PM_{10} CONCENTRATIONS SHORT TERM PREDICTION USING REGRESSION, ARTIFICIAL NEURAL NETWORK AND HYBRID MODELS

AHMAD ZIA UL-SAUFIE MOHAMAD JAPERI

UNIVERSITI SAINS MALAYSIA
2013

PM_{10} CONCENTRATIONS SHORT TERM PREDICTION USING REGRESSION, ARTIFICIAL NEURAL NETWORK AND HYBRID MODELS

by

AHMAD ZIA UL-SAUFIE MOHAMAD JAPERI

Thesis submitted in fulfillment of the requirements for degree of Doctor of Philosophy

JULY 2013

ACKNOWLEDGEMENT

First and above all, I praise Allah, the almighty for providing me this opportunity and granting me capability successfully. This thesis appears in its current from due to the assistance and guidance of several people and organization I would therefore like to offer my sincere thanks to all of them.

I would like to express my greatest appreciation and thanks to my supervisor, Associate Professor Ahmad Shukri Yahaya and my co-supervisor, Professor Dr. Nor Azam Ramli for letting me to be under their supervisions. I really appreciate all the guidance, important suggestion, support, advice, and continuous encouragement in completing my PhD.

Not forgotten my big thanks to all my friends under Clean Air Research Group, Dr. Hazrul, Zul Azmi, Dr Izma, Norrimi, Hasfazilah, Maisarah, Azian, Maher and Nazatul for the cooperation and help during my study.

Lastly and most importantly, I would like to dedicate this thesis to my parents, Mohamad Japeri Hassim and Azizah Awang for their good wishes, continuous encouragement and motivation. For my wife, Wan Nor Aishah Meor Hussain. thank you for always being there for me. My son, Umar Danish and my daughter, Fatimah Tasnim who inspired me to face the challenges and complete this research.

Finally, I wish to express my biggest acknowledgement to Universiti Teknologi Mara for providing me financial support under Skim Latihan Akedemik IPTA (SLAI).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Air Pollution Index
ANOVA	Analysis of Variance
ANN	Artificial Neural Network
ASMA	Alam Sekitar Malaysia Sdn. Bhd.
BAM	Beta Attenuation Monitor
BCG	Bachang
BKE	Butterworth Kulim Expressway
CO	Carbon monoxide
D-W	Durbin Watson
DoE	Department of Environment (Malaysia)
DRM	Direct Reading Monitor
EPA	Environmental Protection Agency
FFBP	Feedforward Backpropagation
GRNN	General Regression Neural Network
GUI	Graphical User Interface
IA	Index of Agreement
ILP	Institut Latihan Perindustrian
JRT	Jerantut
KCH	Kuching
KMO	Kaiser-Meyer Olkin
KTG	Kuala Terengganu
MAAQG	Malaysian Ambient Air Quality Guidelines
MAD	Median Absolute Deviation

MLP	Multi Layer Perceptron
MLR	Multiple Linear Regression
NAE	Normalized Absolute Error
NLI	Nilai
NO_2	Nitrogen Dioxide
O_3	Ozone
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PC	Principal Component
PI	Performance Indicators
PLUS	Projek Lebuhraya Utara Selatan
$PM_{2.5}$	Particulate matter less than 2.5 μm
PM_{10}	Particulate matter less than 10 μm
QR	Quantile Regression
RBF	Radial Basis Function
RR	Robust Regression
R^2	Coefficient of Determination
RH	Relative Humidity
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
SD	Standard Deviation
SLR	Simple Linear Regression
SJY	Seberang Jaya
PRI	Perai
SO_2	Sulphur Dioxide

SSE	Sum of Squares Due to Error
SSR	Sum of Square Due to Regression
SST	Total Sum of Squares
T	Temperature
USEPA	United States Environmental Protection Agency
VIF	Variance Inflation Factor
WHO	World Health Organization

RAMALAN JANGKA PENDEK KEPEKATAN PM_{10} MENGGUNAKAN MODEL REGRESI, MODEL RANGKAIAN NEURAL BUATAN DAN MODEL HIBRID

ABSTRAK

Zarah terampai mempunyai kesan yang signifikan kepada kesihatan manusia apabila kepekatan zarah terampai melebihi garis panduan kualiti udara di Malaysia. Kajian ini hanya akan mengfokuskan kepada zarah terampai yang mempunyai diameter aerodinamik kurang daripada $10\mu m$, dinamakan PM_{10}. Ini memerlukan model berstatistik bagi membuat ramalan kepetakatan PM_{10} pada masa akan datang. Tujuan kajian ini ialah untuk membangunkan dan meramalkan kepekatan PM_{10} pada keesokan hari (D+1), dua hari berikutnya (D+2) dan tiga hari berikutnya (D+3) bagi tiga kategori iaitu kawasan industri (tiga stesen), kawasan bandar (dua kawasan), satu kawasan subkelompok bandar dan satu kawasan rujukan. Kajian ini menggunakan cerapan purata data harian dari tahun 2001 hingga 2010. Tiga kaedah utama telah digunakan dalam membangunkan model ramalan kepekatan PM_{10} iaitu regresi linear berganda, rangkaian neural buatan dan model hibrid. Tiga model regresi telah digunakan iaitu regresi linear berganda (MLR), regresi teguh (RR) dan regresi kuantil (QR). Rangkaian neural rambatan balik (FFBP) dan rangkaian neural regresi umum (GRNN) digunakan dalam rangkaian neural buatan. Model hibrid ialah model yang menggunakan gabungan analisis komponen utama (PCA) dengan semua lima kaedah peramalan iaitu PCA-MLR, PCA-QR, PCA-RR, PCA-FFBP and PCA-GRNN. Keputusan bagi model regresi menunjukkan bahawa RR dan QR lebih baik daripada MLR dan boleh dianggap sebagai kaedah alternatif apabila andaian bagi MLR tidak dapat dipenuhi. Keputusan bagi rangkaian neural buatan menunjukan FFBP lebih baik jika dibandingkan dengan GRNN. Model hibrid memberi keputusan yang lebih baik jika dibandingkan dengan model ramalan tunggal dari segi ketepatan dan ralat. Akhir sekali, sebuah aplikasi peramalan baru dibangunkan untuk membuat ramalan masa hadapan bagi kepekatan PM_{10} dengan menggunakan sepuluh model ramalan yang telah diperolehi dengan purata ketepatan untuk D+1(0.7930), D+2 (0.6926) and D+3 (0.6410). Aplikasi ini akan membantu pihak berkuasa tempatan untuk mengambil tindakan yang wajar bagi mengurangkan kepekatan PM_{10} dan juga sebagai satu sistem amaran awal.

PM_{10} CONCENTRATIONS SHORT TERM PREDICTION USING REGRESSION, ARTIFICIAL NEURAL NETWORK AND HYBRID MODELS

ABSTRACT

Particulate matter has significant effect to human health when the concentration level of this substance exceeds Malaysia Ambient Air Quality Guidelines. This research focused on particulate matter with aerodynamic diameter less than 10 μm, namely PM_{10}. Statistical modellings are required to predict future PM_{10} concentrations. The aims of this study are to develop and predict future PM_{10} concentration for next day (D+1), next two-days (D+2) and next three days (D+3) in seven selected monitoring stations in Malaysia which are represented by fourth different types of land uses i.e. industrial (three sites), urban (three sites), a sub-urban site and a reference site. This study used daily average monitoring record from 2001 to 2010. Three main models for predicting PM_{10} concentration i.e. multiple linear regression, artificial neural network and hybrid models were used. The methods which were used in multiple linear regression were multiple linear regression (MLR), robust regression (RR) and quantile regression (QR), while feedforward backpropagation (FFBP) and general regression neural network (GRNN) were used in artificial neural network. Hybrid models are combination of principal component analysis (PCA) with all five prediction methods i.e. PCA-MLR, PCA-QR, PCA-RR, PCA-FFBP and PCA-GRNN. Results from the regression models show that RR and QR are better than the MLR method and they can act as an alternative method when assumption for MLR is not satisfied. The models for artificial neural network show that FFBP is better than the GRNN. Hybrid models gave better results compared to the single models in term of accuracy and error. Lastly, a new predictive tool for future PM_{10} concentration was developed using ten models for each site with average accuracy for D+1(0.7930), D+2 (0.6926) and D+3 (0.6410). This application will help local authority to take proper action to reduce PM_{10} concentration and as early warning system.

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

Air pollution has significant effect to human health, agriculture and ecosytem (Mohammed, 2012). There are numerous reports pertaining to the effect of air pollution on human health, agriculture crops, forest species and ecosystem. Several large cities in Malaysia have reading of ambient air quality that are increasing and exceeding the national ambient air quality standard (Afroz et al., 2003).

Malaysia has 52 monitoring stations maintained by the Department of Environment Malaysia (2012). All stations provide hourly measurements of particulate matter with aerodynamic diameter less than or equal to 10 µm (PM_{10}), ozone (O_3), sulphur dioxide (SO_2), carbon monoxide (CO) and nitrogen dioxide (NO_2). PM_{10} concentration is chosen because PM_{10} has significant impacts on human health, agriculture and buildings (Lee, 2010).

Fellenberg (2000), Godish (2004) and Tam and Neumann (2004) found that negative health effect were clearly related to PM_{10} such as asthma, nose and throat irritations, allergies, respiratory related illnesses, and premature mortality. Sedek et al., (2006) found that PM_{10} gave negative impact to productivity of short cycle plants such as vegetables.

1.1 AIR POLLUTION IN MALAYSIA

The Department of Environment (DOE) Malaysia uses Air Pollution Index (API) to compare itself with other regional countries. The API was adopted after the Department of Environment Malaysia revised its index system in 1996. The API closely follows the Pollutant Standards Index (PSI) system of the United States (Department of Environment Malaysia, 1996) as shown in Table 1.1. Afroz et al., (2003) reported that the main air pollutant in Malaysia is carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen dioxide, and other particulate matter, with an aerodynamic diameter of less than 10 µm.

Table 1.1: Malaysia Air Pollution Index (API)
(Source: Department of the Environment, Malaysia, 2012)

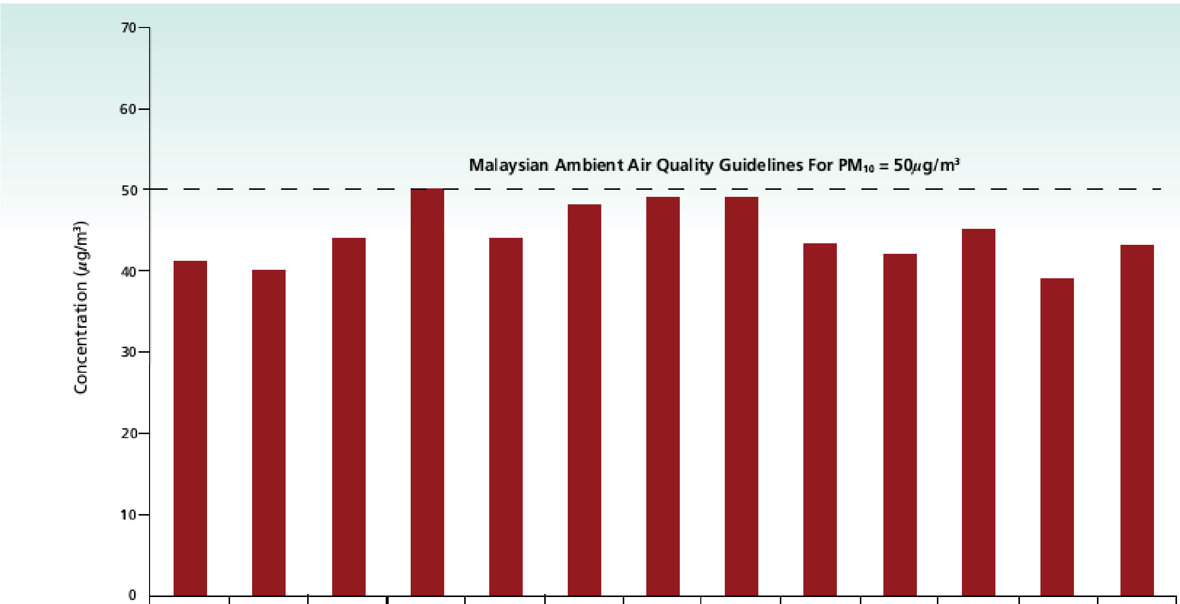
API	Description
$0 < API \leq 50$	Good
$50 < API \leq 100$	Moderate
$100 < API \leq 200$	Unhealthy
$200 < API \leq 300$	Very Unhealthy
> 300	Hazardous

Sansudin (2010), Ramli et al., (2001) and Awang et al., (2000) indicated that PM_{10} is the main contributor to haze events. This means that when the PM_{10} concentration level is higher than Malaysian Ambient Air Quality Guidelines (MAAQG), the government under the National Haze Action Plan can announce warning status for locations with prolonged APIs exceeding 101 for more than 72 hours (Perimula, 2012). Thus, this research was carried out until next three days (72 hours) to predict PM_{10} concentrations. Malaysia's safe concentration for PM_{10} is based on the Department of Environment Malaysia (2002) guidelines, of 150µg/m^3over a 24 hour average, and 50µg/m^3 for 1 year. Table 1.2 shows the relationship between API and PM_{10} concentrations in Malaysia.

Table 1.2: API intervals, description of air quality, and relationship with PM_{10} values (Modified from the Department of Environment, Malaysia (2012))

API	Description	PM_{10} Values ($\mu g/m^3$)
$0 < API \leq 50$	Good	$0 < PM_{10} \leq 75$
$50 < API \leq 100$	Moderate	$75 < PM_{10} \leq 150$
$100 < API \leq 200$	Unhealthy	$150 < PM_{10} \leq 350$
$200 < API \leq 300$	Very Unhealthy	$350 < PM_{10} \leq 420$
$300 < API \leq 500$	Hazardous	$420 < PM_{10} \leq 600$
> 500	Very Hazardous	> 600

The annual average PM_{10} concentrations for Malaysia from 1999 until 2011 is shown in Figure 1.1. The result shows that the average concentration for every year is below the Malaysia ambient air quality guideline for PM_{10} concentrations except for 2002 when the value is equal with Malaysia ambient air quality guideline of 50µg/m^3. Besides that, the Figure 1.1 also show increasing number of monitoring sites from 45 sites in 1999 to 52 sites in 2011.



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Concentration	41	40	44	50	44	48	49	49	43	42	45	39	43
Number of Sites	45	50	50	50	51	51	51	51	51	51	51	52	52

Figure 1.1 Annual Average Concentration of PM_{10} for Malaysia from 1999-2011 (Department of Environment Malaysia, 2012)

This section were discussed annual average concentration of PM_{10} for Malaysia from 2001 until 2010 because the data were used in this study. In 2001, the Department of Environment, Malaysia, stated that overall air quality was good to moderate. Only a few days were identified as unhealthy, because PM_{10} and the ozone were higher than the MAAQG (50 $\mu g/m^3$) for July 2001 of that year (dry season). Klang reported seven days and Kuala Selangor experienced eight unhealthy days in 2001, because PM_{10} was high due to forest fires and other burning activities (Department of Environment Malaysia, 2002). Sabah and Sarawak experienced unhealthy air quality, due to open burning activities from shifting agriculture activities, for June and July 2001 (Department of Environment Malaysia, 2002).

Heil (2007) identified major fires in West Kalimantan during August to November 2002. This caused the number of unhealthy days to increase from three to eight, due to particulate matter from trans-boundary haze pollution in Sarawak (Sansuddin, 2010). The overall air quality in 2002 dropped in comparison to the previous year. However, PM_{10} and the ozone were prevalent as pollutants in Malaysia. In Kuala Selangor unhealthy air quality was caused by high PM_{10} in the air. However, no unhealthy days were reported from the east coast of Malaysia in 2002 (Department of Environment Malaysia, 2003).

The Department of Environment Malaysia (2004) stated that a slightly improved overall air quality was observed compared to the previous year. In Penang, PM_{10} and SO_2 were the main cause of unhealthy days, due to intensive industrial activities in the area. In 2003, trans-boundary haze pollution did not affect the air quality in Sarawak and Sabah such as in previous years (Department of Environment Malaysia, 2004).

In 2004, the Department of Environment, Malaysia stated that PM_{10} was the prevalent pollutant in Malaysia, causing moderate haze in June, August, and September, due to trans-boundary pollution, in the form of forest fires in Sumatra as reported by the ASEAN Specialised Meteorological Centre. Fires in Kalimantan also affected southern Sarawak (Department of Environment Malaysia, 2005).

Several parts of Malaysia experienced haze episodes from mid-May until mid-October 2005, caused by forest and land fires in the Riau Province of Central Sumatra, Indonesia (Sansuddin, 2010 and Md Yusof, 2009). Central, eastern, and northern parts, experienced severe haze between 1st August 2005 and 15th August 2005. However, on 11th August 2005,the air pollution index exceeded 500 in Kuala Selangor and Pelabuhan Klang that was caused by peat land fires in Selangor (Md Yusof, 2009). Other haze episodes affected the overall air quality in Malaysia, between moderate to good levels, in 2005 (Department of Environment Malaysia, 2006).

Hyer and Chew (2010) identified that high particulate events between July and October 2006, was caused by trans-boundary pollution from forest fires in Sumatra and Kalimantan. The Klang Valley recorded that all of its unhealthy air quality days in 2006 (25 days) were caused by PM_{10} as the predominant pollutant, during the South Westerly monsoon (Department of Environment Malaysia, 2007).

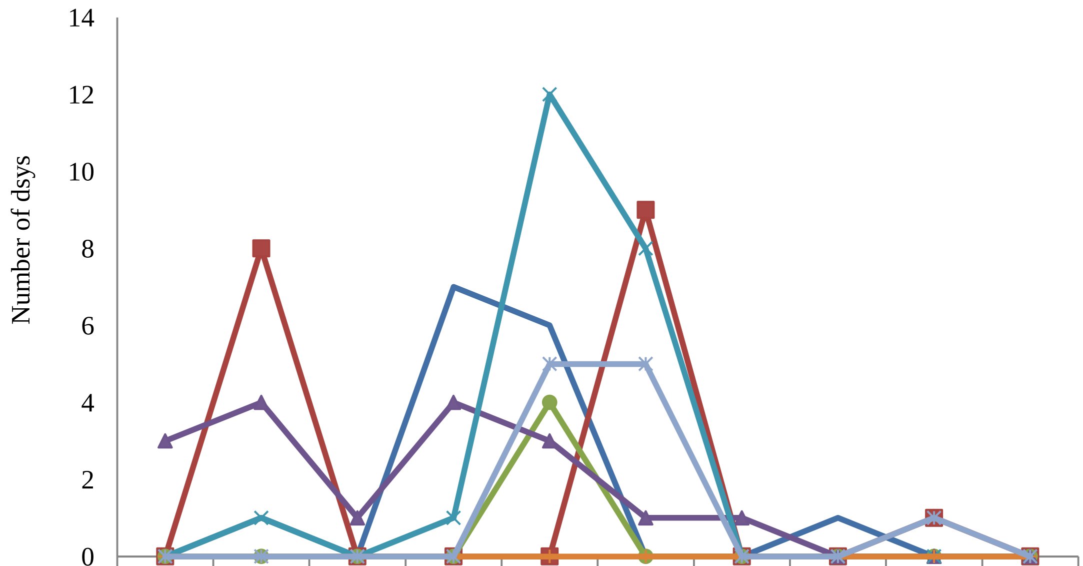
The Malaysian Environment Quality Report (Department of Environment Malaysia, 2008), reported that the overall air quality in 2007 improved significantly compared to the previous year, due to favourable weather conditions (weak to medium La Nina) and no trans-boundary haze pollution. The main pollutant were caused by ground level ozone and PM_{10}.

Sansuddin, (2010) observed a slightly improved air quality days in 2008 compared to the previous year, due to an intensive surveillance programme and preventive measures undertaken by Department of Environment Malaysia. Furthermore, no trans-boundary haze pollution was observed in 2008. PM_{10} and the ozone remained the main pollutant source for unhealthy days recorded in the Klang Valley, Negeri Sembilan, Perak, Kedah, Pulau Pinang, and Johor, in 2008. During this period, the source of PM_{10} comes from peat-land burning during dry periods and emissions from motor vehicles.

In 2009, the mean PM_{10} concentrations slightly increased from $42\mu g/m^3$ in 2008 to $45\mu g/m^3$ in 2009. This was due to peat-land fires and trans-boundary air pollution during hot and dry condition (moderate to strong El-Nino), especially between June and August 2009. However, the annual PM_{10} average was $45\mu g/m^3$, which is below the Malaysian Ambient Air Quality Guideline value of $50\mu g/m^3$ (Department of Environment Malaysia, 2010).

The overall air quality in 2010 was significantly improved ($39\mu g/m^3$) compared to the previous year ($45\mu g/m^3$). Higher PM_{10} values were recorded in several areas of Johor and Melaka in October 2010, due to trans-boundary haze pollution (Department of Environment Malaysia, 2011). However, the annual PM_{10} average in 2010 was only 39 $\mu g/m^3$; which was the lowest value recorded since 1999 (Donham, 2000; Radon et al., 2001).

The number of unhealthy days for the seven selected sites from 2001-2010 is shown in Figure 1.2. The highest number of unhealthy days was recorded in 2002, 2004, 2005 and 2006 because of the high particulate events in those years. The main contributor to the unhealthy days in 2002 was the major fires in the west coast of Kalimantan (Sansuddin, 2010). Trans-boundary pollution from forest and land fires in Sumatra and Kalimantan contributed to the unhealthy days in 2004, 2005 and 2006 (Sansuddin, 2010; Md Yusof, 2009 and Department of Environment Malaysia, 2005 and 2007). For the other years, the unhealthy days were caused by industrial activities, emissions from motor vehicles and open burning from shifting agriculture activities.



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
PRI	0	1	0	7	6	0	0	1	0	0
KCH	0	8	0	0	0	9	0	0	1	0
JRT	0	0	0	0	4	0	0	0	0	0
SJY	3	4	1	4	3	1	1	0	0	0
NLI	0	1	0	1	12	8	0	0	0	0
KTG	0	0	0	0	0	0	0	0	0	0
BCG	0	0	0	0	5	5	0	0	1	0

Figure 1.2 Number of unhealthy days for seven selected sites, 2001 - 2010

1.2 PROBLEM STATEMENT

In Malaysia, the Department of Environment Malaysia (DOE, Malaysia) is the government body responsible for monitoring air quality in Malaysia. Department of Environment Malaysia monitors continuously through 52 stations located in urban, sub-urban, industrial areas and a background area. These monitoring stations are located in strategic locations to detect any significant change of air quality. Malaysia and other countries have guidelines for allowable levels of air pollutant in the air (Department of Environment Malaysia Malaysia, 2012). In Malaysia this is known as the Malaysia Ambient Air Quality Guidelines (MAAQG). In these guidelines the threshold value of PM_{10} for a safe level is at 150 µg/m^3 per 24 hour averaging times and 50µg/m^3 per year.

Short term and chronic human health may occur when the concentration levels of air pollutant exceed the air quality guidelines (QUARG, 1996 ; Lee et al., 2010). Nasir et al., (1998) reported in 1997 (haze episode in Malaysia) the estimated negative effect to health for asthma attacks was 285,277 cases, there were 118,804 cases of bronchitis in children and 3889 cases in adults, and in addition, respiratory hospital admissions (2003 cases) and emergency room visits (26,864 cases). World health Organization, (1998) reported that outpatient treatment for respiratory disease at Kuala Lumpur General Hospital increased from 250 to 800 per day and for outpatient in Sarawak increased between two and three times during the haze episode in 1997. Besides that, Brauer and Jamal (1998) found that haze episode in 1997 also resulted in the increase of asthma, conjuctivitis and acute respiratoty infection.

Md Yusof, (2009) said PM_{10} can primarily cause reduction in visibility by light scattering. Visibility have significant strong correlation with increases in mass concentration of nitrate, elemental carbon element and sulphate (Kim et al., 2006). Therefore, research on effect of PM_{10} to human health and environment has been done by researchers worldwide.

Thus, particulate matter (PM_{10}) has become a challenge to Malaysia's air quality management. One of the most important efforts in PM_{10} monitoring is to develop PM_{10} forecasting models. Statistical modellings could offer good insights in predicting future PM_{10} concentration levels in Malaysia. The aims of this study are to develop and predict future PM_{10} concentration for D+1, D+2 and D+3.

The number of studies for predicting PM_{10} concentration is still limited in Malaysia. This study provides the PM_{10} forecasting models using three main methods i.e. regression, artificial neural network and hybrid models. The methods that were used in regression models were multiple linear regression (MLR), robust regression (RR) and quantile regression (QR), while feedforward backpropagation (FFBP) and general regression neural network (GRNN) were used in artificial neural network. Hybrid models are combination of principal component analysis (PCA) with all five prediction methods i.e. PCA-MLR, PCA-QR, PCA-RR, PCA-FFBP and PCA-GRNN.

This research also developed a new predictive tool for predicting future PM_{10} concentrations in selected areas in Malaysia up to three days in advance. The models could be easily implemented for public health protection to provide early warnings to the respective populations. In addition, the models were useful in helping authorities actuate air pollution impact preventative measures in Malaysia.

1.3 OBJECTIVES

The objectives of this research are given below:

1. To apply multiple linear regression, robust regression and quantile regression to predict PM_{10} concentrations.
2. To apply artificial neural network techniques (ANN) i.e. feedforward backpropagation (FFBP) and general regression neural network (GRNN) to predict PM_{10} concentrations.
3. To create hybrid models by combining regression models and ANN models with principal component analysis (PCA).
4. To determine the most suitable model for predicting future (D+1, D+2 and D+3) PM_{10} concentrations.
5. To develop a new predictive tool for future PM_{10} concentrations prediction in Malaysia.

1.4 SCOPE OF RESEARCH

There are many methods to develop models for prediction of air pollutant concentration data. The most commonly used in air pollutant modelling are multiple linear regression and neural network. Nowadays, hybrid models have become more popular as method for prediction models. All these methods were used in this research to develop and predict future PM_{10} concentration for D+1, D+2 and D+3.

Seven stations have been chosen for this research which is Perai, Jerantut, Kuala Terengganu, Seberang Jaya, Nilai, Bachang and Kuching. Those stations represent four groups that are industrial area (Perai, Nilai and Kuching), urban area (Kuala Terengganu and Bachang), sub-urban area (Seberang Jaya) and a background station (Jerantut). Table 1.3 show the monitoring stations coordinates and basic description.

Table 1.3 : Monitoring stations coordinates and description

ID Code	Monitoring Station	Category	Station Name	Coordinate
CA003	Perai (PRI)	Industry	Sek Keb Taman Inderawasih	N 05^{o} 23.4704' E 100^{o} 23.1977'
CA004	Kuching (KCH)	Industry	Depot Ubat, Kuching	N 01^{o} 33.7696' E 110^{o} 23.3740'
CA007	Jerantut (JRT)	Background	MMS, Batu Embun, Jerantut	N 03^{o} 58.2482' E 102^{o} 20.8891'
CA009	Seberang Jaya (SJY)	Sub-urban	Sek.Keb.Seberang Jaya 2, Perai	N 5^{o} 24.4476' E 100^{o} 24.0403'
CA010	Nilai (NLI)	Industry	Taman Semarak (Phase 2), Nilai	N 02^{o} 49.3001' E 101^{o} 48.6894'
CA034	Kuala Terengganu (KTG)	Urban	Sek.Keb.Chabang Tiga, Kuala Terengganu	N 5^{o} 20.2341' E103° 9.4564'
CA043	Bachang (BCG)	Urban	Sek.Men.Tun Tuah, Bachang	N 02^{o} 12.7850' E 102^{o} 14.0585'

In this research future daily PM_{10} concentration ($PM_{10,D+1}$, $PM_{10,D+2}$ and $PM_{10,D+3}$) were used as dependent variable and seven parameters were chosen as independent variable, that is relative humidity (RH), wind speed (WS ; km/hr), nitrogen dioxide (NO_2 ; ppm), temperature (T ; ^{o}C), PM_{10} (µg/m^3), sulphur dioxide (SO_2 ; ppm) and carbon monoxide (CO ; ppm). Monitoring records used in this research was obtained from the Department of Environment Malaysia from 2001 until 2010.

1.5 THESIS LAYOUT

This thesis consist of six chapters and a brief outline for each chapter are as follows:

Chapter 1 discussed the overview of air pollution in Malaysia, problem statement, objectives and scope of the research.

Chapter 2 summarized the literature review for air pollution in Malaysia, sources of PM_{10} concentration and effect of PM_{10} to human health. This chapter also discussed about the literature review of prediction models for particulate matter in Malaysia and world wide. The importance of statistical analysis in environmental engineering was also explained. Besides that, all types of regression models, artificial neural network models and hybrid models used to predict particulate matter concentration related to this research such as its application in environmental engineering and advantages of all methods were also enlightened.

Chapter 3 described the procedures applied for this research to predict future PM_{10} concentrations. Six main prediction methods have been discussed such as multiple linear regression, quantile regression, robust regression, feedforward backpropagation, general regression neural network and hybrid models. Besides that, performance indicators were also discussed in the last section of this chapter.

Chapter 4 discussed about result from the first until fifth objective in this research. The first two sections of this chapter discussed the characteristics of PM_{10} concentration for all sites. Three regression models were developed i.e. multiple linear regression, quantile regression and robust regression and all findings for these models were discussed. Two models of neural network have also been discussed in the last section of this chapter. Third section on this research discussed about principal component analysis and hybrid models. Then, the best model to predict PM_{10} concentration were obtained and discussed. The final part of this chapter explained about a new predictive tool for PM_{10} concentration in Malaysia.

Chapter 5 discusses the findings from Chapter 4 by comparing the results with the findings of other researchers. This chapter also discusses the PM_{10} concentration model by land use.

Chapter 6 provided the conclusions of this research, significant finding, limitation of study and recommendation for future work.

CHAPTER 2

LITERATURE REVIEW

2.0 PARTICULATE MATTER

There are five criteria pollutants namely carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3) and particulate matter (PM) (Afroz et al., 2003). Particulate Matter (PM) is the most imperative in terms of adverse effects on human health (Mott et al., 2005). 4000 deaths in the London fog in 1952 and 20 deaths in Donora, Pennsylvania in 1948 were recorded. These numbers showed a strong evidence of the impact of the pollutant on human's health (Radojević and Vladimir, 2006). Since then, there have been many research studies about PM; especially regarding particles of less than 10 micrometres (PM_{10}) (Kolehmainen et al., 2000 and Slini et al., 2006).

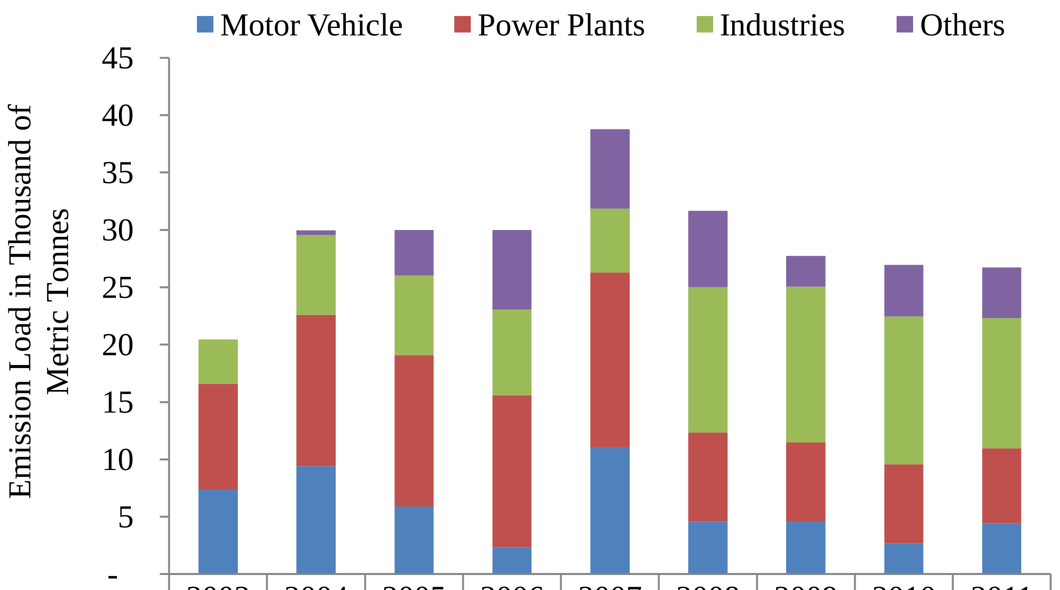
According to the Fierro (2000), particulate matter is made up of things floating around in the air; most of which, are invisible to the naked eye. Particles or particulate matter are a type of air pollution. People's health are most commonly affected by these airborne particulate matter (Mott et al., 2005). It comes from a wide variety of sources, in all sizes, shapes, colours, textures, and chemical compositions, and can remain suspended in the air for periods ranging from a few seconds to a few years (Lynn et al., 1976).

QUARG (1996) defined PM_{10} as particulate matter less than 10 µm aerodynamic diameter or particles which pass through a size selective inlet with a 50% efficiency cut off at 10 µm aerodynamic diameter. Van der Wal and Jansen (2000) defined PM_{10} as inhalable particles with an aerodynamic diameter of approximately 10μm or less. Hence, this research only considers particulate matter less than 10 micrometres, namely PM_{10} because PM_{10} gives the greatest concern to public health, since these particles are small enough to be inhaled (Krewski et al., 2000).

2.1 SOURCES OF PARTICULATE MATTER

Afroz et al., (2003) and Dominick et al., (2012) identified mobile sources, open burning sources and stationary sources as major sources of air pollution in Malaysia. The Department of Environment Malaysia also identified three main sources of air pollution in Malaysia, such as industry (including power stations - or stationary sources), motor vehicles (mobile sources), and open burning (Department of Environment Malaysia, 2001 to 2010). However, Department of Environment Malaysia also included trans-boundary pollution sources as a significant contribution to air pollution emissions (Department of Environment Malaysia, 2004 to 2006).

PM_{10} emission loads by sources (in metric tonnes) from 2003-2011 is shown in Figure 2.1. From Figure 2.1, emission sources of PM_{10} were divided into industry (47.82%), power plants (25.52%), motor vehicles (9.98%), and others (16.68%). However, the percentage contribution of industry and power plants are significantly different for years, such as 2007 and previously, where the highest contribution came from power plants; with more than 9000 metric tonnes per year, but for 2008 onwards, the main contribution of PM_{10} was industry, with a value of more than 12,000 metric tonnes per year. This was caused by a significant increase in industrial air pollution sources.



	2003	2004	2005	2006	2007	2008	2009	2010	2011
Others	-	409	3,943	6,937	6,900	6,667	2,684	4,498	4,392
Industries	3,886	6,937	6,937	7,477	5,569	12,664	13,577	12,895	11,349
Power Plants	9,204	13,201	13,201	13,201	15,289	7,784	6,892	6,880	6,542
Motor Vehicle	7,363	9,413	5,897	2,363	11,005	4,557	4,574	2,691	4,437

Figure 2.1 PM_{10} Emission Loads by source (in metric tonnes), 2003-2011
(Source: Department of Environment Malaysia, 2012)

2.1.1 Motor Vehicles

One of the major contributors of PM_{10} emissions, especially in urban areas, is motor vehicles, including passenger cars, motorcycles, goods vehicles, and buses/taxis. In 2010, the number of registered cars was nearly 20 million; almost double the number of cars a decade ago (Sansuddin, 2010). Figure 2.2 shows the number of registered vehicles in Malaysia from 2004 to 2011 (Road Transport Department, Malaysia, 2012).

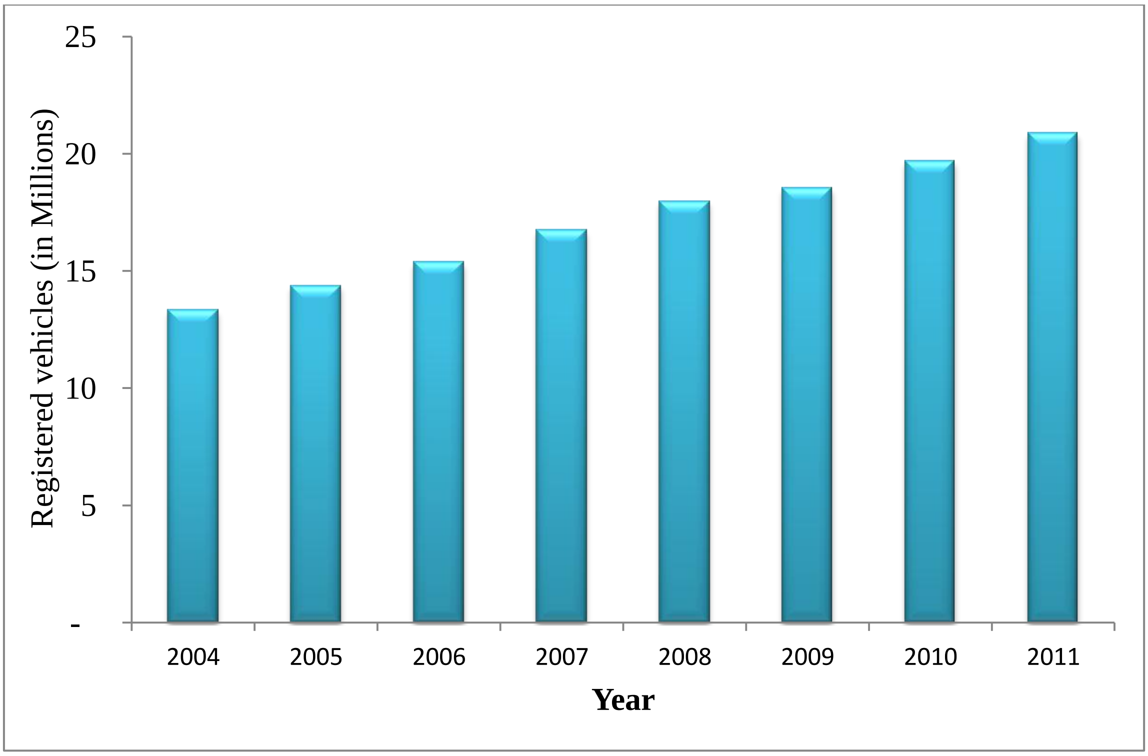


Figure 2.2 Number of registered vehicles in Malaysia from 2004 to 2011
(Source: Road Transport Department, Malaysia, 2012)

Figure 2.3 show the number of registered vehicles in Malaysia by category from 2004 to 2011. Passenger cars and motorcycles increased every year, with average percentage increase of 1% (i.e. an estimated 524,000 passenger cars per year) and 0.85% (i.e., an estimated 452,000 motorcycles per year), respectively. However, goods vehicles showed an increase between 2007 and 2008, but were back to their normal rate between 2009 and 2010. Public transport (taxis and buses) did not show an increase.

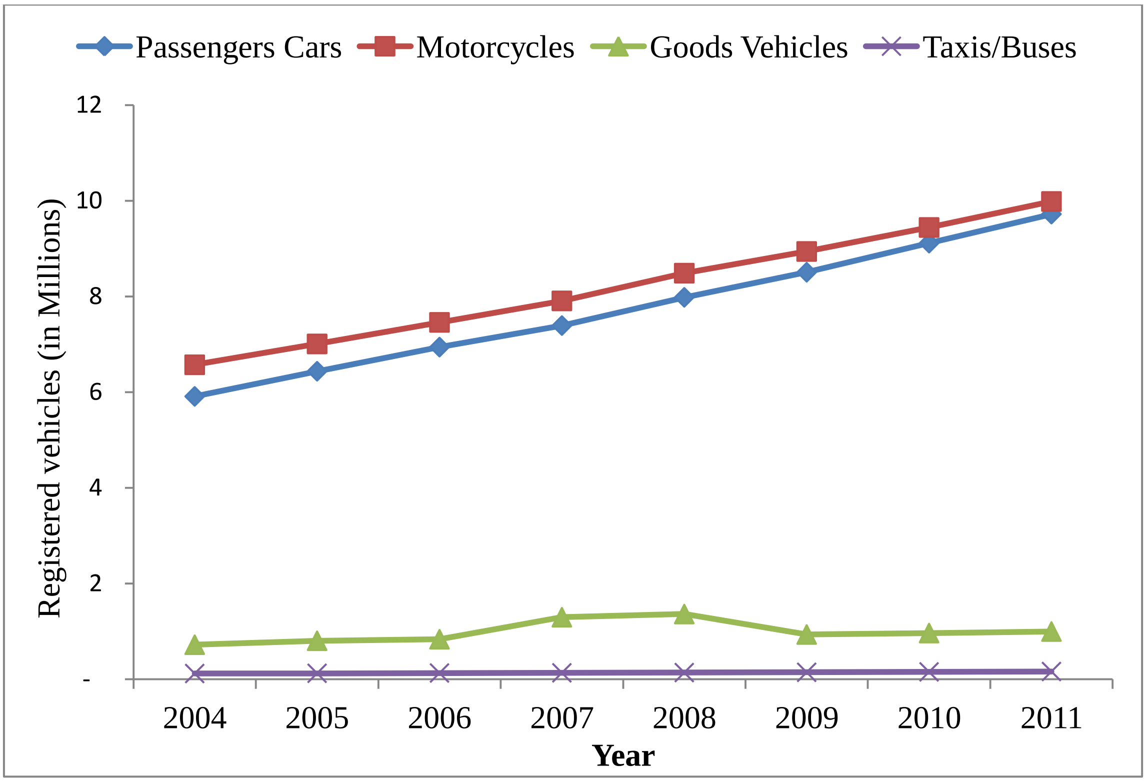


Figure 2.3 Number of registered vehicles in Malaysia by category, from 2004 to 2011. (Road Transport Department, Malaysia, 2012)

2.1.2 Industry / Power plants

The number of industrial related sources contributing to air pollution between 2001 and 2011 are shown in Figure 2.4. In 2010, industrial pollution increased by 71.44%, compared to 2009 with the lowest figure was recorded in 2003. Sansuddin (2010) reported that Malaysia's economic growth was mainly from industries, such as electronics, chemical, and rubber. Combustion processes in industry can cause greenhouse effect, health and natural ecosystem problems (Sansuddin, 2010).

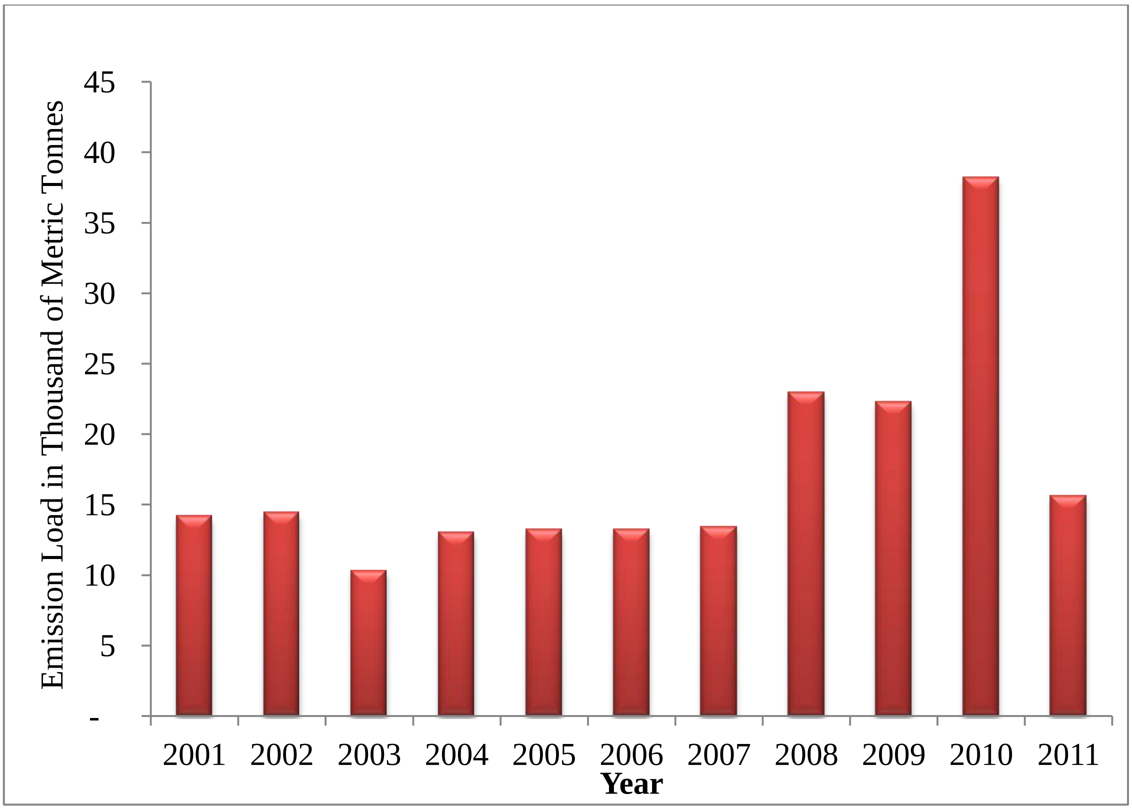


Figure 2.4: Industrial air pollution sources by year (2001 to 2011)
(Source: Department of Environment Malaysia, 2012)

2.1.3 Open Burning / trans-boundary

Haze problems in Malaysia due to trans-boundary pollution from Indonesia (West of Kalimantan) have affected Sarawak in 2002. This resulted in three to 22 unhealthy days between July and September 2002 (Department of Environment Malaysia, 2003). In 2004, forest fires in Sumatra caused a moderate haze in Malaysia (June, August, and September) and fires in Kalimantan also affected southern parts of Sarawak (Department of Environment Malaysia, 2005).

In 2005, the air pollution index exceeded 500 at Kuala Selangor and Pelabuhan Klang, due to trans-boundary air pollution from land and forest fires in the Riau Province of Central Sumatra, Indonesia (Department of Environment Malaysia, 2006). For the years 2009 and 2010 trans-boundary pollution occurred during the hot and dry conditions.

Similar finding were found by other researchers. Manomaiphiboon et. al., (2009) found open burning as one of the major sources of PM_{10} at Chiang Rai, Thailand. Open burning in Thailand include the burning of agricultural waste and forest fires (Sirimongkonlertkun, 2012). Table 2.1 shows summary of international open burning and wildfire episodes. Afroz et al., (2003) and Abas et. al., (2004) stated that open burning have a relation to PM_{10} concentration because PM_{10} is the main pollutant when the open burning or wildfires occur.

Table 2.1: Summary of international open burning and wildfire episode (Modified from Finlay et al., 2012)

Year	Country	Details
1994	Australia	Wildfires in New South Walus in summer of 1993-1994 that burnt over an area of 800 000 hectares and 225 homes were destroyed (NSW Government, 2007).
1997	Indonesia	Over 5 018 000 hectares were burnt in Indonesia (Butler, 2003) and severe adverse health effects in Malaysia , Singapore and Indonesia (Glover and Jessup, 2002).
2003	Canada	Started in British Columbia with 2500 wildfires in 2003 during a period of dry weather and particularly hot season (Filmon, 2003).
2007	USA	Wildfires in Southern California burnt over 410 000 acres, and 1500 homes were destroyed (Flaccus, 2007).
2009	Australia	Known as Black Saturday for Austarlian when over 141 600 hectares were burned with temperatures in Melbourne reaching the hottest in record (46.4°C) and wind speed of 100km/hr (Cameron et. al., 2009).
2010	Russia Federation	Western part of Russian Federation have experienced over 20 000 forest fires over an area of 2800 km^2 at extreme heat (WHO, 2010).

2.2 CHARACTERISTICS OF PARTICULATE MATTER

PM_{10} can come in many shapes and sizes, and can be solid or liquid droplets (EPA, 2000). Particulate matter were categorized into three modes as nucleation mode, accumulation mode and coarse mode by Cambra-Lopez et al., (2010). Particulate matter was classified by size (fine and coarse) and origin (primary and secondary). Figure 2.5 shows a schematic diagram of particle classification, size distribution, formation and elimination processes, modes of distribution, and composition.

Nucleation mode is the smallest diameter and size of group. It is directly emitted from combustion processes and has a short atmospheric lifetime, because it easily coagulates into larger sizes. Accumulation mode is a medium size of diameter and includes the condensation of particles. This mode has a longer lifetime than nucleation, because this particle is too large for Brownian motion and too small to settle from the air rapidly. According to Cambra-Lopez et al., (2010), this mode is usually eliminated from the atmosphere by washout, dry (gravitational settling or inertial impactions of particles) or wet deposition (rain or snow). Coarse mode is a particulate matter with mechanically generated particles. This happens because the particle settles rapidly and is sedimented by gravitational forces.

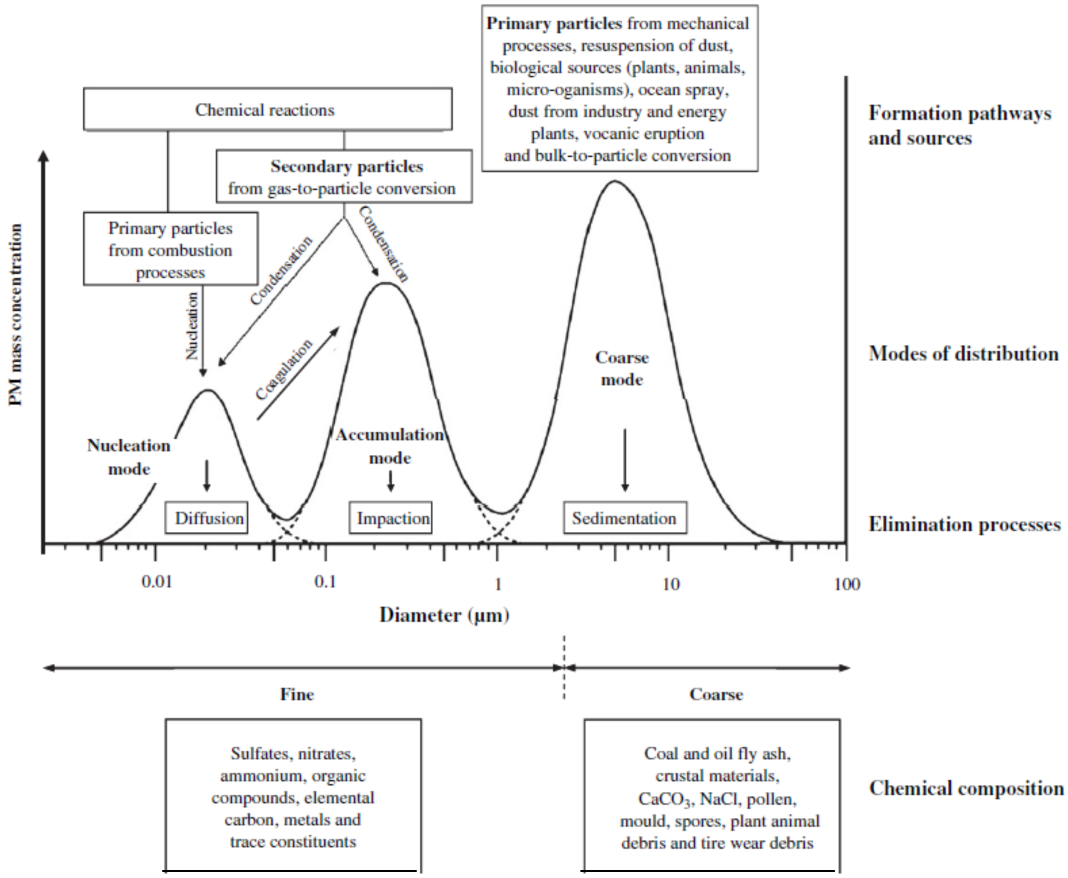


Figure 2.5: Schematic diagram of particle classifications, size distribution, formation and elimination processes, modes of distribution, and composition (Source: Cambra-Lopez et al., 2010)

2.3 EFFECT OF PM_{10} ON HUMANS

Several scientific studies have linked PM_{10} with health problems, such as aggravated asthma, premature mortality, chronic respiratory disease and hospital admission. Figure 2.6 shows particulate matter in the respiratory system. Particulate matter, of more than 10µm, will accumulate in the upper parts of the respiratory system (Figure 2.6(a)). For particulate matter between 1mm and 10µm, the particles will accumulate in the middle part of the respiratory system and the tracheobronchial region (Figure 2.6(b)). Particles less than 1µm will accumulate in the most remote portions of lungs (alveoli or air sacs) as shown in Figure 2.6(c) (BC Lung Association, 2013 and Ramli, 2006).

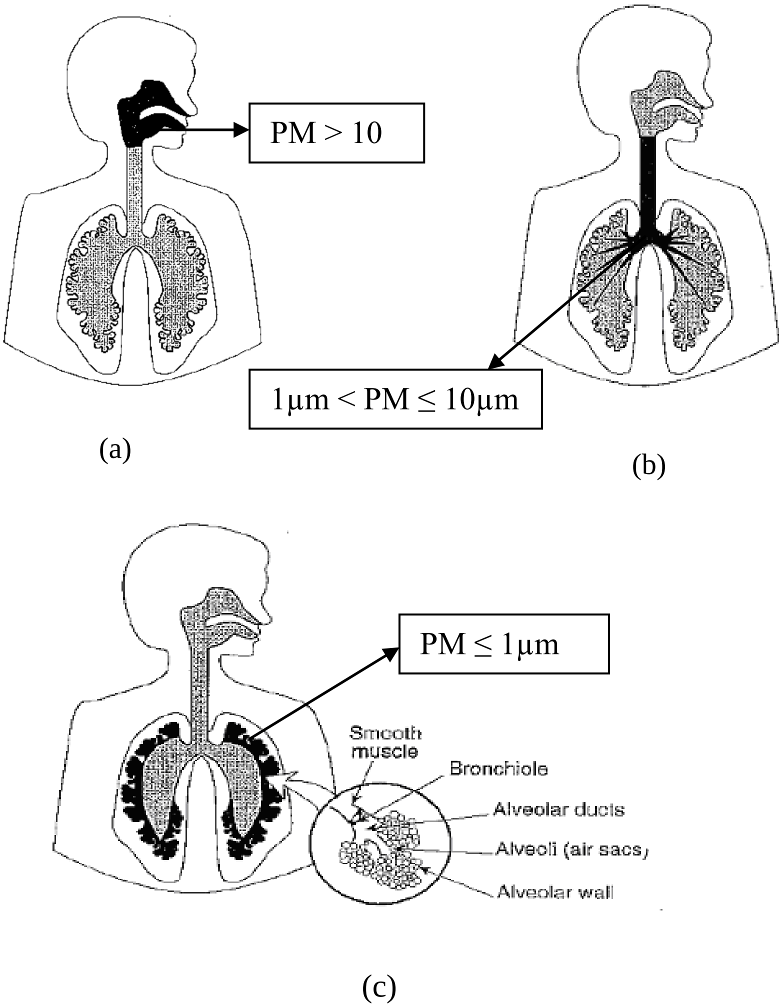


Figure 2.6 Inhalation of Particulate Matter: (a) PM > 10 μm, (b) 1μm < PM ≤ 10μm and (c) PM ≤ 1μm (BC Lung Association, 2013 and Ramli, 2006).

In September 2011, the WHO Global Burden of Disease report said that almost 795,000 premature deaths per year were attributable to PM air pollution in Asian cities (Keefe, 2011). There was relationship between PM_{10} and mortality stated by Feirro (2000). The figure of deaths due to air pollution was 63 in Belgium (1930), 20 in Pennsylvania (1948), 4000 in London (1952), 200 in New York (1953), and 7000 in London (1962). PM_{10} and cardiopulmonary and lung cancer mortality was also discussed by Pope et al., (1995). In addition, Keefe (2011) said that high levels of PM_{10} (> 500 μ/m^3) can cause premature death, such as in London (1952) and studies in the US, Europe, and elsewhere, found an association of PM with mortality at much lower levels (<50 μ/m^3).

Daniels et al., (2004), reported that the mortality rate increases 0.2% per 10μg/m^3 of PM_{10} in a study of 20 cities of USA.

United States Environmental Protection Agency (US EPA, 2000), studied the association between PM_{10} and human health in different levels of air quality index values. Table 2.2 shows a comparison of effects on human health for $PM_{2.5}$ and PM_{10}.

Table 2.2 Comparison of effect on human health for $PM_{2.5}$ and PM_{10}
(Source: Adopted from US EPA, 2000)

API Values	Air Quality Descriptor	Health Concerns*	
		$PM_{2.5}$	**PM_{10}**
0 - 50	Good	None	None
51 - 100	Moderate	None	None
101 - 150	Unhealthy for sensitive groups	People with respiratory or heart disease, the elderly, and children, should limit prolonged exertion.	People with respiratory disease, such as asthma, should limit outdoor exertion.
151 - 200	Unhealthy	People with respiratory or heart disease, the elderly, and children, should avoid prolonged exertion; everyone else should limit prolonged exertion.	People with respiratory disease, such as asthma, should avoid outdoor exertion; everyone else, especially the elderly and children, should limit prolonged outdoor exertion.
201 - 300	Very unhealthy	People with respiratory or heart disease, the elderly, and children should avoid any outdoor activity; everyone else should avoid prolonged exertion.	People with respiratory disease, such as asthma, should avoid any outdoor activity; everyone else, especially the elderly and children, should limit outdoor exertion.
301 - 500	Hazardous	Everyone should avoid any outdoor exertion; people with respiratory or heart disease, the elderly, and children should remain indoors.	Everyone should avoid any outdoor exertion; people with respiratory disease, such as asthma, should remain indoors.

* PM has set two sets of cautionary statement i.e. (1) $PM_{2.5}$ and PM_{10}.