrial
Melaka	Bukit Rambai	WK06	Industrial
Pahang	Jerantut	WK07	Background
Perak	Tasek	WK08	Industrial
Penang	Seberang Jaya	WK09	Suburban
Negeri Sembilan	Nilai	WK10	Industrial
Selangor	Pelabuhan Kelang	WK11	Urban
Pahang	Kuantan	WK14	Suburban
Kedah	Sungai Petani	WK17	Suburban
Johor	Larkin	WK19	Industrial
Perak	Taiping	WK20	Industrial
Kelantan	Kota Bharu	WK22	Urban
Selangor	Shah Alam	WK25	Urban
Kedah	Langkawi	WK32	Suburban
Perlis	Kangar	WK33	Suburban
Terengganu	Kuala Terengganu	WK34	Urban
Penang	USM	WK38	Suburban
Kedah	AlorSetar	WK40	Urban
Melaka	Bandaraya Melaka	WK43	Urban
Johor	Muar	WK44	Suburban
Perak	Tanjung Malim	WK45	Suburban
Perak	Pegoh	WK46	Urban
Negeri Sembilan	Seremban	WK47	Urban

2.2. Extreme Value Theory (EVT)

Extreme value theory (EVT) explores rare events by determining their statistical model, estimation, and prediction. Currently, this method is quite famous and has been used by researchers to analyse data in many other disciplines. As pointed by Coles[12], EVT is a statistical discipline that produces techniques and models for explaining the rarely high or low levels of amounts in catastrophic data (such as heatwaves or widespread

flooding). There are two methods in the application of EVT which are blocks maxima and threshold. In this study, block maxima method was used as the point of concern is daily exceedance of 0.06 ppm, which is a rare event that can cause severe health effects. The block maxima method has a limiting distribution that corresponds to generalized extreme value (GEV) distribution. GEV distribution is a single family of model with distribution functions based on three different distributions namely Gumbel, Fréchet, and Weibull. These three classes of distributions are recognised as type I, II, and III respectively. The GEV for stationary case that assumes all parameters are constant has the following cumulative density function (cdf):

$$F(x; \mu, \sigma, \xi) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} & \xi \neq 0 \\ \exp \left\{ - \exp \left[- \frac{x-\mu}{\sigma} \right] \right\} & \xi = 0 \end{cases} \quad (1)$$

where x is the extreme value data obtained using block maxima method, σ is a scale parameter, ξ is a shape parameter, and μ is a location parameter. The shape parameter, ξ is important in describing the tail behaviour of extreme data to classify the particular type of distribution. More specifically, when $\xi < 0$, the GEV distribution is Weibull distribution which supports $x \leq \mu - \frac{\sigma}{\xi}$; when $\xi > 0$, it corresponds to Fréchet distribution which supports $x \geq \mu + \frac{\sigma}{\xi}$; and when $\xi = 0$,

the GEV distribution is Gumbel distribution which supports $\square$. The model for stationary case is considered as independent series in all parameters. The stationary GEV model to describe the distribution of x is:

$$X \sim GEV(\mu, \sigma, \xi) \quad (2)$$

In the case of non-stationary model, yearly seasonality of maximum observed values in time increments is incorporated in the model. Referring to Coles [12], non-stationary cases are dependent on seasonal effects, or there are trends in the data due to climate change or any other causes. In this study, the trends in several of the GLO data increases doubts about the suitability of model that is assumed as constant distribution. So, two models are suggested; where location parameter is considered to change linearly over the observation period and can be cyclic function over a period of one year. The suitable model for x_t , the weekly maximum value of extreme GLO in weekly t is

$$x_t \sim GEV(\mu(t), \sigma, \xi) \text{ where } \mu(t) = \beta_0 + \beta_1 t \quad (3)$$

This model is considered as linear trend in the location parameter, μ . The parameter β_0 corresponds to mean value while β_1 is the weekly rate of change in weekly maxima of GLO data. The second model is more complex

with yearly cyclic function. The equation of yearly cyclic function is expressed as

$$u(t) = \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{y}\right) + \beta_2 \cos\left(\frac{2\pi t}{y}\right) \quad (4)$$

where β_0 is mean value, β_1 and β_2 represent the harmonic amplitudes, and y is one year (in days). The idea to propose a new model as yearly cyclic function come mainly from the trend in extreme GLO data that shows strong seasonal effects each year.

2.3. Parameter Estimation

In this study, the maximum likelihood estimation (MLE) is used as parameter estimation. This method is widely used by researchers because MLE has high likelihood corresponding to models which give high probability to observed data. Furthermore, this method also has high adaptability to change model structure based on Phalitnonkiat *et al.* [13]. Hence, for stationary GEV model, the log-likelihood where distributions of $x, x_1, \dots, x_m$ are identical independent distribution (iid) is:

$$l(\mu, \sigma, \xi) = -m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] \xi \neq 0 \quad (5)$$

$$l(\mu, \sigma, \xi) = -m \log \sigma - \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left(- \frac{x_i - \mu}{\sigma} \right) \xi = 0 \quad (6)$$

Meanwhile the log-likelihood for non-stationary GEV model, where the distribution of x_t for $t = 1, \dots, m$ is $x_t \sim GEV(\mu(t), \sigma, \xi)$ given that $\mu(t)$ is expressed as in equation (3) and (4), is:

$$l(\mu, \sigma, \xi) = -m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log \left[1 + \xi \left(\frac{x_i - \mu(t)}{\sigma} \right) \right] - \sum_{i=1}^m \left[1 + \xi \left(\frac{x_i - \mu(t)}{\sigma} \right) \right] \xi \neq 0 \quad (7)$$

$$l(\mu, \sigma, \xi) = -m \log \sigma - \sum_{i=1}^m \left(\frac{x_i - \mu(t)}{\sigma} \right) - \sum_{i=1}^m \exp \left(- \frac{x_i - \mu(t)}{\sigma} \right) \xi = 0 \quad (8)$$

2.4. Model Selection

Akaike's information criterion (AIC) and Likelihood ratio test (LRT) are used to select the best model for all extreme GLO for all selected stations. According to Coles [12], LRT is a simple test to identify the best model compared to other models especially for nested models determined by maximum likelihood estimation. It said that, model M_o is a subset of model M_1 and the deviance statistics is denoted as:

$$D = 2 \{ l_1(M_1) - l_o(M_o) \} \quad (9)$$

where $l_1(M_1)$ and $l_o(M_o)$ are the maximized

log-likelihood under models M_o and M_1 , respectively. A test to evaluate the validity of model M_o versus M_1 is done by comparing the values of D and chi-square (χ_k^2) distribution where k is degree of freedom at the level of significance, $\alpha = 0.05$. If the value of D is greater than χ_k^2 , model M_1 will be suggested as the suitable model for the particular data instead of model M_o . Hence, in this study, this test will be used to select the best model between stationary model as mentioned in (2) and two non-stationary models as mentioned in (3) and (4).

After selecting the best model between stationary and two non-stationary models, the next step involved accuracy test for the best non-stationary model as in equation (3) and (4) using AIC. This method is used to choose the non-stationary model which best describes the extreme GLO data in all stations. The AIC formula is

$$AIC = -2l + 2p \quad (10)$$

where p is the parameter of the model and l is the maximized log-likelihood for the non-stationary model. Thus, between these two non-stationary models, the model with the smallest value of AIC is chosen as the best fitted model for extreme GLO data.

2.5. Model Diagnostics

Model diagnostics are important in order to make certain that the final model adopted is a excellent representation of extreme data. Based on Coles [12], quantile and probability plot are used in the diagnostics for the best fitted model. The dotted data that show near to linear line indicate that the model fit well to the stationary or non-stationary model. The probability plots and quantile plots points for diagnostics of stationary model are shown in point (11) and (12), respectively while for non-stationary models the diagnostics points are as in point (13) and (14), respectively.

$$\left\{ \left(\frac{i}{m+1}, \exp \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \right), i = 1, \dots, m \right\} \quad (11)$$

$$\left\{ \left(\mu - \frac{\sigma}{\xi} \left[1 - \left\{ -\log \left(\frac{i}{m+1} \right) \right\}^{-\xi} \right], x_{(i)} \right), i = 1, \dots, m \right\} \quad (12)$$

$$\left\{ \frac{i}{m+1}, \exp \left(-\exp \left(- \left[\frac{1}{\xi} \log \left\{ 1 + \xi \left(\frac{x_t - \mu(t)}{\sigma} \right) \right\} \right] \right) \right), i = 1, \dots, m \right\} \quad (13)$$

$$\left\{ x_i, -\log \left(-\log \left(\frac{i}{m+1} \right) \right) \right\}; i = 1, \dots, m \quad (14)$$

3. Results and Discussion

Table 2 reviews the descriptive statistics of all selected air monitoring stations in Peninsular Malaysia. The results indicate that the highest reading in weekly maximum GLO concentration is recorded at station 19 (WK19) which is in Larkin, Johor with 0.20 ppm. This station is located in an industrial area with higher sources of GLO concentration. The second highest reading is recorded at station 34 (WK34) which is in Kuala Terengganu with 0.18 ppm. Although 0.18 ppm is indicated as the second highest reading, the average value for this station is quite stable around 0.05 ppm which means that this area is still considered safe from exposure to air pollution caused by ozone. The third highest reading of weekly maximum GLO concentration is 0.17 ppm but it is accompanied by a high average reading of 0.09 ppm. This reading is recorded at Shah Alam station (WK25). The increase or decrease in ozone concentrations are influenced by surrounding activities. Overall, the three highest readings for maximum values of weekly GLO concentrations have exceeded the standard recommendation stated by MAAQS and NAAQS.

Based on Abdullah *et al.* [10], the highest readings of ozone concentrations are located in industrial areas. The finding is in line with the results of this study. Meanwhile, urban areas also have higher GLO concentrations because the activities that might affect ozone concentrations are magnified by the larger numbers of populations and vehicles that indirectly cause higher petrol emissions. Furthermore, the study by Abdullah *et al.* [10], is focused on the average value of GLO concentration whereas this study, is focused on weekly maximum GLO concentration. The lowest reading of weekly maximum GLO concentration is recorded in station 22 (WK22) which is located at Kota Bharu, Kelantan where the total amount of GLO is 0.07 ppm. Figure 1 shows the time series for three air monitoring stations, which can give an indication of whether the data have a pattern such as linear, constant or cyclic trend. The results show that further investigations to indicate the best fitted model are important in this analysis. The models are stated as stationary models (constant model) or non-stationary models of either linear or cyclic model.

Table 2. Descriptive summary of extreme ground-level ozone concentration

Code Stations	Mean	Min	Max	SD	Skewness	Kurtosis
WK01	0.05	0.02	0.12	0.02	0.87	0.83
WK02	0.05	0.02	0.10	0.01	0.05	-0.63
WK03	0.06	0.01	0.12	0.02	-0.12	0.71
WK06	0.06	0.02	0.12	0.01	0.66	1.21
WK07	0.04	0.01	0.09	0.01	0.41	0.08
WK08	0.07	0.01	0.16	0.01	0.42	2.10
WK09	0.06	0.02	0.12	0.02	0.21	0.23
WK10	0.07	0.03	0.15	0.02	0.76	1.07
WK11	0.07	0.02	0.13	0.02	0.56	0.21
WK14	0.05	0.01	0.09	0.01	0.09	0.06
WK17	0.06	0.01	0.12	0.02	0.23	0.03
WK19	0.06	0.02	0.20	0.02	0.91	2.08
WK20	0.05	0.01	0.11	0.01	0.56	1.52
WK22	0.04	0.01	0.07	0.01	0.11	-0.21
WK25	0.09	0.04	0.17	0.03	0.44	0.08
WK32	0.05	0.02	0.09	0.01	0.40	-0.14
WK33	0.05	0.01	0.10	0.01	0.37	0.18
WK34	0.05	0.02	0.18	0.02	2.18	7.84
WK38	0.07	0.02	0.12	0.01	0.18	0.11
WK40	0.05	0.01	0.12	0.02	0.51	0.35
WK43	0.06	0.02	0.15	0.02	0.67	1.65
WK44	0.06	0.01	0.14	0.02	0.23	0.51
WK45	0.07	0.02	0.16	0.02	0.67	-0.29
WK46	0.07	0.01	0.12	0.01	0.08	0.69
WK47	0.07	0.02	0.13	0.02	0.10	-0.29

Figure 1. Time series of weekly maxima ground-level ozone concentration (in ppm) in three selected stations, (a) Kemaman (WK02); (b) Jerantut (WK07); (c) Langkawi (WK32)

3.1. Model Selection

LRT is used to compare the two models to determine the conformity of GLO models to either stationary or non-stationary models. Table 3 indicates that each station conforms to either stationary or non-stationary model or both non-stationary models. For example, station WK01 conforms to stationary (constant model) and non-stationary (cyclic model). Station WK02 conforms to both non-stationary either linear or cyclic model. For stations WK01 and WK19, the values of likelihood ratio test or *Dare* “NA”. This means that they are “not applicable” for this testing because there is no result in linear model.

The model that fits best is chosen depending on the value of D which must be greater than chi-square (χ^2_k) distribution where k is degree of freedom at level of significance, $\alpha = 0.05$. This test is used to compare constant and linear models. The second comparison is between constant and cyclic models. Result from the

comparison shows that nine stations conform to stationary model (constant model) while sixteen stations conform to non-stationary model (linear model). From the second comparison, only two stations fit the stationary model (constant model) while the rest of the stations conform to non-stationary model (cyclic model). In order to determine which model is the best fit for GLO data, further testing using AIC method is necessary.

Based on Table 4, the smallest AIC value indicates the best fitted model. There are fourteen (14) air monitoring stations that fit to non-stationary model (linear model) and the remaining stations fit to non-stationary model (cyclic model). After the most excellent model is selected, diagnostic tests for each model are carried out using graphical method which are quantile and probability plot according to the selection of the model. The result is shown in Figure 2 and 3.

Table 3. Model selection by likelihood ratio test for stationary and two non-stationary models

Code Stations	D	M1 (Constant)	M2 (Linear)	D	M1 (Constant)	M3 (Cyclic)
WK01	NA	√		7.33		√
WK02	12.94		√	10.36		√
WK03	6.77		√	4.36		√
WK06	15.94		√	2.31	√	
WK07	86.60		√	7.13		√
WK08	15.63		√	15.11		√
WK09	25.68		√	8.55		√
WK10	19.48		√	0.97	√	
WK11	1.19	√		20.41		√
WK14	20.70		√	17.30		√
WK17	1.93	√		6.23		√
WK19	NA	√		8.83		√
WK20	3.60	√		9.15		√
WK22	14.06		√	20.99		√
WK25	25.06		√	14.18		√
WK32	3.21	√		4.21		√
WK33	57.30		√	39.61		√
WK34	11.66		√	5.91		√
WK38	1.22	√		36.85		√
WK40	115.51		√	75.25		√
WK43	12.59		√	6.66		√
WK44	45.21		√	40.57		√
WK45	3.35	√		5.56		√
WK46	-1.03	√		43.68		√
WK47	19.61		√	60.41		√

Table 4. Model selection by Akaike information criterion (AIC) for the best fitting model

Code Stations	M1 (Constant)	M2 (Linear)	M3 (Cyclic)
WK01	-4712.702		-4716.030
WK02		-4960.846	-4956.272
WK03		-4791.082	-4786.674
WK06	-5037.608	-5051.550	
WK07		-5329.730	-5248.260
WK08		-5081.222	-5078.700
WK09		-4946.972	-4927.838
WK10	-4572.164	-4589.648	
WK11	-4562.152		-4578.566
WK14		-5434.234	-5428.840
WK17	-4662.734		-4664.962
WK19	-4419.868		-4424.692
WK20	-5317.184		-5322.332
WK22		-5598.286	-5603.222
WK25		-4074.528	-4061.646
WK32	-5277.438		-5276.440
WK33		-5095.608	-5075.924
WK34		-4832.134	-4824.388
WK38	-4948.830		-4981.684
WK40		-4985.052	-4942.786
WK43		-4819.454	-4811.530
WK44		-4815.784	-4809.140
WK45	-4150.200		-4151.762
WK46	-5026.230		-5065.910
WK47		-4522.606	-4561.402

Table 5 divides air monitoring stations in Peninsular Malaysia according to the four regions. Three monitoring stations in the North area conform to non-stationary model with cyclic trend. They are located in Sungai Petani, Kedah (WK17), Langkawi, Kedah (WK32), and USM, Penang (WK38). These findings are in line with literature [7], which states that GLO concentrations in Sungai Petani exhibits a seasonal pattern. This pattern refers to the annual

cycle that is consistent with the main GLO concentration occurring in July, and the lowest GLO concentration in December. Another study, as found in [14], also states that several stations show seasonal patterns that contribute to their conformity to cyclic model. One of this station is in Kota Bharu, Kelantan (WK22). The remaining models based on the four regions can be found in Table 5.

Table 5. The twenty-five stations listed according to the four regions

Code Stations	Stations	Non-stationary Model
<i>North</i>		
WK17	Sek. Keb. Bakar Arang, Sungai Petani	Cyclic
WK32	KomplekSukan Langkawi, Kedah	Linear
WK38	Universiti Sains Malaysia, Pulau Pinang	Cyclic
WK03	Sek. Keb. Cederawasih, Taman Inderawasih, Perai	Linear
WK09	Sek. Keb. Seberang Jaya II, Perai	Linear
WK33	Institut Latihan Perindustrian (ILP) Kangar	Linear
WK40	Sek. Men. Agama Mergong, AlorSetar	Linear
<i>West</i>		
WK11	Sek. Men. (P) Raja Zarina, Kelang	Cyclic
WK20	Sek. Men. Keb. Air Puteh, Taiping	Cyclic
WK45	Universiti Pendidikan Sultan Idris, Tanjung Malim	Cyclic
WK46	Sek. Men. Pegoh, Ipoh, Perak	Cyclic
WK06	Sek. Men. Keb. Bukit Rambai	Linear
WK08	Sek. Men. Jalan Tasek, Ipoh	Linear
WK25	Sek. Keb. TTDI Jaya, Shah Alam	Linear
WK43	Sek. Men. Tinggi Melaka, Melaka	Linear
<i>South</i>		
WK01	Sek. Men. Pasir Gudang 2, Pasir Gudang	Cyclic
WK19	InstitutPerguruan Malaysia, Temenggong Ibrahim, Larkin, Johor Bharu	Cyclic
WK47	Sek. Men. Teknik Tuanku Jaafar, Ampangan, Seremban	Cyclic
WK10	Tmn. Semarak (Phase II), Nilai	Linear
WK44	Sek. Men. Teknik Muar, Muar, Johor	Linear
<i>East</i>		
WK07	Pej. Kajicuaca Batu Embun, Jerantut	Cyclic
WK22	Sek. Men. Keb. Tanjung Chat, Kota Bahru	Cyclic
WK02	Sek. Ren. Keb. Bukit Kuang, TelukKalung, Kemaman	Linear
WK14	Sek. Keb. InderaMahkota, Kuantan	Linear
WK34	Sek. Keb. Chabang Tiga, Kuala Terengganu	Linear

Table 6 and 7 show the results of parameter estimates with standard error in parentheses of linear model and cyclic model. Here, parameter estimation value for the best fitted model is obtained from Table 4. The parameter estimation can be substituted with the yearly cyclic function model as in equation (3) and (4) for prediction purposes. The parameter β_1 in the linear model as in Table 6 is the weekly rate of change in weekly maximum GLO concentration. By referring to the value of shape (ξ) parameter in the GEV model in Table 6, negative values indicates that the tail behaviour is Weibull distributions, except in one station WK34 with $\xi \sim 0$ which is Gumbel distribution. This result is also similar to findings in Table

7 which show the values of $\xi < 0$ corresponding to Weibull distribution for all air monitoring stations except one station WK45 which shows Gumbel distribution with the value $\xi \sim 0$. Hence, an example of non-stationary model with linear trends on the location parameter for station WK02 is shown in equation (16). While, non-stationary model with yearly cyclic function on location parameter for station WK01 is mentioned in equation (17).

$$u(t) = 0.045712 + 0.000007t \quad (16)$$

$$u(t) = 0.0447 - 0.0005 \sin\left(\frac{2\pi t}{y}\right) + 0.0006 \cos\left(\frac{2\pi t}{y}\right) \quad (17)$$

Table 6. Parameter estimates and standard errors (in parentheses) of linear models for μ in GEV model applied to GLO data

Code Stations	β_0	β_1	σ	ξ
	MLE (SE)	MLE (SE)	MLE (SE)	MLE (SE)
WK02	0.045712(0.000488)	0.000007(0.000000)	0.014679(0.000002)	-0.272664(0.000092)
WK03	0.052652(0.000000)	0.000005(0.000000)	0.016358(0.000000)	-0.247277(0.000000)
WK06	0.050901(0.000448)	0.000007(0.000000)	0.012888(0.000280)	-0.116909(0.001977)
WK08	0.069890(0.000454)	-0.000007(0.000000)	0.013104(0.000002)	-0.125263(0.000054)
WK09	0.049981(0.000496)	0.000010(0.000000)	0.014555(0.000002)	-0.215518(0.000086)
WK10	0.067207(0.000571)	-0.000010(0.000000)	0.016432(0.000374)	-0.095016(0.002101)
WK14	0.043877(0.000000)	0.000007(0.000000)	0.011164(0.000000)	-0.239500(0.000000)
WK25	0.089467(0.000792)	-0.000015(0.000000)	0.022795(0.000480)	-0.166446(0.002938)
WK32	0.041721(0.000397)	-0.000003(0.000000)	0.011446(0.000245)	-0.150616(0.002112)
WK33	0.046487(0.000446)	-0.000013(0.000000)	0.012876(0.000268)	-0.171316(0.002271)
WK34	0.036697(0.000441)	0.000006(0.000000)	0.012810(0.000348)	0.101173(0.000000)
WK40	0.052130(0.000462)	-0.000019(0.000000)	0.013314(0.000293)	-0.114565(0.002001)
WK43	0.060560(0.000511)	-0.000007(0.000000)	0.014726(0.000315)	-0.111593(0.002002)
WK44	0.056903(0.000522)	-0.000014(0.000000)	0.015050(0.000002)	-0.141686(0.000071)

Table 7. Parameter estimates and standard errors (in parentheses) of cyclic models for μ in GEV model applied to GLO data

Code Stations	β_0	β_1	β_2	σ	ξ
	MLE(SE)	MLE(SE)	MLE(SE)	MLE(SE)	MLE(SE)
WK01	0.0447(0.0006)	-0.0005(0.0007)	0.0006(0.0007)	0.0146(0.0004)	-0.0204(0.0254)
WK07	0.0347(0.0004)	-0.0004(0.0006)	-0.0001(0.0006)	0.0117(0.0003)	-0.1459(0.0213)
WK11	0.0592(0.0006)	-0.0006(0.0008)	-0.0010(0.0008)	0.0168(0.0004)	-0.1165(0.0229)
WK17	0.0565(0.0006)	0.0002(0.0008)	-0.0007(0.0008)	0.0168(0.0004)	-0.2056(0.0186)
WK19	0.0511(0.0007)	0.0002(0.0008)	0.0007(0.0008)	0.0175(0.0005)	-0.0411(0.0222)
WK20	0.0467(0.0004)	-0.0008(0.0006)	-0.0004(0.0006)	0.0113(0.0003)	-0.1342(0.0128)
WK22	0.0343(0.0004)	-0.0004(0.0005)	0.0003(0.0005)	0.0102(0.0002)	-0.2361(0.0198)
WK38	0.0601(0.0005)	0.0005(0.0007)	0.0000(0.0007)	0.0144(0.0000)	-0.2065(0.0133)
WK45	0.0571(0.0008)	0.0001(0.0009)	0.0003(0.0009)	0.0194(0.0006)	0.0298(0.0329)
WK46	0.0602(0.0005)	-0.0006(0.0007)	0.0000(0.0007)	0.0140(0.0000)	-0.2024(0.0011)
WK47	0.0622(0.0007)	0.0006(0.0009)	0.0011(0.0009)	0.0187(0.0000)	-0.2326(0.0131)

3.2. Model Diagnostics

Based on Coles [12], probability plot and quantile plot are important to evaluate the accuracy of the GEV model to ensure that the selected model is best fitted to the data. Each set of plotted points for both plotting's must be near to the line, also called near-linear. The results for probability plots and quantile plots for non-stationary models (linear and cyclic model) are shown in Figure 2 and 3. The plots are constructed by non-stationary parameter values to the corresponding generalized extreme value distribution. In the case of the GEV model, this research is

performed, so the probability plot is invariant to the choice of Gumbel as a reference distribution and the quantile plot is unique to the Gumbel scale choice. Figure 2 displays the residual diagnostic plots for 14 air monitoring stations in linear pattern GEV models, pointing to the validity of the equipped model with perfect plotting points close to the linear axis. Furthermore, the same goes for cyclic trend with 11 air monitoring stations exhibiting perfect plotting points near to the linear line to prove the validity of the fitted model in Figure 3. Finally, this means that all of the model conforms to GEV model.

Figure 2. Probability plots and quantile plots for GEV model (with linear trend) of weekly maxima ground-level ozone concentrations in fourteen air monitoring stations.

Figure 3. Probability plots and quantile plots for GEV model (with cyclic trend) of weekly maxima ground-level ozone concentrations in eleven air monitoring stations

4. Conclusions

The weekly maxima ground-level ozone concentrations in all of the 25 selected air monitoring stations in Peninsular Malaysia conform to both stationary and non-stationary GEV model. However, based on the results of goodness of fit test, non-stationary GEV models are the best fitted model. These non-stationary GEV models allow the location parameters to be expressed in linear trend with time and yearly cyclic trend. Hence, this study shows some improvement from stationary to non-stationary model by allowing the linear trend and inclusion of cyclic trend in one year cycle time. Results also show that Weibull is identified as the significant distribution for weekly maxima GLO concentrations except for two stations as mentioned in results and discussion.

This study is focused on GLO series to determine the best fitted model using non-stationary and stationary models. In view of the important up and down pattern of extreme GLO concentrations specifically in Peninsular Malaysia, it is important to consider the non-stationary nature of GLO data to achieve better estimations of GLO distribution so that it can contribute towards in-depth understanding of the trends of ozone phenomenon. Lastly, for further extension of the current work, other models selection methods and different classifications of data should be considered.

Acknowledgements

We acknowledge the Department of Environment (DoE), Malaysia, for supplying the ground-level ozone data specifically in Peninsular Malaysia, to carry out this study. This study also supported by FRGS grant with code FRGS/1/2019/STG06/UNIMAP/02/4.

REFERENCES

- [1] S. Sharma, P. Sharma, M. Khare, and S. Kwatra, "Statistical behavior of ozone in urban environment," *Sustain. Environ. Res.*, vol. 26, no. 3, pp. 142–148, 2016, doi: 10.1016/j.serj.2016.04.006.
- [2] A. Tobias and M. G. Scotto, "Prediction of extreme ozone levels in Barcelona, Spain," *Environ. Monit. Assess.*, vol. 100, no. 1–3, pp. 23–32, 2005, doi: 10.1007/s10661-005-7057-4.
- [3] I. M. H. Noor, N. M. Noor, and S. Y. Yusof, "Temporal Characterisation of Ground-level Ozone Concentration in Klang Valley," *E3S Web Conf.*, vol. 34, no. April, 2018, doi: 10.1051/e3sconf/20183402047.
- [4] M. T. Latif, E. Z. Abidin, and S. M. Praveena, "The Assessment of Ambient Air Pollution Trend in Klang Valley," *World Environ.*, vol. 5, no. 1, pp. 1–11, 2015, doi: 10.5923/j.env.20150501.01.

- [5] M. I. Z. Mokhtar, N. A. Ghazali, M. Y. Nasir, and N. Suhaimi, "Seasonal variation and statistical distribution of surface ozone concentration in Malaysia: A case study in TanjongMalim," *Adv. Environ. Biol.*, vol. 9, no. 18, pp. 120–131, 2015.
- [6] J. Sentian, F. Herman, C. Y. I. T. Yih, J. Chang, and H. Wui, "Long-term Air Pollution Trend Analysis in Malaysia," vol. 2, no. 4, pp. 309–324, 2019, doi: 10.2495/EI-V2-N4-309-324.
- [7] M. Ismail, "Time-series analysis of ground-level ozone in Muda Irrigation Scheme Area (MADA), Kedah," *J. Sustain. Sci. Manag.*, 2011.
- [8] N. N. L. Mohd Napi, S. Abdullah, A. N. Ahmed, A. A. Mansor, and M. Ismail, "Annual and Diurnal Trend of Surface Ozone (O₃) in Industrial Area," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 498, no. 1, 2020, doi: 10.1088/1755-1315/498/1/012062.
- [9] N. Mohamed Noor, N. I. Mohamad Hasim, and S. Y. Yusof, "Variation of Ground-Level Ozone Concentration in Urbanized Area in Malaysia," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 374, no. 1, 2018, doi: 10.1088/1757-899X/374/1/012097.
- [10] N. Mohamed Noor, N. I. Mohamad Hasim, and S. Y. Yusof, "Variation of Ground-Level Ozone Concentration in Urbanized Area in Malaysia," 2018, doi: 10.1088/1757-899X/374/1/012097.
- [11] M. Y. Nasir, N. A. Ghazali, M. Izwan, Z. Mokhtar, and N. Suhaimi, "Malaysian Journal of Analytical Sciences Fitting Statistical Distribution Functions on Ozone Concentration Data at Coastal Areas (Penyesuaian Fungsi Taburan Statistik pada Data Kepekatan Ozon di Kawasan Pesisiran Pantai)," *Malaysian J. Anal. Sci.*, vol. 20, no. 3, pp. 551–559, 2016, doi: 10.17576/mjas-2016-2003-13.
- [12] S. Coles, *An introduction to statistical modeling of extreme values*. Springer-Verlag, London, 2001.
- [13] P. Phalitnonkiat, W. Sun, M. D. Grigoriu, P. Hess, and G. Samorodnitsky, "Extreme ozone events: Tail behavior of the surface ozone distribution over the U.S.," *Atmos. Environ.*, vol. 128, no. X, pp. 134–146, 2016, doi: 10.1016/j.atmosenv.2015.12.047.
- [14] J. M. Rajab, H. S. Lim, and M. Z. Matjafri, "Monthly distribution of diurnal total column ozone based on the 2011 satellite data in Peninsular Malaysia," *Egypt. J. Remote Sens. Sp. Sci.*, 2013, doi: 10.1016/j.ejrs.2013.04.003.
- [15] Shanmugasundaram, N., Sridharan, S., Swapna, S. "Mitigation and prediction of radiation effects in solar power plants using neuro fuzzy and neural network" *European Journal of Molecular and Clinical Medicine*, 2020, 7(8), pp. 2500–2509,
- [16] Shanmugasundaram, N., Sushita, K., Kumar, S.P., Ganesh, E.N. "Genetic algorithm-based road network design for optimising the vehicle travel distance" *International Journal of Vehicle Information and Communication Systems*, 2019, 4(4), pp. 355–374.