Journal of Engineering Science and Technology
Vol. 10, No.12 (2015) 1560 - 1574

THE MALAYSIA PM_{10} ANALYSIS USING EXTREME VALUE

HASFAZILAH AHMAT*, AHMAD SHUKRI YAHAYA, NOR AZAM RAMLI

Clean Air Research (CARE) Group, School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 P.Pinang, MALAYSIA
Department of Computer and Mathematical Sciences, Universiti Teknologi MARA Pulau Pinang, 13500 Permatang Pauh, P.Pinang, MALAYSIA
*Corresponding Author: hasfazilah.ahmat@gmail.com

Abstract

The study of air quality is closely associated to air pollution. Air pollution is of the main concerns of the authority in view of the fact that it can generate damaging effects to human health, crops and environment. This paper assesses the use of Extreme Value Distributions (EVD) 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 were estimated using the method of Maximum Likelihood Estimator (MLE). The goodness-of-fit of the distribution was determined using six performance indicators namely; the accuracy measures which include Prediction 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). The best distribution was selected based on the highest accuracy measures which are close to 1 and the smallest error measures. The result showed that the Generalized Extreme Value (GEV) distribution was the best fit for daily maximum concentration for PM_{10} for all monitoring stations. The GEV gave the smallest errors (NAE, RMSE and MAE) and the highest accuracy measures (PA, R^2 and IA) when compared to other distributions. The method gave the accuracy of more than 98% in PA, IA and R^2 for all stations. The analysis demonstrated that the estimated numbers of days in which the concentration of PM_{10} exceeded the Malaysian Ambient Air Quality Guidelines (MAAQG) of 150 $\mu g/m^3$ were between ½ and 2 days.

Keywords: Extreme Value Theory, PM_{10}, air pollution, prediction.

Nomenclatures

n	Number of observed data
e_t	Forecast error, $O_t - P_t$
O_t	Observed data
$\overline{O}$	Mean of observation
P_t	Predicted data
$\overline{P}$	Mean of predicted data
σ_O	Standard deviation of Observed data
σ_P	Standard deviation of Predicted data

Greek Symbols

μ	Location parameter
σ	Scale parameter
λ	Shape parameter
Σ	Summation of the expression

Abbreviations

EVD	Extreme Value Distribution
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution
PM_{10}	Particulate Matter of diameter less than 10 micrometre
MLE	Maximum Likelihood Estimator
PA	Prediction Accuracy
R^2	Coefficient of Determination
IA	Index of Accuracy
RMSE	Root Mean Square Error
NAE	Normalized Absolute Error
MAE	Mean Absolute Error
MAAQG	Malaysian Ambient Air Quality Guideline
$\mu g/m^3$	Microgram per cubic metre
CDF	Cumulative Distribution Function

1. Introduction

The study of air quality is closely associated with air pollution. Air pollution is a universal term that refers to the presence of air pollutants in the form of gaseous, liquid or fine particles suspended in air. One of the concerns of the air pollution studies is to compute the concentrations of one or more types of pollutants in space and time in relation to the independent variables, for instance emissions into the atmosphere, meteorological factors and parameters. The Extreme Value Theory (EVT) is one of the most significant statistical disciplines developed for the last few decades for the applied sciences and many other disciplines. The most key 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 [1]. Various studies in different fields have been published for the last couple of years in the applications of the EVT, for example operational risk management [2], Volatile Organic Compound exposures [3], future markets [4], calculation of capital requirement [5], wind speed [6, 7], wave heights [8] and storm [9]. 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 [10, 11].

A widely used method for assessing and estimating the concentrations of air pollution is the Extreme Value Distribution (EVD) [11-20]. In Malaysia, among the studies on air pollution concentrations were that of refs [21, 22].

The study on extreme concentrations is of the concerns of the researchers because the exposure of particulate matter on a higher scale may affect health of sensitive groups such as children, the elderly and individuals with asthma or cardiopulmonary diseases [23, 24]. In addition, it may pose undesirable impact on the environment. It is said to be the major cause of reduced in visibility, resulting in foggy conditions particularly during the dry season [25]

In view of the fact that it can generate damaging effects to human health, crops and environment [26], this study is carried out to attain the best model to predict PM10 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 PM10. Parameters for all distributions are estimated using the method of Maximum Likelihood Estimator (MLE).

2. Methodology

2.1. 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 Peninsular Malaysia. Figure 1 illustrates the three monitoring stations - Pasir Gudang, Johor, Bukit Rambai, Melaka and Nilai, Negeri Sembilan which are classified under industrial by Department of Environment, Malaysia [27].

All the 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 Taman 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 large amount of air pollution [28-30]. In addition, the southern part of Peninsular Malaysia is prone to the trans-boundary smoke from 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 [27, 31].

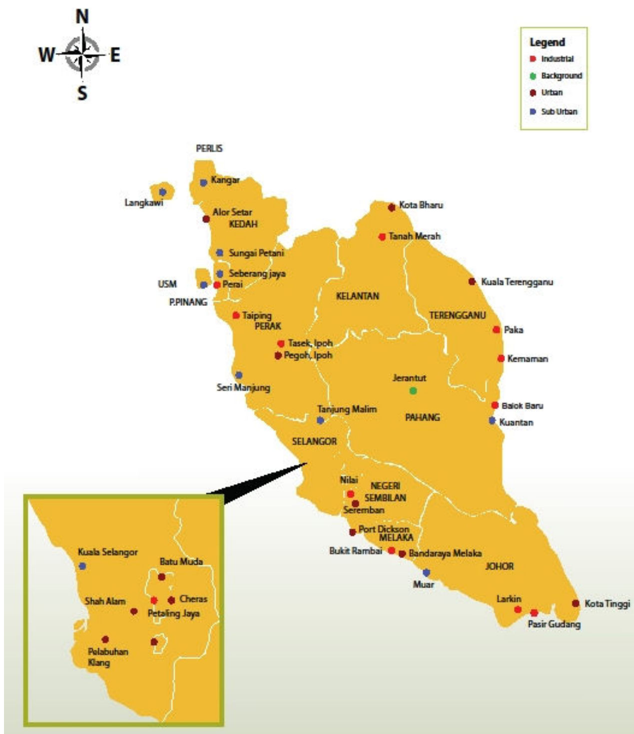


Fig. 1. Location of continuous air quality monitoring stations in Peninsular Malaysia (source: [32]).

The analysis of data with the absence of missing values was completed using a programming language for numerical computation, visualization, and programming package for engineers called MATLAB® [33]

2.2. Probability distribution and parameter estimators

This research undertaken the analysis of PM_{10} data using the EVDs, namely: Gumbel [10], two and three-parameter Weibull [34], Generalized Extreme Value (GEV) [35] and two and three-parameter Generalized Pareto Distribution (GPD) [35, 36]. All the parameters of the distributions were estimated using the method of MLE. Table 1 depicts the Probability Density Function (PDF) of the EVDs and the parameter estimators of each EVD.

2.3. Performance indicators

This study used six performance indicators to select the best distribution to represent the data. The accuracy 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 were the Root Mean Square Error (RMSE), the Normalized Absolute Error (NAE) and the Mean Absolute Error (MAE). Table 2 lists the performance indicators and their formulae used in this study.

Figure 2 depicts the flow of methodology in the process of obtaining the best distribution to predict the numbers of days with concentrations above 150µg/m^3.

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

EVD	Probability Density Function (PDF)	Parameter estimator
2-Gumbel	$f(x;\mu,\sigma)=\frac{1}{\sigma}\times\ldots$ $exp\left[-\frac{x-\mu}{\sigma}-exp\left(-\frac{x-\mu}{\sigma}\right)\right]$	$\sigma=\bar{x}-\frac{\sum_{i=1}^{n}x_i\,exp(-x_i/\sigma)}{\sum_{i=1}^{n}exp(-x_i/\sigma)}$ $\mu=-\sigma\,ln\left(\frac{1}{n}\sum_{i=1}^{n}exp\left(-\frac{x_i}{\sigma}\right)\right)$
2-Weibull	$f(x;\sigma,\lambda)=\frac{\lambda}{\sigma}\left(\frac{x}{\sigma}\right)^{\lambda-1}\times\ldots$ $exp\left[-\left(\frac{x}{\sigma}\right)^{\lambda}\right]$	$\sigma=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\lambda}\right]^{1/\lambda}$, $\frac{1}{\lambda}-\frac{\sum_{i=1}^{n}x_i^{\lambda}\,ln\,x_i}{\sum_{i=1}^{n}x_i^{\lambda}}+\frac{1}{n}\sum_{i=1}^{n}ln\,x_i=0$
3-Weibull	$f(x;\lambda,\sigma,\mu)=\frac{\lambda}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\lambda-1}\times\ldots$ $exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{\lambda}\right]$	$\sigma=\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^{\lambda}\right]^{1/\lambda}$ $\frac{1}{\lambda}-\frac{\sum_{i=1}^{n}(x_i-\mu)^{\lambda}\,ln(x_i-\mu)}{\sum_{i=1}^{n}(x_i-\mu)^{\lambda}}+\ldots$ $\frac{1}{n}\sum_{i=1}^{n}ln(x_i-\mu)=0$ $\frac{\lambda-1}{\lambda}\sum_{i=1}^{n}(x_i-\mu)^{-1}-n\frac{\sum_{i=1}^{n}(x_i-\mu)^{\lambda-1}}{\sum_{i=1}^{n}(x_i-\mu)^{\lambda}}=0$

EVD	Probability Density Function (PDF)	Parameter estimator
GEV	$f(x)=\frac{1}{\sigma}\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)^{-1/\lambda-1}\right]\times \exp\left\{-\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)^{-1/\lambda}\right]\right\}$	$\frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{1-\lambda-(1-(\lambda/\sigma)(x_i-\mu))^{1/\lambda}}{(1-(\lambda/\sigma)(x_i-\mu))}\right)=0$ $-\frac{n}{\sigma}+\frac{1}{\sigma}\times\ldots \sum_{i=1}^{n}\left[\frac{1-\lambda-(1-(\lambda/\sigma)(x_i-\mu))^{1/\lambda}}{(1-(\lambda/\sigma)(x_i-\mu))}\cdot \cdot\left(\frac{(x_i-\mu)}{\sigma}\right)\right]=0$ $-\frac{1}{\lambda^2}\sum_{i=1}^{n}\left[\ln(1-(\lambda/\sigma)(x_i-\mu)). \left\{1-\lambda-[1-(\lambda/\sigma)(x_i-\mu)]^{1/\lambda}\right\} + \frac{1-\lambda-[1-(\lambda/\sigma)(x_i-\mu)]^{1/\lambda}}{(1-(\lambda/\sigma)(x_i-\mu))}\ldots \lambda\left(\frac{x_i-\mu}{\sigma}\right)\right]=0$
2-GPD	$f(x;\lambda,\sigma)=\frac{1}{\sigma}\left[1-\lambda\left(\frac{x}{\sigma}\right)\right]^{1/\lambda-1}$	$\sum_{i=1}^{n}\frac{x_i/\sigma}{1-\lambda(x_i)/\sigma}=\frac{n}{1-\lambda}$, $\sum_{i=1}^{n}\ln(1-\lambda(x_i)/\sigma)=-n\lambda$
3-GPD	$f(x;\lambda,\sigma,\mu)=\frac{1}{\sigma}\left[1-\lambda\left(\frac{x-\mu}{\sigma}\right)\right]^{1/\lambda-1}$	$\mu=x_1$, $\sigma=\frac{\lambda}{\exp(n\lambda)-1}(x_n-\mu)$ $\sum_{i=1}^{n}\left(\exp(n\lambda)+\frac{x_n-x_i}{x_i-\mu}\right)^{-1}=\ldots \frac{n}{\exp(n\lambda)-1}-\frac{1}{\lambda\exp(n\lambda)}$

Table 2. Performance indicators.

Indicators	Equations
PA	$PA=\sum_{t=1}^{n}\frac{(P_t-\overline{P})(O_t-\overline{O})}{(n-1)\sigma_p\sigma_o}$
R^2	$1-\left(\sum_{t=1}^{n}(O_t-P_t)^2\Big/\sum_{t=1}^{n}(O_t-\overline{O})^2\right)$

IA	$1-\dfrac{\sum_{t=1}^{n}\left(P_t-O_t\right)^2}{\sum_{t=1}^{n}\left(\left\lvert P_t-\overline{O}\right\rvert-\left\lvert O_t-\overline{O}\right\rvert\right)^2}$
RMSE	$\sqrt{\dfrac{\sum_{t=1}^{n}\left(O_t-P_t\right)^2}{n}}$
NAE	$\dfrac{\sum_{t=1}^{n}\left(P_t-O_t\right)}{\sum_{t=1}^{n}O_t}$
MAE	$\dfrac{\sum_{t=1}^{n}\left\lvert\left(O_t-P_t\right)\right\rvert}{n}$

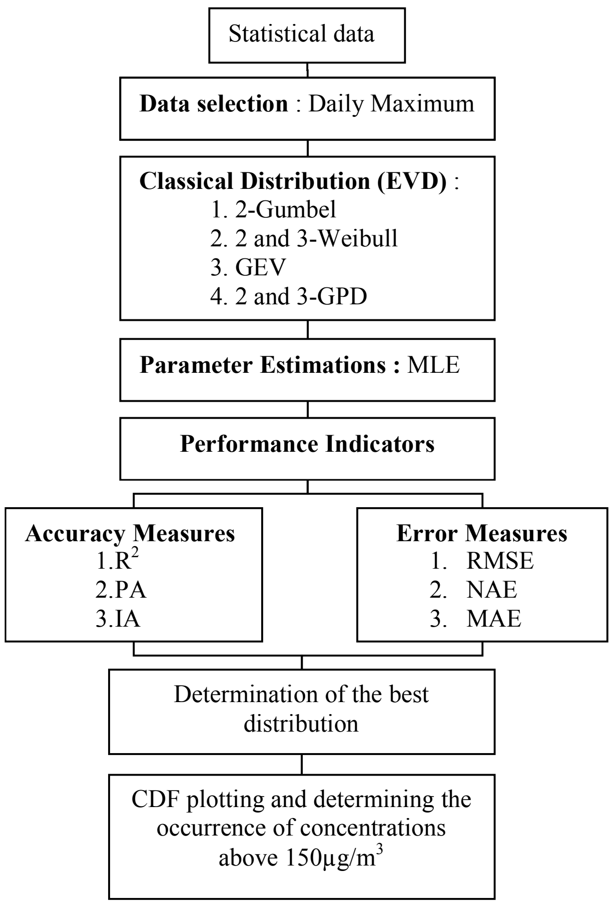


Fig. 2. Flow of methodology.

2.4. Data

Table 3 describes the descriptive statistics of PM_{10} concentration for the monitoring stations. The unit of measurement is microgram per cubic metre ($\mu g/m^3$). 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 $\mu g/m^3$ [38] with the Bukit Rambai station's average recorded slightly above the average of other stations. 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 3. Descriptive statistics of the PM10 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 MAAQG for the yearly average of $50 \mu g/m^3$ as illustrated in Fig. 3.

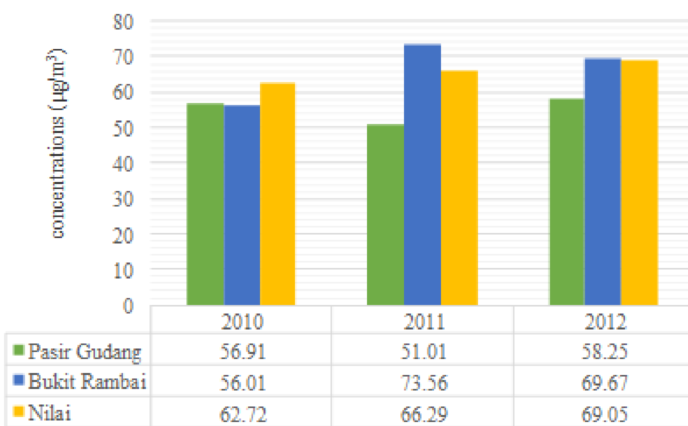
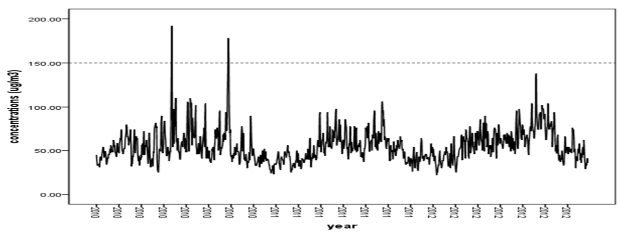
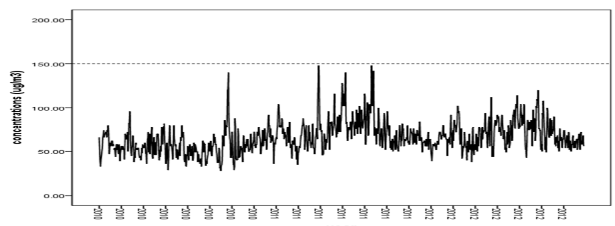


Fig. 3. Annual average concentrations of PM_{10} by monitoring stations, 2010 - 2012.

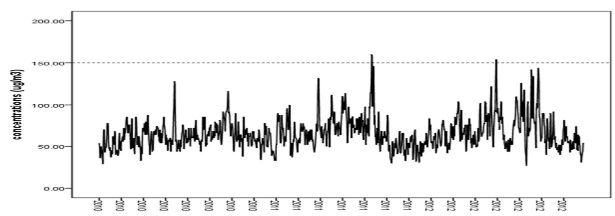
Figure 4 demonstrates the time series plot of PM_{10} concentrations. In general, the country experienced high concentrations of PM_{10} during the second and third-quarter of the year as a result of trans-boundary smoke from the forest fire in Sumatera region during dry season from May to September. 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 [27, 31, 38].



(a)



(b)



(c)

Fig. 4. Time series plot of PM_{10} concentrations in $\mu g/m^3$ for (a) Pasir Gudang, (b) Bukit Rambai and (c) Nilai.

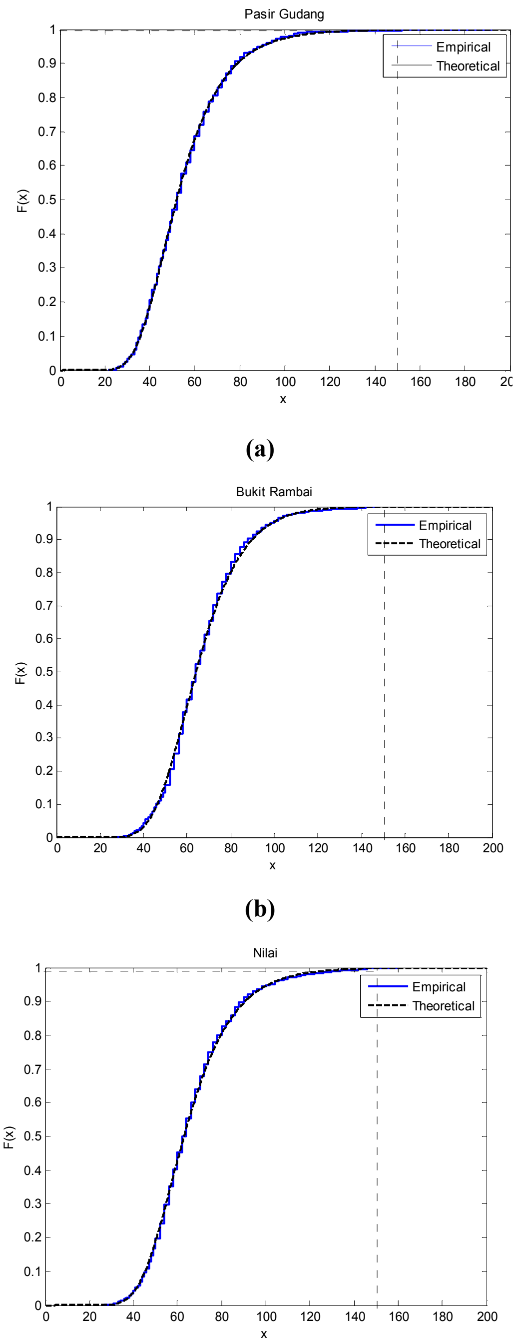
3. Results and Discussion

Table 4 lists the estimates for the location parameter, μ, scale parameter, σ and shape parameter, λ for all distributions using the MLE and their performance indicators.

Based on performance indicators, the distributions were then ranked. The best distribution was 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.

Table 4. Parameter estimates and performance indicators.

Stations	Distributions			Performance Indicators						The best dist.
				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*	GEV
		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	GEV
		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	GEV
		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							



(a)

(b)

(c)

Fig. 4. Cumulative Distribution functions (CDF) of GEV for (a) Pasir Gudang, (b) Bukit Rambai and (c) Nilai.

The Cumulative Distribution Functions (CDF) of the GEV distribution for all three monitoring stations are presented in Fig. 4. From this figure, the probability of the concentrations exceeding the levels of MAAQG of 150 $\mu g/m^3$ was estimated. For Pasir Gudang, the probability was 0.0014 ($F(x)<150 = 0.9986$). The estimated number of days in which PM_{10} concentrations exceeded MAAQG was 0.0014 x 1096 days = 1½ 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 x 1096 days = ½ days. As for Nilai, the probability was 0.0019 ($F(x)<150 = 0.9981$). The estimated number of unhealthy days for three years were 0.0019 x 1096 = 2 days.

Table 5. Comparison of estimated and actual numbers of unhealthy days.

Stations	Predicted no. of unhealthy days	Actual no. of unhealthy days
Pasir Gudang	1½	4
Bukit Rambai	½	0
Nilai	2	3

4. Conclusion

This paper discussed the probability and the numbers of days of the extreme concentrations which exceeded the permitted value of PM_{10} concentrations of 150$\mu g/m^3$ in three monitoring stations in the west coast of Peninsular Malaysia. The MLE was used to estimate the parameters of six distributions under consideration, namely: Gumbel, 2 and 3-parameter Weibull, GEV and 2 and 3-parameter GPD. All the daily maximum data without missing values from 2010 - 2012 were used to analyse the efficiency of the six distributions using two performance indicators, error measures and accuracy measures. The analyses of three accuracy measures, namely PA, R^2 and IA and three error measures - NAE, RMSE and MAE were acquired to indicate the efficiency or the performance indicators of the distributions.

The descriptive statistics showed that the mean concentrations of the three stations exceeded the MAAQG level for the hourly average of 50$\mu g/m^3$ with the maximum reading recorded in Pasir Gudang. 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 years' PM_{10} concentrations data. Six EVDs were compared and it 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 GEV gave the smallest errors (NAE, RMSE and MAE) and the highest accuracy measures (PA, R^2 and IA) when compared to the other distributions. The method gave the accuracy of more than 98% in PA, IA and R^2 for all stations and the smallest errors.

The CDF of observed PM10 and the predicted values obtained from the GEV were fitted and the predicted numbers of days were calculated. The analysis shows that the numbers of days of which the concentrations of PM10 exceeded MAAQG were very minimal in these stations. In general, the air quality in the southern region of Malaysia where the three stations are located was in between of good and moderate except for a few of 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 μg/m^3 for daily concentrations. In the study of air pollutions, the researchers focused on the high concentrations of pollutants as it was detrimental to human health. The GEV may be used to predict the exceedances of future extreme concentrations of PM_{10} and hence, it may help the policy makers in the respective field to plan suitable measures to curb the occurrence of PM_{10} extreme concentrations and eventually may reduce the effects on human health.

Acknowledgement

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

1. Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Bristol: Springer Series in Statistics.
2. Yao, F., Wen, H.; and Luan, J. (2013). CVaR measurement and operational risk management in commercial banks according to the peak value method of extreme value theory. *Math. Comput. Model.*, 58(1-2), 15-27.
3. Su, F.-C., Jia, C.; and Batterman, S. (2012). Extreme Value Analyses of VOC Exposures and Risks: A Comparison of RIOPA and NHANES Datasets. *Atmos. Environ.*, 62, 97-106.
4. Kao, T.; and Lin, C. (2010). Setting Margin Levels in Futures Markets: An Extreme Value Method. *Nonlinear Anal. Real World Appl.*, 11, 1704-1713.
5. Tsai, M.-S.; and Chen, L.-C. (2011). The Calculation of Capital Requirement using Extreme Value Theory. *Econ. Model.*, 28, 390-395.
6. Torrielli, A.; Repetto, M.P.; and Solari, G. (2013). Extreme Wind Speeds from Long-Term Synthetic Records. *J. Wind Eng. Ind. Aerodyn.*, 115, 22-38.
7. Reynolds, A. M. (2012). Gusts within Plant Canopies are Extreme Value Processes. *Phys. A Stat. Mech. its Appl.*, 391, 5059-5063.
8. Petrov, V., Guedes Soares, C.; and Gotovac, H. (2013). Prediction of Extreme Significant Wave Heights using Maximum Entropy. *Coast. Eng.*, 74, 1-10.
9. Reeve, D.T., Randell, D., Ewans, K.C.; and Jonathan, P. (2012). Uncertainty due to Choice of Measurement Scale in Extreme Value Modelling of North Sea Storm Severity. *Ocean Eng.*, 53, 164-176.
10. Kotz, S.; and Nadarajah, S. (2000). *Extreme Value Distributions : Theory and Applications*. London: Imperial College Press.

11. Surman, P.G., Bodero, J.; and Simpson, R.W. (1987). The Prediction of the Numbers of Violations of Standards and the Frequency of Air Pollution Episodes using Extreme Value Theory. *Atmos. Environ.*, 21(8), 1843-1848.
12. Roberts, E.M. (1979). Review of Statistics of Extreme Values with Applications to Air Quality Data. Part II. Applications. *J. Air Pollut. Control Assoc.*, 29(7), 733-740.
13. Horowitz, J.; and Barakat, S. (1979). Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Non-Stationarity. *Atmos. Environ.*, 13(6), 811-818.
14. Smith, R.L. (1989). Extreme Value Analysis of Environmental Time Series : An Application to Trend Detection in Ground-Level Ozone. *Stat. Sci.*, 4(4), 367-393.
15. Dasgupta, R.; and Bhaumik, D.K. (1995). Upper And Lower Tolerance Limits of Atmospheric Ozone Level and Extreme Value Distribution. *Sankhya Indian J. Stat.*, 57(B(2)), 182-199.
16. Kuchenhoff, H.; and Thamerus, M. (1995). Extreme Value Analysis of Munich Air Pollution Data. *Sonderforschungsbereich*, 386, 1-24.
17. Lu, H.; and Fang, G. (2002). Estimating the frequency distributions of PM_{10} and $PM_{2.5}$ by the Statistics of Wind Speed at Sha-Lu, Taiwan. *Sci. Total Environ.,*298, 119-130.
18. Lu, H.-C.; and Fang, G.-C. (2003). Predicting the Exceedances of a Critical PM_{10} Concentration - A Case Study in Taiwan. *Atmos. Environ.*, 37, 3491-3499.
19. Reyes, H.J., Vaquera, H.; and Villasenor, J.A. (2010). Estimation of Trends in High Urban Ozone Levels using the Quantiles of (GEV). *Environmetrics*, 21, 470-481.
20. Quintela-del-Río, A.; and Francisco-Fernández, M. (2011). Analysis of High Level Ozone Concentrations using Nonparametric Methods. *Sci. Total Environ.*, 409, 1123-1133.
21. Hurairah, A., Ibrahim, N.A., Daud, I.B.; and Haron, K. (2005). An Application of a New Extreme Value Distribution to Air Pollution Data. M*anag. Environ. Qual. An Int. J.*, 16(1), 17-25.
22. Md Yusof, N.F.F., Ramli, N.A., Yahaya, A.S., Sansuddin, N., Ghazali, N.A.; and Al Madhoun, W. (2011). Central Fitting Distributions and Extreme Value Distributions for Predicion of High PM_{10} Concentration. *IEEE*, 263-6266.
23. Yu, T.S.I., Wong, T.W., Wang, X.R., Song, H., Wong, S.L.; and Tang, J.L. (2001). Adverse Effects of Low-Level Air Pollution on the Respiratory Health of Schoolchildren in Hong Kong. *J. Occup. Environ. Med.*, 43, 310-316.
24. Yu, I.T., Wong, T.W.; and Liu, H.J. (2004). Impact of Air Pollution on Cardiopulmonary Fitness in School children. *J. Occup. Environ. Med.*, 46, 946-952.
25. Yadav, A.K., Kumar, K., Kasim, M.H.A., Singh, M.P., Parida, S.K.; and Sharan, M. (2003). Visibility and Incidence of Respiratory Diseases During the 1998 Haze Episode in Brunei Darussalam. *Air Qual. Pageoph Top.*,160(1-2), 265-277.

26. Jamal, H.H., Pillay, M.S., Zailina, H., Shamsul, B.S., Sinha, K., Zaman Huri, Z., Khew, S.L., Mazrura, S., Ambu, S., Rasimah, A.; and Ruzita, M.S. (2004). *A Study of Health Impact and Risk Assessment of Urban Air Pollution in the Klang Valley, Malaysia*. Kuala Lumpur.
27. Department of Environment Malaysia. (2013). *Malaysia Environmental Quality Report 2012*.
28. Mohamed Noor, N., Tan, C.Y., Abdullah, M.M. A.-B., Ramli, N.A.; and Yahaya, A.S. (2011). Modelling of PM10 Concentration in Industrialized Area in Malaysia : A Case Study in Nilai. *2011 International Conference on Environment and Industrial Innovation*, 13(23), 18-22.
29. Lee, M.H., Abd. Rahman, N.H., Suhartono, Latif, M.T., Nor, M.E.; and Kamisan, N.A.B. (2012). Seasonal ARIMA for Forecasting Air Pollution Index : A Case Study. *Am. J. Appl. Sci.*, 9(4), 570-578.
30. Yap, X.Q.; and Hashim, M. (2013). A Robust Calibration Approach for PM_{10} Prediction from MODIS Aerosol Optical Depth. *Atmos. Chem. Phys.*, 13, 3517-3526.
31. Department of Environment Malaysia. (2012). *Malaysia Environmental Quality Report 2011*. Kuala Lumpur.
32. Malaysian Meteorological Department. (2013). *Mmd Annual Report 2012*. Petaling Jaya.
33. Chapman, S. (2004). *Matlab Programming For Engineers*. Third Edit. Australia: Thomson.
34. Rinne, H. (2008). *The Weibull Distribution: A Handbook*. Florida: Crc Press.
35. Martins, E.; and Stedinger, J. (2000). Generalized Maximum - Likelihood Generalized Extreme - Value Quantile Estimators For Hydrologic Data. *Water Resour. Res.*, 36(3), 737-744.
36. Singh, V.P.; and Guo, H. (1995). Parameter Estimation for 3-Parameter Generalized Pareto Distribution by the Principle of Maximum Entropy (POME). *Hydrol. Sci. J.*, 40(2), 165-181.
37. Abd-el-hakim, N.S.; and Sultan, K.S. (2004). Maximum Likelihood Estimation from Record-Breaking Data for the Generalized Pareto Distribution. *Int. J. Stat.*, LXII(3), 377-389.
38. Department of Environment Malaysia. (2011). *Malaysia Environmental Quality Report 2010.*