Sains Malaysiana 44(2)(2015): 175–185

PM_{10} Analysis for Three Industrialized Areas using Extreme Value

(Analisis PM_{10} bagi Tiga Kawasan Industri menggunakan Nilai Melampau)

HASFAZILAH AHMAT*, AHMAD SHUKRI YAHAYA & NOR AZAM RAMLI

ABSTRACT

One of the concerns of the air pollution studies is to compute the concentrations of one or more pollutants' species in space and time in relation to the independent variables, for instance emissions into the atmosphere, meteorological factors and parameters. One of the most significant statistical disciplines developed for the applied sciences and many other disciplines for the last few decades is the extreme value theory (EVT). This study assesses the use of extreme value distributions of the two-parameter Gumbel, two and three-parameter Weibull, Generalized Extreme Value (GEV) and two and three-parameter Generalized Pareto Distribution (GPD) on the maximum concentration of daily PM_{10} data recorded in the year 2010 - 2012 in Pasir Gudang, Johor; Bukit Rambai, Melaka; and Nilai, Negeri Sembilan. Parameters for all distributions are estimated using the Method of Moments (MOM) and Maximum Likelihood Estimator (MLE). Six performance indicators namely; the accuracy measures which include predictive accuracy (PA), coefficient of determination (R^2), Index of Agreement (IA) and error measures that consist of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Normalized Absolute Error (NAE) are used to find the goodness-of-fit of the distribution. The best distribution is selected based on the highest accuracy measures and the smallest error measures. The results showed that the GEV is the best fit for daily maximum concentration for PM_{10} for all monitoring stations. The analysis also demonstrates that the estimated numbers of days in which the concentration of PM_{10} exceeded the Malaysian Ambient Air Quality Guidelines (MAAQG) of 150 mg/m^3 are between ½ and 1½ days.

Keywords: Air pollution; extreme value theory (EVT); PM_{10}; prediction

ABSTRAK

Salah satu kebimbangan di dalam kajian pencemaran udara adalah untuk menyukat kepekatan satu atau lebih zarah pencemar di dalam ruang dan masa berhubung dengan pemboleh ubah bebas, sebagai contoh untuk pelepasan ke atmosfera, faktor dan parameter cuaca. Salah satu disiplin statistik yang paling penting untuk sains gunaan dan pelbagai bidang lain untuk beberapa dekad yang lalu adalah Teori Nilai Melampau (EVT). Kajian ini menilai penggunaan taburan nilai melampau dua parameter Gumbel, dua dan tiga parameter Weibull, Nilai Melampau Teritlak (GEV) dan dua dan tiga parameter Taburan Pareto Teritlak (GPD) pada kepekatan maksimum data harian PM_{10} yang dicatatkan dalam tahun 2010 - 2012 di Pasir Gudang, Johor; Bukit Rambai, Melaka dan Nilai, Negeri Sembilan. Parameter untuk semua taburan dianggarkan menggunakan Kaedah Momen (MOM) dan Penganggar Kebolehjadian Maksimum (MLE). Enam petunjuk prestasi iaitu; pengukuran kejituan termasuk Ketepatan Peramalan (PA), Pekali Penentuan (R^2), Indeks Persetujuan (IA) dan pengukuran ralat yang terdiri daripada Ralat Min Punca Kuasa Dua (RMSE), Min Ralat Mutlak (MAE) dan Ralat Mutlak Ternormal (NAE) digunakan untuk mencari kebaikan penyesuaian taburan. Taburan terbaik dipilih berdasarkan pengukuran kejituan tertinggi dan pengukuran ralat yang terkecil. Hasil kajian menunjukkan bahawa GEV adalah taburan terbaik untuk kepekatan maksimum harian bagi PM_{10} di kesemua stesen pemantauan. Analisis juga menunjukkan bahawa anggaran bilangan hari kepekatan PM_{10} melebihi Garis Panduan Kualiti Udara Ambien Malaysia (MAAQG) bagi kepekatan harian PM_{10} iaitu 150 µg/m^3 adalah antara ½ dan 1½ hari .

Kata kunci: Nilai melampau (EVT); pencemaran udara; peramalan teori; PM_1

INTRODUCTION

Air pollution is the common term in reference to the existence of air pollutants in the form of gaseous, liquid or fine particles suspended in air. Previous studies showed that it can generate damaging effects to human health, crops and environment (Jamal et al. 2004). In view of this, the Department of Environment, Malaysia consistently observes the ambient air quality in Malaysia through the continuous monitoring at 52 stations throughout Malaysia (Department of Environment Malaysia 2013). All the strategically located monitoring stations in urban, sub-urban and industrial areas would record any significant changes in air quality that might have detrimental effects on the global environment, crops and human health. The ozone (O3), lead (Pb), carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter of aerodynamic diameter of less than 10 micrometer (PM_{10}) are the six criteria pollutants which are monitored closely by the Department of Environment, Malaysia. These may cause inconvenience to the respiratory system and nervous systems of human being and damages to the vegetation (Jamal et al. 2004). Monitoring data and studies on ambient air quality show that some of the air pollutants in a number of big cities in Malaysia are increasing with time (Afroz et al. 2003; Talib et al. 2002).

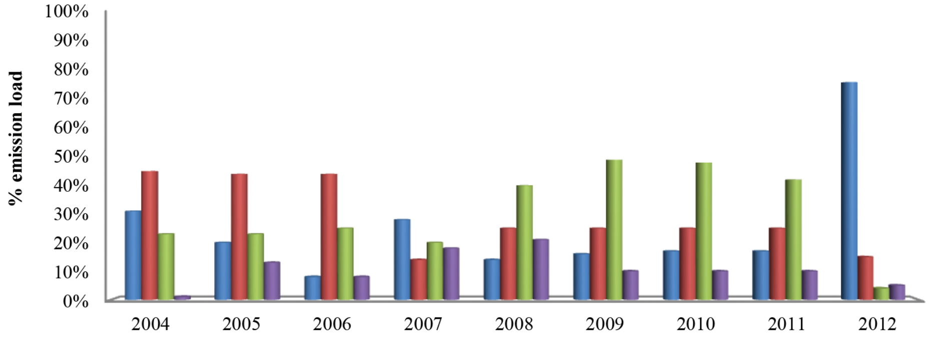
Air pollutants are released to the ambient air by two major sources namely, natural sources and anthropogenic activities. Forest fires and windblown dust are among the natural sources that release pollutants to the air (Department of Environment Malaysia 2013). Stationary and mobile sources are the contributors to anthropogenic activities. Among the major local sources of air pollutions in Malaysia are the increasing numbers of motor vehicle usage (mobile sources) and rapid development of industrial sector (stationary sources). As published in the Environmental Quality Report 2012, the highest contributors of stationary sources were Johor (21.5%) followed by Selangor (18.7%) and Sarawak (11.2%) (Malaysia Environmental Quality Report 2013).

For the past few years, power and industrial plants have been the major source of particulate matter in Malaysia. However, it is significant to note that the trend of main contributor of particulate matter has shifted from the combination of industrial and power plants to a single factor of motor vehicles (76%) in 2012 as shown in Figure 1.

The overall trend of PM_{10} in certain areas in Malaysia had occasionally exceeded the Malaysian Ambient Air Quality Guidelines of 50 $\mu g/m^3$ for PM_{10} due to trans-boundary forest smoke throughout the dry period of May to September (Department of Environment Malaysia 2013).

One of the concerns of the air pollution studies is to compute the concentrations of one or more pollutants' species in space and time in relation to the independent variables, for instance emissions into the atmosphere, meteorological factors and parameters. One of the most significant statistical disciplines developed for the applied sciences and many other disciplines for the last few decades is the extreme value theory. The most important feature of this analysis tool is to compute the unusual or rare (extremes) events such as the minimum or the maximum concentrations, exceedances or frequencies of the data (Coles 2001). Various studies in different fields have been published for the last couple of years in the applications of the extreme value theory, for example operational risk management (Yao et al. 2013), volatile organic compound exposures (Su et al. 2012), future markets (Kao & Lin 2010), calculation of capital requirement (Tsai & Chen 2011), wind speed (Reynolds 2012; Torrielli et al. 2013), wave heights (Petrov et al. 2013) and storm (Reeve et al. 2012). Studies involving natural phenomena such as rainfall, floods, wind speed air pollution, the height of sea waves and corrosion have been of great interest to researchers and scientists for a long period of time (Kotz & Nadarajah 2000; Surman et al 1987).

A widely used method for assessing and estimating the concentrations of air pollution is the extreme value distribution (EVD) (Dasgupta & Bhaumik 1995; Horowitz & Barakat 1979; Kuchenhoff & Thamerus 1996; Lu 2002; Lu & Fang 2003; Quintela-del-Río & Francisco-Fernández 2011; Reyes et al. 2010; Roberts 1979; Smith 1989; Surman et al. 1987). In view of the fact that it can generate damaging effects to human health, crops and environment, therefore, this study is carried out to attain the best model to predict PM_{10} concentration level in Pasir Gudang, Johor; Bukit Rambai, Melaka; and Nilai, Negeri Sembilan which are all located in the Southern region of west coast Malaysia. This study uses six EVDs to fit the distribution of PM_{10}. Parameters for all distributions are estimated using the method of maximum likelihood estimator (MLE).



	2004	2005	2006	2007	2008	2009	2010	2011	2012
🟦 Motor vehicles	31%	20%	8%	28%	14%	16%	17%	17%	76%
🟥 Power plants	45%	44%	44%	14%	25%	25%	25%	25%	15%
🟩 Industrial	23%	23%	25%	20%	40%	49%	48%	42%	4%
🟪 Others	1%	13%	8%	18%	21%	10%	10%	10%	5%

FIGURE 1. Particulate matter (PM) emission load by sources (in percentage), 2004-2012 (Source: Malaysia Environmental Quality Report 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005 & 2004)

MATERIALS AND METHODS

STUDY AREA

The daily maximum data of PM_{10} from January 2010 to December 2012 was furnished by the Department of Environment, Malaysia. The data was collected through a continuous monitoring by Alam Sekitar Sdn. Bhd. (ASMA) from three monitoring stations in the Southern region of west coast Malaysia. The three monitoring stations - Pasir Gudang, Johor; Bukit Rambai, Melaka; and Nilai, Negeri Sembilan are classified under industrial by the Department of Environment, Malaysia (Figure 2) (Department of Environment Malaysia 2013).

All three Pasir Gudang, Bukit Rambai and Nilai monitoring stations are situated at Sek. Men. Pasir Gudang 2, Pasir Gudang, Johor (N01°28.225, E103°53.637); Bukit Rambai, Melaka (N02°15.924, E102°10.554); and Tmn. Semarak (Phase II), Nilai, Negeri Sembilan (N02°49.246, E101°48.877), respectively. Geographically, all the monitoring stations are strategically located in the rapid growth industrial areas resulting in a large amount of air pollution (Lee et al. 2012; Mohamed Noor et al. 2011; Yap & Hashim 2013). In addition, the Southern part of Peninsular Malaysia is prone to the trans-boundary smoke due to forest fires from the Sumatera regions which contributed to the higher PM_{10} concentrations. It general, the air quality in the southern region of Malaysia was in between of good and moderate except for a few of unhealthy days recorded in 2010 - 2012 (Department of Environment Malaysia 2013, 2012).

The analysis of data with the absence of missing values is completed using a programming language for numerical computation, visualization, and programming package for engineers called MATLAB® (Chapman 2004).

PROBABILITY DISTRIBUTION AND PARAMETER ESTIMATORS

This research undertakes the analysis of PM_{10} data using the extreme value distributions, namely: Gumbel (Kotz & Nadarajah 2000), two and three-parameter Weibull (Rinne 2008), Generalized Extreme Value (GEV) (Martins & Stedinger 2000) and two and three-parameter Generalized Pareto Distribution (GPD) (Abd-el-hakim & Sultan 2004; Singh & Guo 1995). All the parameters of the distributions are estimated using the Method of Moments (MOM) and Maximum Likelihood Estimator (MLE). Table 1 shows the probability density function of the EVD and the parameter estimators of each EVD.

PERFORMANCE INDICATORS

This study uses six performance indicators to select the best distribution to represent the data. The accuracy

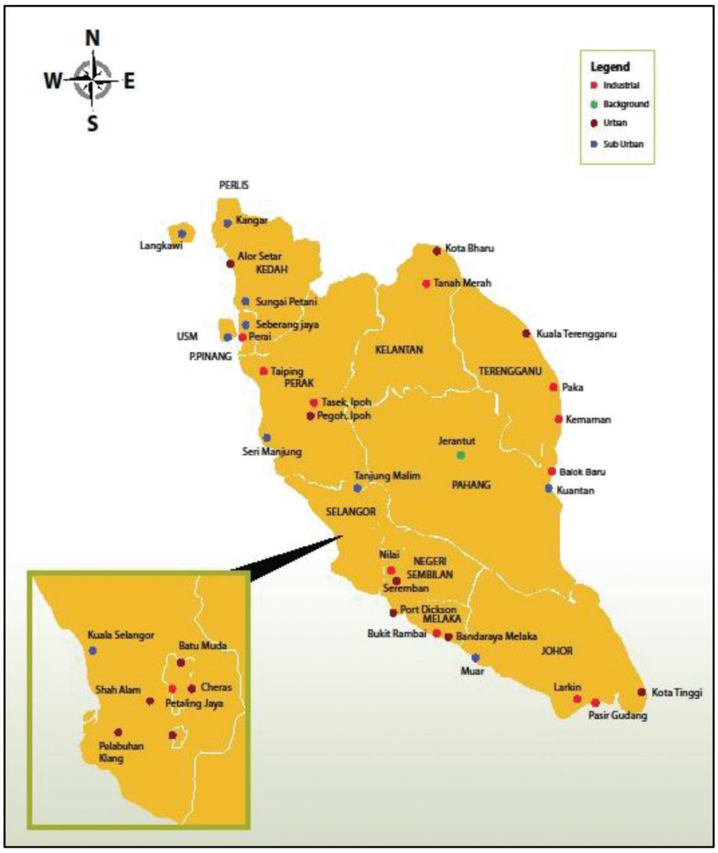


FIGURE 2. Location of continuous air quality monitoring stations in Malaysia
(Source: Malaysia Environmental Quality Report 2013)

TABLE 1. Probability density function (PDF) and its parameter estimators

Dist.	Probability density function	Parameter estimator	
		MOM	MLE
2-Gumbel	$f(x;\mu,\sigma)=\frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma}-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]$	$\mu=\overline{x}-0.45006s$ $\sigma=0.77970s$ (Kotz & Nadarajah 2000)	$\sigma=\overline{x}-\frac{\sum_{i=1}^{n}x_1\exp(-x_i/\sigma)}{\sum_{i=1}^{n}\exp(-x_i/\sigma)},\mu=-\sigma\ln\left(\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\frac{x_i}{\sigma}\right)\right)$
2-Weibull	$f(x;\sigma,\lambda)=\frac{\lambda}{\sigma}\left(\frac{x}{\sigma}\right)^{\lambda-1}\exp\left[-\left(\frac{x}{\sigma}\right)^{\lambda}\right]$	$\lambda=cv^{-1.0852},cv=\frac{s}{\overline{x}},\sigma=\frac{\overline{x}}{\Gamma\left(1+\frac{1}{\lambda}\right)}$ (Bury 1999)	$\sigma=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\lambda}\right]^{1/\lambda},\frac{1}{\lambda}-\frac{\sum_{i=1}^{n}x_i^{\lambda}\ln x_i}{\sum_{i=1}^{n}x_i^{\lambda}}+\frac{1}{4}\sum_{i=1}^{n}\ln x_i=0$
3-Weibull	$f(x;\lambda,\sigma,\mu)=\frac{\lambda}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\lambda-1}\exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{\lambda}\right]$	$\lambda=cv^{-1.0852},\sigma=\frac{\overline{x}}{\Gamma\left(1+\frac{1}{\lambda}\right)}$ $\frac{s}{\overline{x}}=\sqrt{\frac{\Gamma\left(1+\frac{2}{\lambda}\right)-\Gamma^2\left(1+\frac{1}{\lambda}\right)}{\frac{\mu}{\sigma}+\Gamma\left(1+\frac{1}{\lambda}\right)}}$ for $\lambda>2$ (Bury 1999)	$\sigma=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^{\lambda}\right]^{1/\lambda}$, $\frac{1}{\lambda}-\frac{\sum_{i=1}^{n}(x_i-\mu)^{\lambda}\ln(x_i-\mu)}{\sum_{i=1}^{n}(x_i-\mu)}+\frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\mu)=0$ $\frac{\lambda-1}{\lambda}\sum_{i=1}^{n}(x_i-\mu)^{-1}-n\frac{\sum_{i=1}^{n}(x_i-\mu)^{\lambda-1}}{\sum_{i=1}^{n}(x_i-\mu)^{\lambda}}=0$
GEV	$f(x)=\frac{1}{\sigma}\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)^{-1/\lambda-1}\right]$ $.\exp\left\{-\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)^{-1/\lambda}\right]\right\}$	$\mu=\overline{x}-\frac{\sigma}{\lambda}\left[1-\Gamma(1+\lambda)\right]$ $\sigma=\frac{s\lvert\lambda\rvert}{\left\{\Gamma(1+2\lambda)-\left[\Gamma(1+\lambda)\right]^2\right\}^{3/2}}$ $C_s=sign(\lambda)$ $\frac{-\Gamma(1+3\lambda)+3\Gamma(1+\lambda)\Gamma(1+2\lambda)-2\left[\Gamma(1+\lambda)\right]^3}{\left\{\Gamma(1+2\lambda)-\left[\Gamma(1+\lambda)\right]^2\right\}^{3/2}}$ (Martins & Stedinger 2000)	$\frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{1-\lambda-\left(1-(\lambda/\sigma)(x_i-\mu)\right)^{1/\lambda}}{\left(1-(\lambda/\sigma)(x_i-\mu)\right)}\right)=0$ $-\frac{n}{\sigma}+\frac{1}{\sigma}\sum_{i=1}^{n}\left[\frac{1-\lambda-\left(1-(\lambda/\sigma)(x_i-\mu)\right)^{1/\lambda}}{\left(1-(\lambda/\sigma)(x_i-\mu)\right)}\left(\frac{(x_i-\mu)}{\sigma}\right)\right]=0$ $-\frac{1}{\lambda^2}\sum_{i=1}^{n}\left[\ln\left(1-(\lambda/\sigma)(x_i-\mu)\right)\left\{1-\lambda-\left[1-(\lambda/\sigma)(x_i-\mu)\right]^{1/\lambda}\right\}\right.$ $\left.+\frac{1-\lambda-\left[1-(\lambda/\sigma)(x_i-\mu)\right]^{1/\lambda}}{\left(1-(\lambda/\sigma)(x_i-\mu)\right)}\lambda\left(\frac{x_i-\mu}{\sigma}\right)\right]=0$
2-GPD	$f(x;\lambda,\sigma)=\frac{1}{\sigma}\left[1-\lambda\left(\frac{x}{\sigma}\right)\right]^{1/\lambda-1}$	$\lambda=\frac{1}{2}\left(\frac{\overline{x}^2}{s^2}-1\right)\ \sigma=\frac{1}{2}\overline{x}\left(\frac{\overline{x}^2}{s^2}+1\right)$ (Bermudez & Kotz 2010)	$\sum_{i=1}^{n}\frac{x_j/\sigma}{1-\lambda(x_i)/\sigma}=\frac{n}{1-\lambda},\sum_{i=1}^{n}\ln\left(1-x(x_i)/\sigma\right)=-n\lambda$
3-GPD	$f(x;\lambda,\sigma,\mu)=\frac{1}{\sigma}\left[1-\lambda\left(\frac{x-\mu}{\sigma}\right)\right]^{1/\lambda-1}$	$\mu=\overline{x}-\frac{\sigma}{(1+\lambda)}\ \sigma=s(1+\lambda)(1+2\lambda)^{0.5}$, $C_s=\frac{2(1-\lambda)(1+2\lambda)^{1/2}}{1+3\lambda}$ (Oztekin 2005)	$\mu=x_1,\ \sigma=\frac{\lambda}{\exp(n\lambda)-1}(x_n-\mu)$ $\sum_{i=1}^{n}\left(\exp(n\lambda)+\frac{x_n-x_i}{x_i-\mu}\right)^{-1}=\frac{n}{\exp(n\lambda)-1}-\frac{1}{\lambda\exp(n\lambda)}$

Notation: μ is the location parameter, s is the scale parameter and λ is the shape parameter. $\overline{x}$ = sample mean, s = standard deviation, C_s = skewness

measures are the Prediction Accuracy (PA), Coefficient of Determination (R^2) and Index of Agreement (IA). The accuracy value is between 0 and 1 and as the value approaches 1, the model is appropriate. On the other hand, as the value of error measures approaching 0, the model is deemed to be the best model. The error measures used in this study are the Root Mean Squared Error (RMSE), the Normalized Absolute Error (NAE) and the Mean Absolute Error (MAE) (Junninen et al. 2004; Lu and Fang 2003; Yahaya and Ramli 2008). Table 2 lists the performance indicators and their formulae used in this study.

DATA

Table 3 describes the descriptive statistics of PM_{10} concentration for the monitoring stations. The unit of measurement is microgram per cubic meter (µg/m^3). All the three average readings of the PM_{10} concentrations were slightly above the stipulated Malaysian Ambient Air Quality Guidelines (MAAQG) for the yearly average of 50 µg/m^3 (Department of Environment Malaysia 2011) with the average for Bukit Rambai was a little above the other stations' averages. All the data from the three stations were skewed to the right - above 1, an indication of the existence of the extreme concentrations during 2010 – 2012.

TABLE 2. Performance indicators

Indicators	Equations
Prediction accuracy	$PA=\sum_{t=1}^{n}\frac{\left(P_t-\bar{P}\right)\left(O_t-\bar{O}\right)}{(n-1)\sigma_p\sigma_o}$
Coefficient of determination	$1-\frac{\sum_{t=1}^{n}\left(O_t-P_t\right)^2}{\sum_{t=1}^{n}\left(O_t-\bar{O}\right)^2}$
Index of accuracy	$1-\frac{\sum_{t=1}^{n}\left(P_t-O_t\right)^2}{\sum_{t=1}^{n}\left(\lvert P_t-\bar{O}\rvert-\lvert O_t-\bar{O}\rvert\right)^2}$
Root mean square error	$\sqrt{\frac{\sum_{t=1}^{n}\left(O_t-P_t\right)^2}{n}}$
Normalized absolute error	$\frac{\sum_{t=1}^{n}\left(P_t-O_t\right)}{\sum_{t=1}^{n}O_t}$
Mean absolute error	$\frac{\sum_{t=1}^{n}\lvert\left(O_t-P_t\right)\rvert}{n}$

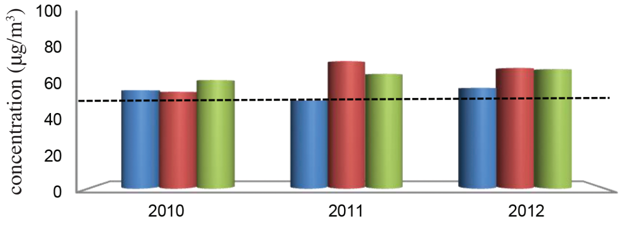
Notation: n = number of observations, O_t = Observed values, $\bar{O}$ = Mean of observed values = $\frac{1}{n}\sum_{t=1}^{n}O_t$, $\bar{O}$ = Mean of observed values = $\frac{1}{n}\sum_{t=1}^{n}O_t$, P_t = Predicted values, $\bar{P}$ = Mean of predicted values = $\frac{1}{n}\sum_{t=1}^{n}P_t$

TABLE 3. Descriptive statistics of the PM_{10} data

		Pasir Gudang	Bukit Rambai	Nilai
N	Valid	1096	1093	1095
	Missing	0	3	1
Mean		55.3887	66.4437	66.0192
Median		52.0000	64.0000	62.0000
Std. Deviation		18.61816	17.65014	19.03342
Variance		346.636	311.527	362.271
Skewness		1.623	1.018	1.260
Kurtosis		6.023	2.169	2.731
Minimum		22.00	28.00	27.00
Maximum		192.00	148.00	160.00
Percentiles	50	52.0000	64.0000	62.0000
	75	64.0000	76.0000	76.0000
	95	90.0000	98.0000	102.0000

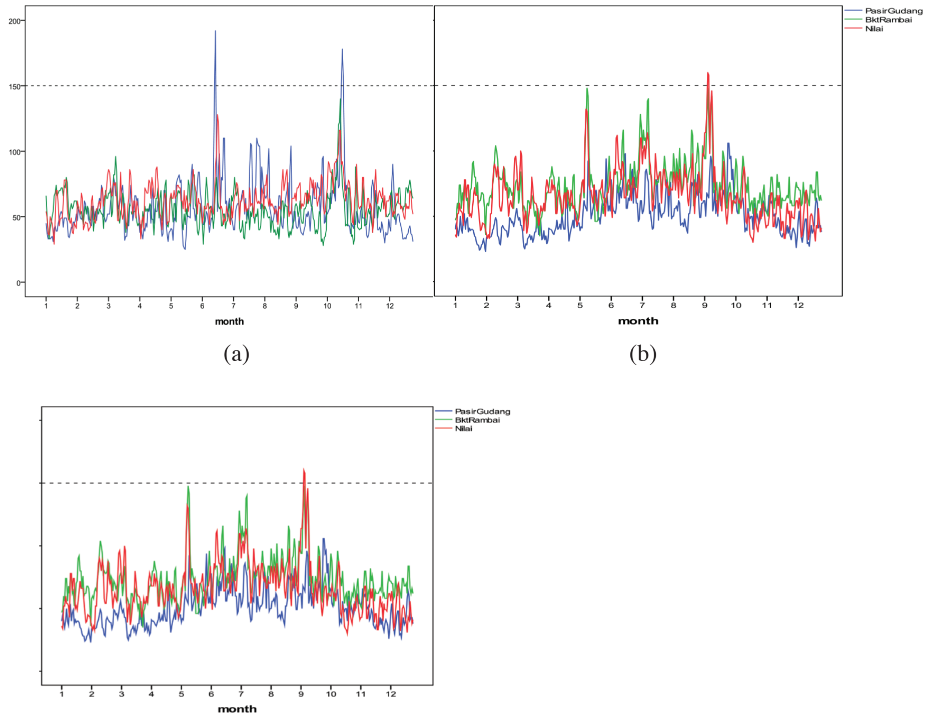
The trend of annual average of PM_{10} concentrations in 2010 - 2012 showed that the levels exceeded the Malaysian Ambient Air Quality Guidelines (MAAQG) for the yearly average of 50 µg/m^3 is shown in Figure 3.

Figure 4 shows the time series' plot of PM_{10} concentrations. In general, the country experiences the high concentrations of the PM_{10} during the second and third-quarter of the year as a result of trans-boundary smoke from the forest fire in Sumatra region during dry season from May to September. In 2010, the air quality in the Southern part of Peninsular Malaysia particularly in Johor, Melaka and Negeri Sembilan deteriorated and recorded the increase in PM_{10} concentrations (Department of Environment Malaysia 2012, 2011, 2010).



	2010	2011	2012
🟦 Pasir Gudang	56.91	51.01	58.25
🟥 Bukit Rambai	56.01	73.56	69.67
🟩 Nilai	62.72	66.29	69.05

FIGURE 3. Annual average concentrations of PM$_{10}$ by monitoring stations, 2010 - 2012



(a)

(b)

(c)

FIGURE 4. Monthly time series plot of PM10 concentrations in (a) 2010, (b) 2011 and (c) 2012

RESULT AND DISCUSSION

Tables 4 and 5 list the estimates for the location parameter, µ, scale parameter, σ and shape parameter, λ for all distributions using the method of moment and maximum likelihood estimator and their performance indicators.

Based on performance indicators, the distributions were then ranked. The best distribution is selected based on the highest accuracy measures and the smallest error measures. It is significant to note that for all the three stations under consideration, the best distribution was the GEV distribution. The best estimator for two stations, Pasir Gudang and Bukit Rambai was the MLE while for Nilai station, the best estimator was the MOM.

Cumulative distribution functions (CDF) of the GEV distribution for all three monitoring stations are presented in Figure 5. From this figure, the probability of the concentrations exceeding the levels of MAAQG of 150 µg/m3 was estimated. For Pasir Gudang, the probability was 0.0014 ($F(x) < 150 = 0.9986$). The estimated numbers of days in which PM$_{10}$ concentrations exceeded MAAQG was 0.0014 × 1096 days = 1 ½ days. In the case of Bukit Rambai, the probability was 0.0005 ($F(x) < 150 = 0.9995$). The predicted number of unhealthy days was 0.0005 × 1096 days = ½ days. As for Nilai, the probability was 0.0019 ($F(x) < 150 = 0.9989$). The estimated number of unhealthy days for three years was 0.0011 × 1096 = 1 day (Table 7).

TABLE 4. Parameter estimates and performance indicators using MOM

Stations	Distributions			Performance Indicators					
				NAE	PA	R^2	RMSE	IA	MAE
Pasir Gudang	2-Gumbel	μ σ	47.01 14.52	0.304	0.850	0.721	19.757	0.759	16.826
	2-Weibull	σ λ	61.78 3.26	0.063	0.957	0.913	5.518	0.978	3.514
	3-Weibull	μ σ λ	34.02 22.45 1.15	0.063	0.980	0.959	4.511	0.986	3.510
	GEV	μ σ λ	46.85 13.15 -0.07	0.034	0.990	0.978	4.163	0.985	1.869
	2-GPD	σ λ	272.80 3.93	0.193	0.715	0.510	14.082	0.837	10.690
	3-GPD	μ σ λ	35.43 21.45 0.07	0.080	0.958	0.916	11.113	0.940	4.457
Bukit Rambai	2-Gumbel	μ σ	58.50 13.76	0.240	0.890	0.791	18.024	0.782	15.952
	2-Weibull	σ λ	73.09 4.21	0.049	0.967	0.933	4.600	0.983	3.255
	3-Weibull	μ σ λ	36.13 34.11 1.83	0.180	0.959	0.917	12.860	0.890	11.947
	GEV	μ σ λ	58.58 14.14 0.02	0.020	0.987	0.973	3.413	0.991	1.349
	2-GPD	σ λ	504.02 6.59	0.165	0.680	0.462	14.141	0.816	10.966
	3-GPD	μ σ λ	44.63 27.56 0.26	0.268	0.673	0.452	70.502	0.430	17.830
Nilai	2-Gumbel	μ σ	57.45 14.84	0.261	0.868	0.752	19.865	0.769	17.202
	2-Weibull	σ λ	72.99 3.86	0.061	0.958	0.915	5.588	0.978	4.024
	3-Weibull	μ σ λ	40.47 27.89 1.36	0.183	0.971	0.940	13.242	0.900	12.095
	GEV	μ σ λ	57.39 14.46 -0.02	0.013	0.994	0.986	2.182	0.997	0.885
	2-GPD	σ λ	430.15 5.52	0.182	0.673	0.452	15.419	0.810	11.996
	3-GPD	μ σ λ	43.91 25.97 0.17	0.164	0.842	0.708	33.803	0.724	10.801

TABLE 5. Parameter estimates and performance indicators using MLE

Stations	Distributions			Performance Indicators					
				NAE	PA	R^2	RMSE	IA	MAE
Pasir Gudang	2-Gumbel	μ	65.74	0.283	0.850	0.721	25.789	0.786	15.649
		σ	29.44						
	2-Weibull	σ	61.75	0.071	0.963	0.925	5.669	0.979	3.946
		λ	2.94						
	3-Weibull	μ	20.91	0.265	0.980	0.959	15.117	0.872	14.649
		σ	38.98						
		λ	1.96						
	GEV	μ	46.94	0.013	0.993	0.985	2.264	0.996	0.707
		σ	13.60						
		λ	0.04						
	2-GPD	σ	67.95	1.376	0.584	0.341	353.337	0.111	76.211
		λ	-0.35						
	3-GPD	μ	22.00	0.354	0.974	0.946	27.008	0.795	19.631
		σ	67.95						
		λ	-0.35						
Bukit Rambai	2-Gumbel	μ	75.99	0.151	0.890	0.791	16.760	0.870	9.997
		σ	22.79						
	2-Weibull	σ	73.15	0.057	0.972	0.942	4.955	0.982	3.800
		λ	3.75						
	3-Weibull	μ	25.38	0.334	0.959	0.917	22.712	0.738	22.165
		σ	46.29						
		λ	2.44						
	GEV	μ	58.80	0.014	0.995	0.989	1.750	0.998	0.934
		σ	14.66						
		λ	-0.05						
	2-GPD	σ	93.98	5.799	0.203	0.041	6616.86	0.002	385.282
		λ	-0.63						
	3-GPD	μ	28.00	0.341	0.964	0.928	28.979	0.758	22.658
		σ	93.98						
		λ	-0.63						
Nilai	2-Gumbel	μ	76.49	0.183	0.868	0.752	19.808	0.849	12.059
		σ	25.40						
	2-Weibull	σ	73.05	0.070	0.964	0.927	5.882	0.978	4.601
		λ	3.44						
	3-Weibull	μ	24.87	0.314	0.971	0.940	21.245	0.786	20.755
		σ	46.50						
		λ	2.27						
	GEV	μ	57.56	0.013	0.993	0.984	2.327	0.996	0.823
		σ	14.71						
		λ	0.00						
	2-GPD	σ	90.48	3.732	0.251	0.063	3233.06	0.006	246.374
		λ	-0.56						
	3-GPD	μ	27.00	0.342	0.964	0.927	28.956	0.778	22.579
		σ	90.48						
		λ	-0.56						

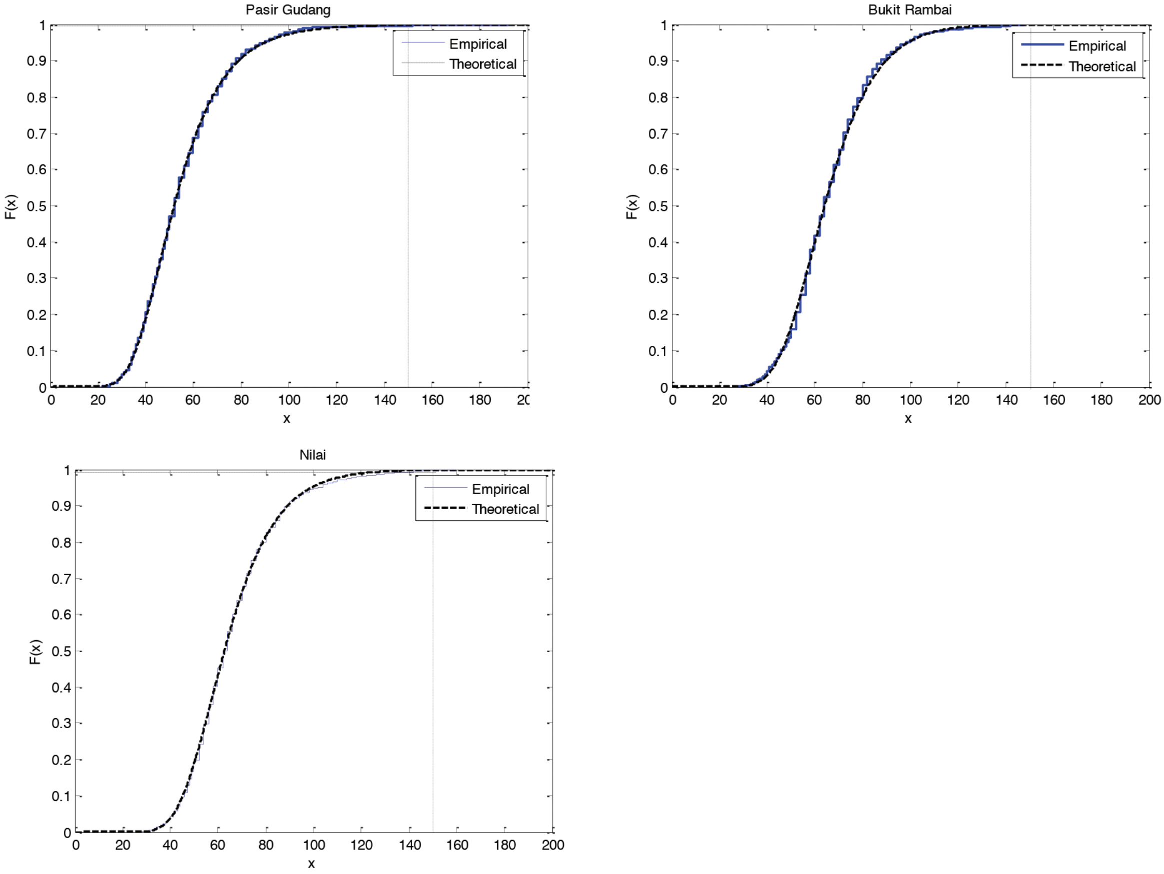


FIGURE 5. Cumulative distribution functions (CDF) of GEV for the three stations

TABLE 6. The best distribution

	MOM	MLE	The best distribution
Pasir Gudang	GEV	GEV	GEV - mle
Bukit Rambai	GEV	GEV	GEV - mle
Nilai	GEV	GEV	GEV - mom

TABLE 7. Comparison of estimated and actual number of unhealthy days

Stations	Predicted no. of unhealthy days	Actual no. of unhealthy days
Pasir Gudang	1½	4
Bukit Rambai	½	0
Nilai	1	3

CONCLUSION

In the study of air pollutions, the researchers focused on high concentrations of pollutants as it is detrimental to human health. The descriptive statistics show that the mean concentrations of the three stations exceeded the Malaysian Ambient Air Quality Guidelines (MAAQG) level for the hourly average of 50 μg/m^3. In general, the country experienced the high concentrations of the PM$_{10}$ during the second and third-quarter of the year as a result of trans-boundary smoke from the forest fire in Sumatera region during dry season from May to September as demonstrated in the three year PM$_{10}$ concentrations data. Six extreme distributions were compared using two different estimators, namely, MOM and MLE. The results showed that the GEV distribution was the most appropriate distribution for daily maximum density of PM$_{10}$ for all the monitoring stations under study. The best parameter estimator for two stations, Pasir Gudang and Bukit Rambai, was the MLE while for Nilai, the best estimator was using MOM. The CDF of observed PM_{10} and the predicted values obtained from the GEV were then fitted. The analysis showed that the number of days of which the concentrations of PM_{10} exceeded daily MAAQG were very minimal, between ½ -1½ days in these stations. In general, the air quality in the Southern region of Peninsular Malaysia where the three stations are located was in between of good and moderate except for a few unhealthy days recorded in 2010 – 2012. To conclude, the GEV had an advantage over the other distributions since it provides better performance indicators in estimating the number of days that exceeded the specified levels of MAAQG of 150 $\mu g/m^3$ for daily concentrations. Thus, the GEV may be used to predict the exceedances of future extreme concentrations of PM_{10} in Malaysia.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Environment, Malaysia for providing the air quality data in this study and USM for the RUI grant: 814165.

REFERENCES

Abd-el-hakim, Nagi, S. & Khalaf S Sultan. 2004. Maximum likelihood estimation from record-breaking data for the generalized pareto distribution. *International Journal of Statistics* LXII 3: 377-389.

Afroz, R., Mohd Nasir Hassan & Noor Akma Ibrahim. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research* 92(6): 71-77.

Bermudez, P. de Zea & Kotz, S. 2010. Parameter estimation of the generalized pareto distribution-Part II. *Journal of Statistical Planning and Inference* 140(6): 1374-1388.

Bury, K. 1999. *Statistical Distributions in Engineering*. London: Cambridge University Press.

Chapman, S. 2004. *MATLAB Programming for Engineers*. 3rd ed. Australia: Thomson.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Bristol: Springer series in statistics.

Dasgupta, R. & Bhaumik, D.K. 1995. Upper and lower limits of atmospheric ozone level and extreme value distribution. *Sankhya: The Indian Journal of Statistics* 57(B2): 182-199.

Horowitz, J. & Barakat, S. 1979. Statistical analysis of the maximum concentration of an air pollutant: Effects of autocorrelation and non-stationarity. *Atmospheric Environment (1967)* 13(6): 811-818.

Jamal, H.H., Pillay, M.S., Zailina, H., Shamsul, B.S., Sinha, K., Zaman Huri, Z., Khew, S.L., Mazrura, S., Ambu, S., Rahimah, A. & Ruzita, M.S. 2004. *A Study of Health Impact & Risk Assessment of Urban Air Pollution in Klang Valley, Malaysia*. Kuala Lumpur: UKM Pakarunding Sdn Bhd.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. & Kolehmainen, M. 2004. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment* 38(6): 2895-2907.

Kao, T.C. & Lin, C.H. 2010. Setting margin levels in futures markets: An extreme value method. *Nonlinear Analysis: Real World Applications* 11(6): 1704-1713.

Kotz, S. & Nadarajah, S. 2000. Extreme-Value Distributions: Theory and Applications. London: Imperial College Press.

Kuchenhoff, H. & Thamerus, M. 1996. Extreme value analysis of Munich air pollution data. Environmental and Ecological Statistics 3: 127-141.

Lee, Muhammad Hisyam, Nur Haizum Abd. Rahman, Suhartono, Mohd Talib Latif, Maria Elena Nor & Nur Arina Bazilah Kamisan. 2012. Seasonal ARIMA for forecasting air pollution index: A case study. *American Journal of Applied Sciences* 9(4): 570-578.

Lu, H.C. 2002. The statistical characters of PM10 concentration in Taiwan area. *Atmospheric Environment* 36(3): 491-502.

Lu, H.C. & Fang, G.C. 2003. Predicting the exceedances of a critical PM10 concentration - A case study in Taiwan. *Atmospheric Environment* 37(8): 3491-3499.

Malaysia Environmental Quality Report 2012. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2011. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2010. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2009. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2008. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2007. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2006. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2005. Putrajaya: Department of Environment Malaysia.

Malaysia Environmental Quality Report 2004. Putrajaya: Department of Environment Malaysia.

Martins, E.S. & Stedinger, J.R. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research* 36(3): 737-744.

Mohamed Noor, Norazian, Cheng Yau Tan, Mohd Mustafa Al-Bakri Abdullah, Nor Azam Ramli & Ahmad Shukri Yahaya. 2011. Modelling of PM10 concentration in industrialized area in Malaysia : A case study in Nilai. In *2011 International Conference on Environment and Industrial Innovation,* 13: 18-22. Singapore: IACSIT Press.

Oztekin, T. 2005. Comparison of parameter estimation methods for the three-parameter generalized pareto distribution. *Turk. J. Agric. For.* 29: 419-428.

Petrov, V., Guedes Soares, C. & Gotovac, H. 2013. Prediction of extreme significant wave heights using maximum entropy. *Coastal Engineering* 74(4): 1-10.

Quintela-del-Río, A. & Francisco-Fernández, M. 2011. Analysis of high level ozone concentrations using nonparametric methods. *The Science of the Total Environment* 409(2): 1123-1133.

Reeve, D.T., Randell, D., Ewans, K.C. & Jonathan, P. 2012. Uncertainty due to choice of measurement scale in extreme value modelling of North Sea storm severity. *Ocean Engineering* 53(10): 164-176.

Reyes, H.J., Vaquera, H. & Villasenor, J.A. 2010. Estimation of trends in high urban ozone levels using the quantiles of (GEV). *Environmetrics* 21: 470-481.

Reynolds, A.M. 2012. Gusts within plant canopies are extreme value processes. *Physica A: Statistical Mechanics and Its Applications* 391(11): 5059-5063.

Rinne, H. 2008. *The Weibull Distribution: A Handbook*. Florida: CRC Press.

Roberts, E.M. 1979. Review of statistics of extreme values with applications to air quality data Part I . Review. *Journal of the Air Pollution Control Association* 29(6): 632-637.

Singh, V.P. & Guo, H. 1995. Parameter estimation for 3-parameter generalized pareto distribution by the principle of maximum entropy (POME). *Hydrological Sciences* 40(2): 165-181.

Smith, R.L. 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Sciences* 4(4): 367-393.

Su, F.C., Jia, C. & Batterman, S. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. *Atmospheric Environment* 62(12): 97-106.

Surman, P.G., Bodero, J. & Simpson, R.W. 1987. The prediction of the numbers of violations of standards and the frequency of air pollution episodes using extreme value theory. *Atmospheric Environment* 21(8): 1843-1848.

Talib, M.L., Rozali, M.O., Norela, S., Ahmad Daud, M.N. & Permata, N.J. 2002. Air quality in several industrial areas in Malaysia. In *Proceedings of the Regional Symposium on Environment and Natural Resources*, edited by Omar, R., Ali Rahman, Z., Latif, M.T., Lihan, T. & Adam, J.H. April 10-11. Renaissance Hotel, Kuala Lumpur. pp. 703-710.

Torrielli, A., Repetto, M.P. & Solari, G. 2013. Extreme wind speeds from long-term synthetic records. *Journal of Wind Engineering and Industrial Aerodynamics* 115(4): 22-38.

Tsai, M.S. & Chen, L.C. 2011. The calculation of capital requirement using extreme value theory. *Economic Modelling* 28(1): 390-395.

Yahaya, Ahmad Shukri & Nor Azam Ramli. 2008. Modelling of carbon monoxide concentration in major towns in Malaysia: A case study in Penang, Kuching and Kuala Lumpur. USM Short Term Grant. Penang: Universiti Sains Malaysia.

Yao, F., Wen, H. & Luan, J. 2013. CVaR measurement and operational risk management in commercial banks according to the peak value method of extreme value theory. *Mathematical and Computer Modelling* 58(1-2): 15-27.

Yap, X.Q. & Hashim, M. 2013. A robust calibration approach for PM_{10} prediction from MODIS aerosol optical depth. *Atmospheric Chemistry and Physics* 13(3): 3517-3526.

Hasfazilah Ahmat*, Ahmad Shukri Yahaya & Nor Azam Ramli
Clean Air Research Group, School of Civil Engineering
Engineering Campus, Universiti Sains Malaysia
14300 Pulau Pinang
Malaysia

Hasfazilah Ahmat*
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
13500 Permatang Pauh, Pulau Pinang
Malaysia

*Corresponding author; email: hasfazilah.ahmat@gmail.com

Received: 26 March 2014
Accepted: 3 August 2014