Sains Malaysiana 41(11)(2012): 1335–1344

Assessing and Comparing PM_{10} Pollutant Behaviour using Functional Data Approach

(Menilai dan Membanding Perilaku Pencemar PM_{10} dengan Pendekatan Data Fungsi)

NORSHAHIDA SHAADAN, SAYANG MOHD DENI & ABD AZIZ JEMAIN

ABSTRACT

This study highlights the advantage of functional data approach in assessing and comparing the PM_{10} pollutant behaviour as an alternative statistical approach during and between the two extreme haze years (1997 and 2005) that have been reported in Selangor, state of Malaysia. The aim of the study was to improvise the current conventional methods used in air quality assessment so that any unforeseen implicit information can be revealed and the previous research findings can be justified. An analysis based on the daily diurnal curves in place of discrete point values was performed. The analysis results provided evidences of the influence of the change in the climate (due to the El-Nino event), the different levels of different emission sources and meteorological conditions on the severity of the PM_{10} problem. By means of the cummulative exceedence index and the functional depth method, most of the monitoring stations for the year 2005 experienced the worst day of critical exceedences on the 10th of August, while for the year 1997 it occurred between 13th and 26th September inclusively at different dates among the stations.

Keywords: Air quality; behaviour; exceedences; functional data; PM_{10} pollutant

ABSTRAK

Kajian ini memperlihatkan kelebihan pendekatan data fungsi dalam menilai dan membanding perilaku pencemar PM_{10} sewaktu dan antara dua tahun jerebu yang paling ekstrim (1997 dan 2005) yang pernah dilaporkan berlaku di negeri Selangor, Malaysia sebagai satu pendekatan alternatif. Tujuan kajian adalah untuk membuat penambahbaikan terhadap kaedah konvensional yang biasa digunakan dalam penilaian dan penaksiran kualiti udara bagi membolehkan maklumat tersirat yang tidak disedari dicungkil dan dapatan daripada kajian lepas dapat diperjelaskan. Analisis dilakukan ke atas data lengkung harian PM_{10} berdasarkan tempoh mengikut jam bagi menggantikan penggunaan nilai-nilai diskrit titik. Keputusan analisis kajian telah mempamerkan bukti terhadap pengaruh perubahan iklim (ekoran daripada peristiwa El-Nino), perbezaan aras kepekatan berpunca daripada berlainan sumber pelepasan dan keadaan meteorologi terhadap keterukan masalah pencemar udara PM_{10}. Dengan menggunakan indeks kumulatif lebihan dan kaedah kedalaman fungsi, kebanyakan stesen pemantau bagi tahun 2005 didapati mengalami hari yang paling teruk lebihan kritikal kepekatan PM_{10} pada 10 Ogos manakala pada tahun 1997 berlaku pada tarikh yang berlainan di antara 13 dan 26 September di stesen pemantau yang berbeza.

Kata kunci: Data fungsian; kualiti udara; lebihan; pencemar PM_{10}; perilaku

INTRODUCTION

The quality of the life performance of human beings depends on various factors. Clean air is one of the crucial factors but nowadays the atmospheric environment is never perfectly clean due to the various aspects such as urbanisation and industrialisation. Air pollution has become a global problem. In considering its harmful effects in many countries in the world including Malaysia, a monitoring system and various strategies have been established and implemented to assess and evaluate the air quality. The study of air pollutant behaviour, as part of the air quality assessment process is extremely important in order to search for a wise strategy for mitigation. However, the assessment in the current practice has been based on discrete point approach including summary statistics, the box plot, histogram and the analysis of trend. These approaches are found to be inadequate in understanding the underlying process of the pollutant and hence prevent the implicit information from being revealed (Gao 2007). As an example, ADB and CAI-Asia (2006) stated that due to the employment of average values, the exceedences on a daily basis cannot be identified as the information is hidden by the annual averaging of the daily values. In spite of the exceedences and exposure of PM_{10} which are often discussed, the important concept related to the degree of severity to the environment is another concern in this study.

PM_{10}, a class of particulate matter less than 10 micrometer has been identified as being a common problem in many countries all over the world. In Malaysia, PM_{10} is known as one of the prevailing dominant pollutants in the country. Previous researchers have stated that this harmful pollutant is found significantly associated to haze (Afroz et al. 2003; Awang et al. 2000) which has become a repeated typical problem in the country. High concentration levels of it are often recorded during the dry season, especially in urbanised areas for example in the Klang Valley which is located in the Selangor state of Malaysia. Geographical positions, high industrial and commercial activities, high density populations, heavy vehicular activities and stable atmospheric conditions (prevailing winds) were among the related factors that contributed to the problem (Afroz et al. 2003; DoE 2005). Previous research has discussed that there is a linkage between changes in climate, for example the El-Nino phenomenon and the occurrences of haze in the Southeast Asia region (Tangang et al. 2011). According to Field et al. (2009), a severe haze disaster occurred in 1997 during the extreme El-Nino event. Besides El-Nino, the changes in the meteorological conditions and the emission characteristics of the dominant sources (traffic influence and transboundary source) were also the driven factors that affected the PM_{10} concentration behaviour (Laakso et al. 2003). Thus, the aim of this study was to improvise and enhance the current conventional approach in assessing the PM_{10} behaviour using the functional data approach. Data in many scientific experiments are recorded repeatedly through time or space and have been seen to arise as a continuous process. Examples of such a kind of observations are hourly records of the temperature data, the heights of children over a wide range of ages and daily records of precipitation. Nowadays, it is possible to analyze these types of data by transforming them into a continuous function. The data can be considered as observations varying over a continuum. Thus, the data at hand are no longer a set of discrete time point values but a continuous curve or termed as functional data. A new modern statistical methodology to analyze functional data is referred to as Functional Data Analysis (FDA) (Ramsay & Silverman 1997). The FDA concentrates on the tools or statistical methodologies that are used to explore, gain insights, model and test the hypothesis with such data.

Specifically, the study will focus on the PM_{10} problem in the Selangor state of Malaysia. The first objective was to compare the PM_{10} behaviour between two extreme haze years (1997 and 2005) in order to provide the evidence of the influence of a change in climate and of the different levels of emission from the various sources. The second objective was to detect the critical exceedence day and determine the worst day of critical exceedences.

Materials and Methods

Data and Locations

The secondary monitoring records of the hourly PM_{10} concentration measured in μgm^{-3} were used in this study to assess and compare the behaviour of the pollutant during the two extreme haze years (1997 and 2005). The records were obtained from Department of Environment Malaysia (DoE) – Alam Sekitar Malaysia (ASMA) covering January to December at seven air quality monitoring stations that were available during the study period. The information of the stations/sites category (DoE 2008) and the location is represented in Table 1. The units of measurement for longitude and latitude are in degree (°). Due to the station's availability, the data from the five stations for the year 1997: ST1, ST2, ST3, ST4 and ST5 were used in the analysis. Meanwhile, data from all the seven stations including ST6 and ST7 for the year 2005 were also analysed. In the case of ST4 and ST5 for the year 1997, the data collected began from the month of June and July, respectively.

The summary statistics of the data for the year 1997 and 2005 at each station is tabulated in Table 2 to preliminarily describe the characteristics of the data. Overall, at all stations, the average, median and the standard deviation of PM_{10} level were higher in the year 1997 compared with 2005; meanwhile, a larger value of maximum PM_{10} level was recorded in 2005 compared with in 1997. However, information based on the conventional statistics (summary statistics) given in Table 2, is found limited particularly in visualising the whole structure and pattern of PM_{10} over the entire period of time (say for the whole one day period). Thus, to enhance the current approach, a functional data (curve) approach is used as an alternative method of analysis in this study.

Figure 1 shows the map of the seven monitoring stations involved. The data are considered reliable since they have undergone a standard quality control procedure imposed by the DoE. In addition, the recorded data for the year 2000 are also considered as the control reference to allow for comparable phenomena based on the report by DoE (2005) and Ling et al. (2010).

TABLE 1. The monitoring stations used in the study

Monitoring Stations	Category	Latitude (°N)	Longitude (°E)
ST1: Gombak	Residential	3.26	101.65
ST2: Kelang	Residential	3.01	101.41
ST3: Petaling Jaya	Industrial	3.11	101.64
ST4: Kajang	Residential	2.99	101.74
ST5: Shah Alam	Residential	3.10	101.56
ST6: Kuala Selangor	Residential	3.33	101.26
ST7: Cheras	Traffic	3.11	101.72

TABLE 2. The summary statistics of PM_{10} (μgm^{-3}) at the different stations

Year	Station	Started date	Average	Median	Standard deviation	Minimum	Maximum
1997	ST1	January	92.38	70	77.22	1	750
	ST2	January	102.8	85	68.09	8	699
	ST3	January	85.99	73	55.64	8	607
	ST4	Jun	85.32	68	64.18	10	536
	ST5	July	81.02	60	68.16	8	585
2005	ST1	January	52.32	46	41.76	5	739
	ST2	January	78.47	62	68.53	5	994
	ST3	January	64.22	56	47.23	5	806
	ST4	January	50.58	43	41.43	5	950
	ST5	January	64.57	54	56.81	5	995
	ST6	January	61.61	49	59.85	5	995
	ST7	January	56.01	47	47.67	5	879

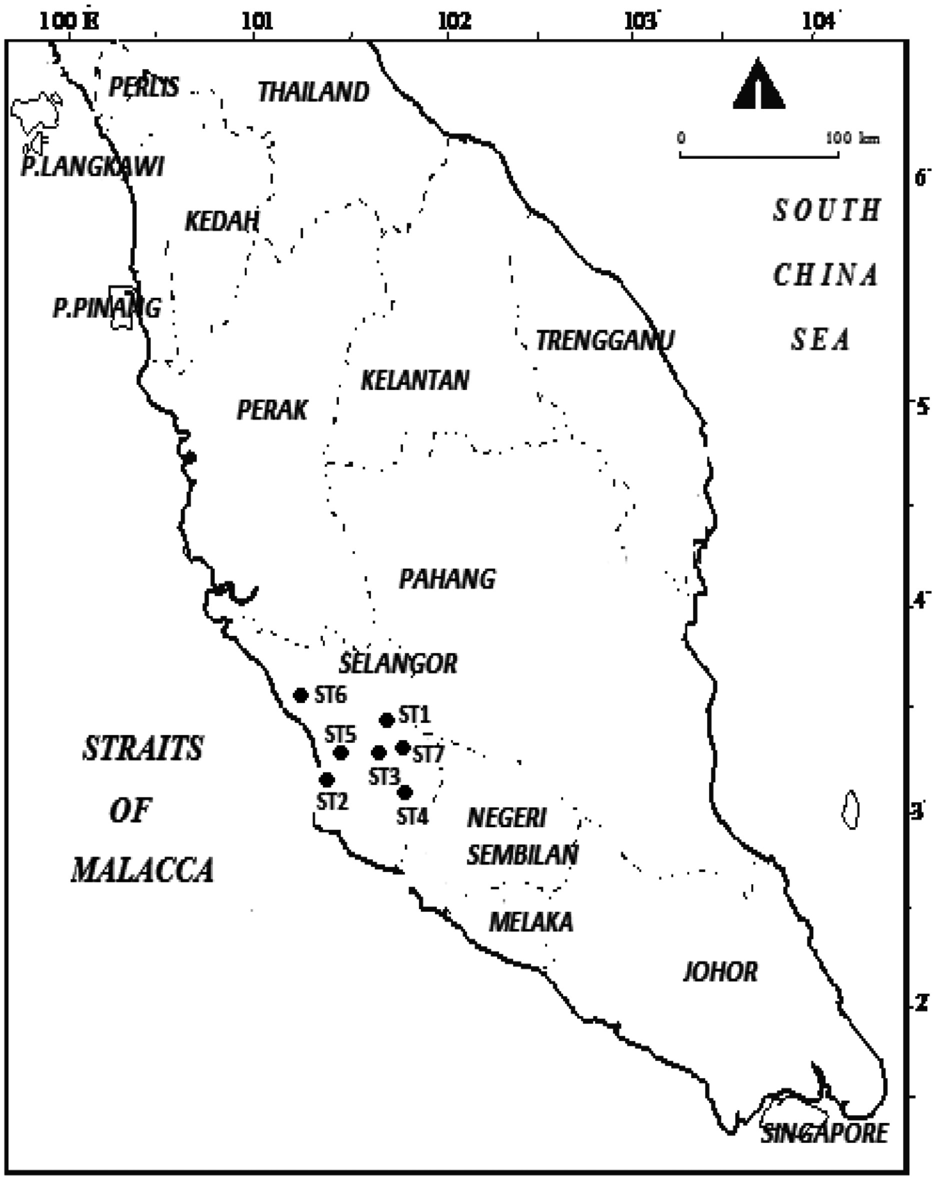

FIGURE 1. Map of the monitoring stations involved
in the study which are located in Selangor

Before the analysis is carried out, the incomplete records due to missing values are treated using an imputation technique namely the nearest neighbour method. The missing percentage for the data set by the different years and stations ranged between 1% and 11%. All the data sets had less than 5% missing rate, except for the data sets from ST4 and ST3 in 1997 with 11% and 6.2% missing, respectively. The nearest neighbour imputation method is considered and used in this study due to its simple scheme and it has also been used in the study by Azmi et al. (2010) for the data set involving the year 1997 until 2006. The employment of this technique is also supported by the research conducted by Junninen et al. (2004). Moreover, the recorded missing rate for the data set was considerably low. In this study, the analysis is conducted using the R software which is available online.

Assessing The Daily Diurnal Behaviour

In this study, a modern approach namely the functional data analysis that has been popularised by Ramsay and Silverman (1997) is considered as the alternative over the current conventional approach in order to observe for any new insight possibilities, to justify previous research findings and also to reveal any unforeseen implicit information. Compared to the conventional approach, the data studied are curves instead of the discrete point values. The fundamental assumption for using this approach is that the observed data follow a continuous process over a certain period of time. Due to the appropriateness of the functional data approach for air quality data, the hourly PM_{10} concentration from the monitoring stations are treated as functional objects or curves in this study. The analysis is conducted based on the 365 daily diurnal curves for each year. To establish the daily curves over the 24 h period [t_1, t_{24}], discrete point values of the PM_{10} hourly data (x) must be treated as a functional form or curve which is defined as follows:

$$\{x_i(t_j);\ i = 1, \ldots, 365;\ j = 1, \ldots, 24\}. \tag{1}$$

INVESTIGATING THE INFLUENCE OF CHANGE IN CLIMATE

Another concern of the study was to investigate whether the diurnal behaviour is influenced by the change in climate and meteorological variation. With respect to the objective, the analysis of identifying the shifting in the levels of the PM_{10} concentration between the two years is conducted. For that purpose, the analysis focusing on a robust location estimator for the centre of the distribution namely the trimmed mean is conducted. The concept of functional trimmed mean is similar to the conventional trimmed mean. Outliers in the data set are trimmed first to obtain the average value with respect to some amount (percentage - α) of the most outlying data. In the context of functional data, to compute the functional trimmed mean, curves are ranked according to decreasing value of their depth (Fraiman & Muniz 2001). Further explanation on the functional depth is discussed in the next section. For a given $\alpha = 0.10$ in this case, to compute the trimmed mean, the less 10% deepest curves are ignored or trimmed. The functional trimmed mean of n-[αn] curves (the remaining curves) is computed using the following equation:

$$\hat{\mu}_{\text{Trimmed}} = \frac{1}{n-[\alpha n]} \sum_{i=1}^{n-[n\alpha]} x_i\left(t_j\right). \tag{2}$$

A bootstrap confidence set for trimmed mean is then used to investigate upon the shifting in the PM_{10} level that might be due to the influence of the change in climate and meteorological variation. A bootstrap confidence set for trimmed mean CS ($\hat{\mu}_{\text{Trimmed}}$) at the confidence level β is defined as the set of curves $c(t)$ which has the same distribution that $\hat{\mu}_{\text{Trimmed}}$ and such that:

$$CS\ (\hat{\mu}_{\text{Trimmed}}) = \{\ c(t) : d\ (\hat{\mu}_{\text{Trimmed}}(t), c(t)\ ) < D_\beta\ \}, \tag{3}$$

where D_β is such that Pr (d ($\hat{\mu}_{\text{Trimmed}}$ (t), $c(t)$ < D_β) = β and d is a functional distance. The computation of d is associated with L_2 distance where d (($\hat{\mu}_{\text{Trimmed}}(t)$, $c(t)$) = $\| \hat{\mu}_{\text{Trimmed}}(t) - c(t) \|_2$.

For this study, the confidence set is established based on 500 bootstrap samples with smoothing parameter $\gamma = 0.05$ and confidence level $\beta = 0.95$. The detailed procedure and further explanation on the bootstrap approach to obtain a bootstrap confidence set can be referred in Febrero et al. (2007).

DETECTING THE DAY OF EXCEEDENCES

In addition, the assessment and comparison of the behaviour in the concept of exceedences are also highlighted in this study. Unlike the conventional approach, the frequency of the exceedence day is obtained based on the continuous nature of the PM_{10} pollutant process. Consequently, the idea in comparing the behaviour in terms of the severity of the PM_{10} condition is proposed by means of an index namely the Exceedence Index (EI). The development of the index is based on the ratio of the sum of the number of hours in a day that exceeds the threshold value of the 1 h average standard limit with the total number of hours per day, written as follows:

$$EI_{ij} = \frac{\text{frequency of } t_j \text{ with } x_i(t_j) > 200\ \mu\text{gm}^{-3}}{24}, \tag{4}$$

where the range of EI_{ij} is $0 \leq EI_{ij} \leq 1$ for $j = 1, \ldots, 24$.

Since Malaysia does not have the 1 h average standard yet, thus, the Japan Air Quality Standard (JAQS) is used as a proxy in this study (MOE 2011). The value of $E = 1$, indicates that the concentration has exceeded the permissible limit at all entire hours of the day; therefore, this day is called as the critical exceedence day. This particular day can be represented by a curve that lies totally above the JAQS line. An estimate measure for the duration of exposure is proposed to be based on the Cumulative Exceedence Index (*CEI*), which represents the length (proportion) of time in a day that people are potentially exposed to the unsafe condition of PM_{10}. By employing *EI* and *CEI*, the behaviour in terms of the condition of the PM_{10} problem could be assessed and compared between the two years among the stations in the study area.

THE RANKING OF THE CRITICAL EXCEEDENCE DAYS

In the context of the objective to identify the worst day of critical exceedences, the extended notion of depth for functional data which was introduced by Fraiman and Muniz (2001) is used. Originally, the concept of depth was introduced to measure the centrality of a data point within a data cloud in multivariate analysis. The concept of depth provides the ideas of order statistics or rank for which points in the Euclidean space are ordered from the centre to the outward. The nearer the point to the centre, the higher the depth; the further away the point from the centre, the lower the depth. The Fraiman and Muniz (FM) functional depth is used to measure the centrality of a given curve, x_i within a set of curves, $x_1, x_2, \ldots, x_n$ (Febrero et al. 2007). Thus, the corresponding FM functional depth of a curve (x_i) is given by the following equations:

$$D_{\mathrm{FM}}\left(x_i\right) = \int_a^b D\left(x_i\left(t\right)\right) dt, \tag{5}$$

where $D(x_i(t))$ is the FM depth measure at specific time t and is given by:

$$D\left(x_i\left(t\right)\right) = 1 - \left|\frac{1}{2} - F_{n,t}\left(x_i\left(t\right)\right)\right|, \tag{6}$$

and $F_{n,t}(x_i(t))$ is the empirical cumulative distribution function of the values of the curves $x_1(t), x_2(t), \ldots, x_n(t)$ at a given time point $t \in [a = 1, b = 24]$, such that

$$F_{n,t}\left(x_i\left(t\right)\right) = \frac{1}{n}\sum_{k=1}^{n} I\left(x_k\left(t\right) \leq x_i\left(t\right)\right), \tag{7}$$

where $I(.)$ is an indicator function which equals to 1 when the property (.) holds and equals 0 otherwise. The depth and outlyingness are opposite notions. The lower the depth, the more the outlying will be a curve from its centre. In this study, the depth measure obtained from Equation (5) is used to rank the days of the critical exceedences and the day with the lowest depth is identified as the worst day. The deepest curve is the one which attains the maximum $D_{FM}(x)$ while the most outlying curve is the one with the minimum $D_{FM}(x)$. In the context of air pollution data particularly for the case of the critical exceedence days which represent the severest condition of PM_{10} within the 24 h period, it is believed that the further away the curve from the centre indicates the higher the magnitude and the intensity level when compared with the vast majority of the considered curves. Therefore, the least deep curve is the curve having the largest magnitude and lies outside the range of the rest; consequently, this curve represents the worst day of the critical exceedences.

RESULTS AND DISCUSSION

Figure 2 shows that the PM_{10} behaviour differed between the haze years (1997 and 2005) and the non haze year (2000). Two significant groups of curve patterns across the hours during the haze years existed. A group of curves with consistent behaviour was observed at the lower level of the PM_{10} concentration. On the other hand, a sparser curve at the higher level indicated extreme

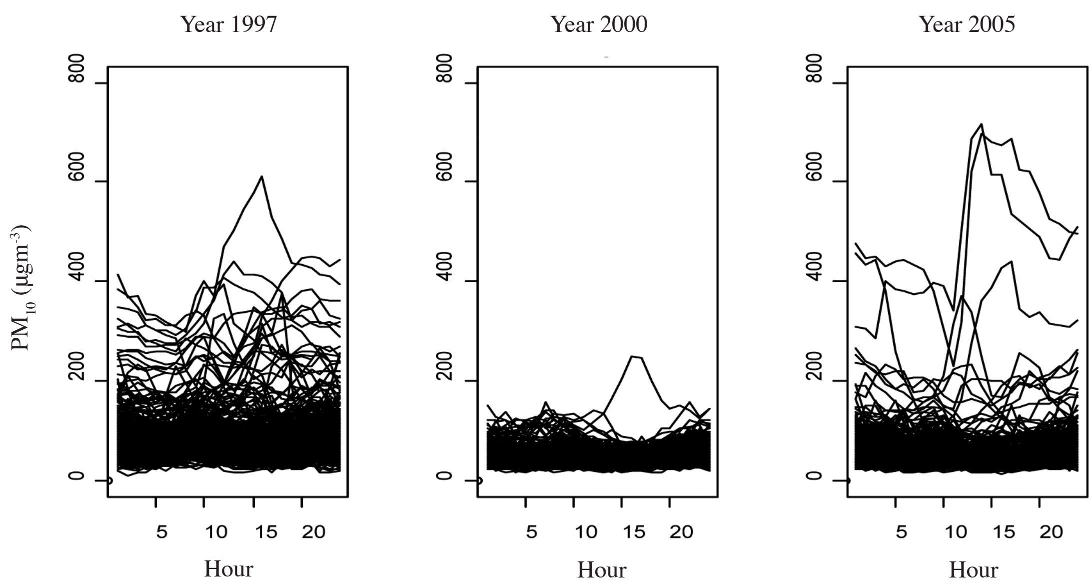

FIGURE 2. The daily diurnal PM_{10} curves

behaviour. Thus, the obvious difference in the behaviour of the diurnal daily curves might be attributed to the different dominant emission source. The changes in the meteorological condition caused the occurrence of the maximum peak at around 15:00 in the afternoon for all the three years whereby the peak in the year 2005 was the highest.

Figure 3 exhibits the overall daily diurnal behaviour of PM_{10} with a similar daily maximum occurring between 8.00 and 10:00 am. Rationally, the condition might be due to the morning rush hour as the majority of the stations are located in the residential areas, one in the industrial and one in the traffic area. During this time, the number of motor vehicles is expected to increase as people have started their day time activity (adults go to work while children go to school). The significant shift between the trimmed mean curves was shown whereby the year 1997 had the highest, followed by the year 2005. The self explanatory independent trimmed mean concentration curves with different interval widths indicated that 1997 was the worst year due to the El-Nino phenomena as compared with the year 2000 and 2005. The findings supported the study of Field et al. (2009) which indicated that the severity of the PM_{10} problem was influenced by the change in the climate and the different levels of emission source. The results obtained from Figures 2 and 3 are among the implicit information that cannot be revealed if using the current conventional approach, such as the point summary statistics and aggregated data. In addition, the results using the 95% confidence set for the trimmed mean has successfully justified the discussion and previous research findings regarding the linkage between the changes in climate (due to the phenomenon of El Nino) and the occurrences of severe PM_{10} level.

Generally, the number of exceedence days in the year 1997 was always more than that in the year 2005 in the area as shown in Table 3. In addition, the number of days with more than half of the hours in a day that people were potentially exposed to the unsafe condition of PM_{10} in the year 1997 was approximately four times more than that in the year 2005. The number of critical exceedence days also exhibited a similar behaviour.

Based on the proposed statistics in Table 3, the functional objects representing the critical exceedence days are shown in Figure 4. Obviously, it could be seen that the PM_{10} problem in the year 1997 was more severe than that in the year 2005 with a greater number of critical exceedence days.

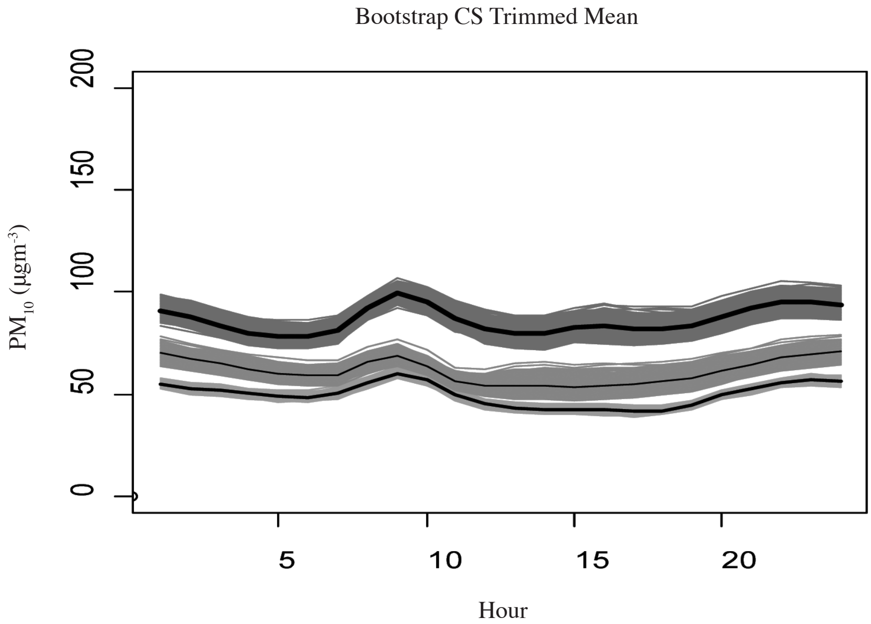

FIGURE 3. The bootstrap confidence set for trimmed mean with middle black line (thickest - 1997, thicker - 2000, light - 2005)

TABLE 3. The Cumulative Exceedence Index (*CEI*) and the number of exceedence days

Year	1997						2005					
Cum. Exceedence Index (CEI_{ij})	0.5	0.6	0.7	0.8	0.9	1	0.5	0.6	0.7	0.8	0.9	1
Number of days	18	11	10	9	8	7	5	3	2	2	2	2

(Note: CEI=1 indicates the critical exceedence day)

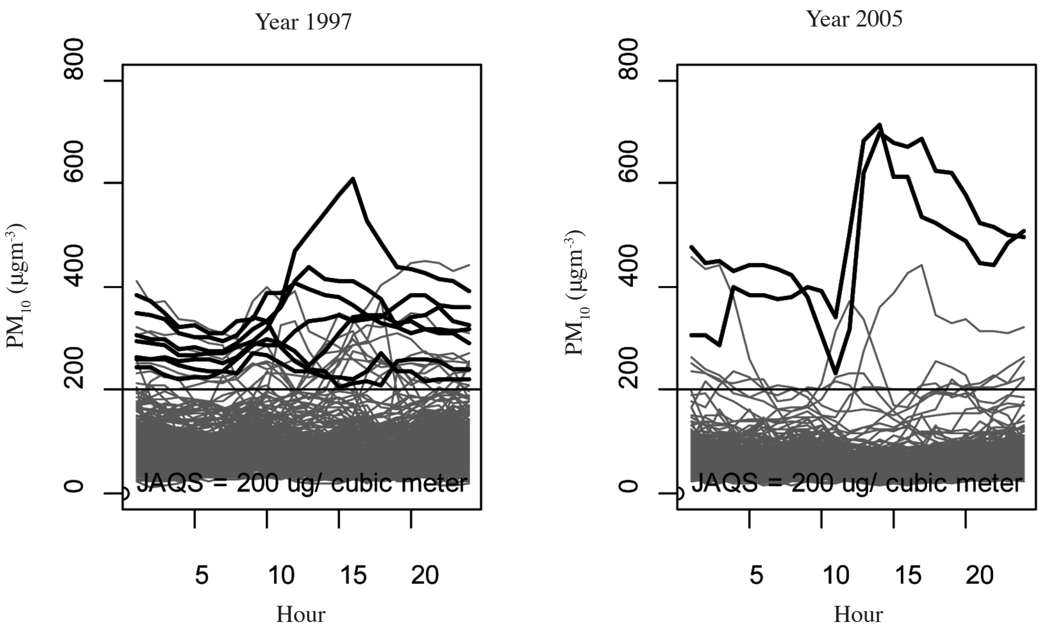

FIGURE 4. The critical exceedence day curves during the year 1997 and 2005

Further analysis of analysing the *CEI* and the exceedence days across the available stations was conducted. Table 4 shows that at average, there were six critical exceedence days at each station in 1997 and two days in 2005.

In terms of the number of critical exceedence days, the most severe station in the area for the year 1997 was ST1 (Gombak) while for the year 2005 it was ST2 (Klang). In addition, if the evaluation was to be based on the frequency of exceedence days with respect to at least 50 percent of the day a monitoring site was possibly exposed to the unsafe condition of PM_{10}, then for both years, ST2 was found to be the most severe station.

By transforming the observed data into functional forms based on the statistics obtained in Table 4, the visualisation of the severity among the available stations in the years 2005 and 1997 is shown in Figure 5. The most number of critical exceedence day curves were observed at ST1 in 1997 and ST2 in 2005. Moreover, different severity patterns were also observed among the different stations. Unlike 1997, the behaviour in 2005 was more inconsistent among the stations.

The corresponding date of the day of critical exceedences depicted in Figure 5 at different stations in the study area for year 1997 and 2005 is shown in Table 5. The critical exceedence days experienced in the area were between day 13th September and 27th September inclusively in the year 1997 and between day 10th August and 11th August in the year 2005. The results showed that the occurrence of the severe PM_{10} problem was during September for the year 1997 and during August for the year 2005. Both months fell under the dry season (the Southwest Monsoon). The worst day of critical exceedences occurred

TABLE 4. The Cumulative Exceedence Index (*CEI*) and the number of exceedence days by stations

Year	Cum. Exceedance Index CEI_{ij}	Stations						
		ST1	ST2	ST3	ST4	ST5	ST6	ST7
1997	0.5	20	25	13	13	14	NA	NA
	0.6	15	16	11	11	8	NA	NA
	0.7	14	13	8	8	8	NA	NA
	0.8	11	10	7	7	7	NA	NA
	0.9	10	8	5	5	5	NA	NA
	1	8	7	5	5	4	NA	NA
2005	0.5	3	8	5	3	8	5	3
	0.6	3	6	3	2	3	2	3
	0.7	2	4	3	2	3	2	3
	0.8	2	3	2	2	3	2	3
	0.9	2	3	2	2	2	2	2
	1	2	3	2	2	2	1	2

(Note: CEI=1 indicates the critical exceedence day)

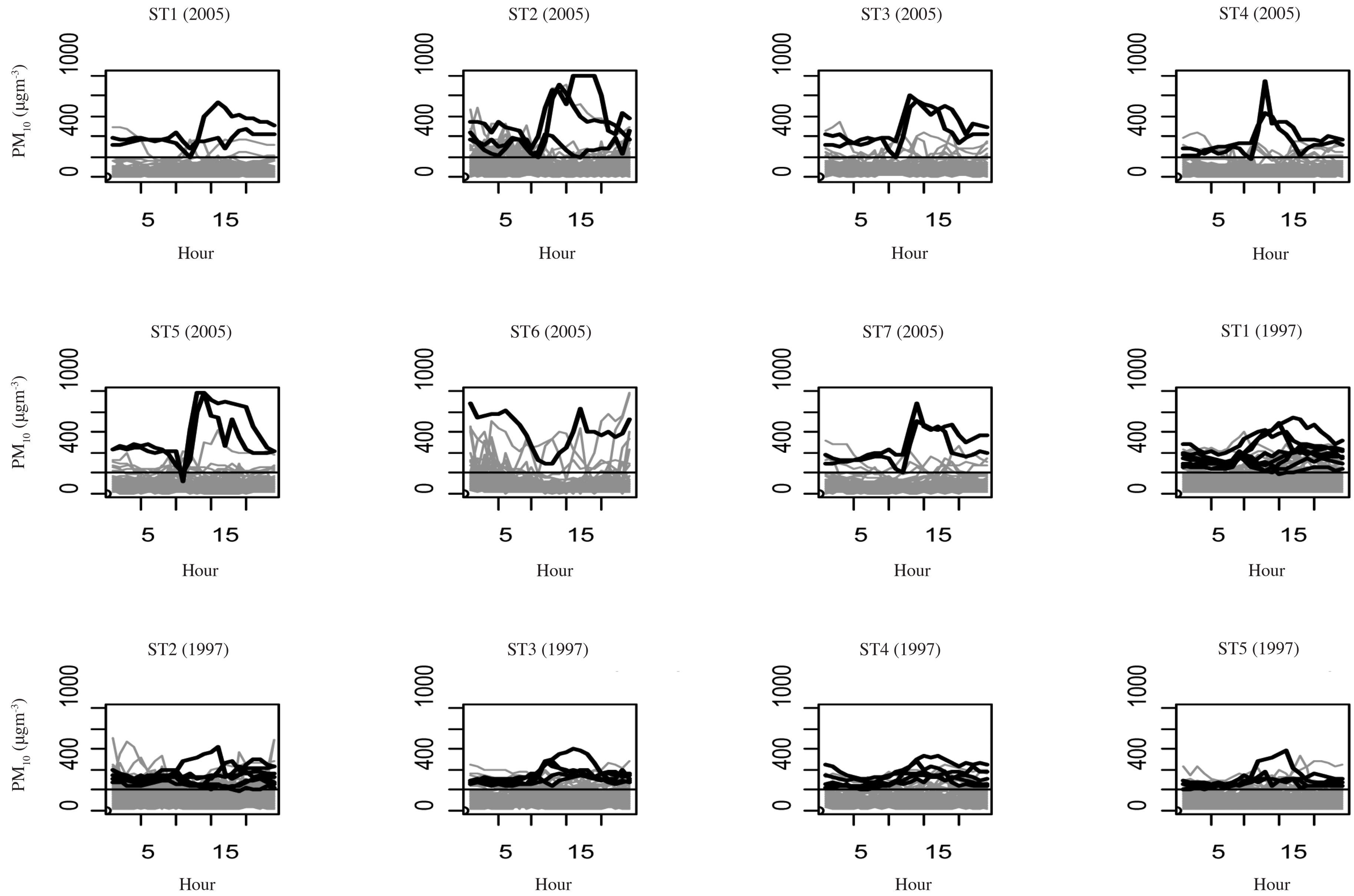

FIGURE 5. The critical exceedence days according to stations

TABLE 5. Date of the day of critical exceedences and the ranking

Station	Year 1997								Station	Year 2005		
	Date									Date		
ST1	13/9	14/9	15/9	19/9	24/9	25/9	26/9	27/9	ST1	10/8	11/8	
Depth	0.899	0.679	0.571	0.766	0.758	0.856	0.652	0.81	Depth	0.870	0.63	
Rank	1	6	8(*)	4	5	2	7	3	Rank	1	2(*)	
ST2	13/9	14/9	15/9	19/9	25/9	26/9	27/9		ST2	7/8	10/8	11/8
Depth	0.848	0.658	0.801	0.758	0.755	0.705	0.689		Depth	0.833	0.667	0.667
Rank	1	7(*)	2	4	3	5	6		Rank	1	3.5(*)	3.5(*)
ST3	13/9	14/9	15/9	19/9	25/9				ST3	10/8	11/8	
Depth	0.807	0.715	0.652	0.743	0.787				Depth	0.804	0.696	
Rank	1	4	5(*)	3	2				Rank	1	2(*)	
ST4	13/9	14/9	15/9	19/9	25/9				ST4	10/8	11/8	
Depth	0.804	0.683	0.678	0.678	0.865				Depth	0.783	0.717	
Rank	2	3	4.5(*)	4.5(*)	1				Rank	1	2(*)	
ST5	14/9	19/9	25/9	26/9					ST5	10/8	11/8	
Depth	0.712	0.821	0.772	0.696					Depth	0.891	0.609	
Rank	3	1	2	4(*)					Rank	1	2(*)	
NA									ST6	11/8		
									Depth	-		
									Rank	(*)		
NA									ST7	10/8	11/8	
									Depth	0.783	0.717	
									Rank	1	2(*)	

NA - The station does not operate yet (*) - The worst critical exceedence day

on similar days in ST1 and on the 15th of September, 1997 in ST3. Meanwhile, ST4 experienced two worst days of exceedences which occurred on the 15th and 19th September. On the other hand, the worst day at ST5 was on the 26th of September. In the year 2005, 11th August was identified as the worst day of exceedence in almost all of the stations. In addition, ST2 experienced two consecutive worst day of critical exceedence which occurred on the 10th and 11th of August, respectively.

CONCLUSION

The study highlighted the advantages of using the functional based approach for assessing and comparing the behaviour of the PM_{10} pollutant during two extreme haze years (1997 and 2005) in Selangor, state of Malaysia. The results from the analysis revealed implicit information regarding the existence of two significantly different behaviours of the PM_{10} within the year and between the years of study which would not have been visualised if the conventional approach was used. The employment of the confidence set for functional trimmed mean has provided the evidence to justify the previous research findings on the linkage between the change and climate and the severity of PM_{10} level. The results also provided the evidence that the diurnal behaviour was driven by the different dominant emission sources and other meteorological conditions. The new insight of the possibility to investigate the critical day of exceedences which incorporated the concept of the continuous nature of PM_{10} process was triggered based on the results of the behaviour of PM_{10} curves with high PM_{10} level that lay above the standard. The proposed *CEI* used in this study and the application of functional depth had successfully detected the critical exceedence days. The results showed that the occurrences of the critical exceedence days were found during the Southwest Monsoon; September for the year 1997 and August for the year 2005. By means of the functional depth method, it was identified that the worst day of critical exceedences to the PM_{10} pollution occurred at different dates among the stations in 1997 (between 13th and 27th September) while in the year 2005, most of the stations experienced the worst day of critical exceedences on the 10th of August. Despite the fact that the year 1997 was proven to be more severe than the year 2005, the PM_{10} behaviour for the year 2005 had shown a more inconsistent pattern. It is hoped that this alternative statistical approach can be proposed as a new paradigm in the methodology and procedure for future air quality assessment. The results obtained were based on the proposed statistics employed which could be used as an alternative measurement for further analysis, for example in the spatial prediction or hot spot identification of severity; not only for PM_{10} but also for other types of pollutants such as ozone, CO and NOx. The study could also be extended to the usage of the functional depth method in identifying outlier or abnormal behaviour in the air quality data set.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Environment Malaysia and Assoc. Prof. Dr. Mohd Talib Latif from University Kebangsaan Malaysia (UKM) for providing the information and data; UKM for the allocation of the Research University Grant [UKM-AP-2011_19] and the anonymous reviewers for the constructive comments.

REFERENCES

ADB & CAI-Asia. 2006. *Country Synthesis Report on Urban Air Quality Management (Malaysia)*. Philippines: Asian Development Bank (ADB) & Clean Air Initiative for Asian Cities (CAI-Asia) Centre.

Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research 92*: 71-77.

Awang, M., Jaafar, A.B., Abdullah, A.M., Ismail, M., Hassan, M.N., Abdullah, R., Johan, S. & Noor, H. 2000. Air quality in Malaysia: Impacts, management issues and future challenges. *Respirology* 5: 183-196.

Azmi, S.T., Latif, M.T., Ismail, A.S., Juneng, L. & Jemain, A.A. 2010. Trend and status of air quality at three different monitoring stations in Klang Valley, Malaysia. *Air Quality Atmospheric Health 3*: 53-64.

DoE (Department of Environment Malaysia). 2005. *Malaysian Environmental Quality Report 2005*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

DoE (Department of Environment Malaysia). 2008. *Malaysian Environmental Quality Report 2008: Air Quality*. Available online at: http://apps:doe.gov.my, Accessed on 25 May 2012.

Febrero, M., Galeano, P. & Manteiga, W.G. 2007. A functional analysis of NOx levels: Location and scale estimation and outlier detection. *Computational Statistics* 22: 411-427.

Field, R.D., Werf, G.R.V. & Shen, S.S.P. 2009. Human implication of drought-induced biomass burning in Indonesia since 1960s. *Nature Geoscience* 2: 185-188.

Fraiman, R. & Muniz, G. 2001. Trimmed means for functional data. *Test* 10: 419-440.

Gao, H.O. 2007. Day of week effects on diurnal ozone, NOx cycles and transportation emissions in Southern California. *Transportation Research Part D* 12: 292-305.

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: 2895-2907.

Laakso, L., Hussein, T., Aarnio, P., Komppula, M., Hiltunen, V., Viisanen, Y. & Kulmala, M. 2003. Diurnal and annual characteristics of particle mass and number concentrations in urban, rural and arctic environments in Finland. *Atmospheric Environment* 37: 2629-2641.

Ling, L.O.H., Ting, K.H., Shaharuddin, A., Kadaruddin, A. & Yaakob, M.J. 2010. Urban growth and air quality in Kuala Lumpur city, Malaysia. *Environment Asia* 3(2): 123-128.

MOE (Ministry of the Environmental Government of Japan). *The Environmental Quality Standards in Japan – Air Quality*. Available online at: http://www.env.go.jp/en/air/aq/aq.html, Accessed on 15 July 2011.

Ramsay, J.O. & Silverman, B.W. 1997. *Functional data analysis*. New York: Springer.

Tangang, F., Latif, M.T. & Juneng, L. 2011. *The roles of climate variability and climate change on smoke haze occurrences in Southeast Asia region*. Available online at: http:// www2. Ise.ac.uk/ideas/publications/reports/pdf/sr004/num.pdf, Accessed on 20 January 2011.

Norshahida Shaadan* & Sayang Mohd Deni
Center for Statistical and Decision Science Studies
Faculty of Computer & Mathematical Sciences
40450 UiTM Shah Alam
Selangor D.E.
Malaysia

Abd Aziz Jemain
DELTA, School of Mathematical Sciences
Faculty of Science & Technology
Universiti Kebangsaan Malaysia (UKM)
43600, Bangi
Selangor D.E.
Malaysia

*Corresponding author; email: shahida@tmsk.uitm.edu.my

Received: 4 October 2011
Accepted: 30 May 2012