MODELING OF PARTICULATE MATTER DURING HAZE POLLUTION USING LINEAR REGRESSION METHOD: A CASE STUDY IN WESTCOAST OF PENINSULAR MALAYSIA

SITI NADHIRAH BINTI REDZUAN

FACULTY OF CIVIL ENGINEERING TECHNOLOGY
UNIVERSITI MALAYSIA PERLIS
2022

MODELING OF PARTICULATE MATTER DURING HAZE POLLUTION USING LINEAR REGRESSION METHOD: A CASE STUDY IN WESTCOAST OF PENINSULAR MALAYSIA

by

SITI NADHIRAH BINTI REDZUAN

Report submitted in partial fulfilment
of the requirements for the degree
of Bachelor of Engineering

UNIVERSITI MALAYSIA PERLIS

AUGUST 2022

ACKNOWLEDGEMENT

First and foremost, all praises to Allah for the strength and His blessings in completing this thesis report. I wish to express my sincere thanks to the authority of Civil Engineering Technology Faculty, Universiti Malaysia Perlis for providing me opportunity for my final year project. Gratitude to my supervisor, PM. Ts. Dr. Norazian Binti Mohamed Noor for her supervision, constant support, who expertise, friendly advice and sharing knowledge regarding my thesis writing. Besides that, I would like to thank to Ts. Dr. Sara Yasina Binti Yusuf and Madam Zaity Syazwani Binti Mohd Odli who help me to improvise my report, as my research Examiners. Furthermore, I would like to thank Department of Environment Malaysia (DOE) for providing the air quality data as a secondary to my research.

Additionally, I thank my fellow air pollution teammates Syaza Ezzati Binti Baidrulhisham, Nurin Haziqah Binti Hamidi and Nur Ain Binti Jamal for the good teamwork, stimulating discussion and support to finish this thesis report. Not to forget my classmates those show a concern and help me in order to complete this research.

Lastly, I would like to thank my parents Mr. Redzuan Bin Saidin and Mrs. Fadzilah Binti Mat Esa for supporting me spiritually throughout my life and also thanks to my beloved family which always give moral support to me. My spouse Amirul Hazim Bin Rozahan deserves a lot of credit for always being there for me through good times and bad. Finally, I would want to express my gratitude to everyone who was directly or indirectly engaged in my study by contributing ideas and exchanging expertise.

APPROVAL AND DECLARATION SHEET

This project report titled Modeling of Particulate Matter During Haze Pollution Using Linear Regression Method: A Case Study in Westcoast of Peninsular Malaysia was prepared and submitted by Siti Nadhirah binti Redzuan (Matrix Number: 181130714) and has been found satisfactory in terms of scope, quality, and presentation as partial fulfilment of the requirement for the Bachelor of Environmental Engineering with Honours in Universiti Malaysia Perlis (UniMAP).

Checked and Approved by

(PM. TS. DR. NORAZIAN BINTI MOHAMED NOOR)
Project Supervisor

Faculty of Civil Engineering Technology
Universiti Malaysia Perlis

August 2022

PEMODELAN JIRIM ZARAH SEMASA PENCEMARAN JEREBU MENGGUNAKAN KAEDAH REGRESI LINEAR: KAJIAN KES DI PANTAI BARAT SEMENANJUNG MALAYSIA

ABTRAK

Malaysia telah menghadapi peristiwa jerebu rentas sempadan setiap tahun di mana udara mengandungi bahan zarahan, terutamanya PM_{10}, yang menjejaskan kesihatan manusia dan alam sekitar. Oleh itu, adalah penting untuk membangunkan model ramalan PM_{10} untuk maklumat awal dan amaran amaran kepada pihak yang bertanggungjawab agar mereka dapat mengurangkan dan merancang langkah berjaga-jaga semasa kejadian tersebut. Kajian ini bertujuan untuk menyiasat statistik deskriptif variasi tahap PM_{10} semasa kejadian zarahan tinggi dan menganalisis gas surih dan parameter cuaca yang dikaitkan dengan PM_{10} di Pasir Gudang, Melaka dan Petaling Jaya pada tahun 1997, 2005, 2013 dan 2015. Data sekunder setiap jam telah diperolehi daripada Jabatan Alam Sekitar (JAS). Korelasi Pearson digunakan untuk mencari korelasi antara PM_{10} dan parameter bagi setiap tempat. Analisis korelasi Pearson menunjukkan bahawa variasi dalam kepekatan PM_{10} secara amnya bertepatan dipengaruhi oleh CO dan SO_2. Regresi Linear Berbilang dan Regresi Kuantil digunakan untuk membangunkan dan membandingkan model yang paling sesuai untuk ramalan PM_{10} semasa kejadian jerebu merentas sempadan. Tiga model Regresi Linear Berganda (MLR) berperingkat yang berbeza untuk meramal kepekatan PM_{10} kemudiannya dibangunkan berdasarkan tiga jam ramalan yang berbeza, iaitu t_{-24}, t_{-48}, dan t_{-72}. Bagi Regresi Kuantil, tiga nilai persentil telah dipilih untuk ramalan PM_{10} iaitu 0.25, 0.50 dan 0.75. Daripada perbandingan antara MLR dan QR, model $PM_{10\ t-24}$ dengan nilai persentil 0.75 adalah model QR terbaik untuk meramalkan PM_{10} semasa kejadian jerebu merentas sempadan berbanding model $PM_{10\ t-48}$ dan $PM_{10\ t-72}$.

MODELING OF PARTICULATE MATTER DURING HAZE POLLUTION USING LINEAR REGRESSION METHOD: A CASE STUDY IN WESTCOAST OF PENINSULAR MALAYSIA

ABSTRACT

Malaysia has been facing transboundary haze events every year in which the air contains particulate matter, particularly PM_{10}, that affects human health and the environment. Therefore, it is crucial to develop a PM_{10} forecasting model for early information and warning alerts to the responsible parties in order for them to mitigate and plan precautionary measures during such events. This study aimed on investigating the variation descriptive statistic of PM_{10} level during high particulate event and analyzing trace gases and the weather parameters that associated with PM_{10} at Pasir Gudang, Melaka and Petaling Jaya in 1997, 2005, 2013 and 2015. Hourly secondary data were acquired from the Department of Environment (DOE). Pearson correlation was used to find the correlation between PM_{10} and parameters for each place. Pearson correlation analysis showed that variations in PM_{10} concentrations coincides influenced by CO and SO_2. Multiple Linear Regression and Quantile Regression were used to develop and compare the best-fitted model for PM_{10} prediction during transboundary haze events. Three different stepwise Multiple Linear Regression (MLR) models for predicting the PM_{10} concentration were then developed based on three different prediction hours, namely t_{-24}, t_{-48}, and t_{-72}. For Quantile Regression, three percentile values were chosen for PM_{10} prediction which were 0.25, 0.50 and 0.75. From the comparison between MLR and QR, the $PM_{10\ t-24}$ model with percentile value of 0.75 was the best QR model to predict PM_{10} during transboundary haze events compared to $PM_{10\ t-48}$ and $PM_{10\ t-72}$ models.

TABLE OF CONTENT

CHAPTER 3 METHODOLOGY

CHAPTER 4 RESULT AND DISCUSSION

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

API	Air Pollution Index
CO_2	Carbon Dioxide
DOE	Department of Environmental
IA	Index of Agreement
MAAQG	Malaysian air ambient quality guideline
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
NEM	North-East Monsoon
NO_2	Nitrogen Dioxides
NO_3	Nitrate
NO	Nitric Oxides
NO_x	Nitrogen Oxides
O_3	Ozone
PM_{10}	Particulate Matter with Diameter 10 μm
QR	Coefficient of determination
RMSE	Root Mean Square Error
R^2	Quantile regression
SO_2	Sulphur Dioxide
SWM	South-East Monsoon
UVB	Ultraviolet
UV	Ultraviolet
USEPA	United States Environmental Protection Agency

CHAPTER 1

INTRODUCTION

1.1 Research Background

Recently, air quality has emerged as a significant environmental concern on a global scale (Latif et al., 2014; Abdullah et al., 2019). Because of this, a lot of human activities now considerably pollute the air in these nations by releasing hazardous particles (usually known as PM10 or PM2.5) and gases (including ozone and nitrogen dioxide). Approximately seven million premature deaths globally occur each year as a result of poor air quality, with growing nations like Indonesia, India, and China suffering the greatest effects, according to the World Health Organization (WHO News). Unhealthy environments have a close correlation with the detrimental impacts on human health, which in turn lowers life pleasure. It is significant to note that some groups of people, like the elderly, kids, and those with heart or lung conditions (such as asthmatics), have been proven to be more vulnerable to breathing in ambient pollution (Stanek & Brown, 2019). In addition, air pollution also has negative effects on visibility, materials, buildings, artwork, and ecosystems (Ilyas et al. 2009; Mage et al. 1996; Riga-Karandinos and Saitanis, 2005). The ability of air pollutants to lodge in the bronchi and lungs and create health issues, particularly particulate matter (PM) smaller than approximately 10 micrometers, or PM_{10}, should have attracted a lot of attention (Nawrot et al., 2007).

Unchecked forest fires that originated on Sumatra, an Indonesian island, extended during the burning season, resulting in a territorial fog scenario that had an impact on all of the ASEAN nations, but Malaysia. Past haze outbreaks that afflicted Malaysia between June and September in the years 1983, 1990, 1991, 1994, 1997, 1998, and 2005 (Mahmud, 2008) have been well-documented in terms of their causes and effects (Keywood et al., 2003). However, particulate matter (PM_{10}) was the primary air pollutant characteristic during Southeast Asia's biggest haze epidemic in 1997 (DoE, 1997). Small, evenly dispersed particles with sizes between 0.1 and 1.0 mm that are scattered at high magnifications but scarcely noticeable to the unassisted eye are referred to as haze (Soleiman et al., 2003). On the other hand, cloudiness is also referred to by (Cheng et al., 2013) as tiny, suspended strong or fluid particles.

In several Southeast Asian countries, notably Indonesia, Malaysia, Singapore, and Brunei, as well as, to a lesser extent, Thailand, Vietnam, and the Philippines, haze pollution is a common occurrence (Apichai, 2015). Haze event, sometimes referred to as the high particle event (HPE), is one of the main causes of air pollution in Malaysia's west coast. The issue is mostly the result of massive tracts of Indonesian forest being burned. During the southwest monsoon season, which typically lasts from June to September, high particulate event was common.

The burning of an estimated 45,600 km^2 of vegetation on the Indonesian islands of Kalimantan and Sumatra was blamed for the smoke haze pollution that severely harmed substantial portions of Southeast Asia in the second half of 1997 (Heil, 2001). When the fires were put out and the area's air quality was restored in 1997, the effect of the fires in Kalimantan and Sumatra on the ambient air quality became apparent by July, peaked in September and October, and then began to wane by November. For the first time since 1997, 34 regions in the nation recorded poor air quality status on September 15, 2015. Putrajaya, Kuala Lumpur, Selangor, Negeri Sembilan, and Melaka all closed their schools as a result of the API reading surpassing 200. (DOE, 2015). Haze, or "the pollution that people perceive," as Hyslop (2009) puts it, not only immediately depresses people's moods owing to the shade of the gloomy sky. It's also a symptom of elevated PM_{10} levels, which have been linked to adverse effects on the environment, biological systems, and general health (Tie et al., 2009). (Solomon et al., 2007).

The requirement for a reliable model to forecast particle concentrations during high particle events derives from the negative effects of haze episodes on both people and the environment. Air pollution modelling (APM) is being created and used to improve knowledge, research, assessment, and control of the quality of the atmospheric environment and the dispersion of hazardous pollutants. Modeling air pollutants is important because it may be used to assess the link between the sources of pollution and how they affect the quality of the surrounding air. The ability to advise the authorities of the present and most probable future trend is a requirement for an efficient air quality control system. In this research, the use of linear regression models to forecast PM_{10} concentration during high particulate events was undertaken.

1.2 Problem Statements

The haze event, also known as the high particulate event (HPE), typically takes place during the Southwest monsoon season as a result of high concentrations of atmospheric particle matter, especially PM_{10}, which are primarily brought on by anthropogenic activities like open forest, construction sites, landfills, agriculture, wildfires, and waste burning. As particulate matter emissions from local traffic and industrial sources rise, haze's impact on both the outside air and the internal environment gets worse. In Malaysia, high particulate events are a significant cause of air pollution. In order to estimate the average PM_{10} concentration during HPE, it is crucial to construct linear regression models for forecasting PM_{10} concentration.

Particulate matter modeling research was sparse during Malaysia's high-particulate event. Most of the studies modeled PM_{10} concentration during non-high particulate event using linear regression compared to high particulate event. Besides that, as machine learning only provides anticipated values, such as ANN models, it does not provide information on how PM_{10} relates to trace gases and whether parameters.

Additionally, linear regression is more straightforward to use, analyze, and train using. The linear regression models Multiple Linear Regression and quantile regression were used in this research to estimate PM_{10} concentrations during high particulate events. The models' worth was then evaluated.

1.3 Research Objectives

It is particularly important to ensure that this research will contribute to the environmental engineering field of studies. Therefore, this research was conducted with the three main objectives, which are:

- To analyze the trend and association of PM_{10} concentration with other trace gases and weather parameter during high particulate event.
- To develop regression models, consist of Multiple Linear Regression (MLR) and Quantile Regression (QR) of PM_{10} concentration during high particulate event.
- To validate the regression models of PM_{10} concentration using several performance indicators.

1.4 Scope of Research

In this study, study regions were chosen using daily, hourly, and monthly PM_{10} monitoring records from three stations: Pasir Gudang, Melaka, and Petaling Jaya. These three places were picked because they are close to Sumatera, Indonesia, which was the source of transboundary air pollution brought on by high particulate events in Southeast Asia, and because they are on the west coast of peninsular Malaysia. The years when Malaysia had high particulate events—1997, 2005, 2013, and 2015—were when the data were gathered. PM_{10} was selected as the study's main pollutant since it has been recognized as a significant atmospheric contaminant.

Firstly, by using descriptive statistics, PM_{10} measurement data from HPE (1997, 2005, 2013, 2013) were reported. A boxplot, histogram, and time series plot were employed as graphical presentations to observe the trend of PM_{10} during high particulate events. To ascertain the relationship between PM_{10} and the trace gases and whether parameter in these three research regions, Pearson Correlation was performed.

Quantile regression (QR) and multiple linear regression (MLR) were the two types of linear models that were created for prediction purposes. Eighty percent of the data were used to create the model and the other twenty percent of the data were utilized for model validation. Quantile regression was performed with percentile values of 0.25, 0.50, and 0.75. Several performance measures, including mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and index of agreement (IA), were used to validate the model. The computed performance metric was used to choose the optimal linear model.

1.5 Significance of The Project

One of the major environmental issues in Malaysia is air pollution. A portion of the pollution is caused by air pollutants, mainly PM_{10} particles. Additionally, PM_{10} has the potential to be harmful to human health due to its ease of passage through the nose, the body's primary defense mechanism, and its ability to clog lung tissues. We can raise awareness of PM_{10} mostly in relation to the potentially harmful effects they can have on the human body since this study is intended to assess the trend of PM_{10} during high particulate events.

In addition, the characteristics of PM_{10} during the high particle event time in the chosen regions were carefully investigated. Since the trace gases and whether parameter in the study areas will be obtained from specific governmental or regulatory entities, this information may be used as a starting point for authorities to design a risk management strategy or it can be used to further the research by other investigators.

However, since quantile regression is a variation of standard linear regression that estimates the conditional median of the outcome variable and can be used when the assumptions of linear regression are violated, the results achieved may give benefit to everyone. PM_{10} particles may go deep into the lungs, they can have a negative effect on your respiratory and cardiovascular systems. Lastly, since we were aware of the PM_{10} concentration in the areas impacted by haze, human exposure to PM_{10} could be decreased.

CHAPTER 2

LITERATURE REVIEW

2.1 Air Pollutant in Malaysia

With rapid economic growth, urbanization, and a high number of cars on the road, air pollution is a global health problem, and Malaysia is not an exception to this. Malaysia is subject to significant transboundary air pollution, which shows up as sporadic "haze" outbreaks linked to local agricultural land burning. Air pollution is a significant problem in Malaysia on both a local and international level. Numerous health repercussions of air pollution include headaches, coughing, and breathing difficulties among commuters as well as increased throat pain and respiratory illnesses, including asthma, in children and the elderly. The local economy also suffers because air pollution affects shipping, tourism, medical expenses, hospitalization costs, and missed income opportunities (Zahari et al. 2016).

Since it is one of the most important factors affecting people's quality of life and standard of living, air pollution is a major issue that needs to be tackled right away and severely by all relevant agencies and governments. Air pollution is becoming a major environmental issue worldwide in Malaysia as a result of the growing number of transportations (mobile sources), trans-boundary pollutants from surrounding countries, and industrial operations (stationary sources). These are Malaysia's main causes of air pollution (Mutalib et al. 2013). When air pollution enters the atmosphere, it may harm the built and natural surroundings as well as lead to acute and chronic illnesses in humans and other living things. (Moustris et al. 2010). Poor air quality can cause irritation of the nose, throat, eyes, and skin as well as headaches, exhaustion, dizziness, and breathing difficulties (Xie et.al. 2009).

One of the trickiest problems in the world is preventing pollution from its source. To offer residents with up-to-date information on the concentrations of major pollutants, the Malaysian Department of Environment (DOE) has been regularly monitoring air quality and gathering data (Dragomir, 2010). As soon as the lack of compliance is discovered, the information can be used to alert or forewarn planners or decision-makers in the event that adverse health effects result. PM_{10} and surface ozone (O_3) are two important air pollutants that have been highlighted as one of the main issues with a high potential for harming human health, especially in Malaysia's urban and suburban regions (Mahiyudin et.al. 2013).

Public perception of risks is frequently too low when considering air pollution's harmful impacts. This is important because, if they feel that their health is in danger, individuals are more inclined to take action to reduce the adverse effects of air pollution. Sadly, there is often a mismatch between popular perception and reality when it comes to air pollution (Schmitz et.al. 2018). Recent research found that the public in a large portion of Asia has little knowledge of the long-term impacts of air pollution on health, that health authorities are not a reliable source of information, and that seasonal fluctuations in air quality are what pique the people's interest (Manisalidis et.al. 2020).

Limits for air quality pollutants were set by Malaysia's Department of the Environment (DOE) in 1989. The Malaysian Air Quality Guidelines (RMG) defined concentration limits for several air pollutants that might have a detrimental effect on the health and welfare of the general people (Abd Rani et al., 2018). Malaysia uses the Air Pollutant Index to gauge the quality of the country's general air (API). When reporting air quality, the API uses ranges of values that are simple to interpret rather than exact concentrations of air contaminants. According to the action criteria of the National Haze Action Plan, this score also indicates the impact on human health, which can range from favorable to unfavorable. The Pollutant Standard Index (PSI) methodology used by the United States Environmental Protection Agency (US-EPA) and Malaysia is quite similar to each other1 (DoE, 2022). The air pollutants included in the API calculations were ground level ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), and particulate matter smaller than 10 millimeters (PM_{10}). Each sub-greatest index's value serves as the basis for the API value. The values of the API are used to notify or caution the public about possible health dangers. As the foundation for determining the API value, the ambient standards adopted in the Malaysia Ambient Air Quality Guidelines (MAAQG) are displayed in Table 1.1.

Table 1.1 Malaysian Ambient Air Quality Guidelines (DOE, 2016)

Pollutants	Averaging Time	Malaysia Guidelines	
		ppm	(μg/m³)
Ozone (O^3)	1 hour 8 hours	0.10 0.06	200.0 120.0
Carbon Monoxide (CO)	1 hour 8 hours	30.0 9.0	35.0 -
Nitrogen Dioxide (NO_2)	1 hour 24 hours	0.17 0.04	320.0 10.0
Sulphur Dioxide (SO_2)	1 hour 24 hours	0.13 0.04	350.0 105.0
Particulate Matter (PM_{10})	24 hours 12 months	-	150.0 50.0
Total Suspended Particulate (TSP)	24 hours 12 months	-	260.0 90.0
Lead (Pb)	3 months	-	1.5

2.2 Haze

For the past few decades, Southeast Asia (SEA), notably Malaysia, has experienced haze virtually every year. Haze is associated with high levels of air pollution; it reduces visibility and has a detrimental effect on human health in the affected SEA nations (Latif et.al. 2018). Haze is defined as a region of the atmosphere with a high concentration of tiny particles (0.1-1.0 m in diameter) that block horizontal view and give the atmosphere a distinctive opalescent look (Gao and Niu, 2007). Haze is more likely to form in Malaysia every year between the months of January and February and June and August (Hanafi et.al. 2018). This is a result of dry weather spells. When sunlight strikes minute airborne pollutants, haze is the result. Particles can absorb some light. Before it reaches an observer, further light is dispersed away. More pollutants result in more light being absorbed and scattered, which lessens the brightness and color of what we can see (EPA, 2021). 2015 was the most recent time it occurred. According to table 1.2 below, the air pollution index (API) readings in a number of Malaysian districts were over the harmful limit. The Malaysian Air Pollution Index standard was developed by the Department of Environment (DOE), and it is as follows: good (0-50), moderate (51-100), unhealthy (101-200), very unhealthy (201-300), and dangerous (201-300). in excess of 300. Haze has afflicted Malaysia for the past seven years, and it is thought that peat and forest fires in Indonesia are to blame. Malaysia's haze history may be traced back to 1983, when the country's first record haze occurrence that impacted daily life. Forest fires in Sumatra were believed to have caused extremely foggy weather conditions in the country in 1991. Three years later, during the month of September, an even more severe haze occurrence occurred that lasted for over a month. Forest fires in Kalimantan and Southern Sumatra have been identified as the primary caused the problem.

When haze returned to Malaysia in 1997, the dry weather and steady atmospheric conditions, together with emissions from regional sources of pollution including cars, factories, and open waste burning, made the problem worse. This haze occurrence is considered one of the worst since it occurred at the same time as El Nino, which prolonged the dry season that year (Keywood, 2003). The second half of 1997 saw the international air pollution catastrophe known as the Southeast Asian Haze, which had a negative impact on Southeast Asia's health and atmospheric visibility. The 1997 haze, which is regarded as the worst Southeast Asian haze phenomenon ever, is estimated to have cost $9 billion in total, mostly because it disrupted air travel and commercial operations and caused health problems (Tacconi and Luca, 2016). The 1997 fires in Kalimantan and Sumatra had an impact on ambient air quality by July, which peaked in September and October before waning by November as the delayed monsoonal rain put out the flames and improved the local air quality.

Malaysia experienced an air pollution issue in 2005, largely brought on by fires in neighboring Indonesia. When forest fires on the Indonesian island of Sumatra spread haze throughout Malaysia in August 2005, the air quality in several regions and the nation's capital, Kuala Lumpur, reached dangerous levels. The Malaysian government issued emergency declarations, closed schools, and initiated crisis discussions with Indonesian authorities in the impacted areas. During the dry season, farmers frequently burn brush and woodland to make way for agricultural land. The worst smog to affect Malaysia since 1997 has occurred there in 2005. (BBC News, 2005).

Health officials urged residents to stay indoors with the doors locked on August 10, 2005, when air quality in Kuala Lumpur, the capital of Malaysia, was at an extremely low level. To protect children from the smog, some schools were closed. As a result of dangerously high levels of Kuala Selangor neighborhood and Port Klang, the 12th biggest port in the world, in August 11. (defined as a value greater than 500 on the Air Pollution

Index or API). Since Sarawak was declared under a state of emergency for comparable reasons during the September 1997 haze, Malaysia has not had a state of emergency.

The Southeast Asian Haze of 2013 was a haze issue that mostly affected Brunei, Indonesia, Malaysia, Singapore, and Southern Thailand in the months of June and July 2013. A lot of Sumatra and Borneo were burned extensively during the hazy era. The haze, according to NASA's Terra and Aqua satellite images, was mostly caused by smoke from fires in Indonesia's Riau region (Show and Chang, 2016).

The Southeast Asian haze of 2013 was significant for setting pollution records in Singapore and numerous regions of Malaysia. On June 21, 2013, Singapore's 3-hour Pollution Standards Index recorded a record high of 401, breaking the previous mark of 226 set during the Southeast Asian Haze in 1997. (BBC News, 2014). When the Air Pollution Index (API) in Muar, Johor, reached 746 at 7 a.m. on June 23, it was about 2.5 times above the minimal range of the Hazardous category, leading to the declaration of emergency in the two towns (which was later revoked on June 25 in the morning) (The Star, 2013).

Southeast Asian nations Brunei, Indonesia (particularly the islands of Sumatra and Borneo), Malaysia, Singapore, southern Thailand, Vietnam, Cambodia, and the Philippines were all affected by the 2015 Southeast Asian haze, which was an air pollution disaster. From at least late June to at least the end of October, Indonesia was impacted by the haze, which in September began to affect other nations internationally. It was the most recent instance of the Southeast Asian haze, a persistent problem that occurs throughout the region's dry seasons with different intensity. Specifically in the Indonesian islands of Sumatra and Kalimantan, slash-and-burn forest management techniques resulted in forest fires that started small but quickly spread during the dry season. 2015, The New York Times.

The issue affected more than 28 million people in Indonesia alone, and more than 140,000 individuals there suffered respiratory illnesses. A 2016 Harvard-Columbia University research found that the haze was responsible for over 100,000 extra fatalities, the most of which (> 90,000) occurred in Indonesia (Koplitz et al, 2016). Later, however, Indonesian, Singaporean, and Malaysian health officials denied the assertion. There is evidence that the haze brought on by the Indonesian forest fires is making Singaporeans worse with haze-related conditions like acute conjunctivitis and upper respiratory diseases. The haze forced the closure of schools in Indonesia, Malaysia, and Singapore; in Malaysia alone, approximately four million pupils were impacted. The Kuala Lumpur Marathon in Malaysia and the 2015 FINA Swimming World Cup in Singapore were two events that were hampered or perhaps cancelled as a result of the smog. In the final days of October 2015, heavy rainfall in Sumatra and Kalimantan greatly decreased the size and quantity of flames and improved the air quality in most impacted districts.

Haze outbreaks often take place during the dry season (southwest monsoon), when Sumatra and the southern half of Kalimantan encounter drier-than-average conditions, leading to extensive and uncontrollable forest fires. The southwest monsoon (summer), which starts in May and lasts for three to four months, generally ends in August, is characterized by low-level southwesterly winds. Major haze outbreaks in Malaysia and the greater Southeast Asian region are virtually certain to occur around this time due to long-term and large-scale forest fires in Sumatra and Kalimantan, with primary winds blowing northwestward (Tangang, 2010).

Table 1.2: Air Pollutant Index (API) in Malaysia (DOE, 2021)

API	Status	Health Effect	Health Advice
0-50	Good	Low pollution with no negative health effects	There are no limitations on public outdoor activities. Keep up a healthy lifestyle.
51-100	Moderate	Mild pollution that has no negative effects on health	No rules about what the public can do outside. Keep a healthy way of life.
101-200	Unhealthy	Worsens the health of high-risk individuals, i.e., those with heart and lung issues	For those at high risk, few outside activities are available. People should engage in less dangerous outdoor activities

201-300	Very Unhealthy	People with heart and lung conditions have worse health and have a lower tolerance for exercise. impact the public health	Elderly and at-risk Individuals are recommended to limit their exposure to intense outdoor activities and remain indoors.
>300	Hazardous	Hazardous to high-risk people and ublic health	Outdoor activities are not permitted for the elderly or those at high risk. The general public is urged to avoid outdoor activities.
>500	Emergency	Hazardous to high-risk people and ublic health	The general public is instructed to abide by National Security Council directives and to consistently pay attention to media announcements.

2.3 Effect of the haze

The impact of the transboundary smoke haze on Southeast Asia's public health and country economies has led to it being a significant issue. The smog has a negative impact on the health of both our nation and our neighbors. These health effects have had a considerable negative impact on the health economy, including an increase in the price of hospital admissions and prescription drugs, an increase in the cost of disease over time, which extends to the price of sick days taken, and the loss of economic prospects. Haze occurrences have increased the number of patients seeking treatment in hospitals for rhinitis, upper respiratory infections, asthma, and chronic obstructive pulmonary disease. Haze occurrences resulted in an increase in respiratory mortality of 19%. The negative effects of haze on health are more likely to affect children and the elderly. Malaysian estimates place the annual inpatient cost from haze outbreaks at about USD 91,000. (Latif et.al. 2018).

Health impacts of haze are numerous. Haze particles can induce coughing, wheezing, shortness of breath, as well as a feeling of tiredness and weariness. (Chung, 2020). In people who already have heart or lung issues, the consequences of haze are amplified. On the other hand, the adverse health consequences of a brief exposure to haze are often momentary and do not cause long-term health issues. The most hazardous component of the haze is the tiny particulate matter (PM) that is prevalent in the air. According to Carolyn Payus et al. (2013), during the fog periods, airborne PM_{10} was discovered to be the true toxin while the other air quality measurements remained below the acceptable sound limits. Indonesian land clearing causes fires to start during haze, which then accumulates and affects tourism, transportation, biodiversity, and health issues throughout the region. Therefore, exposure to environments with high particulate matter concentrations can have negative effects on human health, agricultural output, plant species, and biological systems (Afroz et. al., 2003).

Table 2.1: Effects of the various pollutants (Ghorani, 2016)

Pollutants	Health
Carbon monoxide	• Reduction in the ability of the circulatory system to transport oxygen • Impairment of performance on tasks requiring vigilance • Aggravation of cardiovascular disease
Nitrogen dioxide	• Increased susceptibility to respiratory pathogens – more likely to suffer lung infections
Ozone	• Decrement in lung infection • Coughing, chest discomfort • Increased asthma attacks
Sulphur dioxide/ particulate matter	• Increased prevalence of chronic respiratory disease • Increased prevalence of acute respiratory disease

2.4 Particulate Matter (PM_{10})

Environmental health issues related to air pollution are growing more complex and are impacting both emerging and industrialized nations. According to several studies, particulate matter (PM_{10}) is one of the main pollutants that damage both human health and the environment (Mekdes, 2011). Particulate matter is made up of minuscule particles that are either solid or liquid and are suspended in the air. PM_{10} and $PM_{2.5}$ are the two most prevalent primary types of particulate matter (PM_{10}), where PM_{10} refers to particles with an aerodynamic diameter of less than 10 microns and $PM_{2.5}$ refers to particles with a diameter of less than 2.5 m. (EPA, 1997). $PM_{2.5}$, also known as fine particulate matter, also encompasses ultrafine particles with a diameter of less than 0.1 m, with the majority of $PM_{2.5}$ coming from ignition-related sources.

The complex blend of gaseous and particle elements that make up air pollution are all dangerous to human health. Despite the fact that the makeup of air pollution varies greatly depending on its source, studies conducted throughout the globe consistently show that air pollution is a significant modifiable risk factor for increased morbidity and death. Additionally, clinical studies have shown that particulate matter (PM_{10}) air pollution has a greater negative effect on health than do gaseous molecules. Humans are subject to a wide range of detrimental health effects from PM_{10}, particularly on the cardiovascular system. Exposure to PM air pollution, both acute and chronic, is linked to a higher risk of dying from cardiovascular disorders, such as ischemic heart disease, heart failure, and ischemic/thrombotic stroke (Hamanaka and Mutlu, 2018).

As a substantial endocrine disruptor, particulate matter has been linked to the emergence of metabolic diseases including obesity and diabetes mellitus, both of which are risk factors for cardiovascular disease. As epidemiological evidence for the detrimental effects of particulate matter (PM_{10}) air pollution on human health gains acceptance, new studies are shedding light on the processes by which PM_{10} harms humans. Understanding how PM_{10} negatively affects human health is crucial for preventing and limiting the negative health impacts of this prevalent environmental problem. Due of its alleged adverse effects on human health, contamination caused by PM_{10} is a matter of growing public concern. This is true because PM_{10} from burning sources, such as forest fires, is more damaging than PM_{10} from crustal or other non-burning source.

2.5 Sources of particulate matter PM_{10} exposure

In Malaysia, particulate matter (PM_{10}) is a significant contaminant, especially in urban settings. The worry over this pollutant's impact on human health is pertinent and essential because its level was also seasonally substantial throughout most of Malaysia. The atmosphere can produce secondary particles from gaseous precursors such sulphur dioxide, nitrogen oxides, ammonia, and non-methane volatile organic compounds or particles can be directly radiated into the atmosphere (primary PM) (WHO, 2013). Primary PM and precursor gases can come from both natural (non-anthropogenic) and artificial (anthropogenic) sources. Chemical reactions involving gaseous pollutants result in secondary particles being formed in the atmosphere. Fine PM contains the bulk of secondary particles.

High PM levels were seen during the southwest monsoon in Peninsular Malaysia's central and southern regions as well as Malaysian Borneo, which correlated with Indonesia's biomass burning. According to the long-term investigation, Malaysia's PM_{10} pollution was mostly caused by local and transnational sources, particularly in metropolitan areas. Even though there are small but noticeable long-term decreasing trends in PM_{10} pollution, special attention must be paid to short-term pollution episodes, particularly those related to transboundary pollution during extreme weather conditions like the El Nino event, to protect the health of a larger population (Sentian et.al. 2018).

2.6 Effects of Particulate Matter (PM_{10}) to Human Health and The Environment

PM_{10} has long been acknowledged as a pollutant that has harmful impacts on both people and the environment (Betha et al., 2013; Martin, 2011; Norela et al. 2008). In Malaysia, PM_{10} is the main pollutant. A severe sickness brought on by PM_{10}'s effects on human health is possible. Adults and elderly individuals are often those who are most impacted by PM_{10}. Where a particle settles when it is absorbed into the respiratory system is largely determined by its size. Particulate matter larger than roughly 10 micrometers, or PM_{10}, is typically filtered in the nose and throat by cilia and mucus, but PM_{10} can collect in the bronchi and lungs and cause health issues. Changes in visibility are particulate matter's main consequences (haze). Therefore, before getting worse, this pollutant needs specific care.

According to medical study, PM_{10} exacerbates cardiovascular and respiratory diseases. It stresses out the cardiovascular system by making the heart and lungs work harder to provide oxygen, hastens the natural aging process of the lungs, and accelerates the loss of lung capacity, damages the lungs even after minor irritation symptoms go away, and may even be a factor in the development of cancer.

Haze occurrences have influenced an increase in hospital visits for asthma, rhinitis, upper respiratory infections, and chronic obstructive pulmonary disease treatments. Haze occurrences led to a 19% increase in respiratory mortality. The health effects of haze are more likely to affect children and elderly people. In Malaysia, the annual inpatient cost from haze events was estimated to be approximately USD 91,000. (Latif et.al. 2018).

Inhaled particulate matter (PM_{10}), or particles having an aerodynamic diameter of less than 10 micrometers, is a pertinent public health issue that has been extensively documented in epidemiological research conducted across the world. Hospitalizations and fatalities from respiratory and heart disorders are among the consequences of exposure to PM_{10}, especially in the most vulnerable population groups, such as young children and the elderly (Ostro et.al, 2009).

If particle concentrations are excessively high, asthma attacks in children may occur. Particulate matter may trigger or exacerbate asthma episodes for unknown reasons. On the day of admission, PM_{10} mostly exacerbates already inflamed airways, but the delayed impact may show that exposure to some pollutants makes people more susceptible to other triggers, such viral infections. Children who have asthma are more vulnerable to the effects of air pollution than their non-asthmatic classmates or kids who only have cough symptoms (Nastos et.al. 2010).

2.6.1 Relationship of PM_{10} with other pollutants and weather parameters

Weather conditions in specific places also have an impact on PM_{10} concentration. The association between PM_{10} and whether parameter has gotten more attention and has been carefully studied by various researchers. In Malaysia, PM_{10} is the most prevalent pollutant. Trace gases and whether parameters play a significant role in affecting ambient air quality of an urban environment. Whether parameters and trace gases are major determinants on urban air quälity (Dey et al., 2017; Zhang et al., 2018; Manju et al., 2018). Temperature, relative humidity, and wind speed and direction are considered major factors because they can affect the dispersion process, removal mechanisms, and particle structure in the atmosphere (Goyal and Chalapati Rao, 2007; Zhang et al., 2015), and thus play a significant role in controlling air pollutant concentration levels.

Previous research at the Klang Valley area was a developed industrial and economic zone in the Malaysia and heavy traffic flow was the main reason for the elevated atmospheric pollutants in these areas (Azmi et al., 2010). According to national standards, daily particulate matter ($PM_{2.5}$ and PM_{10}) exceeds the optimal level during the dry season but is slightly lower during the monsoon. Weather parameters influence has been extensively studied in China (Li et al., 2014; Zhang et al., 2015; Yin et al., 2016; He et al., 2017) and India (Jayamurugan et al., 2013; Ojha et al., 2015; Manju et al., 2018), but it is underestimated in Bangladesh, despite the fact that climate variability in weather forecasting may have variable effects.

2.7 Statistical Modelling

The connection between the response variable (in this case, PM_{10}) and the variables must be linear in order for a linear regression to be valid. Several statistical models have been developed for predicting air quality. Deterministic statistical models are therefore more suited for predicting over broad regions, such as significant urban agglomerations and regional air quality. Regression analysis is a vital statistical technique used in the majority of scientific inquiries. The main objective of linear regression is to create precise statistical models for predicting PM_{10} concentrations based on weather and air quality data, as well as to assess the effectiveness of the models. However, this model requires certain input data, such as emissions and weather conditions. Although statistical models can link input variables (predictors) and output variables (predictions), they are unable to provide an explanation for why these correlations occur.

According to Ahmad Zia et al. (2012), statistical modeling that could provide useful insights in the short-term forecasting of future air pollution levels to forecasting PM_{10} concentration in Pulau Pinang, robust regression models for predicting PM_{10} concentration in an industrial area were discovered (industrial area). The local environmental authorities will thus employ this prediction algorithm to execute daily air pollution forecasts. This data may be used by public health experts to decide what actions to take to reduce air pollution.

To identify the pollutant association status over particulate matter (PM_{10}) in East Malaysia, Ku Yusof et al. (2019) did a study concerning the use of artificial neural networks (ANN) and multiple linear regressions (MLR) combined with sensitivity analysis (SA). Both fully completed models that were chosen, and the ANN performed better in both hazy and non-hazy conditions. Additionally, during hazy and non-hazy days, respectively, ANN and MLR both predicted the variables UVB and carbon monoxide. Precise forecasts were necessary to aid any relevant agency in focusing on the pollutant that primarily caused fluctuations in PM_{10}, especially during the haze season.

Multiple regression models were also used by Vlachagianni et al. (2011) to anticipate NOx and PM_{10} concentrations in Athens and Helsinki. They compared the projections for the two cities and evaluated the model performance using a range of statistical assessment metrics. To see if there were any notable improvements that would warrant the additional processing expense, the forecasts from the MLR model were contrasted with those from an Artificial Neural Network (ANN) model. The best key parameters for both cities were the concentrations of NO(x) and PM10 at night, together with wind speed and the Monin-Obukhov length.

According to Azid et al. (2015), the use of chemometrics approaches can provide valuable information on the geographical variability of a significant and complicated set of air quality data. Hierarchical agglomerative cluster analysis (HACA), discriminant analysis (DA), principal component analysis (PCA), and multiple linear regressions (MLR) methods can be used to review the air quality more clearly and create a novel air quality monitoring network for better management of air pollution.

According to Dunn et al. (2012), obesity prevalence in Malaysia is unequally distributed by gender and ethnic group. Quantile regression was used in the study to evaluate the influence of socioeconomic disparity in explaining these differences. The findings show that obesity inequality is unlikely to be eliminated even with rising wealth and educational levels brought on by economic growth. This leads to the conclusion that greater efforts are needed to change Malaysians' lifestyle patterns in order to reduce obesity disparity and the overall level of obesity. Thus, a far more thorough understanding of how explanatory factors affect the dependent variable may be obtained using the quantile regression technique (Kang and Liu, 2014).

According to Shaziayani et al. (2021), the aim of their study is to determine the best loss function between quantile regression (QR) and ordinary least squares (OLS) using boosted regression tree (BRT) for the prediction of PM_{10} concentration in Alor

Setar, Klang and Kota Bharu, Malaysia. Model comparison statistics using coefficient of determination (R^2), prediction accuracy (PA), index of agreement (IA), normalized absolute error (NAE) and root mean square error (RMSE) show that Quantile Regression (QR) is slightly better than ordinary least squares (OLS) with the performance of R^2 (0.60–0.73), PA (0.78–0.85), IA (0.86–0.92), NAE (0.15–0.17) and RMSE (9.52–22.15) for next-day predictions in BRT model.

In conclusion, the majority of research found that MLR was superior to other models in evaluating the significance of PM_{10} concentration and whether parameter. The majority of previous research focused on the seasonal changes in PM_{10}, such as the cold and hot seasons. The study's inability to create a model employing MLR during a high-particulate event was a problem.

As a result, statistics are crucial for the analysis and interpretation of data since they allow for the use of analysis results as prediction tools, which has become the main goal of environmental engineering. In order to pick the optimal distribution for future prediction, the performance indicators root means squared error (RMSE), index of agreement (IA), mean absolute error (MAE), and coefficient of determination (R^2) will be employed to evaluate the performance of the probability distributions. This research will thus use multiple linear regression (MLR) and quantile regression (QR) to estimate PM_{10} concentration during haze in Pasir Gudang, Melaka, and Petaling Jaya. Short-term control tools may then be devised.

CHAPTER 3

METHODOLOGY

3.1 Research Flow

Figure 3.1 illustrates the research activities for this study. Hourly data set of air pollutants and weather parameter during high particulate event were collected from DOE Malaysia. The hourly PM_{10} concentrations records in 1997, 2005, 2013 and 2015 were used due to the high level of PM_{10} concentration were recorded in atmosphere during that particular period. The location of monitoring station was at Pasir Gudang, Melaka and Petaling Jaya. The air pollutant parameters used in these studies were PM_{10}, O_3, CO, SO_2, NO_2, NO and NO_x. In addition, weather parameters, such as: wind speed, temperature, UVB and humidity also used.

The data were first organized by day, month, and year. Then, using descriptive statistics such as boxplot, time series plot, and histogram, the trend of PM_{10} was examined during HPE. The relationship between the high concentration of PM_{10} and the trace gases and meteorological characteristics in these three research regions was then ascertained using Pearson Correlation Analysis. Finally, MLR and QR models with values of 0.25, 0.50, and 0.75 were created. Several performance measures, including as MAE, RMSE, R^2, and IA, were used to assess the models' performance.

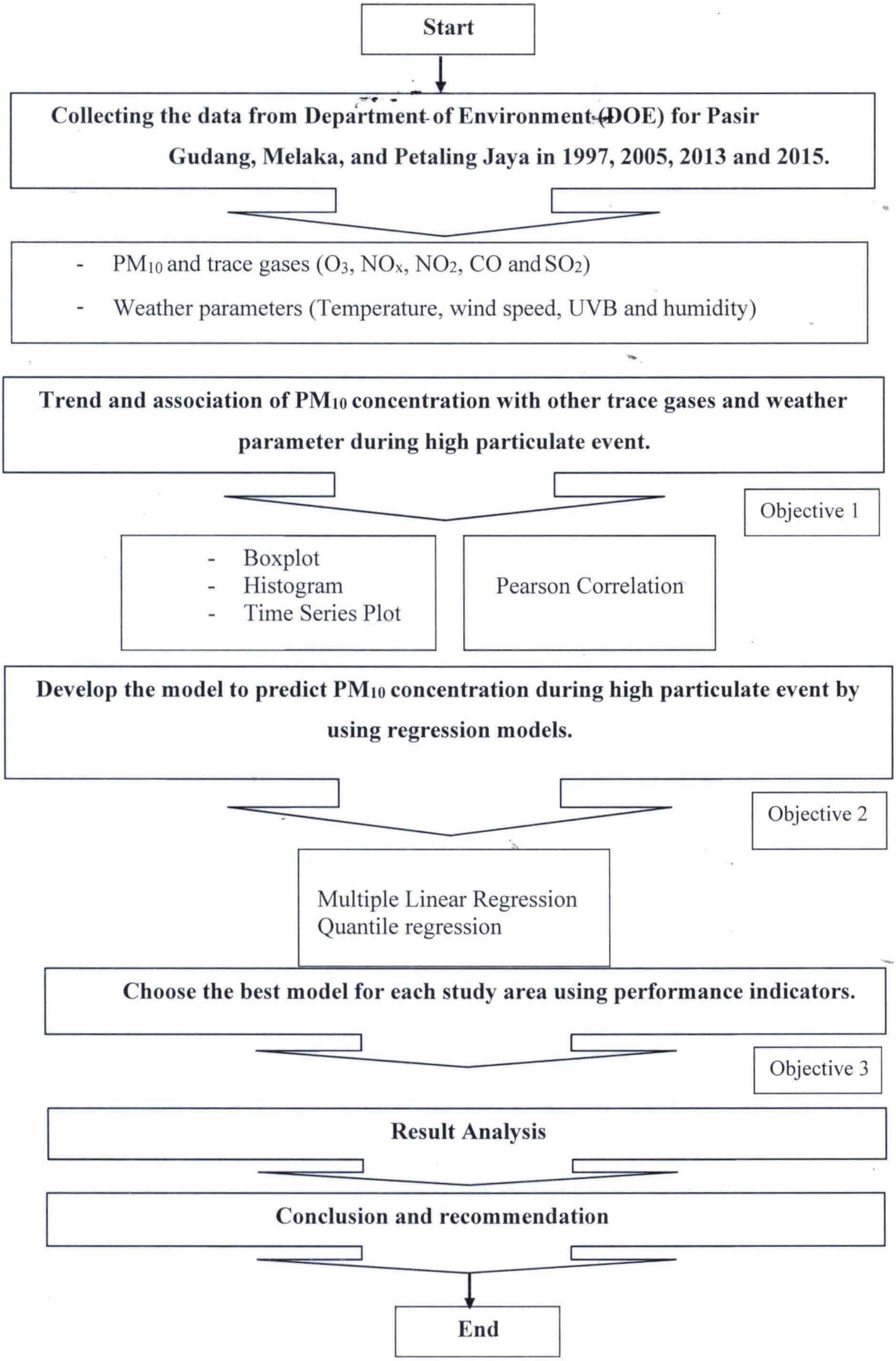


Figure 3.1: Flowchart of research work

3.2 Location of Study Area

Three primary air-observing stations were used in this study: Pasir Gudang, Melaka, and Petaling Jaya. Table 3.1 displays the checking site descriptions for each of the three research regions that were looked at during the whole study. The neighborhoods of all the places were the same. According to Carolyn Payus et al. (2013), transboundary smoke emission causes haze occurrences in various research locations. In addition, these research locations are directly affected by transboundary flow and are situated in the southern region of Peninsular Malaysia's west bank, close to Indonesia. There has been a few research done on the PM_{10} levels that have been recorded in Pasir Gudang, Melaka, and Petaling Jaya. According to Ya'acob and Mar Iman (2020), this shows that industrial locations have a greater amount of air pollution.

3.2.1 Pasir Gudang

East of Johor Bahru, in Johor, Pasir Gudang is the region's most well-known industrial city. Pasir Gudang's proximity to the Straits of Johor makes it a hub for the shipping and logistics, shipbuilding, and petrochemical sectors. Also carried out in the region are heavy businesses like palm oil distribution and storage. In 1918, Pasir Gudang was a collection of 5 villages before growing into a significant industrial metropolis. Four further settlements were created by 1920. Like Kampung Air Biru and Kampung Pasir Merah, which are now the location of Pasir Gudang's port, many of the villages were situated where significant infrastructure now stands. The Malaysian state of Johor contains the city of Pasir Gudang. The first city in Malaysia to have an air quality and dangerous airborne pollutants monitoring system is Pasir Gudang, Johor. Sekolah Menengah Pasir Gudang 2, Pasir Gudang, Johor is where the monitoring stations are located (N01°28.225, E103°53.637). Geographically speaking, the monitoring stations are well placed in fast expanding industrial areas, which causes considerable air pollution (Lee et al., 2012; Mohamed Noor et al., 2011; Yap & Hashim, 2013).

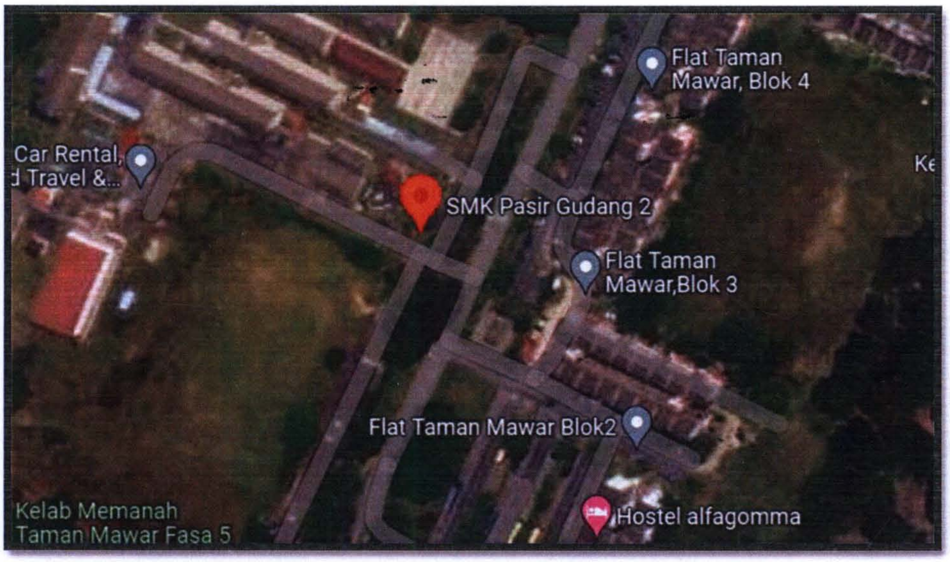


Figure 3.2: Location monitoring station in Pasir Gudang (Source: Google map, 2022)

3.2.2 Melaka (Residential area)

The capital of the coastal state of Malacca, located in southwest Malaysia, is Malacca City (also called Melaka). Jonker Street, Chinatown's main avenue, is famed for its night market and antique stores. A nearby Chinese temple from the 17th century with several prayer chambers and elaborate embellishments is called Cheng Hoon Teng. The Kampung Kling Mosque, built with Javanese influences in the 18th century, has a green, three-tiered roof on top. The most significant trade port in Southeast Asia throughout the 16th century was Malacca. The Portuguese made tremendous profits from the particularly profitable spice trade that went through the port, which attracted frequent visits from traders from India, the Arab world, and Europe. Sekolah Menengah Kebangsaan Bukit Rambai, Melaka, is home to an air quality monitoring station, which is physically situated at latitude N 02o 15'510" and longitude E 102o 10' 364". The Strait of Malacca borders the 1664 km2 that make up Melaka's main city. The population increased to 842,500 in 2012 over the research period.

Figure 3.3: Monitoring station in Melaka (Source: Google map, 2022)

3.2.3 Petaling Jaya (Industrial area)

Petaling Jaya, sometimes known as "PJ" by residents, is a city in the Malaysian state of Selangor's Petaling District. It is a component of the Greater Kuala Lumpur region and was first built as a satellite city for Kuala Lumpur, the Malaysian capital. On June 20, 2006, Petaling Jaya was given city status. The Petaling Jaya air monitoring station is situated in Sekolah Kebangsaan Bandar Utama in Petaling Jaya, Selangor (N03° 06.612', E101° 42.274'; S1).

Figure 3.4: Location monitoring station in Petaling Jaya (Source: Google map, 2022)

Table 3.1: Study area description

Study area	Air monitoring station	Surrounding land use
Pasir Gudang	Sekolah Menengah Pasir Gudang 2, Pasir Gudang, Johor	• Heavy industrial areas • Commercia land • Transportation and logistics
Melaka	Sekolah Menengah Kebangsaan Bukit Rambai, Melaka	• Agriculture • Residential area and housing
Petaling Jaya	Sekolah Kebangsaan Bandar Utama, Petaling Jaya; Selangor	• Heavy traffic particulars during the morning hour • Industrial area and housing

3.3 Air Pollutant Dataset

Data on the continuous hourly air quality for 1997, 2005, 2013, and 2015 were received from the Department of Environment's Air Quality Division. Eleven factors were employed in this study, and they were split into two groups. Six air pollutants were included in one group: ground-level ozone O_3, PM_{10} (g/m^3), NO_x (ppm), SO_2 (ppm), NO_2 (ppm), and NO_x (ppm). While weather variables include temperature (°C), humidity (%), UVB (W/m^2), wind speed (m/s), and wind direction (°). These data were sorted in Microsoft Excel before being subjected to an SPSS analysis.

3.4 Descriptive Statistic

This analysis is crucial to extract details from data sets that contain information about how the data were gathered, arranged, summarized, and presented. This research also examined descriptive statistics for meteorological parameters. In order to determine how PM_{10} interacts with other air contaminants and meteorological variables during high particle events, these characteristics were included in the study. Measures of central tendency (mean, median, mode, and form) and measures of dispersion are among the tests that were carried out (range, variance, and standard deviation). Table 3.2 below displays the condensed formulae for computing the descriptive statistic.

Table 3.2: Descriptive statistics test (Goos & Meintrup, 2015)

Descriptive Statistics	**Formula**	**Description**
Mean	$\bar{x} = \frac{\sum x}{n}$	x = Observations given n = Total number of observations
Median	Odd: $x = \frac{n+1}{2}$ Even: Median = n/2 and n/2+1	n = Total number observation
Standard Deviation	$s = \sqrt{\frac{\sum(x - \bar{x})^2 - \bar{x}^2}{n-1}}$	x = Observations given x̄ = Mean n = Total number of observations

3.5 Graphical Presentation

3.5.1 Boxplot

The distribution of the data set was shown using a boxplot. It is frequently used to show the distribution properties of a collection of datasets as well as the level of scores (Thirumalai, 2017). Additionally, the box plot will interpret descriptive statistics including the data's mean, maximum, and lowest values. The lower quartile (Q1), median (Q2), upper quartile (Q3), and outliers are displayed using this straightforward method for dividing the data into quartiles. The overview of data dispersion it offers is excellent. It is a system that has been institutionalized for illustrating how information is distributed in light of the five numerical categories of least, first quartile, medium, third quartile, and greatest. greatest. In the simplest box plot, the first quartile through the third quartile is included in the focus rectangle (the interquartile reach or IQR). A segment inside the rectangle designates the center, while "stubbles" above and below the case designate the portions of the least and greatest. The greatest and lowest numbers, on the other hand, are displayed in the whiskers section. At the right whisker's end, the highest score is displayed after removing outliers.

The box plot has been set up using the SPSS to visualize the attributes of the data clearly and effectively. The box plot (Figure 3.5) is a graph that summarizes all data, including the information's emphasis, shape, and distribution. The top and lower whiskers show the lower and upper 25 percent of scores, which are outside the center 50 percent. The middle 50% of all scores, or the range between the 25th and 75th percentiles, is known as the interquartile range (IQR) (McLeod, 2019). The interquartile range is discussed in the box plot section (IQR). The term "whiskers" refers to the lines extending from the container. The lowest and highest values in the arrangement of information excluding exceptions are displayed on this line. Additionally, the highest and lowest estimates of the information and midpoint are displayed in this container.

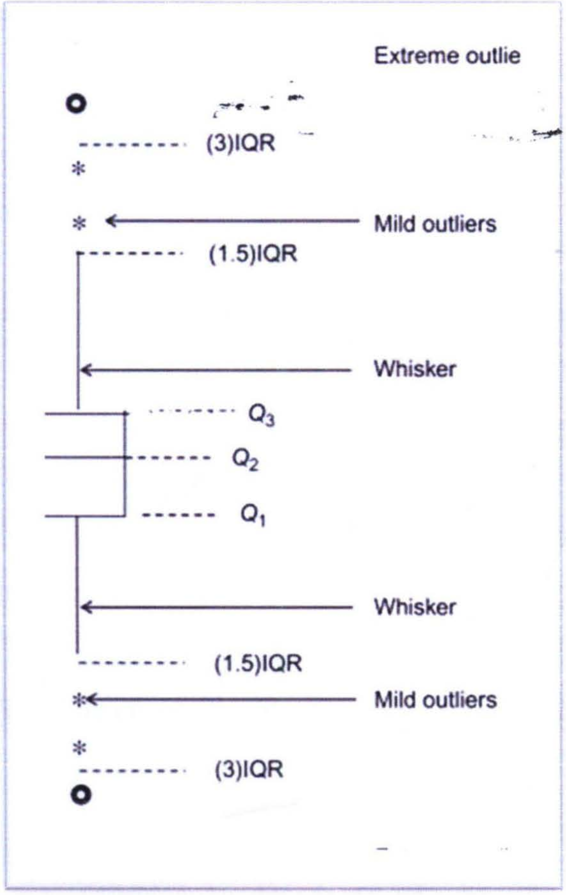


Figure 3.5: Example of box plot Source: Mathematical Statistics with Applications in R (Ramachandran & Tsokos, 2015)

Besides that, box plot also highlights the extreme value and outlier that was used in this study as indication of haze episodes. The extreme value and outliers are represented by the star dots. Outliers' values are considered if it is outside of the box which is it less than Q1 or higher than Q3 by times with 1.5 of IQR where this range is the difference between the first and third quartile vales. (Q3 - Q1). The extreme values are used in a broad sense, encompassing both the occurrence of extraordinary values and the exceedance of a particular threshold level. In this study, extreme values of PM_{10} were included because the chosen research period were during the event of haze or during high particulate event.

3.5.2 Time Series Plot

A set of data gathered throughout time is a time series plot. identifies time-series plots over history by plotting perception on the y-axis versus spread over the same time on the x- axis. The use of time series graphs in statistical analysis is crucial for spotting trends or patterns in the data. It is challenging to recognize any pattern or trend when the same variable is continuously recorded throughout time. Some component stood out when all points tothe same data that was graphically shown. Because a pattern may be employed in the present, time series plots make it straightforward to create one. The data trend toward PM_{10} concentration in this study was shown using trace gases such as O_3, NO_2, NO_x, SO_2, and CO. When the PM_{10} concentration started to build, the trend was seen. It is possible to quantify and determine the precise time by comparing the smaller and higher peaks.

3.5.3 Histogram

A histogram is a graphical representation that organizes a group of data points into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins. It is useful for the visualization of the distribution of data. With the use of a histogram, the median, distribution, and variations in data can be found out in this stud. Histogram tells us about the skewness of data plotted. The future performance of this research can be predicted by using histogram.

3.6 Pearson Correlation

The measure of association, sometimes referred to as the Pearson's correlation coefficient statistical test, evaluates the relationship between two variables. Pearson The most effective technique for calculating the relationship between two variables of interest is correlation. It offers details on the strength and direction of the relationship's association. The relationship between PM_{10} and other pollutants, such as wind speed, temperature, humidity, UVB, and wind direction— all of which are classified as meteorological parameters—was determined using Pearson correlation. The two variables X and Y are measured using the Pearson correlation analysis, which provides a number between +1 and -1, with 1 denoting a positive correlation, 0 denoting no connection, and -1 denoting a negative correlation. It is frequently employed as a gauge of how linearly dependent two variables are on one another.

To see the relation or association of PM_{10} with the other pollutants, the results can be read through Pearson Correlation. This Pearson Correlation will show the value of correlate parameters with the weather conditions at this research study areas. Pearson Correlation is made using SPSS software.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Where:

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of values of the y-variable

From the calculated r value, the degree of correlation can be identified. If the correlation coefficient is higher than 0.6, hence it was declared as high correlation: if the correlation coefficient is between 0.3 and 0.6, hence it was declared as moderate correlation: if the correlation coefficient is smaller than 0.3, then it was declared as low correlation (Gogtay & Thatte, 2017). Because the correlation is an effect size, the absolute value of "r" of the correlation can be described informally using the following guide. Table 3.3 shows the description of correlation using the following guide for the absolute value of "r" (Mukaka, 2012):

Table 3.3: Shows the description of correlation related to the value of *r*

Value of *r*	**Description**
0.00 – 0.19	Very weak
0.20 – 0.39	Weak
0.40 – 0.59	Moderate
0.60 – 0.79	Strong
0.80 – 1.0	Very Strong

3.7 Regression models

In order to create the regression models, 80% of the data set was used, and the remaining 20% was used for model validation in the future. Multiple Linear Regression and Quantile Regression arethe two categories of models that have been produced. It is done with the intention of comparing the performances of the standard model MLR and the QR model since MLR had been widely utilized to model the linear parameters. For both the MLR and QR models, i.e., the next day (t_{-24}), the following two days (t_{-48}), and the following three days (t_{-72}) were predicted.

In this research, linear regression was used to predict hourly mean PM_{10} concentrations. Linear regression was also used to determine the relationship between PM_{10} concentration and whether parameter (wind speed, wind direction, ambient temperature, and humidity) and other gases (SO_2, NO_2, NO_x, O_3 and CO).

3.7.1 Multiple Linear Regression (MLR)

MLR tries to simulate the connection between two or more independent variables and a dependent variable by fitting a linear equation to the observed data. MLR is one of the most used forecasting techniques. Multiple linear regression analysis is a statistical approach that may be used to examine the relationship between a single dependent (criterion) variable, Y, and a number of independent (predictor) variables, x's. Equation (8) depicts a random response Y based on a multiple regression model's independent variables x1, x2,..., xk (Wackerly, 2002). These experiments were therefore carried out once utilizing hourly PM_{10} concentration data from three monitoring stations in Pasir Gudang, Melaka, and Petaling Jaya. According to James, Witten, Hastie, and Tibshirani (2015), MLR may also demonstrate the variability relationship between the independent variable and the dependent variable by fitting the observed data into a linear equation (Eq. 3.2):

$$Y_i = y + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon \qquad i = 1,2, \ldots, n \qquad (3.2)$$

Where;

- $y, \beta_1, \ldots, \beta_k$ are unknown parameters
- ϵ is an error term factor

3.7.2 Quantile Regression (QR)

Target's conditional median is calculated using quantile regression. When the prerequisites for linear regression—namely, linearity, homoscedasticity, independence, or normality—are not satisfied, the quantile regression method is applied. A certain value in the features variables may yield any quantile (percent) using quantile regression, which is not just limited to computing the median. When a home's price is found to be in the 25th quantile, for instance, there is a 25% chance that the real price will be less than the estimate and a 75% chance that the price would be greater. Comparable in structure to the linear regression model is the quantile regression model equation for the quantile. By minimizing by minimizing the median absolute deviation, the optimum quantile regression line is discovered.

In this research, quantile regression will be used to predict hourly mean PM_{10} concentrations by using 0.25, 0.50 and 0.75 percentile value. Quantile regression also will be used to associate PM_{10} concentration with trace gases and weather parameter.

$$Q\tau\,(y_i) = \beta_0(\tau) + \beta_1(\tau)\, x_{i1} + \ldots + \beta_p(\tau)\, x_{ip} \qquad i=1,\ldots,n$$

Where;

3.7.2.1 $y, \beta_1, \ldots, \beta_k$ are unknown parameters

3.7.2.2 p is the number of regressor variables

3.8 Performance Indicator

Performance metrics were used to evaluate how well the regression models predicted the PM_{10} concentration values at the research sites. In order to illustrate the healthiness for each selected conveyance, there should be at least four different sorts of execution pointers. The performance measures used in this study are mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and index of agreement (IA). The MAE and RMSE of a good model should be close to zero. The fit of the prediction model is expressed using the coefficient of determination (R^2) and agreement index. R^2 and IA calculations must fall between 0 and 1, and values closer to 1 show that the model is quite accurate in predicting the observed data.

Performance Index	Equation	Description
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n} \lvert p_i - O_i}{n}$	When value of MAE is closer to zero indicates better method.
Root-mean-square deviation (RMSE)	$RMSE = \frac{1}{n-1}\sum_{i=1}^{n}(P_i - O_i)^2$	When value of RMSE is closer to zero indicates better method.
Coefficient of determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \overline{O})}{n.S_p.S_O}\right)$	When value of R^2 is closer to one indicates better method.
Index of Agreement (IA)	$IA = \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}\lvert P_i - \bar{O}\rvert + \lvert O_i - \bar{O}\rvert^2}\right]$	When value of IA is closer to one indicates better method.

where;

n = total number of annual measurements of particular site.

P_i = predicted values of one set of annual monitoring record

O_i = observed values of one set of annual monitoring record

$\bar{P}$= mean of the predicted values of one set of annual monitoring record

$\bar{O}$= mean of the observed values of one set of annual monitoring record.

S_p= standard deviation of the predicted values of one set of annual monitoring record

S_O = standard deviation of the observed values of one set.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the results and discussion according to the specified objective in this research. Firstly, the characteristics of PM_{10} concentration during high particulate event in the three difference study areas i.e. Pasir Gudang, Melaka and Petaling Jaya in year 1997, 2005, 2013 and 2015 were discussed. In addition, the concentration of PM_{10} with the trace gases such as NO_2, NO, NO_X, CO, SO_2 and weather parameter i.e. wind speed, temperature, UVB and humidity were investigated. Then, the associations of PM concentration with the weather parameters were identified using Pearson Correlation Analysis. Finally, the regression models of MLR and quantile were developed according to the study areas and the best model for each area were chosen by calculating several performance measurements.

4.2 Distribution of data

Table 4.1 illustrates the descriptive statistics for PM_{10} concentration at Pasir Gudang, Melaka and Petaling Jaya respectively in 1997, 2005, 2013 and 2015. According to Malaysia Ambient Air Quality Guidelines (MAAQG), the guideline for 24 hour averaging time of PM_{10} was 150 $\mu g/m^3$. The annual average of PM_{10} concentration during the Southwest Monsoon had exceeded the standard set which was 50 $\mu g/m^3$ while the PM_{10} concentration during the Northeast Monsoon was below the acceptable level for both stations (Abdul Hamid et al. 2018). Mean of PM_{10} concentration in Pasir Gudang for 1997 was 47.7 $\mu g/m^3$; in 2005 was 46.59 $\mu g/m^3$; in 2013 was 51 $\mu g/m^3$ and in 2015 was 64.8 $\mu g/m^3$. Then, the maximum of PM_{10} concentration in Pasir Gudang for 1997 was 268 $\mu g/m^3$; in 2005 was 116 $\mu g/m^3$; in 2013 was 462 $\mu g/m^3$ and in 2015 was 351 $\mu g/m^3$. The maximum of PM_{10} concentration at Pasir Gudang in 2013 was recorded as the highest value compared to all study areas during these four years period.

Moreover, the maximum of PM_{10} concentration was observed at Melaka in 2013 with the value of 577 μg/m^3 was the highest PM_{10} concentration recorded, if compared to Pasir Gudang and Petaling Jaya. The mean concentration of PM_{10} concentration in Melaka for 1997, 2005, 2013 and 2015 were 71.7, 83.3, 79.2 and 69.7 μg/m^3 respectively. PM_{10} concentration was recorded higher than the MAAQG for 12 month averaging time that was 50 μg/m^3. This is because due to the haze episode that occurred during summer season from June to August. In addition, the unfavorable weather conditions of hot and dry season during that time are one of the main factors for high concentration of PM_{10} (Czernecki et.al, 2016).

Furthermore, the mean concentration of PM_{10} concentration in Petaling Jaya for 1997, 2005, 2013 and 2015 were 69.4, 64.3, 48.4 and 60.5 μg/m^3 respectively. The trend of PM_{10} concentration in Petaling Jaya were fluctuated and higher concentration can be observed in 1997 which was 69.4 μg/m^3. Next, the maximum for PM_{10} concentration in Petaling Jaya for 1997, 2005, 2013 and 2015 were 393, 494, 372 and 472 μg/m^3 respectively and in 2015 was the highest PM_{10} concentration reading recorded which was 472 μg/m^3.

However, as stated in the Malaysian Ambient Air Quality Guidelines, the mean PM_{10} concentration levels for Pasir Gudang, Melaka, and Petaling Jaya were below the threshold (150 g/m3) (MAAQG). In all of the studied regions throughout these four years, the highest figure for the maximum PM_{10} concentration was found in Melaka in 2013. From 1997 to 2015, Pasir Gudang exhibited an increase in concentration. As a consequence, Pasir Gudang had the lowest amounts, with Petaling Jaya having the second-lowest concentrations. In 2005, Melaka had the highest average concentration. The standard deviation shows that there were less variations and a narrower range of PM_{10} concentration data in Pasir Gudang. Additionally, the fact that all study regions had medians that were lower than the means shows that PM_{10} concentrations in those areas were skewed to the right, suggesting that an extreme particle event most likely happened.

Table 4.1: Descriptive Statistic for air pollutant dataset at Pasir Gudang, Melaka and Petaling Jaya in 1997, 2005, 2013 and 2015

		Pasir Gudang				Melaka				Petaling Jaya			
		1997	**2005**	**2013**	**2015**	**1997**	**2005**	**2013**	**2015**	**1997**	**2005**	**2013**	**2015**
Total Value, N	**Valid**	8631	8715	8745	8710	8337	8669	8669	8759	8222	8727	8659	8591
	Missing	129	45	15	50	423	91	91	1	538	33	101	169
Mean		47.7	46.59	51	64.8	71.7	83.3	79.2	69.7	69.4	64.3	48.4	60.5
Median		33.0	44.0	45.0	54.0	46.0	78.0	72.0	58.0	49.0	56.0	43.0	49.0
Standard deviation		39.9	13.7	38.4	36.1	61.6	27.4	42.8	41.5	55.1	40.7	29.3	50.1
Minimum		11.0	19	10	27	13.0	29	32	24	20	20	17.0	5
Maximum		268.0	116	462	351	415.0	268	577	338	393	494	372.0	472

Figure 4.1 showed the box plots for PM_{10} concentration in Pasir Gudang, Melaka and Petaling Jaya. Generally, it indicates that the measurement data were skewed to the right, and it indicates a distribution with a tail extending towards more positive value for the year of 1997, 2005, 2013 and 2015 at Pasir Gudang, Melaka and Petaling Jaya. Hence, it signified that the occurrence of extreme values and outlier for the data sets. These values were due to the high particulate event that had been experienced by Malaysia in those years. As we can see, PM_{10} concentrations at the studied areas were exceeding the recommended value of MAAQG, hence the number of unhealthy days was recorded higher in these years compared to non-haze years. The recommended value for PM_{10} over a 24-hour period, as stated by the MAAQG, was 150 $\mu g/m^3$.

In all of the research locations, 2013 saw the largest PM_{10} fluctuation. The haze phenomenon that occurred between June 2013 and October 2013—which was supposed to have the same effects as the smog in 1997—was to blame for this. The 1997 haze outbreak and the 2013 haze outbreak were the two years with the greatest number of Air Pollutant Index (API) readings. Both haze outbreaks severely affected various portions of Malaysia (Koh and Ho, 2022). Pasir Gudang was not affected by the haze in 2005, while Melaka was less affected than Petaling Jaya. The effects of haze and the greatest PM_{10} concentration in 2015 were nearly same in all locations. Melaka had the highest percentage of PM_{10}. To put it briefly, the entire portion had surpassed the recommended limit of 150 g/m^3.

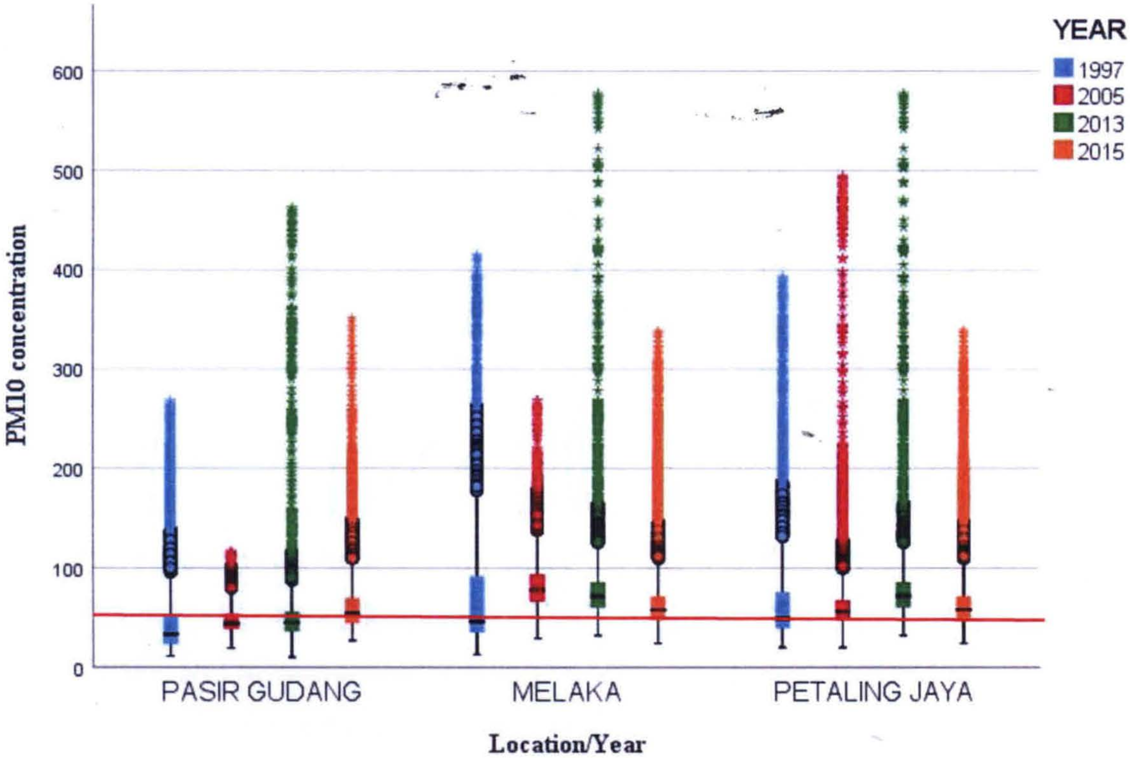


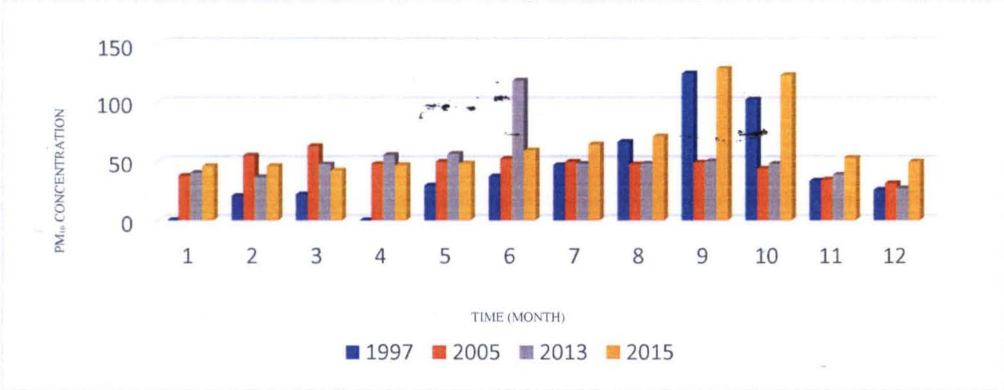
Figure 4.1: Box plot for Pasir Gudang, Melaka and Petaling Jaya in 1997, 2005, 2013, and 2015

4.1 Monthly Variation of PM_{10} concentration

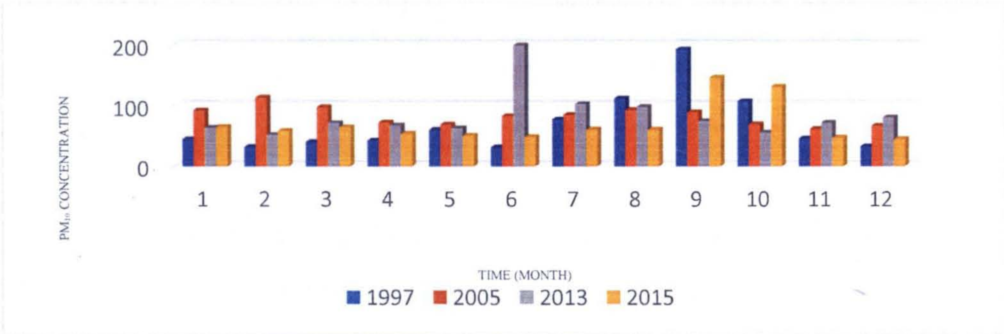
For a period of 12 months of averaging, the MAAQG recommends a PM_{10} concentration of 50 g/m^3. The straight line represented the required MAAQG value for the 12-month average period (50 g/m^3). Figure 4.2 displays the Pasir Gudang, Melaka, and Petaling Jaya monthly average plot for the years 1997, 2005, 2013, and 2015. High levels of PM_{10} concentration were present during all years with hazy weather. The monthly average plot in Pasir Gudang, Melaka, and Petaling Jaya during the dry season, which is from June to September, surpassed the advised value. Additionally, Melaka had the greatest concentration while Pasir Gudang showed the lowest. Two monsoon regimes, the Southwest Monsoon (May to August) and the Northeast Monsoon (November to February), have the most impact on Peninsular Malaysia (Tangang et al., 2012). Peninsular Malaysia experiences much more rainfall in the north and east than the rest of Malaysia does during the NEM. During the SWM, Peninsular Malaysia sees a dry, hot season with less rain and clearer skies.

PM_{10}, on the other hand, might have negative effects on the ecosystem. High PM concentrations in the air are a primary contributor to poor vision and foggy conditions, especially during the dry season. The three areas' greatest monthly average PM_{10} concentrations were found from June through October, during the summer monsoon season. Due to the south-westerly winds that will transport the pollution from our neighboring nation to Malaysia, this monsoon, also known as the Southwest Monsoon, is the most opportune time for transboundary air pollution to occur.

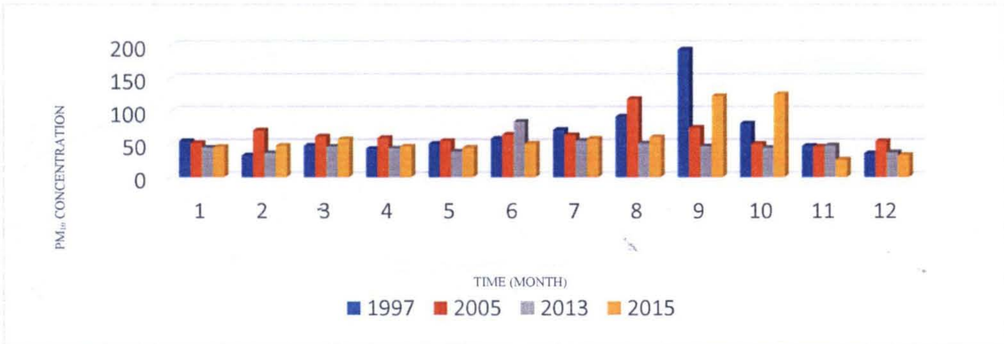
As a consequence, for all three research regions, the trend of PM_{10} concentration was high in September. This is because the region had a number of days with poor air quality, which were mostly brought on by transboundary pollution during the south-west monsoon (Wen et al., 2016). The greatest PM_{10} concentration, meanwhile, was found in 2013 in Melaka. Due to transboundary pollution, Malaysia suffered brief intervals of mild to moderate haze from June through October. This was a result of the hot and dry weather, and Melaka saw the worst episode of haze on that month compared to past years. The proximity of Melaka to Sumatera in the southwest has intensified the high PM concentration during the high particulate event. Additionally, it was clearly demonstrated in 1997 and 2015 that significant levels of PM_{10} concentration were observed from September to October. Conclusion: Because PM_{10} concentrations in the study regions were greater than the MAAQG-recommended limit of 50 g/m^3, more unwell days were observed in these years than in years without haze.



a)



b)



c)

Figure 4.2: Monthly average plot for (a) Pasir Gudang, (b) Melaka and (c) Petaling Jaya in 1997, 2005, 2013, and 2015

4.2 Variation of PM_{10} concentration

Figure 4.3 to 4.5 show the daily time series plot for all the sites. The reference lines indicate the recommended value of MAAQG for daily 24 hours averaging time (50 $\mu g/m^3$). PM_{10} concentrations were fluctuated, and higher concentration can be observed mostly from day 146 to 250 for 2013. This is because in 2013 Southeast Asian haze was notable for causing record high levels of pollution in several parts of Malaysia. The Air Pollution Index in several parts in Malaysia reached a record high of 401 on 21 June 2013, surpassing the previous record of 226 set during the 1997 Southeast Asian Haze (BBC News, 2014).

Figure 4.3 illustrates the time series plot for Pasir Gudang in 1997, 2005, 2013 and 2015. In 1997, no trend of occurrence of HPE. The highest daily average for 1997 was 101 $\mu g/m^3$ at day 173. While for year 2005, higher concentration of PM_{10} noticeable starting day 233 until 348. The highest daily concentration was found in day 297 with 253 $\mu g/m^3$. For year 2013, a short peak was observed starting from day 162. The highest daily average of PM_{10} concentration was recorded in 2013 at day 353 with the value of 394 $\mu g/m^3$. Lastly, in 2015 the period of HPE was same as in 2005. Among the twelve months in 1997, 2005, 2013 and 2015, June to October recorded the higher concentration that exceeded the recommended value of MAAQG.

Figure 4.4 shows the time series plot for Melaka for 1997, 2005, 2013 and 2015. In 1997, the peak of PM_{10} concentration started at day 234 with the value 135 $\mu g/m^3$. Meanwhile, in 2005 bimodal peak was observed from day 29 to 86 and day 172 to 226. In 2013, the highest concentration of PM_{10} was observed in day 174 with the value 504 $\mu g/m^3$ and this value was the highest concentration recorded in all the study areas. Lastly, in 2015, high concentration of PM_{10} were detected starting day 238 (111 $\mu g/m^3$) to 306 (44 $\mu g/m^3$). Similar to before, among twelve months in all years, higher concentrations of PM_{10} were observed during June to October that exceeded the endorsed value of MAAQG.

The daily time series plot for Petaling Jaya in 1997, 2005, 2013 and 2015 was shown in Figure 4.5. Petaling Jaya also had quite high number of PM_{10} readings that exceeded the MAAQG. The highest PM_{10} concentration in 1997 was found out to be 462 $\mu g/m^3$. In 2005, two peaks were observed. The smaller peak was detected at day 48 and the higher peak was measured at day 223 with the value 462 $\mu g/m^3$. In 2013, a noticeable peak was observed on day 175 (344 $\mu g/m^3$) whereas for 2015, the highest observation was 320 $\mu g/m^3$.

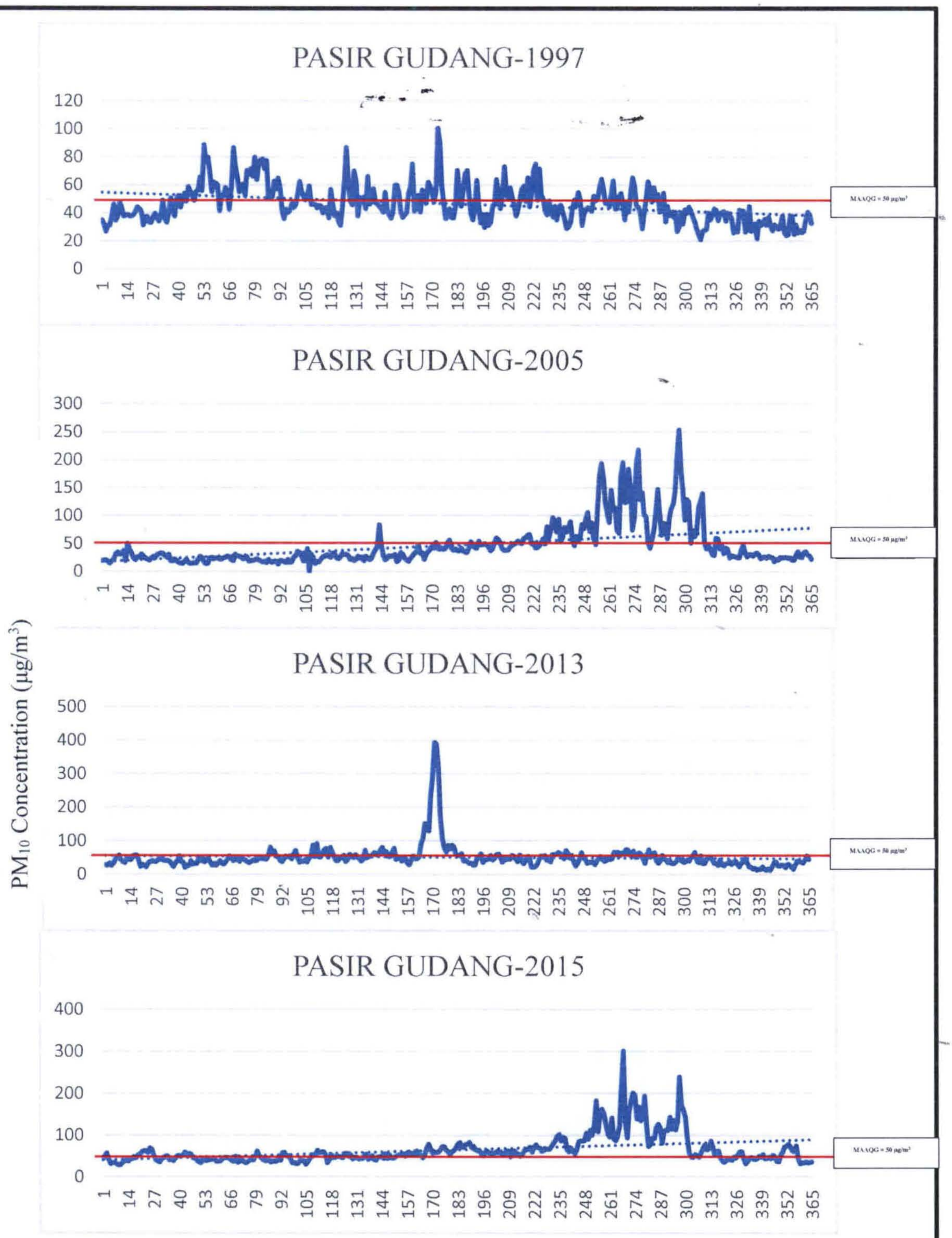


Figure 4.3: Time Series Plot for Pasir Gudang in 1997, 2005, 2013 and 2015

Time (Day)

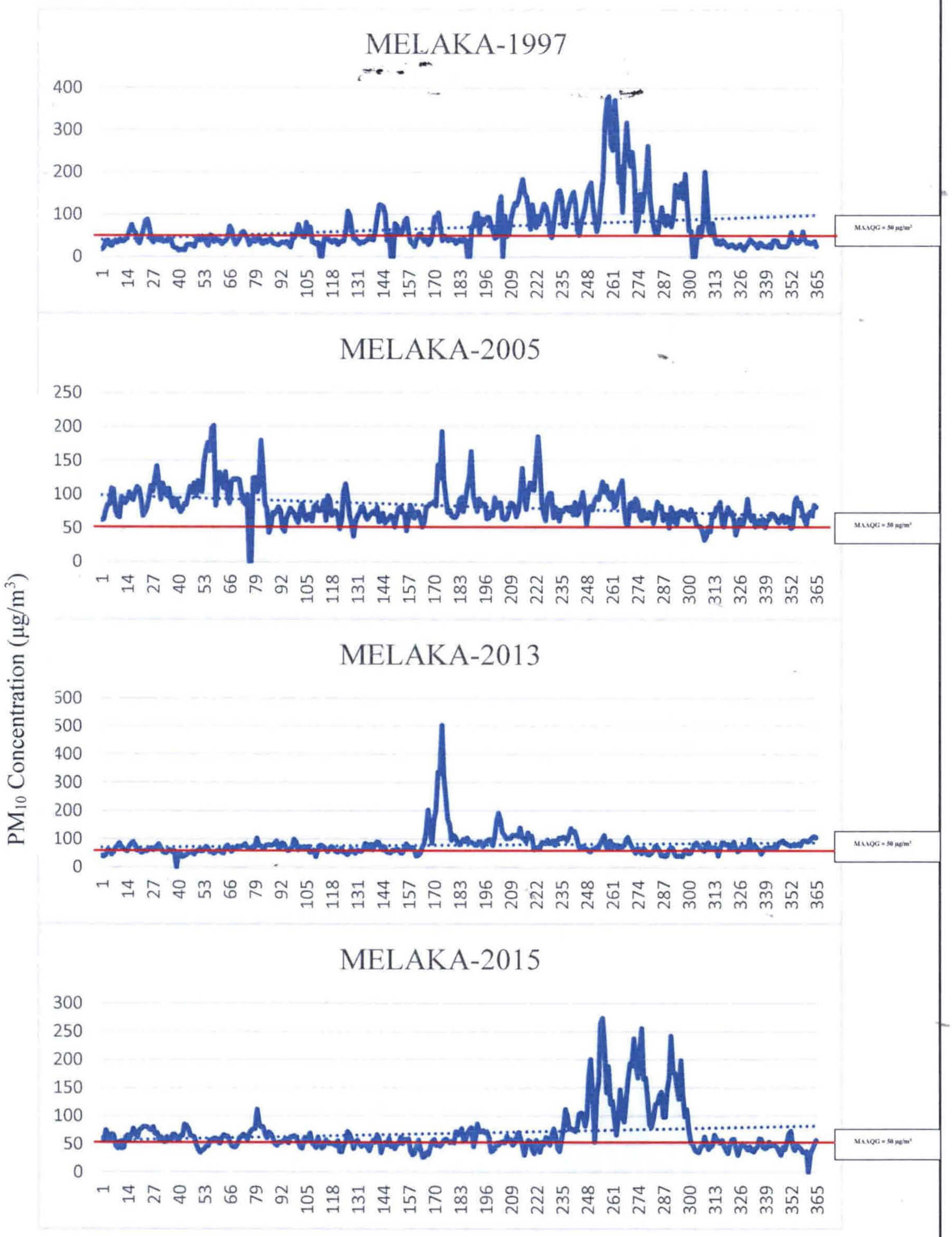


Figure 4.4: Time Series Plot for Melaka in 1997, 2005, 2013 and 2015

Time (Day)

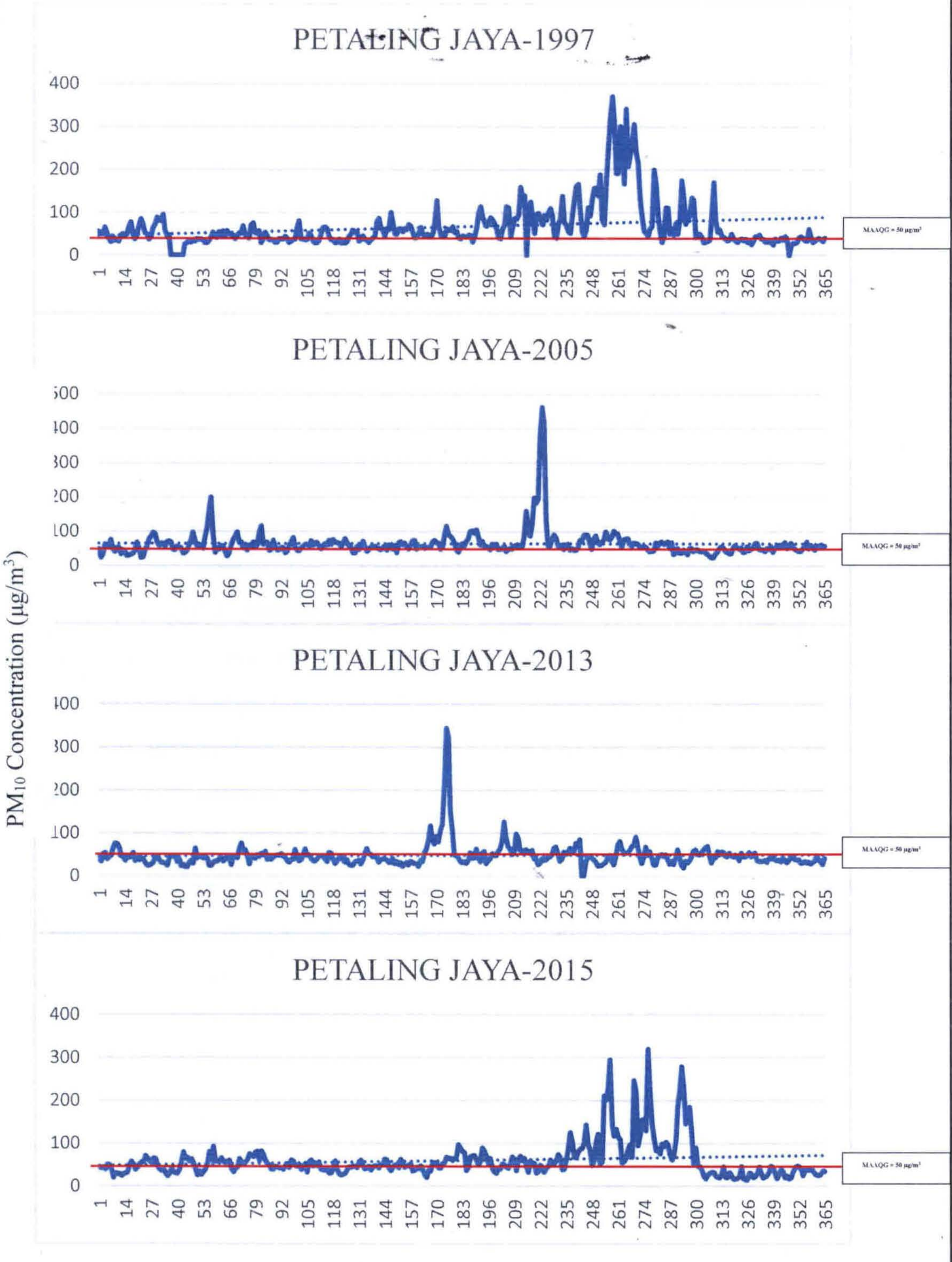


Figure 4.5: Time Series Plot for Petaling Jaya in 1997, 2005, 2013 and 2015
Time (Day)

4.3 Association of PM_{10} with other air pollutants and whether parameter during HPE

The Pearson Correlation matrix for air pollutants and whether variables in Pasir Gudang, Melaka, and Petaling Jaya is shown in Table 4.2. The air pollutant of interest in this investigation was PM_{10}; as a result, the correlation between the level of PM_{10} and other parameters and trace gases was examined. In general, there was a substantial correlation between CO and SO_2 levels and PM_{10} concentration in Pasir Gudang, Melaka, and Petaling Jaya.

In Pasir Gudang, there is a moderate and weak correlation between PM_{10} concentration and CO, SO_2, and NO_2 with r values of 0.575, 0.211, and 0.214, respectively, and a negative correlation with wind speed with r = -0.070. The most correlated meteorological variables in Melaka, however, were SO_2 and CO, with r values of 0.730 and 0.303, respectively. Figure 4.6 show heat matrix of association PM_{10} with trace gases and weather parameter at Pasir Gudang, Melaka and Petaling Jaya. Among these three places, Melaka was recording the highest value correlation of CO with PM_{10} concentration. This finding demonstrates how the concentration of PM_{10} is influenced by CO, SO_2, and surrounding industrial areas as well as by vehicle emissions in the downtown region.

Overall, across all research study areas, CO showed the greatest association with PM_{10} concentration, with a r value more than 0.5. Other than that, there was a poor correlation between PM_{10} level and all other pollutants and meteorological variables, with a value of less than 0.3. In addition, there is very minimal relationship with the weather parameter.

Table 4.2: Correlation between the parameter at Pasir Gudang, Melaka and Petaling Jaya

Station	Parameter	WS	TEMP	NO_x	SO_2	NO_2	O_3	CO	PM_{10}
Pasir Gudang	Wind speed	1	0.408**	-3.96**	0.110**	-0.371**	0.540**	-0.290**	-0.070**
	Temperature		1	-0.197**	-0.017**	-0.063**	0.539**	-0.086**	0.135**
	NO_x			1	0.152**	0.682**	-0.417**	0.362**	0.096**
	SO2				1	0.214**	0.027**	0.108**	0.211**
	NO_2					1	-0.293**	0.258**	0.214**
	O_3						1	-0.131**	0.114**
	CO							1	0.575**
	PM_{10}								1
Melaka	Wind speed	1	0.477**	-0.152**	-0.174**	-0.131**	0.363**	-0.106**	-0.115**
	Temperature		1	-0.106**	-0.130**	-0.063**	0.512**	-0.095**	-0.056**
	NO_x			1	-0.074**	0.660**	-0.314**	0.162**	0.045**
	SO_2				1	-0.006	-0.009	0.148**	0.303**
	NO_2					1	-0.179**	0.172**	0.100**
	O_3						1	-0.056**	0.081**
	CO							1	0.730**
	PM_{10}								1
Petalin g Jaya	Wind speed	1	0.587**	-0.307**	-0.216**	-0.080**	0.512**	-0.265**	-0.080**
	Temperature		1	-0.237**	-0.118**	0.106**	0.695**	-0.186**	0.031**
	NO_x			1	0.339**	0.554**	-0.355**	0.511**	0.095**
	SO_2				1	0.178**	-0.004	0.347**	0.247**
	NO_2					1	-0.073**	0.447**	0.156**
	O_3						1	-0.165**	0.091**
	CO							1	0.505**
	PM_{10}								1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

	WS	T	NO_x	SO_2	NO_2	O_3	CO	PM_{10}
WS	1	0.408	-3.96	0.11	-0.371	0.54	-0.29	-0.07
T		1	-0.197	-0.017	-0.063	0.539	-0.086	0.135
NO_x			1	0.152	0.682	-0.417	0.362	0.096
SO_2				1	0.214	0.027	0.108	0.211
NO_2					1	-0.293	0.258	0.214
O_3						1	-0.131	0.114
CO							1	0.575
PM_{10}								1

Scale: 1, 0.8, 0.6, 0.4, 0.2, 0, -2, -4

a)

	WS	T	NO_x	SO_2	NO_2	O_3	CO	PM_{10}
WS	1	0.477	-0.152	-0.174	-0.131	0.363	-0.106	-0.115
T		1	-0.106	-0.13	-0.063	0.512	-0.095	-0.056
NO_x			1	-0.074	0.66	-0.314	0.162	0.045
SO_2				1	-0.006	-0.009	0.148	0.303
NO_2					1	-0.179	0.172	0.1
O_3						1	-0.056	0.081
CO							1	0.73
PM_{10}								1

Scale: 1, 0.8, 0.6, 0.4, 0.2, 0, -1, -2

b)

	WS	T	NO_x	SO_2	NO_2	O_3	CO	PM_{10}
WS	1	0.587	-0.307	-0.216	-0.08	0.512	-0.265	-0.08
T		1	-0.237	-0.118	0.106	0.695	-0.186	0.031
NO_x			1	0.339	0.554	-0.355	0.511	0.095
SO_2				1	0.178	-0.004	0.347	0.247
NO_2					1	-0.073	0.447	0.156
O_3						1	-0.165	0.091
CO							1	0.505
PM_{10}								1

Scale: 1, 0.8, 0.6, 0.4, 0.2, 0, -0.2, -0.4

c)

Figure 4.6: Pearson correlation matrix of PM10 with the trace gases and weather parameter (a) Pasir Gudang, (b) Melaka and (c) Petaling Jaya

4.4 Model development

4.4.1 Multiple Linear Regression (MLR)

In order to create the MLR models, 80% of the data set was utilized for model creation and 20% for model validation. MLR and quantile are two different kinds of models that have been established. It is done with the intention of comparing the performances of the standard model (MLR) with the quantile model since MLR had been widely utilized to represent the linear parameters. The forecast was created for the following day (t_{-24}), the following two days (t_{-48}), and the following three days (t_{-72}).

All study regions' MLR models were created using parameters that have a reasonably high ($r > 0.5$) correlation with PM_{10} levels. (See Section 4.5, Table 4.2.) For each of the research locations, the MLR model is shown for 24 hours, 48 hours, and 72 hours in Table 4.3. Overall, O_3 and SO_2 concentrations showed favorable influences, whereas NO_2 concentrations showed negative influences. Additionally, a relatively low coefficient for meteorological variables, particularly temperature, was found. This shows that using MLR, less correlation between meteorological factors and PM_{10} level was found. At Pasir Gudang, Melaka, and Petaling Jaya, SO_2 is a significant predictive variable.

The MLR model was validated across all research areas using a variety of performance indicators. The validation findings are presented for each research area in Table 4.3. With an average R^2 value of roughly 0.8, MLR generally performs rather well in forecasting PM_{10} concentration in all studied regions. Regarding Pasir Gudang, Lastly, MLR models provide superior prediction to other locations. Petaling Jaya had the QR models' R^2 score of 0.6432 that were least well-fitted. While Pasir Gudang had the highest R^2 score with a reading of 0.9574.

Table 4.3: MLR model for PM10 concentration prediction in Pasir Gudang, Melaka and Petaling Jaya

MLR Method	Prediction Day	Model
Pasir Gudang	**Next day:** **(PM_{10+24})**	0.782 + 0.762 PM_{10} – 0.097 WS + 0.222 T – 39.544 NOx + 45.075 SO_2 + 193.460 NO_2 + 65.241 O_3 + 4.904 CO
	Next 2- day: **(PM_{10+48})**	-2.663+ 0. .668 PM_{10} – 0.038 WS + 0.614 T – 29.444 NOx + 106.516 SO_2 + 144.560 NO_2 – 15.720 O_3 + 2.553 CO
	Next 3- day: **(PM_{10+72})**	-6.709+ 0. 603 PM_{10} – 0.180 WS + 0.997 T – 35.375 NOx + 152.795 SO_2 + 104.844 NO_2 – 53.335 O_3 + 0.607 CO
Melaka	**Next day:** **(PM_{10+24})**	14.933 + 0.663 PM_{10} – 0.492 WS – 0.022 T + 54.068 NO_x + 840.512 SO_2 – 32.539 NO_2 + 108.158 O_3 + 10.147 CO
	Next 2- day: **(PM_{10+48})**	28.849 + 0.541 PM_{10} – 0.366 WS – 0.190 T + 34.670 NO_x + 998.280 SO_2 – 142.880 NO_2 + 105.562 O_3 + 10.235 CO
	Next 3- day: **(PM_{10+72})**	38.876 + 0.444 PM_{10} – 0.389 WS – 0.253 T + 11.231 NO_x + 1238.200 SO_2 – 85.471 NO_2 + 101.377 O_3 + 10.018 CO
Petaling Jaya	**Next day:** **(PM_{10+24})**	15.143 + 0.677 PM_{10} - 0.633 WS +0.140 T – 2.975 NO_x + 17.075 SO_2 - 14.996 NO_2 + 94.115 O_3 + 1.557 CO
	Next 2- day: **(PM_{10+48})**	25.242 + 0.548 PM_{10} - 0.352 WS +0.239 T + 10.134 NO_x + 217.742 SO_2 – 76.798 NO_2 + 40.925 O_3 – 2.607 CO
	Next 3- day: **(PM_{10+72})**	36.692 + 0.456 PM_{10} + 0.104 WS +0.038 T + 6.061 NO_x + 445.166 SO_2 – 56.348 NO_2 – 17.150 O_3 – 4.787 CO

Table 4.3 summarize the models developed using MLR for the next day, the next two days, and the next three days. The performances of each model for all research areas will be evaluated in the next section.

4.4.2 Quantile Regression

To illustrate the predicted and observed values of PM_{10} concentration Figure 4.7 shows the scatter plot of the predicted and observed values of PM_{10} at all study areas.

Table 4.4: MLR model for PM_{10} concentration prediction for Pasir Gudang

Quantile Method	Prediction Day	Model
Quantile = 0.25	**Next day: (PM_{10+24})**	-6.830 + 0.628 PM_{10} - 0.166 WS + 0.647 T – 10.003 NOx – 63.538 SO_2 + 78.150 NO_2 – 38.006 O_3 - 0.587 CO
	Next 2- day: (PM_{10+48})	-10.308 + 0.511 PM_{10} - 0.186 WS + 1.016 T – 13.174 NOx – 21.247 SO_2 + 15.550 NO_2 – 130.403 O_3 – 1.185 CO
	Next 3- day: (PM_{10+72})	-15.196 + 0.451 PM_{10} - 0.238 WS + 1.305T + 4.565 NOx – 24.089 SO_2 - 39.366 NO_2 – 150.481 O_3 – 0.692 CO
Quantile = 0.50	**Next day: (PM_{10+24})**	-3.589 + 0.763 PM_{10} - 0.138 WS + 0.4797 T – 10.074 NOx 41.173 SO_2 + 113.223 NO_2 + 5.514 O_3 + 0.698 CO
	Next 2- day: (PM_{10+48})	-4.886 + 0.689 PM_{10} - 0.156 WS + 0.772T + 9.849 NOx + 23.200 SO_2 - 33.444 NO_2 – 65.279 O_3 – 1.472 CO
	Next 3- day: (PM_{10+72})	-6.001+ 0.626 PM_{10} - 0.187 WS + 0.978 T – 1.172 NOx + 24.126 SO_2 - 71.864 NO_2 – 1017.845 O_3 – 1.572 CO
Quantile = 0.75	**Next day: (PM_{10+24})**	1.129 + 0.950 PM_{10} - 0.133 WS + 0.201 T – 4.788 NOx + 37.697 SO_2 + 131.604 NO_2 + 40.443 O_3 + 3.026 CO

	Next 2- day: **(PM_{10+48})**	-0.487 + 0.909 PM_{10} - 0.180 WS + 0.501T – 12.698 NOx + 86.767 SO_2 + 46.017 NO_2 – 36.432 O_3 – 0.006 CO
	Next 3- day: **(PM_{10+72})**	0.171+ 0.849 PM_{10} – 0.141 WS + 0.666 T – 23.149 NOx + 111.324 SO_2 - 48.084 NO_2 – 78.628 O_3 – 0.140 CO

Table 4.5: MLR model for PM_{10} concentration prediction for Melaka

Quantile Method	Prediction Day	Model
Quantile = 0.25	**Next day:** **(PM_{10+24})**	19.702 + 0.589 PM_{10} - 0.463 WS - 0.075 T + 72.756 NOx - 406.468 SO_2 -15.151 NO_2 + 89.399 O_3 - 0.947 CO
	Next 2- day: **(PM_{10+48})**	29.306 + 0.487 PM_{10} - 0.649 WS - 0.086 T + 89.406 NOx – 615.999 SO_2 - 121.952 NO_2 + 77.502 O_3 -3.763 CO
	Next 3- day: **(PM_{10+72})**	43.099 + 0.403 PM_{10} - 0.790 WS – 0.430 T + 61.227 NOx – 562.053 SO_2 - 12.214 NO_2 + 104.4019 O_3 - 4.419 CO
Quantile = 0.50	**Next day:** **(PM_{10+24})**	20.669 + 0.734 PM_{10} - 0.411 WS - 0.0225 T + 38.094 NOx + 62.819 SO_2 - 33.682 NO_2 + 95.478 O_3 + 4.227 CO
	Next 2- day: **(PM_{10+48})**	35.763 + 0.617 PM_{10} - 0.673 WS – 0.211 T + 22.675 NOx – 112.197 SO_2 - 142.081 NO_2 + 50.855 O_3 -2.166 CO
	Next 3- day: **(PM_{10+72})**	43.622 + 0.548 PM_{10} - 0.712 WS -0.396 T +9.634 NOx - 229.339 SO_2 - 71.320 NO_2 + 79.418 O_3 - 0.370 CO

Quantile = 0.75	**Next day: (PM_{10+24})**	19.657 + 0.844 PM_{10} - 0.314 WS - 0.290 T + 4.206 NOx + 848.972 SO_2 - 90.704 NO_2 + 84.913 O_3 + 12.468 CO
	Next 2- day: (PM_{10+48})	30.634 + 0.723 PM_{10} - 0.399 WS – 0.224 T - 19.839 NOx +1168.840 SO_2 - 181.055 NO_2 + 70.452 O_3 +9.001 CO
	Next 3- day: (PM_{10+72})	33.385 + 0.667 PM_{10} – 0.333 WS – 0.051 T - 50.444 NOx +1163.855 SO_2 - 177.544 NO_2 + 7.130 O_3 +7.262 CO

Table 4.6: MLR model for PM_{10} concentration prediction for Petaling Jaya

Quantile Method	Prediction Day	Model
Quantile = 0.25	**Next day: (PM_{10+24})**	13.073+ 0.369 PM_{10} – 0.661 WS + 0.215 T – 21.902 NOx + 53.574 SO_2 + 84.108 NO_2 + 48.262 O_3 +2.356 CO
	Next 2- day: (PM_{10+48})	21.213+ 0.245 PM_{10} – 0.416 WS + 0.204 T – 27.867 NOx + 87.038 SO_2 + 15.211 NO_2 + 26.648 O_3 +1.569 CO
	Next 3- day: (PM_{10+72})	27.485+ 0.181PM_{10} – 0.305 WS + 0.087 T –20.139 NOx + 132.561 SO_2 + 3.873 NO_2 + 22.854 O_3 +0.686 CO
Quantile = 0.50	**Next day: (PM_{10+24})**	16.373+ 0.619 PM_{10} – 0.481 WS + 0.069 T – 4.107NOx – 264.092 SO_2 + 42.570 NO_2 + 79.951 O_3 + 1.165 CO

	Next 2- day: (PM_{10+48})	$27.699 + 0.443\ PM_{10} - 0.119\ WS + 0.057\ T - 15.147\ NOx - 90.114\ SO_2 - 24.742\ NO_2 + 47.681\ O_3 - 0.019\ CO$
	Next 3- day: (PM_{10+72})	$27.803 + 0.381 PM_{10} - 0.153\ WS + 0.237\ T - 9.583 NOx + 12.878\ SO_2 - 57.231\ NO_2 - 1.317\ O_3 - 0.888\ CO$
Quantile = 0.75	**Next day: (PM_{10+24})**	$23.945 + 0.912\ PM_{10} - 0.340\ WS - 0.160\ T + 15.740\ NOx - 361.891\ SO_2 + 4.155\ NO_2 + 109.390\ O_3 - 2.106\ CO$
	Next 2- day: (PM_{10+48})	$32.361 + 0.754\ PM_{10} + 0.030\ WS - 0.054\ T + 7.199\ NOx - 132.909\ SO_2 - 89.982\ NO_2 + 75.423\ O_3 - 3.121\ CO$
	Next 3- day: (PM_{10+72})	$37.106 + 0.634 PM_{10} + 0.343\ WS + 0.003\ T + 5.725 NOx - 46.191\ SO_2 - 129.878\ NO_2 - 0.631\ O_3 - 2.699\ CO$

4.7 Model validation

By computing the performance indicators, the model's performances in both the development phase and the validation process may be assessed. To illustrate the predicted and observed values of PM_{10} concentration level, Figure 4.7 shows the QR with the percentile value of 0.75 scatter plot of the predicted and observed values of PM_{10} concentration level at all study areas.

The Pasir Gudang model's performance indicator shows that the QR is superior to the MLR. This is because the models' efficiency for the next day (t_{-24}) is fairly excellent as a result of their superior forecast accuracy (R^2). The R^2 is 0.9590, which is close to 1, with the remaining values being 0.8960 (t_{-48}) and 0.8234 (t_{-72}). The QR with the 0.9648 for IA (t_{-24}) and the R^2 value is a reasonable model for Melaka (0.9313). It still demonstrates that (t_{-24}) is a better forecast day than others.

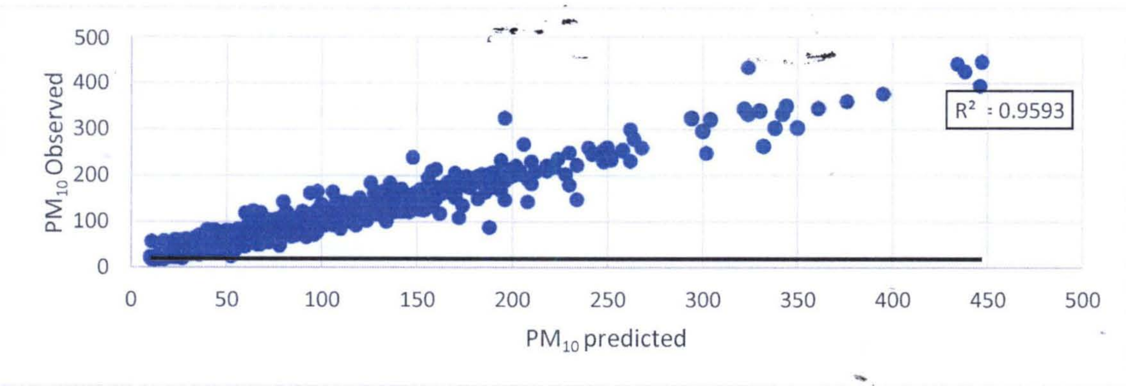
On the other hand, the performance of prediction in Petaling Jaya for QR shows the highest value of IA 0.9402 and also the R^2 0.8477 with the larger difference value compared to the MLR model. Thus, from Table 4.9, all the models show the QR was better and efficiency to predict the PM_{10} concentration for future plan on haze episode in three locations which are Pasir Gudang, Melaka and Petaling Jaya due to the less of error prediction and better in the performance for predict the next day of PM_{10} concentration. While the next two day and next three day is more inefficiency due to the larger difference value of MLR and QR for all the performance indicators.

Table 4.6: Performance Indicator for PM_{10} concentration

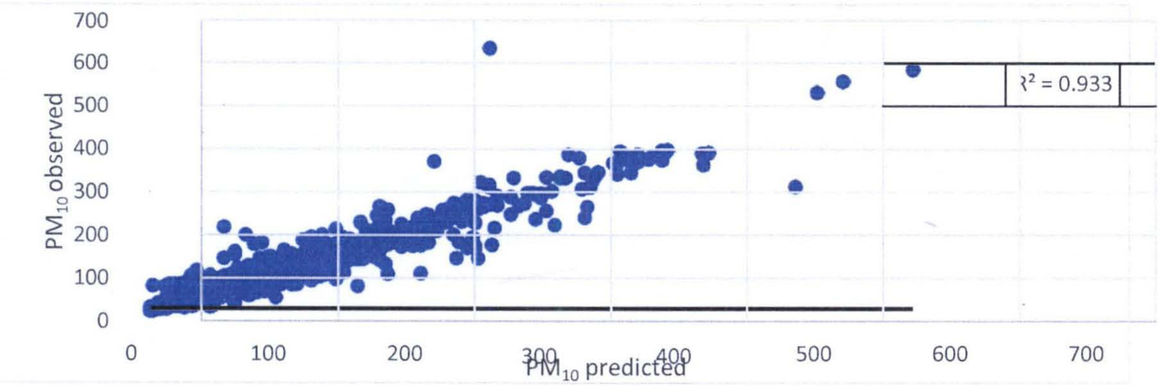
Area	Method		Time	MAE	RMSE	R2	IA
Pasir Gudang	MLR		t_{+24}	5.1103	8.8998	0.9574	0.9786
			t_{+48}	7.8344	13.6104	0.8953	0.9426
			t_{+72}	9.8603	17.0323	0.8199	0.9002
	QR	0.25	t_{+24}	10.4284	16.8980	0.95332	0.9083
		0.50		5.2537	9.8915	0.9575	0.9725
		0.75		8.5804	10.3296	0.9590	0.9763
		0.25	t_{+48}	12.9812	22.0149	0.8785	0.8219
		0.50		7.7282	14.3681	0.8910	0.9347
		0.75		10.1688	13.4344	0.8960	0.9567
		0.25	t_{+72}	14.1221	24.5201	0.7975	0.7639
		0.50		9.5325	17.6273	0.8154	0.8920
		0.75		12.0039	16.4257	0.8234	0.9298
Melaka	MLR		t_{+24}	8.9307	14.4313	0.9313	0.9656
			t_{+48}	13.0490	20.8531	0.84960	0.9162
			t_{+72}	16.4779	25.5475	0.7631	0.8576

	QR	0.25	t+24	16.5609	25.7681	0.9303	0.8703
		0.50		9.5023	14.4703	0.9370	0.9656
		0.75		13.0533	16.3806	0.9327	0.9648
		0.25	t+48	45.3932	52.9326	0.8120	0.5988
		0.50		12.7690	22.2740	0.8410	0.9011
		0.75		16.6684	21.6554	0.8500	0.9285
		0.25	t+72	22.7082	37.0225	0.7349	0.6704
		0.50		15.2397	26.3561	0.7645	0.8459
		0.75		19.2472	25.5602	0.7681	0.8911
	MLR		t+24	10.7237	19.4529	0.8483	0.9264
			t+48	14.6795	25.8775	0.7376	0.8398
			t+72	22.7322	41.8505	0.6432	0.4640
	QR	0.25	t+24	17.9423	32.4801	0.8310	0.7261
		0.50		11.0787	21.4725	0.8482	0.9036
		0.75		14.4401	19.9354	0.8477	0.9402
		0.25	t+48	21.3532	38.8354	0.7340	0.5596
		0.50		15.1969	29.1241	0.7390	0.7723

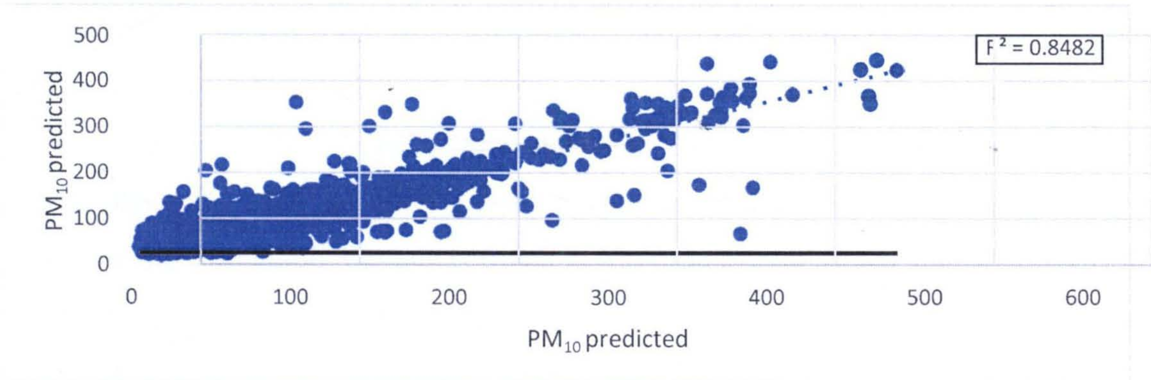
		0.75		17.4244	24.8754	0.7363	0.886
		0.25	t_{+72}	22.7322	41.8505	0.6432	0.4640
		0.50		16.8178	32.3383	0.6446	0.6897
		0.75		19.2922	28.5155	0.6396	0.8229



(a) Pasir Gudang



(b) Melaka



(c) Petaling Jaya

Figure 4.7: Scatter plot of the relationship between observed and predicted value of PM_{10} concentration using quantile regression with percentile value of 0.75 (a) Pasir Gudang, (b) Melaka, (c) Petaling Jaya

CHAPTER 5

CONCLUSION AND RECOMMENDATION

This chapter was concluded the research findings based on the data analysis that had been discussed in Chapter 4 and the whole final year project, focusing on the main findings and contributions of the project. Further work had been suggested in the light of the findings and constraints of this project. Section 5.1 conclude the project based on the objective that had been outlined and section 5.2 describes the main recommendations of the project for the future work, which was the work that could not be done in this project due to data, time and resources constraints or was out of the scope of this project.

5.1 Conclusion

This analysis could be considered successful since all the objectives are achieved. The statistical characteristics for PM_{10} and weather parameters had been explored by using the descriptive statistics, box plot, monthly average, and time series plot. The association of PM10 and trace gases including weather parameter was test using non-parametric correlation. Pearson correlation was used to analyze the relation of PM_{10} with one independent variable. From the time series plot, it can be concluded that during June to October, PM_{10} concentration in Pasir Gudang, Melaka and Petaling Jaya recorded the higher concentration that exceeded the recommended value of MAAQG that is more than 50 $\mu g/m^3$. During the dry season the reading concentration for PM_{10} is the highest value for this research. The highest concentration of PM_{10} mostly in June to September because of the open burning from Sumatera. The cooperation between the government can reduce the open burning that occurred for every year can affect to human health and environment and also human activity during the highest concentration of PM_{10}. The presence of high level of PM_{10} in the atmosphere is a major cause visibility especially during the dry season. These high values of PM_{10} concentration will affect by the pollutant in the atmosphere and became the major air pollutant in Malaysia known as high particulate event. Meanwhile, the distribution of PM_{10} concentration is skewed to the right due to the positive values for all sites for four years. For the weather parameters, all of these are also more to the positive value except humidity.

For the second objective was developed regression models of PM_{10} concentration for three study areas which were Pasir Gudang, Melaka and Petaling Jaya during HPE by using Multiple Linear Regression and Quantile Regression. 24 hours, 48 hours and 72 hours predictions was made and from 24 hours prediction with 0.75 percentile value for all study areas, the result in Pasir Gudang was the best fit model (MAE = 8.5804, RMSE = 10.3296, R^2 = 0.9590 and IA=0.9763). Petaling Jaya shows the weakest prediction with R^2 = 0.8483. Meanwhile, R^2 value for Melaka was 0.9313.

Lastly, the third objective for this research validated the regression models of PM_{10} concentration using several Performance Indicator. For each of the locations, it has their own parameters factor that significant with the air pollutant concentration. From these significant parameters, it can be developing the model to achieve the objectives three. After that, this result (model) was chosen based on the performance indicators to get the better and efficiency model. The good model was used in future for predict the PM_{10} concentration in Pasir Gudang, Melaka and Petaling Jaya where the QR for next day prediction is better than next two days and next three day due to less of error and higher of performance.

5.2 Recommendation for The Further Research

This research is important to predict the air pollution in Malaysia for future. However, from this research there are still to improve for further and the following suggestions are:

5.2.1 Other advance modelling technique such as Artificial Neural Network (ANN) can be used to predict PM_{10} concentration during high particulate event.

5.2.2 Hybrid model that is combination of two methods is known can improve the accuracy of prediction model. For next study, combination of the feature selection analysis such as Support Vector Machine (SVM) with the regression method or Artificial Neural Network (ANN) can be implemented to evaluate the performances of the hybrid model.

REFERENCES

Abdullah, S., Mansor, A. A., Napi, N., Mansor, W., Ahmed, A. N., Ismail, M., & Ramly, Z. (2020). Air quality status during 2020 Malaysia Movement Control Order (MCO) due to 2019 novel coronavirus (2019-nCoV) pandemic. *The Science of the total environment*, *729*, 139022.

Abdullah, S.; Ismail, M.; Fong, S.Y.; Ahmed, A.N. (2016). Evaluation for Long Term PM_{10} Concentration Forecasting Using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models. *Environment Asia*, *9*, 101–110.

Abdullah, S.; Ismail, M.; Samat, N.N.A.; Ahmed, A.N. Modelling Particulate Matter (PM_{10}) Concentration in Industrialized Area: A Comparative Study of Linear and Nonlinear Algorithms. ARPN J. *Engineering and Applied Science*, 13, 8227–8235.

Abdul Rahman, H. (2013). Haze phenomenon in Malaysia: Domestic or transboundary factor. *3rd International Journal Conference on Chemical Engineering and its Applications* (597-599).

Abdul-Wahab, S.A.; Bakheit, C.S.; Al-Alawi, S.M. (2005) Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting Its Concentrations. *Environmental Model Software*, *20*, 1263–1271.

Ab Manan, N., Manaf, M.R.A. and Hod, R. (2018). The Malaysia Haze and its health economy impact: A literature review. *Malaysian Journal of Public Health Medicine*, *18(1)*, 38-45.

Afroz R, Hassan MN, Ibrahim NA. (2003). Review of air pollution and health impacts in Malaysia. *Environmental Research; 92(2)*,71-77.

Ahmat, H., Yahaya, A. S., & Ramli, N. A. (2015). PM_{10} analysis for three industrialized areas using extreme value. *Sains Malaysian*, *44(2)*, 175-185.

Allu, S. K., Srinivasan, S., Maddala, R. K., Reddy, A., & Anupoju, G. R. (2020). Seasonal ground level ozone prediction using multiple linear regression (MLR) model. *Modeling Earth Systems and Environment*, *6(4)*, 1981-1989.

Azid, Azman & Juahir, Hafizan & Amran, Mohammad & Suhaili, Zarizal & Osman, Romizan & Muhamad, Asyaari & Abidin, Ismail & Sulaiman, Nur & Mohd Saudi, Ahmad Shakir. (2015). Spatial air quality modelling using chemometrics techniques: A case study in Peninsular Malaysia. *Malaysian Journal of Analytical Sciences. 19.* 1415 - 1430.

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

Billet, S., Garçon, G., Dagher, Z., Verdin, A., Ledoux, F., Cazier, F., Courcot, D., Aboukais, and Shirali, P. (2007). Ambient particulate matter ($PM_{2.5}$): Physicochemical characterization and metabolic activation of the organic fraction in human lung epithelial cells (A549*). Environmental Research, 105(2),* 212–223.

Czernecki, B., Półrolniczak, M., and Kolendowicz, L. (2017). Influence of the atmospheric conditions on PM_{10} concentrations in Poznań, Poland *Journal of Atmospheric Chemistry*, *74*, 115–139.

Department of Environment (DOE). Chronology of haze episodes in Malaysia. Ministry of Science, Technology and the Environment, Malaysia: Kuala Lumpur (2015).

Department of Environment (DOE). A Guide to Pollution Index in Malaysia. Ministry of Science, Technology and the Environment, Malaysia: Kuala Lumpur (2021).

Delfino RJ, Brummel S, and Wu J. (2009). The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003, *Occupational and Environmental Medicine ;66,* 189-197.

Goyal SK and Chalapati Rao CV. (2007) Assessment of atmospheric assimilation potential for industrial development in an urban environment: Kochi (India). *The Science of the Total Environment. 376*, 27–39.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. Monographs on statistics and applied probability, *International Journal of Forecasting, 33(3)*, 143.

Heil, Angelika & Goldammer, J. (2001). Smoke-Haze Pollution: A Review of the 1997 Episode in Southeast Asia. *Regional Environmental Change. 2*. 24-37.

Ilyas SZ, Khattak AI, Nasir SM, Qurashi T, Durrani R (2009). Air pollution assessment in urban areas and its impact on human health in the city of Quetta, Pakistan. *Clean Technologies and Environmental Policy 12(3).* 291-299. https://doi.org/10.1007/s10098-009-0209-4

Ivanova, and V. (2020). The Anthropogenic Air Pollution and Human Health. *Journal of IMAB - Annual Proceeding (Scientific Papers) 26(2).*3057-3062. https://doi.org/10.5272/jimab.2020262.3057

Jamalani, Asrul & Ku Yusof, Ku & Azid, Azman & Samsudin, Mohd & Amin, Siti & Abd Rani, Nurul & Abdullah Sani, Muhamad Shirwan. (2018). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models over particulate matter (PM_{10}) variability during haze and non-haze episodes: A decade case study. *Malaysian Journal of Fundamental and Applied Sciences. 15.* https://doi.org/10.11113/mjfas.v15n2.1004

Kelly, F.J. and Fussell, J.C. (2015). Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental geochemistry and health, 37(4)*, 631-649.

Latif, M. T., Othman, M., Idris, N., Juneng, L., Abdullah, A. M., Hamzah, W. P., Khan, M. F., Sulaiman, N. M. N., Jewaratnam, J., Aghamohammadi, N., Sahani, M., Xiang, C. J., Ahamad, F., Amil, N., Darus, M., Varkkey, H., Tangang, F., & Jaafar, A. B. (2018). Impact of regional haze towards air quality in Malaysia: *A Review Atmospheric Environment 177,* 28-44. https://doi.org/10.1016/j.atmosenv.2018.01.002

Lee, C.C., Ballinger, T.J., Domino, N.A., (2012). Utilizing map pattern classification and surface weather typing to relate climate to the air quality index in Cleveland, Ohio. *Atmospheric Environment. 63*, 50–59.

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., Bates & K. H. (2019). Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proceedings of the National Academy of Sciences, 116(2),* 422–427. https://doi.org/10.1073/pnas.1812168116

Lu H. C. (2003), Estimating the emission source reduction of PM_{10} in central Taiwan. *Air Quality Atmosphere & Health, 8(4),* 805 – 814. https://doi.org/1016/j.chemosphere.2003.10.012

Lu, X;Zhang, L; Wang, L; Gao, M; Li, K; Zhang, X; Yue, X and Zhang, Y. (2020) *Environmental Science & Technology Letters* 7 *(4),* 240-247. https://doi.org/10.1021/acs.estlett.0c00171

Manju, A., Kalaiselvi, K., and Dhananjayan, V. (2018). Spatio-seasonal variation in ambient air pollutants and influence of meteorological factors in Coimbatore, Southern India. *Air Quality Atmosphere and Health, 11*, 1179–1189. https://doi.org/10.1007/s11869-018-0617-x

Mohamad, Jamilah & Kiggundu, Amin. (2007). The rise of the private car in Kuala Lumpur, Malaysia: Assessing the policy options. *IATSS Research. 31*. 69-77.

https://doi.org/1016/S0386-1112(14)60185-0

Mukaka M.M. (2012). Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J.; 24:69–71. Munir, S., Chen, H., and Ropkins, K., (2012). Characterising the temporal variations of ground level ozone and its relationship with traffic-related air pollutants in the UK: a quantile regression approach. Sustainable development and planning, accepted April, 2012.

Mohamed Noor, N, Mohamad Hasim, N. I. & Yusof, S. Y. (2018). Variation of Ground-Level Ozone Concentration in Urbanized Area in Malaysia. *IOP Conference Series: Materials Science and Engineering, 374*. https://doi.org/10.1088/1757-899X/374/1/01209

N.A. Ghazali, N.A. Ramli, A.S. Yahaya, N.F.F.M.D. Yusof, N. Sansuddin, and W.A. Al Madhoun., (2010). Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. *Environment Monitoring and Assessment, 165*, 475–489. https://doi/10.1007/s10661-009-0960-3

Nastos, P.T., Paliatsos, A.G., Anthracopoulos, and M.B. (2010). Outdoor particulate matter and childhood asthma admissions in Athens, Greece: a time-series study. *Environmental Health, 9*, 45. https://doi.org/10.1186/1476-069X-9-45

Nazif A, Mohammed NI, Malakahmad A, and Abualqumboz MS. (2019). Multivariate analysis of monsoon seasonal variation and prediction of particulate matter episode using regression and hybrid models. *International Journal of Environmental Science and Technology, 16*. 2587–600. https://doi.org/10.1186/1476-069X-9-45

Oltra, C. and Sala, R., (2018). Perception of risk from air pollution and reported behaviors: a cross sectional survey study in four cities. *Journal of Risk Research, 21(7),* 869-884. https://doi.org/10.1080/13669877.2016.1264446

Ostro, B.; Roth, L.; Malig, B.; and Marty, M. (2009). The effects of fine particle components on respiratory hospital admissions in children. *Environmental Health Perspectives. 117*, 475–480. https://doi.org/10.1289/ehp.11848

Ul-Saufie, A.Z.; Yahya, A.S.; Ramli, N.A.; Hamid, and N.A. (2011). Comparison Between Multiple Linear Regression and Feed Forward Back Propagation Neural Network Models for Predicting PM_{10} Concentration Level Based on Gaseous and Meteorological Parameters. *International Journal of Sustainability Science and Management, 12 (1)*, 42–49.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Hamid, H. A. (2012). Performance of multiple linear regression model for longterm PM_{10} concentration prediction based on gaseous and meteorological parameters. *Journal of Applied Sciences, 12*, 1488-1494. https://doi.org/0.3923/jas.2012.1488.1494

Usmani, R.S.A., Saeed, A., Abdullahi, and A.M. (2020). Air pollution and its health impacts in Malaysia: a review. Air Quality Atmosphere and Health 13, 1093–1118. https://doi.org/10.1007/s11869-020-00867-x

Park, J.; Chang, and S. (2021). A Particulate Matter Concentration Prediction Model Based on Long Short-Term Memory and an Artificial Neural Network. *International Journal of Environmental Research and Public Health, 18*, 6801. https://doi.org/10.3390/ijerph18136801

P. H. Chou, K.N. Chiang, and Steven Y. Liang. (2019) Reliability Assessment of Wafer Level Package using Artificial Neural Network Regression Model, *Journal of Mechanics, 35*, 829–837. https://doi.org/10.1017/jmech.2019.20

R.San José, A. Baklanov, R.S. Sokhi, K. Karatzas, and J.L. Pérez. (2008). Air Quality Modeling, *Encyclopedia of Ecology,* 111-123.

https://doi.org/10.1016/B978-008045405-4.00201-9

R. Win Lee, Fine particulate matter measurement and international standardization for air quality and emissions from stationary sources, (2010). *Fuel, 89*, 874-882. https://doi.org/10.1016/j.fuel.2009.03.023

Richard A. Dunn, Andrew K.G. Tan & Rodolfo M. Nayga (2012) Obesity inequality in Malaysia: decomposing differences by gender and ethnicity using quantile regression, *Ethnicity&Health, 17(5)*, 493-511, https://doi.org/10.1080/13557858.2012.661407

Shaziayani, W.N., Ul-Saufie, A.Z., and Ahmat, H. (2021). Coupling of quantile regression into boosted regression trees (BRT) technique in forecasting emission model of PM_{10} concentration. *Air Quality, Atmosphere & Health, 14*, 1647–1663. https://doi.org/10.1007/s11869-021-01045-3

Tangang, Fredolin & Latif, Mohd Talib & Juneng, Liew. (2010). The roles of climate variability and climate change on smoke haze occurrences in Southeast Asia region. *LSE Research Online, 4,*36-49.

Wong, L.P., Alias, H., Aghamohammadi, N., Ghadimi, A. and Sulaiman, N.M.N., (2017). Control measures and health effects of air pollution: A survey among public transportation commuters in Malaysia. *Sustainability, 9(9),* 1616.

https://doi.org/10.3390/su9091616

Yap, X.Q. and Hashim, M. (2013). A robust calibration approach for PM_{10} prediction from MODIS aerosol optical depth. *Atmospheric Chemistry and Physics, 13(3),* 3517-3526. https://doi.org/10.5194/acp-13-3517-2013

Zaman NAFK, Kanniah KD, Kaskaoutis DG, and Latif MT. (2021). Evaluation of Machine Learning Models for Estimating $PM_{2.5}$ Concentrations across Malaysia. *Applied Sciences, 11(16),* 732. https://doi.org/10.3390/app11167326

APPENDICES

Descriptive Statistics for Year 1997

STUDY AREA	PARAMETER		WS (m/s)	T (°c)	UVB	H (%)	NO_X	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)	PM_{10} (ppm)
PASIR GUDANG	N	VALID	8716	8702	8661	0	7827	8598	7813	8253	8494	8631
		MISSING	44	58	99	8760	933	162	947	507	266	129
	MEAN		8.4997	26.8365	139.8687		.0209	.0131	.0117	.0192	.58745	47.7212
	MEDIAN		7.4000	26.6000	1.0000		.0140	.0090	.0100	.0170	.45000	33.0000
	STANDARD DEVIATION		4.97371	2.50371	234.27507		.02229	.01199	.00899	.01412	.423010	39.87401
	MINIMUM		.70	21.50	.00		.00	.00	.00	.00	.000	11.00
	MAXIMUM		27.40	33.30	1136.00		.25	.11	.07	.11	2.790	268.00
MELAKA	N	VALID	8474	6728	8421	0	6411	8379	6411	7320	8437	8337
		MISSING	286	2032	339	8760	2349	381	2349	1440	323	423
	MEAN		6.1619	26.8436	185.6037		.0125	.00722	.00807	.01870	.62922	71.699
	MEDIAN		5.3000	26.4000	8.0000		.0100	.00600	.00700	.01500	.45000	46.000
	STANDARD DEVIATION		4.01254	3.81672	285.33736		.00906	.004841	.004865	.015666	.500179	61.5892
	MINIMUM		.60	20.00	1.00		.00	.001	.000	.000	.000	13.0
	MAXIMUM		21.70	37.50	1229.00		.10	.030	.049	.095	3.780	415.0
PETALING JAYA	N	VALID	8743	8744	0	0	8270	8653	8270	0	0	8222
		MISSING	17	16	8760	8760	490	107	490	8760	8760	538
	MEAN		3.289	27.061			.1017	.0121	.0311			69.42
	MEDIAN		2.900	26.200			.0910	.0120	.0290			49.00
	STANDARD DEVIATION		1.9541	3.1440			.06142	.00381	.01705			55.054
	MINIMUM		.1	20.7			.00	.00	.00			20
	MAXIMUM		25.9	37.7			.53	.03	.12			393

Descriptive Statistics for Year 2005

STUDY AREA	PARAMETER		WS (m/s)	T (°c)	UVB	H (%)	NO_X	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)	PM_{10} (ppm)
PASIR GUDANG	N	VALID	8708	8709	6313	8709	7971	8692	7970	8237	8610	8715
		MISSING	52	51	2447	51	789	68	790	523	150	45
	MEAN		6.337	27.336	123.773	80.990	.0242	.0055	.0117	.0143	.6640	46.59
	MEDIAN		5.500	26.800	9.000	84.000	.0160	.0050	.0100	.0110	.5400	44.00
	STANDARD DEVIATION		3.6298	2.7487	192.8832	11.1602	.02496	.00456	.00768	.01252	.39542	13.700
	MINIMUM		.9	21.6	.0	40.0	.00	.00	.00	.00	.00	19
	MAXIMUM		24.2	35.0	875.0	96.0	.23	.02	.06	.12	2.88	116
MELAKA	N	VALID	8701	7940	6482	8705	8259	8623	8258	8109	8628	8669
		MISSING	59	820	2278	55	501	137	502	651	132	91
	MEAN		2.99	25.90	140.97	77.69	.01779	.00637	.01005	.0182	.561	83.34
	MEDIAN		2.70	25.20	2.00	81.00	.01600	.00600	.00900	.0150	.540	78.00
	STANDARD DEVIATION		1.651	3.543	220.100	11.824	.011267	.002659	.005108	.01423	.2242	27.383
	MINIMUM		1	20	0	36	.000	.001	.000	.00	.1	29
	MAXIMUM		9	36	958	94	.118	.022	.087	.08	2.2	268
PETALING JAYA	N	VALID	8737	8738	0	8692	8270	8726	8270	8149	8721	8727
		MISSING	23	22	8760	68	490	34	490	611	39	33
	MEAN		3.623	27.859		70.6949	.0645	.0062	.0326	.015	1.683	64.32
	MEDIAN		3.300	26.800		73.0000	.0590	.0060	.0310	.007	1.620	56.00
	STANDARD DEVIATION		2.1024	3.9305		12.61017	.03605	.00231	.01412	.0185	.7606	40.704
	MINIMUM		.7	20.6		30.00	.00	.00	.00	.0	.2	20
	MAXIMUM		13.6	38.6		92.00	.31	.02	.11	.1	9.2	494

Descriptive Statistics for Year 2013

STUDY AREA	PARAMETER		WS (m/s)	T (°c)	UVB	H (%)	NO_X	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)	PM_{10} (ppm)
PASIR GUDANG	N	VALID	6883	7338	0	7515	8207	8760	8207	8306	8697	8745
		MISSING	1877	1422	8760	1245	553	0	553	454	63	15
	MEAN		5.37	29.21		79.66	.0273	.0048	.0156	.0165	.6184	50.99
	MEDIAN		4.70	28.80		82.00	.0190	.0040	.0140	.0130	.5400	45.00
	STANDARD DEVIATION		3.281	2.900		12.454	.02557	.00370	.01076	.01471	.41388	38.430
	MINIMUM		1	23		38	.00	.00	.00	.00	.09	10
	MAXIMUM		19	39		100	.20	.02	.09	.11	4.41	462
MELAKA	N	VALID	3630	3091	0	3041	8048	8454	8048	7031	8447	8669
		MISSING	5130	5669	8760	5719	712	306	712	1729	313	91
	MEAN		6.2727	27.2889		77.12	.0202	.00385	.0097	.0212	.6741	79.15
	MEDIAN		5.7000	26.7000		78.00	.0170	.00300	.0090	.0170	.6300	72.00
	STANDARD DEVIATION		3.38110	2.47967		13.065	.01457	.002765	.00535	.01684	.46609	42.838
	MINIMUM		1.00	22.00		36	.00	.000	.00	.00	.18	32
	MAXIMUM		18.70	34.10		99	.12	.015	.04	.12	10.56	577
PETALING JAYA	N	VALID	8650	8667	0	8671	8017	8659	8017	8236	8643	8659
		MISSING	110	93	8760	89	743	101	743	524	117	101
	MEAN		4.756	28.631		69.260	.0586	.0037	.0282	.0133	1.1199	48.429
	MEDIAN		4.200	28.100		70.000	.0520	.0040	.0260	.0060	1.0800	43.000
	STANDARD DEVIATION		2.4767	2.8223		16.4757	.03297	.00162	.01174	.01669	.48712	29.3081
	MINIMUM		1.0	22.6		28.0	.00	.00	.00	.00	.09	17.0
	MAXIMUM		22.0	36.4		99.0	.30	.01	.09	.12	5.67	372.0

Descriptive Statistics for Year 2015

STUDY AREA	PARAMETER		WS (m/s)	T (°c)	UVB	H (%)	NO_X (ppm)	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)	PM_{10} (ppm)
PASIR GUDANG	N	VALID	8721	5519	0	8722	8148	8677	8145	8265	8730	8710
		MISSING	39	3241	8760	38	612	83	615	495	30	50
	MEAN		5.89	28.824		72.61	.0208	.00426	.0136	.0180	.673	64.80
	MEDIAN		5.50	28.600		75.00	.0150	.00400	.0120	.0160	.540	54.00
	STANDARD DEVIATION		3.735	2.5507		11.804	.01854	.002519	.00836	.01320	.3679	36.086
	MINIMUM		1	23.6		34	.00	.001	.00	.00	.1	27
	MAXIMUM		18	35.6		94	.17	.016	.05	.10	3.1	351
MELAKA	N	VALID	3194	6309	0	6344	8163	8736	8163	8064	8746	8759
		MISSING	5566	2451	8760	2416	597	24	597	696	14	1
	MEAN		5.44	27.87860		75.1280	.0207	.00320	.0114	.0227	.6949	69.69
	MEDIAN		5.70	27.50000		76.0000	.0170	.00300	.0100	.0190	.6300	58.00
	STANDARD DEVIATION		2.381	2.520488		13.66850	.01594	.001627	.00657	.01654	.36534	41.473
	MINIMUM		1	21.600		38.00	.00	.000	.00	.00	.00	24
	MAXIMUM		11	34.200		100.00	.18	.018	.05	.11	2.79	338
PETALING JAYA	N	VALID	8723	8725	0	8430	8187	8125	8187	8232	8163	8591
		MISSING	37	35	8760	330	573	635	573	528	597	169
	MEAN		5.02	28.55		70.72	.0610	.00359	.0280	.0179	1.258	60.53
	MEDIAN		4.60	27.90		73.00	.0570	.00300	.0270	.0100	1.140	49.00
	STANDARD DEVIATION		2.448	2.966		15.968	.03433	.002516	.01239	.01869	.6602	50.067
	MINIMUM		1	23		39	.00	.000	.00	.00	.0	5
	MAXIMUM		19	37		99	.19	.033	.11	.12	4.9	472