Extreme Value Distribution for Prediction of Future PM_{10} Exceedences

Noor Faizah Fitri Md Yusof, Nor Azam Ramli, Ahmad Shukri Yahaya

Clean Air Research Group

Environmental and Sustainable Development Section

School of Civil Engineering, Universiti Sains Malaysia

Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia.

noorfaizah@eng.usm.my

***Abstract*- Central fitting distribution (CFD) such as Weibull, gamma and lognormal distribution can give a good result for fitting the mean concentration of air pollutants data. However, it cannot precisely fit the high concentration region. Therefore, extreme value distributions (EVD) that are Gumbel and Frechet distributions were used in this research to fit the high particulate event in Seberang Perai, Penang from 2002 to 2006 to reduce the predicting error. The cfd (Weibull, gamma and lognormal distributions) and evd (Frechet and Gumbel distributions) were used to fit the daily maximum concentration. The best distribution that can fit the data was selected based on performance indicators. Furthermore, the exceedences of a critical PM_{10} concentration over the Malaysian Ambient Air Quality Guidelines were estimated using the best distributions. The results of performance indicators show that the extreme value distribution gives better fit to the actual high PM_{10} concentration than the central fitting distribution. The exceedences over a high particulate event were successfully predicted. In 2002, the exceedences is 291 days, 224 days in 2003, 151 days in 2004, 156 days in 2005 and 9 days in 2006.**

Keywords*- central fitting distribution; Gumbel distribution; Frechet distribution; method of moments; daily maximum concentration

I. INTRODUCTION

In recent years, statistical analysis and probability distributions have been used widely in the analysis of air pollution data to understand the current situation of air quality and to predict future air quality. There are many types of probability distributions that have been used to fit air pollutant data, such as the Weibull distribution [1,2], the gamma distribution [3,4] and the lognormal distribution [2,5]. Weibull, gamma and lognormal distributions can give a good result for fitting the mean concentration of air pollutants data. However, it cannot precisely fit the pollutants data when the concentration is high. Extreme value distribution is usually used for fitting the high concentration of air pollutants data.

The extreme value theory (EVT) which is used in storm, flood, wind, sea waves, and earthquake estimation, dates back to the pioneering works by Frechet in 1927 and Fisher and Tipett in 1928 [6]. This theory was extensively developed by Gumbel in 1958 following the extremal type theorem originated by Ginedenko in 1943. The EVT concerns probability calculations and the statistical inference associated with the extreme values of random processes [7].

The EVT has mostly been applied in hydrology for the statistical treatment of floods and droughts [8]. Besides, EVT has also been widely used in wind speed [9,10], health [7,11] and air pollution studies [11,3].

Gumbel distribution was applied to fit CO data in India in order to predict violations of air quality standards at urban road intersections [8]. The results showed that the Gumbel distribution gave satisfactory performance for predictions of extreme air pollution events. In addition, the extreme value theory was successfully used to fit the monthly maximum data and high concentration data of air pollutant concentration over a specific percentile in China [3]. Furthermore, research conducted in Switzerland to model indoor radon distributions in using EVT showed that the EVT is relevant in areas characterized by high mean concentrations, while lognormal distributions seem to be more relevant in small or medium concentration areas [12].

Therefore, the aim of this study is to compare between the central fitting distribution (CFD) that is the Weibull, gamma and lognormal distributions with the extreme value distributions (EVD) that are Gumbel and Frechet distributions.

II. EXPERIMENTAL SET UP

A. Area of Study

The station selected for this research is Seberang Perai (SP). SP is situated in the north part of Peninsular Malaysia and is categorized as an industrial area. Therefore, PM10 concentration is expected to originate mostly from the industrial emission as well as the vehicles emission. Penang is a small state in Peninsular Malaysia but the estimated population density is high. For every square kilometer, there are 1274 inhabitants in Penang. Fig. 1 displays the location of the monitoring site in SP and its description is as in Table 1.

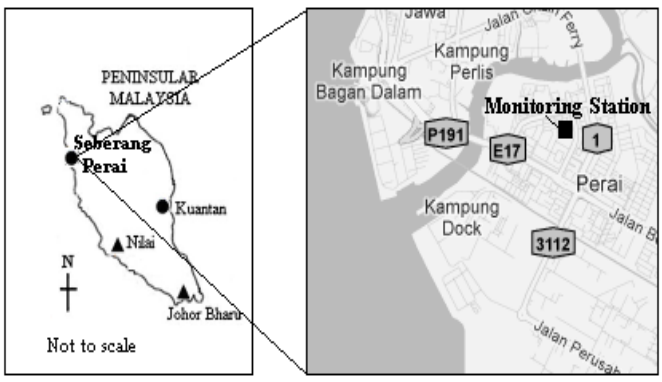

Fig. 1 Location of monitoring site

TABLE 1 MONITORING SITE DESCRIPTION

Site	Location	Coordinate	State Area (km^2)	Population Density (inhabitant/ km^2)
Seberang Perai, Penang	Industrial area	N 05° 23.4704 E 100° 23.1977	1,031	1274

B. The Data

The air quality monitoring stations in Malaysia are strategically located in residential, urban, and industrial areas to detect any significant change in the air quality which may be harmful to human health and the environment. SP station is located in a heavily industrialized area of Seberang Perai. The samples of PM10 were collected by using continuous particulate monitor BAM 1020 (Met One Instruments, Inc.). This instrument automatically measures and records hourly PM10 concentration levels (in milligrams or micrograms per cubic meter) using the industry proven principle of beta-ray attenuation. The data recorded are regularly subjected to standard quality control processes and quality assurance procedures by the Department of Environment (DoE), Malaysia. In order to achieve the aim of this study, the maximum daily PM10 concentrations were selected for year 2002 until 2006. Therefore, the total number of data for 1 year is 365.

C. The Methodology

Fig. 2 illustrates the flow to obtain the best distribution that can represent PM10 daily maximum concentration data. First, the input data was prepared by selecting the maximum concentration for each day in year 2002 to 2006. This data were then used to be fitted with the cfd and evd. For cfd, Weibull, gamma and lognormal distribution were applied, meanwhile for evd, Gumbel and Frechet distribution were used. Method of moments (MoM) was used to estimate parameters for the cfd and evd. In order to select the best distribution that can fit well the input data, performance indicators (PI) that are mean absolute error (MAE), normalized absolute error (NAE), prediction accuracy (PA), coefficient of determination (R2) and index of agreement (IA) were used. For MAE and NAE, values that are closer to zero indicate the best distribution. Conversely, for PA, R2 and IA, values closer to 1 indicate the best distribution. Result of PI for cfd and evd were compared and the best distribution was selected. Finally, the predicted exceedences were estimated with using the best distribution.

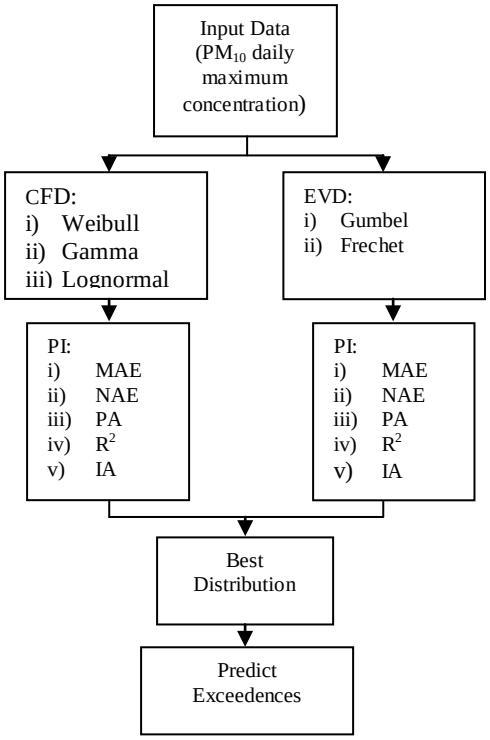

Fig. 2 The flow of the methodology

III. THE WEIBULL, GAMMA AND LOGNORMAL DISTRIBUTIONS

The two parameters Weibull, gamma and lognormal cumulative distribution function (cdf) and probability density function (pdf) with parameters α and β is given in Table 2 [3]. The distributions parameters α and β were estimated with using the method of moments (MoM).

TABLE 2 CDF AND PDF EQUATIONS FOR WEIBULL, GAMMA AND LOGNORMAL DISTRIBUTIONS

Distributions	Cdf equations, F(x)	Pdf equations, f(x)
Weibull	$1-\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], x>0, \alpha>0, \beta>0$	$\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], x>0, \alpha>0, \beta>0$
Gamma	$\int_0^x \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\frac{x}{\beta}\right), x\geq 0, \alpha>0, \beta>0$	$\frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\frac{x}{\beta}\right), x\geq 0, \alpha>0, \beta>0$
Log-normal	$\frac{1}{2\pi}\int_{-\infty}^{\frac{\ln x-\beta}{\alpha}} e^{\frac{-x^2}{2}}dx,\ x>0, \alpha>0, \beta>0$	$\frac{1}{x\alpha\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\beta}{\alpha}\right)^2\right], x>0, \alpha>0, \beta>0$

A. The MoM for Weibull Distribution

The average or expectation of a function of a random variable x can be found by weighting the function by its density or mass function. This procedure is called the method of moments [6]. Since the basic Weibull model has two parameters, estimation of the parameters can be obtained using the sample mean and sample variance.

Using the expression for the mean and variance, β was obtained by the solution of Equation (1).

$$\frac{s^2}{\bar{x}} = \frac{\Gamma\left(1+\frac{2}{\beta}\right)}{\Gamma\left(1+\frac{1}{\beta}\right)^2} - 1 \tag{1}$$

α is then calculated by the following equation:

$$\alpha = \frac{\bar{x}}{\Gamma\left(1+\frac{1}{\beta}\right)} \tag{2}$$

B. The MoM for Gamma Distribution

The equations for this method are as follows [13]:

$$\beta = \frac{1}{cv^2} \tag{3}$$

Where cv = $\frac{s}{\bar{x}}$ (the coefficient of variation)

α is the solution of:

$$\alpha\beta = \bar{x} \tag{4}$$

The values of α and β in this distribution can also be calculated by Equation (5) and (6).

$$\alpha = \frac{s^2}{\bar{x}} \tag{5}$$

$$\beta = \left(\frac{\bar{x}}{s}\right)^2 \tag{6}$$

C. The MoM for Lognormal Distribution

In this method, α and β were obtained directly from Equation (7) and (8), [13].

$$\alpha = \sqrt{\ln(s^2 + (\bar{x})^2) - 2\ln(\bar{x})} \tag{7}$$

$$\beta = \ln(\bar{x}) - \frac{\alpha^2}{2} \tag{8}$$

IV. The Extreme Value Distribution (EVD)

A. The Gumbel Distribution

The Gumbel distribution was extensively developed and applied to flood flows by Gumbel in 1954 and 1958. This distribution results from any underlying distribution of the xi's of the exponential type [6].

The probability density functions for the Gumbel distribution is as follows [6];

$$f(x) = \frac{1}{\beta}\exp\left[-\frac{x-\delta}{\beta} - \exp\left(-\frac{x-\delta}{\beta}\right)\right], \quad -\infty < x < \infty, -\infty < \delta < \infty, \beta > 0 \tag{9}$$

The cumulative distribution function for the Gumbel distribution is as follows;

$$F(x) = \exp\left[-\exp\left(\frac{x-\delta}{\beta}\right)\right], \quad x > 0, -\infty < \delta < \infty, \beta > 0 \tag{10}$$

The location parameter, δ is the mode of the distribution [6],

$$\frac{df(x)}{dx} = 0 \text{ for } x = \delta \tag{11}$$

The parameter β is a measure of dispersion, and it only depends on the variance of Xmax. The moment generating function is found to be;

$$M_{X_{\max}}(x) = \exp(\delta x)\Gamma(1-\beta x),\ x < 1/\beta \tag{12}$$

Therefore, the mean, E(Xmax) and variance, Var(Xmax) of Xmax are as follows;

$$\mathrm{E}(X_{max}) = \mu = \delta + n_e \beta \tag{13}$$

$$\mathrm{Var}(X_{max}) = \sigma^2 = \frac{\pi^2\beta^2}{6} \tag{14}$$

Where ne = 0.5772 (Euler constant)

As a result, from Equation (13) and (14), β and δ were obtained by the following equations;

$$\beta = \frac{\sqrt{6}}{\pi}\sigma \tag{15}$$

and,

$$\delta = \mu - n_e\ \beta = \mu - \frac{n_e\sqrt{6}}{\pi}\sigma \tag{16}$$

B. The Frechet Distribution

The Frechet distribution was first developed and applied to flood flows by Frechet in 1927. The probability density function of the Frechet distribution is as follows [6];

$$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1}\exp\left[-\left(\frac{\beta}{x}\right)^{\alpha}\right], \quad x > 0, \alpha > 0, \beta > 0 \tag{17}$$

The cdf form of the Frechet distribution is as shown below;

$$F(x) = \exp\left[-\left(\frac{\beta}{x}\right)^{\alpha}\right], \quad x > 0, \alpha > 0, \beta > 0 \tag{18}$$

The scale and shape parameters (α and β) in the Frechet distribution was also estimated by using the method of moment (MoM).

In this method, the coefficient of variation needs to be identified first. The coefficient of variation (cv) is the ratio of the sample standard deviation to the sample mean. The formula to estimate the standard deviation and sample mean are as in Equation (19) and (20), respectively;

Standard deviation,

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(n-1)}} \tag{19}$$

$$\text{Mean, } \bar{x} = \sum_{i=0}^{n} \frac{x_i}{n} \tag{20}$$

$$\text{cv} = \frac{s}{\bar{x}}$$

α is obtained by the following equation;

$$\text{cv} = \sqrt{\frac{\Gamma\left(1-\frac{2}{\alpha}\right)}{\Gamma^2\left(1-\frac{1}{\alpha}\right)} - 1} \quad , \alpha > 2 \tag{21}$$

β is the solution of;

$$\beta = \frac{\bar{x}}{\Gamma\left(1-\frac{1}{\alpha}\right)} \tag{22}$$

V. RESULTS AND DISCUSSION

A. Time Series Plot

Fig. 3 shows the time series plot of daily maximum PM10 concentration at the study site for year 2002 to 2006 and Table 3 shows the descriptive statistics. Malaysia that is located at the equatorial, experience uniformed temperature, high humidity and copious rainfall. Changes of wind flow patterns determine the seasons in this country. The wind over the country is generally light and variable. However, there are some uniform periodic changes in the wind flow patterns that describe the four seasons experienced by the country namely, northeast monsoon (November to March), transitional period (April to May), southwest monsoon (June to September), and another transitional period (October to November). PM10 concentration during southwest monsoon is expected to be high in the area of study as the effect of dry weather condition. Transboundary sources aggravate this situation. A study conducted on chemical characterization of the haze in Brunei had found that particulate matter was the dominant pollutants during haze episodes [14]. In the case of haze event in 1997, the particles come from biomass burning in Indonesia for clearing vegetated (forest and grassland) areas. The wild fires significantly increase the input of organic aerosol components to the atmosphere [15]. Therefore PM10 concentration is high during haze episodes.

TABLE 3 THE DESCRIPTIVE STATISTICS FOR DAILY MAXIMUM OF PM10 CONCENTRATION AT SEBERANG PERAI FOR 2002 TO 2006

Descriptive Statistics	2002	2003	2004	2005	2006
N	365	365	365	365	365
Mean	128	145	161	144	80
Median	122	139	159	154	74
Std. Deviation	41	37	40	71	28
Variance	1682	1375	1639	5105	787
Minimum	64	17	68	31	41
Maximum	540	362	421	425	222
Range	476	345	353	394	181

The maximum concentration was recorded during the southwest monsoon (Jun to September) in 2002, 2005 and 2006 as indicated in Fig. 2. In 2003 and 2004, the maximum concentration occurs in the month of February and January, respectively, which falls under the northeast monsoon. However, Table 4 explains that the maximum mean concentration occurs either during the southwest monsoon or during the transition period. Therefore, the maximum concentration in 2003 and 2004 does not represent the overall situation in that year and this might be because of outliers due to ad hoc event, e.g. open burning.

The time series plot for 2005 shows a sudden decrease of PM10 concentration caused by relocation of monitoring site by the DoE Malaysia. The new station is located at about 2 km radius from the old station. The mean for PM10 daily maximum concentration in 2005 for January until July is 196 µg/m3, and from August to December is only 72 µg/m3. The effect of relocation of the monitoring stations is clearly seen after 2005 when PM10 concentration was reduced significantly in 2006. The ranges of PM10 daily maximum concentration from 2002 to 2005 were between 345µg/m3 to 476µg/m3, but in 2006 the range had been reduced to only 181µg/m3.

TABLE 4 THE MONSOONAL MEAN OF DAILY MAXIMUM PM10 CONCENTRATION

***MONSOON**	**Year**				
	2002	2003	2004	2005	2006
1	118.34	133.38	151.76	142.43	72.38
2	122.18	141.66	161.13	175.67	74.68
3	141.98	160.60	169.46	150.23	89.88
4	132.94	144.58	178.00	62.35	90.52

*1 Northeast monsoon, 2 Transition (Apr-May), 3 Southwest monsoon, 4 Transition (Oct)

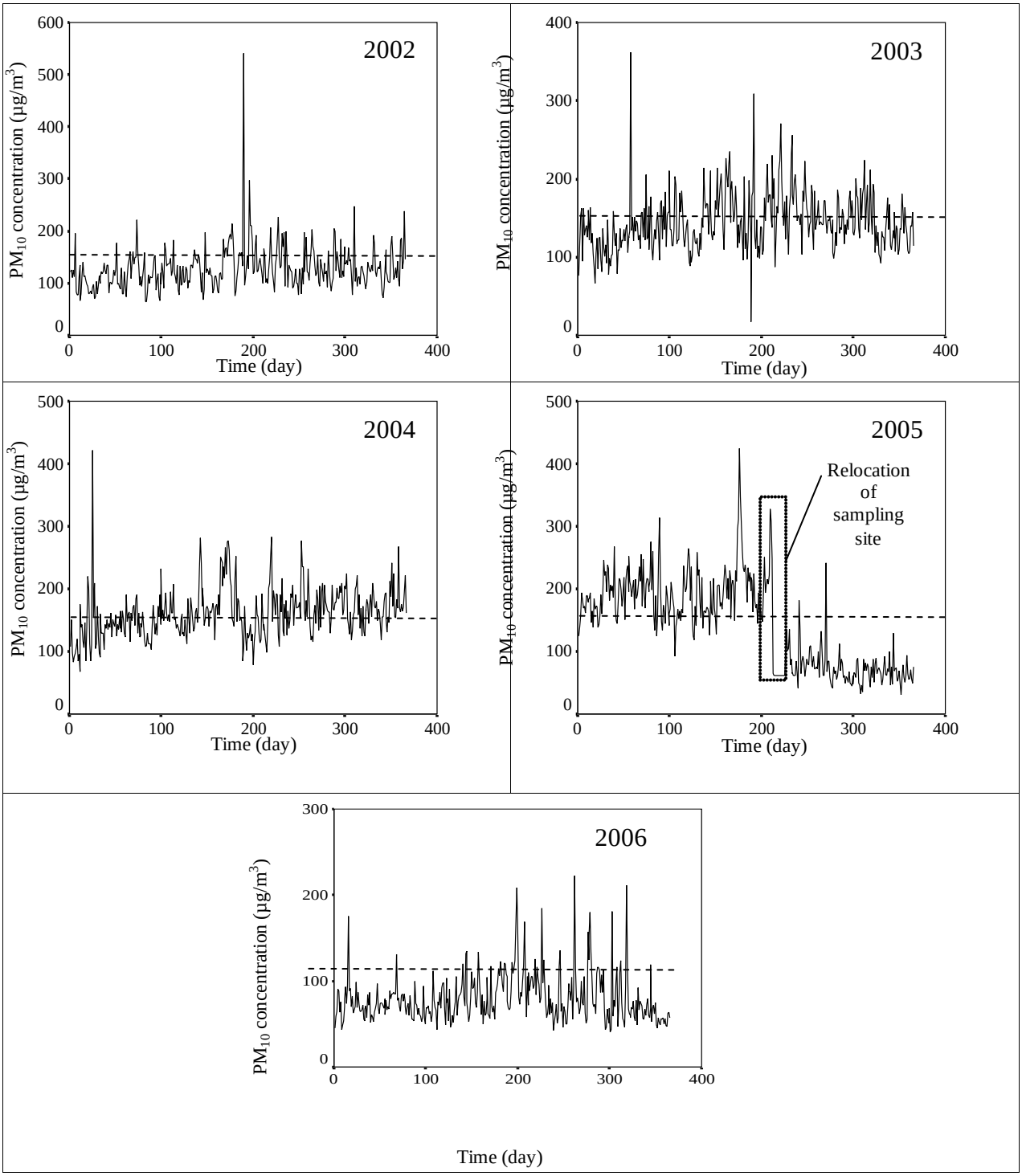

Fig. 3 The time series plot for daily maximum of PM_{10} concentration at Seberang Perai for 2002 to 2006

TABLE 5 PARAMETERS FOR WEIBUL, GAMMA, LOGNORMAL AND EXTREME VALUE DISTRIBUTION USING MOM

Distributions	CFD						EVD			
	Weibull		Gamma		Lognormal		Gumbel		Frechet	
Parameters	α	β	α	β	α	β	β	δ	α	β
2002	3.46	142	13.1	9.76	0.312	4.80	32.0	110	4.93	110
2003	4.42	159	9.49	15.3	0.252	4.94	28.9	128	5.89	128
2004	4.53	177	10.1	15.9	0.247	5.05	31.6	143	5.99	143
2005	2.12	162	35.5	4.05	0.470	4.86	55.7	112	3.60	114
2006	3.13	89.6	9.82	8.16	0.340	4.33	21.9	67.5	4.60	67.7

TABLE 6 GOODNESS OF FIT CRITERIA

Year	Distributions		MAE	NAE	PA	R^2	IA
2002	CFD	Weibull	9.1733	0.0716	0.8898	0.7874	0.9417
		Gamma	6.6611	0.0520	0.9253	0.8516	0.9612
		Lognormal	5.5267	0.0431	0.9387	0.8762	0.9683
	EVD	Gumbel	4.9803	0.0389	0.9431	0.8845	0.9706
		Frechet	4.8310	0.0377	0.9625	0.9213	0.9793
2003	CFD	Weibull	6.4169	0.0443	0.9568	0.9104	0.9779
		Gamma	3.0576	0.0211	0.9827	0.9604	0.9912
		Lognormal	2.2677	0.0157	0.9871	0.9691	0.9934
	EVD	Gumbel	2.0950	0.0145	0.9889	0.9726	0.9943
		Frechet	5.4840	0.0379	0.9776	0.9505	0.9876
2004	CFD	Weibull	9.3759	0.0581	0.9629	0.9221	0.9763
		Gamma	2.7837	0.0172	0.9825	0.9601	0.9906
		Lognormal	2.5561	0.0158	0.9870	0.9689	0.9932
	EVD	Gumbel	2.8473	0.0176	0.9889	0.9725	0.9943
		Frechet	6.1556	0.0381	0.9776	0.9505	0.9877
2005	CFD	Weibull	13.4575	0.0936	0.9780	0.9512	0.9889
		Gamma	15.9995	0.1113	0.9683	0.9324	0.9838
		Lognormal	19.9595	0.1388	0.9505	0.8984	0.9744
	EVD	Gumbel	17.2847	0.1202	0.9632	0.9226	0.9811
		Frechet	30.7316	0.2137	0.8823	0.7743	0.9333
2006	CFD	Weibull	7.0631	0.0881	0.9303	0.8607	0.9639
		Gamma	4.6765	0.0583	0.9682	0.9323	0.9838
		Lognormal	3.3117	0.0413	0.9829	0.9608	0.9912
	EVD	Gumbel	3.0801	0.0384	0.9846	0.9641	0.9921
		Frechet	2.5359	0.0316	0.9945	0.9836	0.9953

B. Parameters for CFD and EVD

Table 5 displays parameters obtained for CFD and EVD using MoM. The best distribution that can fit the daily maximum concentration was selected based on goodness of fit criteria in Table 6.

From the result of goodness of fit with using mean absolute error (MAE), normalised absolute error (NAE), Prediction Accuracy (PA), coefficient of determination (R2) and Index of Agreement (IA) in Table 6, it is clear that the EVD fits the high concentration better than the CFD, except for 2005. For MAE and NAE, value closer to zero indicates a better distribution, whereas for PA, R2 and IA, value closer to 1 indicates a better distribution. In this study, 2002 and 2006 show Frechet as the best distribution and 2003 and 2004 are better fit with the Gumbel distribution.

C. cdf, pdf and Probability of Exceedences

Fig. 4 illustrates the cdf plot for 2002 to 2006 with using the best distributions. It shows that the EVD can fit the actual data very well except for the year 2005 where the best distribution is Weibull. This is due to inconsistency of data recorded when Seberang Perai station was relocated as mentioned in 3.1 The Time Series Plot. PM10 concentration decrease when the station was transferred to a new site. Starting from August 2005, the daily maximum concentrations exceed MAAQG only twice (Fig. 3).

From Fig. 4, the exceedences of PM10 concentration that is more than the Malaysian Ambient Air Quality Guideline (MAAQG) was estimated. For the year 2002, the probability for PM10 concentration more than 150μg/m3 is equal to 0.7973 (F(x>150) = 0.7973). Therefore, the exceedences or number of days that PM10 concentration is more than MAAQG is 291 days. The probability of exceedences for 2003 is 0.6148 (F(x>150) = 0.6148) and the number of days that exceed MAAQG is 224 days. For 2004, the probability of exceedences is 0.4153 (F(x>150) = 0.4153) with 151 days exceed the MAAQG.

For 2005 and 2006, the probability of exceedences is 0.4286 (F(x>150) = 0.4286) and 0.0254 ((F(x>150) = 0.0254) respectively. Thus, the number of days that exceed MAAQG is 156 days for 2005 and 9 days for 2006.

The exceedences obtained in this study show a decreasing trend. However, it does not indicate that air quality in the study area is improving. The reason for this is the data that were used is the daily maximum data, not the daily average data. Thus, the maximum concentration might occur only once a day, due to ad hoc event such as open burning. Hence, the data are not the indication of overall air quality of the study area. Furthermore, it is obvious that the decreasing trend is also because of relocation of sampling site.

pdf plots by using method of moment for EVD show almost similar distribution with long tail to the right (Fig. 5). This pattern indicates that there are dominant sources that contribute to high PM10 concentration in Seberang Prai and increasing with time. However, this is not peculiar as it is known that Seberang Prai is a heavily industrialized area and this sector is believed to have become a major contributor to air pollution problem.

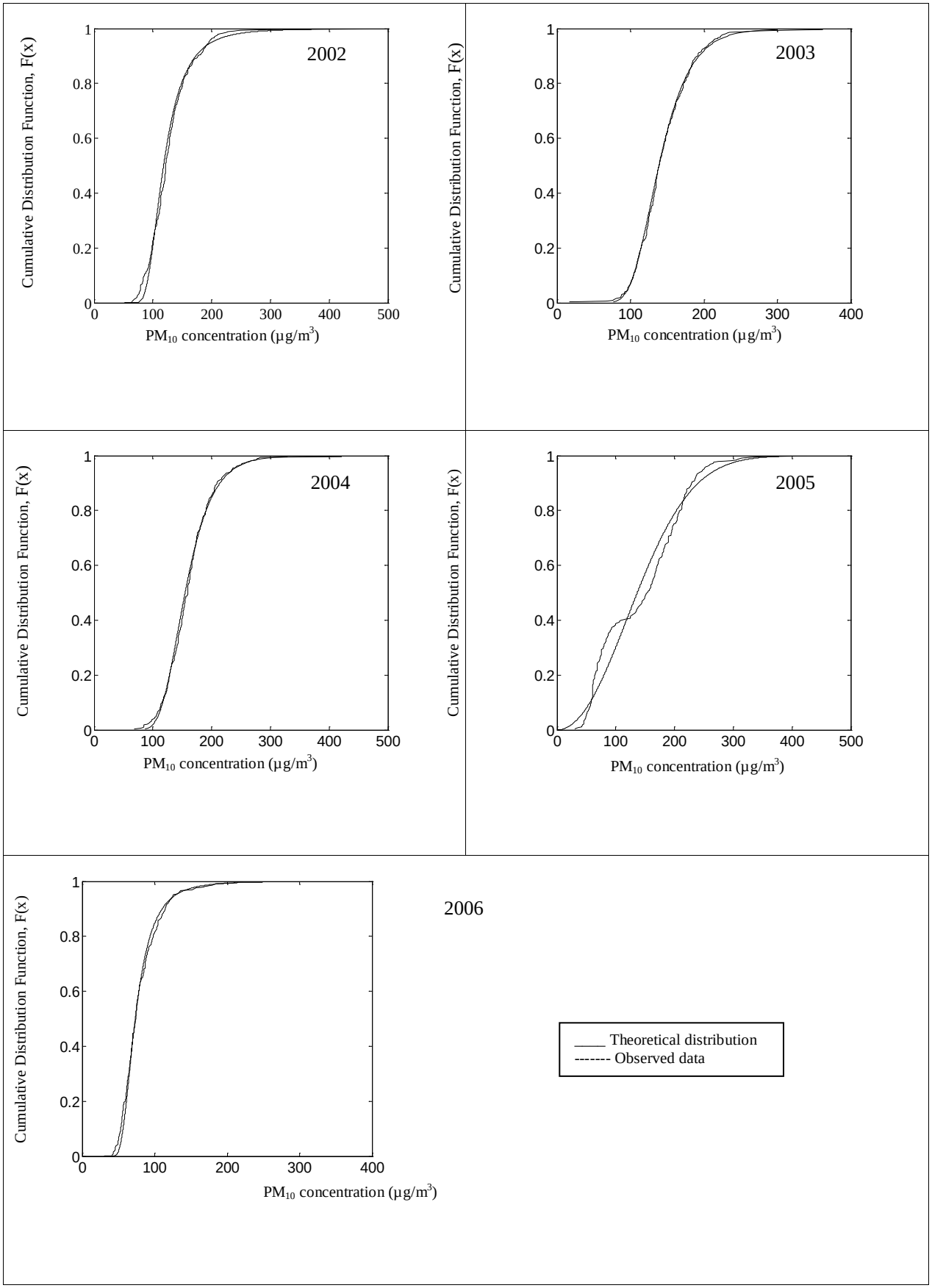

Fig. 4 cdf plots for 2002 to 2006

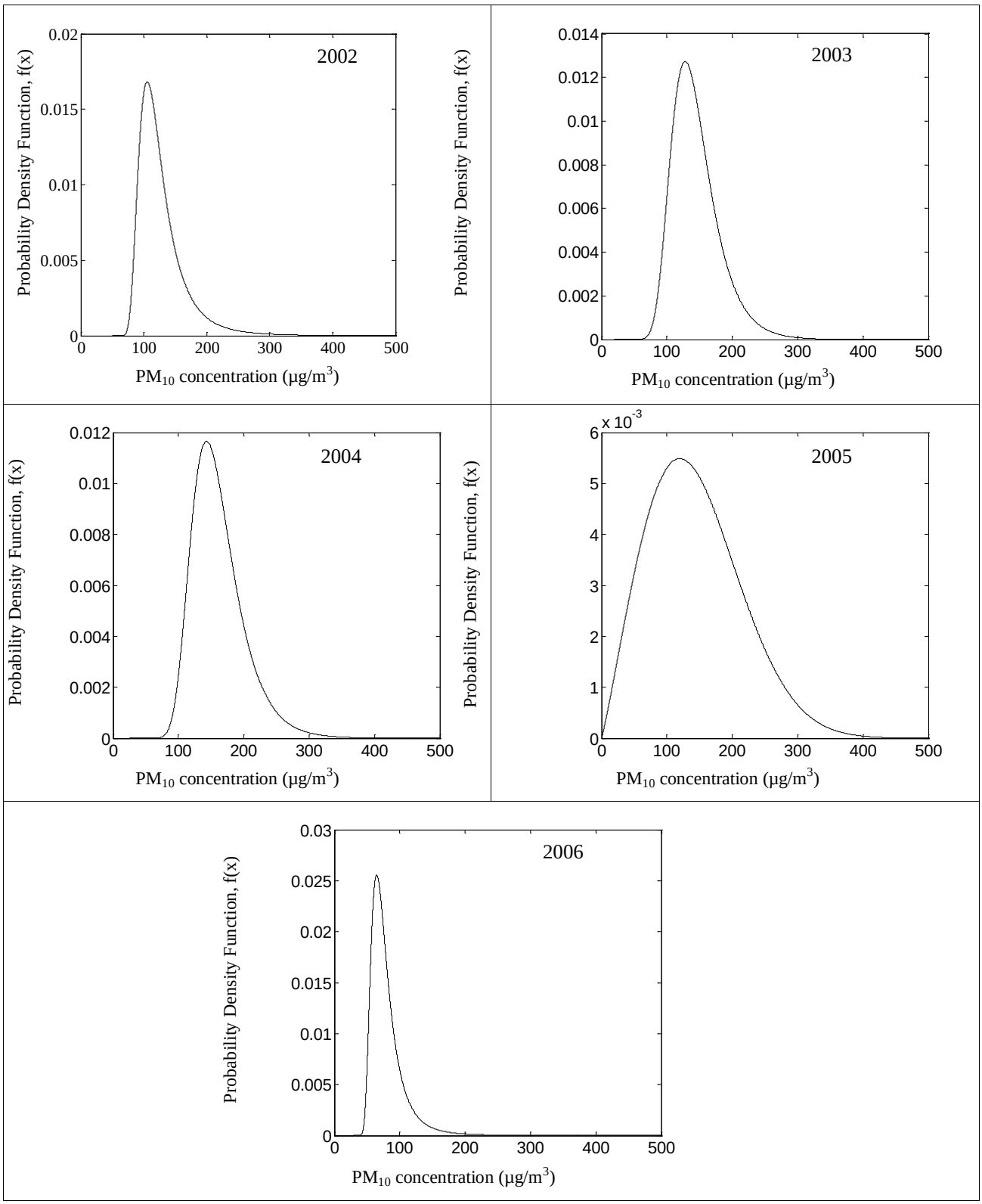

Fig. 5 pdf plots for 2002 to 2006

VI. CONCLUSIONS

In environment, particularly air pollution, researchers are more concern on high pollutants concentration because it can affect human health as well as the ecosystem. This study proved that EVD gives better fit than the CFD for the daily maximum PM10 concentration from 2002 to 2006 except for 2005 when there is inconsistency of data recorded due to relocation of sampling site. Therefore, the prediction for future air quality for high PM10 concentration was more accurate. It was found out that the exceedences or number of days when PM10 concentration is over the MAAQG for 2002 is 291 days, 2003 is 224 days, 2004 is 151 days, 2005 is 156 days and 2006 is 9 days.

REFERENCES

[1] Wang, X. & Mauzerall, D. L., Characterizing Distributions of Surface Ozone and Its Impact on Grain Production in China, Japan and South

Korea: 1990 and 2020. Atmospheric Environment. 38, p. 4383-4402, 2004

[2] Lu, H. C., The Statistical Characters of PM_{10} Concentration in Taiwan Area. Atmospheric Environment. 36, p. 491-502, 2000

[3] Lu, H. C., Estimating the Emission Source Reduction of PM_{10} in Central Taiwan. Chemosphere, 54, p. 805-814, 2003

[4] Karaca, F., Alagha, O., Erturk, F., Statistical characterization of atmospheric PM_{10} and $PM_{2.5}$ concentrations at a non-impacted suburban site of Istanbul, Turkey. Chemosphere. 59, p.1183-1190, 2005

[5] Hitzenberger, R. and Tohno, S., Black carbon (BC) concentrations and size distributions at two urban sites (Uji, Japan and Vienna, Austria). Journal of Aerosol Science, 31,Suppl.p.112-113, 2000

[6] Kottegoda, N. T. & Rosso, R., Statistics, Probability and Reliability for Civil and Environmental Engineers. Singapore: McGraw-Hill Book Co., 1998

[7] Leong, Y.P., Sleigh, J.W. and Torrance, J.M., Extreme value theory applied to postoperative breathing pattern. British Journal of Anaesthesia, 88, p.61-4., 2002

[8] Sharma, P., Khare, M., Chakrabarti, S.P., Application of extreme value theory for predicting violations of air quality standards for an urban road intersection. Transportation Research Part D, 4, p.201-216, 1999

[9] Whalen, T.M., Savage, G.T., Jeong, G.D., An evaluation of the self-determined probability-weighted moment for estimating extreme wind speeds. Journal of Wind Engineering and Industrial Aerodynamic, 92, p.219-239,2004

[10] Xiao, Y.Q., Li, Q.S., Li, Z.N., Chow, Y.W., Li, G.Q., Probability distributions of extreme wind speed and its occurrence interval. Engineering Structures, 28, p.1173-1181, 2006

[11] Park, H.W., and Sohn, H., Paramater estimation of the generalized extreme value distribution for structural health monitoring. Probabilistic Engineering Mechanics, 21, p.366-376, 2006

[12] Tuia, D., and Kanevski, M., Indoor radon distribution in Switzerland: lognormality and Extreme Value Theory. Journal of Environmental Radioactivity, 99, p.649 – 657, 2008

[13] Evans, M., Hastings, N., and Peacock , B., Statistical distributions (3^{rd} edition) New York: Wiley, 2000

[14] Muraleedharan, T. R., Radojevic, M., Waugh, A. and Caruana, A., Chemical characterization of the haze in Brunei Darussalam during the 1998 episode. Atmospheric Environment. 34, p. 2725–2731, 2000

[15] Abas, M. R. B., Rahman, N. A., Omar, N. Y. M. J., Maah, M. J., Samah, A.A., Oros, D. R., Otto, A., Simoneit, B. R. T., Organic composition of aerosol particulate matter during a haze episode in Kuala Lumpur, Malaysia. Atmospheric Environment, 38, p.4223-4241, 2004