COMPARISON OF PARAMETER ESTIMATORS OF LOGNORMAL DISTRIBUTION FOR PREDICTING PM_{10} AND $PM_{2.5}$ CONCENTRATIONS

MUHAMMAD UTHMAN BIN OMAR

UNIVERSITI SAINS MALAYSIA

2024

COMPARISON OF PARAMETER ESTIMATORS OF LOGNORMAL DISTRIBUTION FOR PREDICTING PM_{10} AND $PM_{2.5}$ CONCENTRATIONS

by

MUHAMMAD UTHMAN BIN OMAR

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

November 2024

② ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful. Alhamdulillah, all praises to Allah for giving me the blessings, strength and courage to complete this research and thesis.

First and foremost, I would like to express my gratitude and appreciation to my supervisor Dr Hazrul bin Abdul Hamid for his continuous encouragement, guidance, support and patience throughout my study.

A deep appreciation to my beloved wife, Nik Afaf Madihah binti Nik Ismail for her endless support and patience. Whenever I felt discouragement or down, she was always there to cherish and encourage me. And to my family who always prays for my success.

And to my friend Mohamad Faiz bin Ahmad Johari, thank you so much for your help and tips in this study journey.

And to those who have helped me during this journey, I thank you all and pray that all blessings be upon you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS

c_s	Skewness
CO	Carbon monoxide
erf	Error function
erf^{-1}	Inverse error function
k	Kurtosis,
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxide
O_3	Ozone
R^2	Coefficient of determination
s	Standard deviation
s^2	Variance
SO_2	Sulphur dioxide
SO_3	Sulfur trioxide
$\bar{x}$	Mean or expected value
μ	Location parameter
σ	Scale parameter
μg/m^3	Micrograms per cubic meter
μm	Micrometer or micron

LIST OF ABBREVIATIONS

AD	Anderson–Darling test
AF	Atrial fibrillation
AIC	Akaike's Information Criterion
API	Air Pollution Index
ARI	Acute respiratory infections
CAQM	Continuous air quality monitoring stations
CDF	Cumulative distribution function
COVID-19	Coronavirus Disease 2019.
CPCB	Central Pollution Control Board, India
DCCs	Daycare centers
DIO	Diet-induced obesity
DOE	Department of Environment, Malaysia
EQMP	Environmental Quality Monitoring Programme
IA	Index of agreement
ICA	Imperialist competitive algorithm
IIT Madras	Indian Institute Of Technology Madras
KS	Kolmogorov–Smirnov test
LN	Lognormal distribution
MAAQS	Malaysia Ambient Air Quality Standard
MAAQS-IT	Malaysia Ambient Air Quality Standard – Interim Target
MAQM	Manual air quality monitoring stations
MCAQM	Mobile continuous air quality monitoring stations
MCO	Movement control order
MLE	Maximum likelihood estimator
MoM	Method of moments
NAE	Normalized absolute error
PA	Prediction accuracy
PDF	Probability density function
PI	Performance indicator
PM	Particulate matter
$PM_{0.1}$	Particulate matter 0.1 microns or less in aerodynamic diameter

$PM_{2.5}$	Particulate matter 2.5 microns or less in aerodynamic diameter
PM_{10}	Particulate matter 10 microns or less in aerodynamic diameter
PNC	Particulate number concentrations
PWM	Probability weighted moments
RMSE	Root mean squared error
SPSS	Statistical Package for the Social Sciences software
T1D	Type 1 diabetes
UMVUE	Uniformly minimum variance unbiased estimator
WHO	World Health Organisation

LIST OF APPENDICES

PERBANDINGAN PENGANGGAR PARAMETER BAGI TABURAN LOGNORMAL UNTUK MERAMAL KEPEKATAN PM_{10} DAN $PM_{2.5}$

ABSTRAK

PM_{10} and $PM_{2.5}$ merupakan di antara pencemar udara utama di Malaysia dan telah mengakibatkan kesan buruk terhadap kesihatan manusia, terutamanya kepada kanak-kanak, ibu mengandung dan warga emas. Bagi memahami taburan PM_{10} dan $PM_{2.5}$ dengan lebih baik, pemodelan statistik menggunakan taburan lognormal telah digunakan kerana sifat semulajadinya yang pencong positif dan dianggap sebagai taburan yang paling sesuai bagi jirim zarahan di Malaysia. Taburan lognormal mempunya dua parameter iaitu parameter lokasi dan parameter skala. Penganggaran parameter adalah langkah penting bagi mendapatkan ramalan terbaik kerana nilai parameter lokasi dan parameter skala boleh mempengaruhi ralat dan ketepatan ramalan. Dalam kajian ini, empat penganggar iaitu kaedah momen (MoM), penganggar kebolehjadian maksimum (MLE), momen berwajaran kebarangkalian (PWM), dan penganggar varians minimum seragam tidak bias (UMVUE), telah digunakan untuk menganggar parameter lokasi dan parameter skala. Data setiap jam bagi kepekatan PM_{10} dan $PM_{2.5}$ dari 2017 hingga 2020 di empat lokasi stesen pengawasan yang berbeza klasifikasi iaitu Jerantut sebagai latarbelakang, Sungai Petani sebagai pinggir bandar, Alor Setar sebagai bandar, dan Perai sebagai industri telah digunakan. Dengan menggunakan lima petunjuk prestasi iaitu ralat punca kuasa dua min (RMSE), ralat mutlak ternormal (NAE), ketepatan ramalan (PA), pekali penentuan (R^2), dan indeks persetujuan (IA), penganggar MoM terbukti sebagai penganggar terbaik bagi semua stesen. Dapatan ini diuji dengan lebih lanjut bagi melihat samada saiz sampel boleh mempengaruhi prestasi penganggar parameter. 30

set nombor rawak telah dijana menggunakan Penjana Nombor Rawak Mersenne Twister dengan nilai N = 50, 100, 500, 1000, 5000, dan 10000. Berdasarkan keputusan petunjuk prestasi, penganggar yang sesuai tidak dapat diputuskan apabila nilai N adalah kecil. Apabila nilai N menjadi besar iaitu 1000, 5000, dan 10000, penganggar MoM menjadi penganggar yang paling sesuai. Penemuan ini adalah selari dengan keputusan dari data yang diperoleh dari Jabatan Alam Sekitar (JAS) di mana bilangan data secara jam adalah 8760, yang boleh dianggap sebagai nilai N yang besar. Kesimpulannya, kaedah MoM merupakan penganggar yang paling berkesan dan sesuai bagi meramal kepekatan PM_{10} dan $PM_{2.5}$. Penemuan ini dapat membantu kajian masa hadapan dalam menentukan penganggar yang perlu digunakan ketika menyesuaikan taburan lognormal.

COMPARISON OF PARAMETER ESTIMATORS OF LOGNORMAL DISTRIBUTION FOR PREDICTING PM_{10} AND $PM_{2.5}$ CONCENTRATIONS

ABSTRACT

PM_{10} and $PM_{2.5}$ are among the dominant air pollutants in Malaysia and have caused severe adverse effects on human health, especially on children, pregnant mothers, and senior citizens. To better understand the distribution of PM_{10} and $PM_{2.5}$, statistical modelling using the lognormal distribution was used because of its positively-skewed nature and is regarded as the most appropriate distribution for particulate matter in Malaysia. The lognormal distribution has two parameters namely location and scale parameters. Parameter estimation is a crucial step in getting the best prediction since the value of location and scale parameters can affect the error and accuracy of the prediction. In this study, four different estimators namely the method of moments (MoM), maximum likelihood estimator (MLE), probability weighted moments (PWM), and uniformly minimum variance unbiased estimator (UMVUE) were used to estimate the location and scale parameters. Hourly data of PM_{10} and $PM_{2.5}$ concentrations from 2017 to 2020 on four different classifications of monitoring stations namely Jerantut as background, Sungai Petani as suburban, Alor Setar as urban, and Perai as industrial were used. By using five different performance indicators namely the root mean squared error (RMSE), the normalized absolute error (NAE), the prediction accuracy (PA), the coefficient of determination (R^2), and the index of agreement (IA), the method of moments turned out to be the best estimator for all stations. This result was further tested to see whether the sample size can affect the performance of the parameter estimators. 30 sets of random numbers were generated using the Mersenne Twister Pseudorandom Number Generator with N=50, 100, 500,

1000, 5000, and 10000. Based on the performance indicators, there were inconsistencies in determining the appropriate estimator when the N value was small. As the N value became larger i.e. 1000, 5000, and 10000, the method of moments turned out to be the most appropriate estimator. This finding is consistent with the secondary data obtained from the Department of Environment (DOE) as the number of hourly data is 8760, which can be regarded as a large N value. Thus, it can be concluded that the method of moments is the most effective and appropriate estimator to predict PM_{10} and $PM_{2.5}$ concentrations. This finding can help future studies on determining the estimator to be used when fitting lognormal distribution.

CHAPTER 1

INTRODUCTION

1.1 Air Pollution

Air pollution has been a critical issue worldwide. It has caused detrimental effects, not just on human health, but also on the economy and agriculture (Latif et al., 2018; Manan et al., 2018). Many studies have been done to understand the behaviour of air pollution. One way to understand the behaviour is by using statistical modelling.

In Malaysia, the Department of Environment (DOE) is the government agency that is responsible for monitoring the status of ambient air quality. The air quality status is determined by the calculation of six major pollutants including particulate matter 10 microns or less in aerodynamic diameter (PM_{10}), particulate matter 2.5 microns or less in aerodynamic diameter ($PM_{2.5}$), ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2) and sulphur dioxide (SO_2). It is then reported as the Air Pollution Index (API) and categorized into five categories as presented in Table 1.1. (Department of Environment, 2020).

Table 1.1 Air Pollution Index

API	AIR QUALITY STATUS
0 – 50	Good
51 – 100	Moderate
101 – 200	Unhealthy
201 – 300	Very Unhealthy
> 300	Hazardous

Under the new Environmental Quality Monitoring Programme (EQMP), DOE improved the previous environmental quality monitoring system. The new system includes data collection for air, river water and marine water quality for the whole country, and reports real-time data on air, river and sea conditions. Pakar Scieno TW

Sdn Bhd (PSTW) has been appointed by DOE to conduct the EQMP starting 2017 to 2032 (Department of Environment, 2018).

In mid-April 2017, the number of continuous air quality monitoring stations (CAQM) was increased from 56 to 65, to improve the air quality monitoring network. There are also 14 manual air quality monitoring stations (MAQM) and 3 mobile continuous air quality monitoring stations (MCAQM). These stations are strategically placed in urban, suburban, industrial and rural areas. Figure 1.1 shows the location of the monitoring stations in Malaysia (Department of Environment, 2018; Usmani et al., 2020).

Figure 1.1 Air Quality Monitoring Stations in Malaysia
(Source: (Usmani et al., 2020))

Beginning on 16th August 2018, the particulate matter 2.5 microns or less in aerodynamic diameter ($PM_{2.5}$) was included in the API calculation making it the sixth major pollutant monitored (Department of Environment, 2018). Since then, the API reading has had drastic increases and caused an increase in the number of unhealthy days as compared to the previous year due to the inclusion of $PM_{2.5}$ in the calculation.

Air pollutants can be categorized as primary pollutants and secondary pollutants. Primary air pollutants are those that are released directly into the air from their sources. Examples of primary air pollutants are carbon monoxide (CO), nitrogen oxide (NO_x) and particulate matter (PM). Secondary air pollutants are those that are not directly emitted from their polluting sources but formed in the atmosphere from chemical interactions between primary air pollutants, or the reaction of primary air pollutants with pre-existing atmospheric constituents. Examples of secondary air pollutants are tropospheric ozone (O_3), sulfur trioxide (SO_3) and acid rain (Goswami & Chandra, 2019).

Air pollution comes from many sources such as mobile and stationary sources. Mobile sources refer to any moving objects that emit pollutants into the atmosphere. These include vehicles and transportation like cars, trains, planes and cargo ships. While stationary sources refer to the fixed sites or places that emit pollutants. Stationary sources include power plants, industrial facilities and factories (Kulshreshtha, 2019; Usmani et al., 2020).

In Malaysia, major sources of air pollution are industrial sources and motor vehicle emissions. As of December 2020, the number of industrial sources that were emitting air pollutants was 13,776 as compared to 11,126 in 2019. Selangor had the highest pollution sources (21%), followed by Johor (18%) and Perak (11%) as shown in Figure 1.2. The least number of pollution sources are Putrajaya and Labuan (Department of Environment, 2020; Usmani et al., 2020).

Figure 1.3 and Figure 1.4 show the number of in-use vehicles in 2017 and 2018, and the number of in-use vehicles in 2019 and 2020 respectively. The number of passenger cars, motorcycles and goods vehicles show a great increment over four years, while the number of taxis and buses shows a low decrement. Surely the great increment of in-use vehicle numbers will affect air quality status.

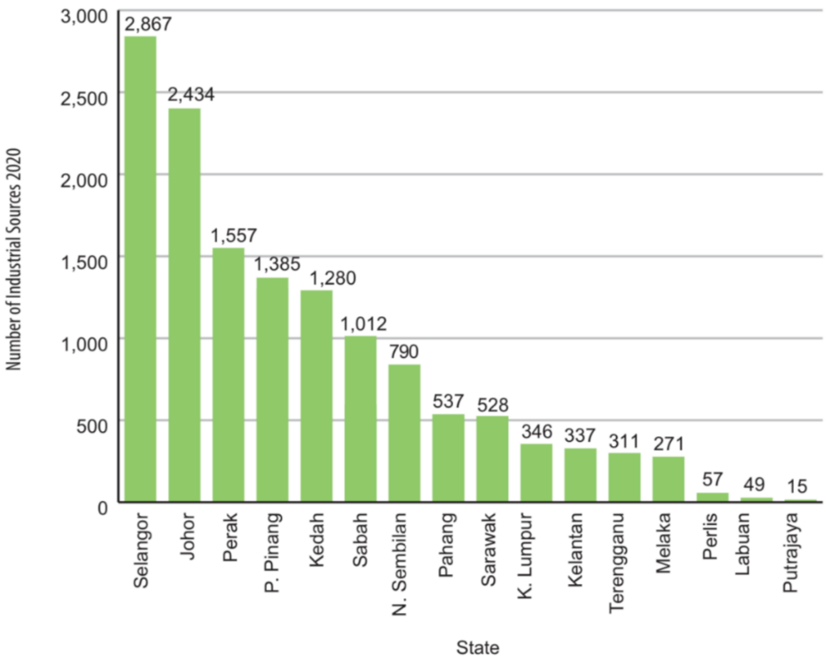


Figure 1.2 Industrial Air Pollution Sources by State in the Year 2020
(Source: (Department of Environment, 2020))

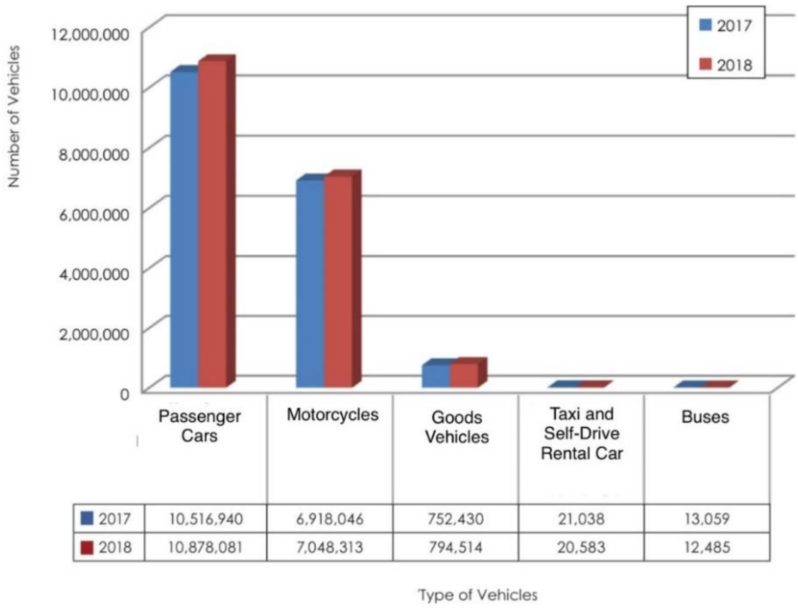
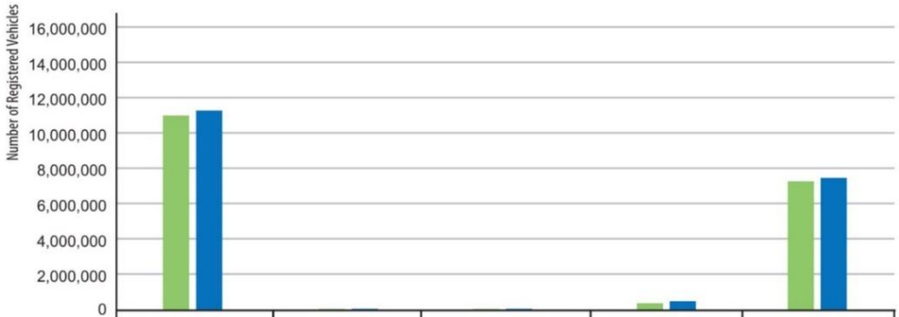


Figure 1.3 Number of In-Use Vehicles in 2017 and 2018
(Source: (Department of Environment, 2018))



	MOTORCAR	TAXI	BUS	GOODS VEHICLE	MOTORCYCLE
2019	11,166,702	19,696	10,886	764,358	7,356,468
2020	11,461,510	18,738	10,071	799,097	7,525,840

Figure 1.4 Number of In-Use Vehicles in 2019 and 2020
(Source: (Department of Environment, 2020))

Figure 1.5 to Figure 1.10 show the air quality status for all stations in Klang Valley, Northern Region, Southern Region of The West Coast Peninsular Malaysia, East Coast Peninsular Malaysia, Sarawak, Sabah and Labuan for the year 2020. Overall air quality status for all stations was good and moderate. There are several unhealthy days recorded with Balok Baru - 9 days; Putrajaya and Rompin - 4 days; Cheras - 3 days; Petaling Jaya, Balik Pulau, Port Dickson, Temerloh and ILP Miri - 2 days; and Shah Alam, Tanjung Malim and Melaka - 1 day. Unfortunately, Balok Baru station recorded very unhealthy readings for one day due to the high release of pollutants from the Gebeng Industrial Area, Kuantan (Department of Environment, 2020).

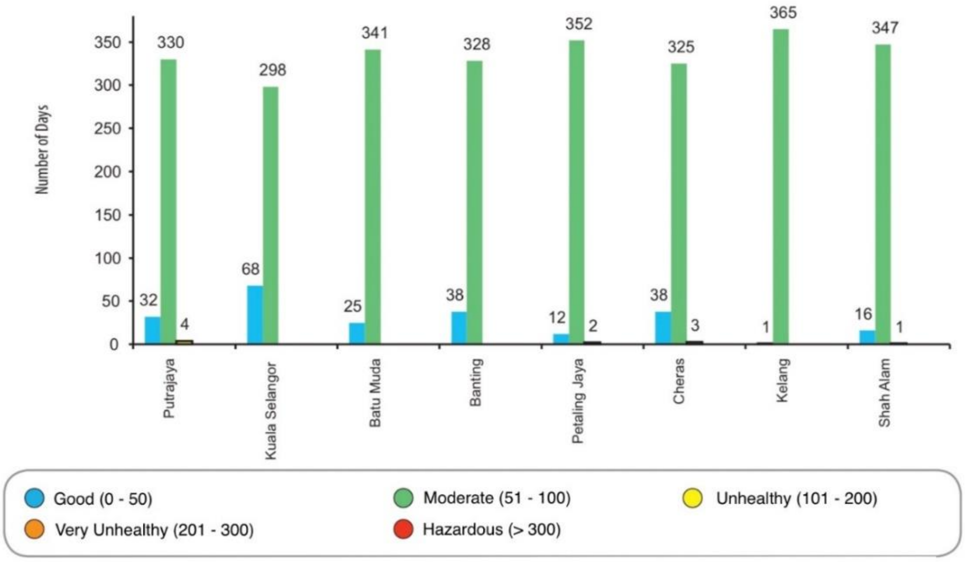


Figure 1.5 Air Quality Status, Klang Valley, 2020
(Source: (Department of Environment, 2020))

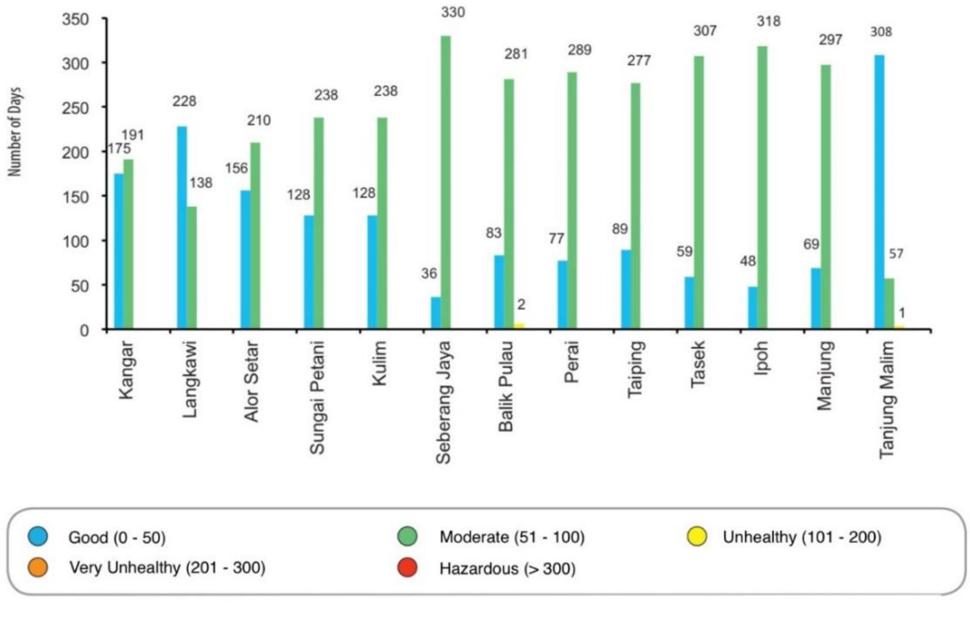


Figure 1.6 Air Quality Status, Northern Region of The West Coast Peninsular Malaysia, 2020
(Source: (Department of Environment, 2020))

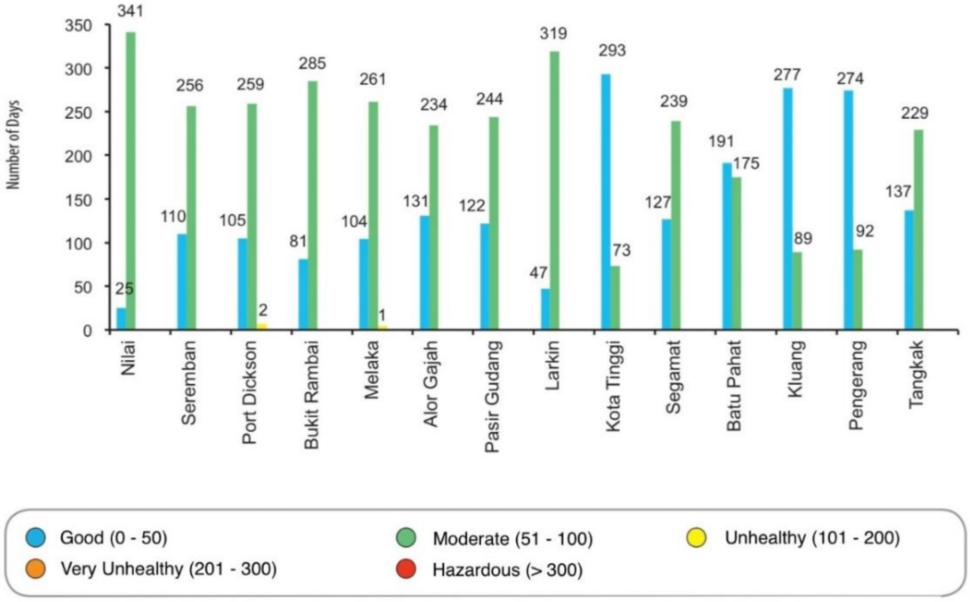


Figure 1.7 Air Quality Status, Southern Region of The West Coast Peninsular Malaysia, 2020
(Source: (Department of Environment, 2020))

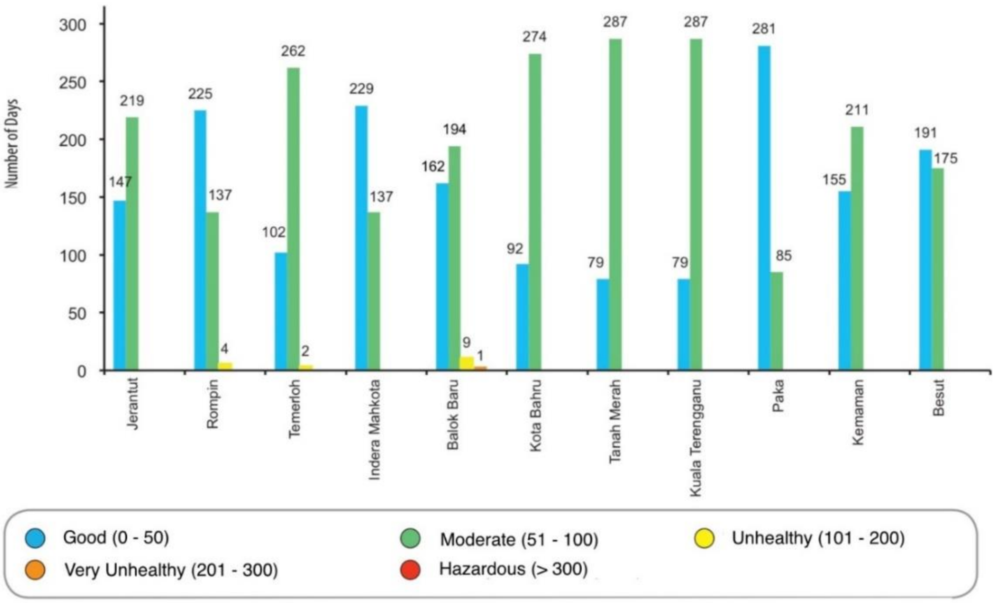


Figure 1.8 Air Quality Status, East Coast Peninsular Malaysia, 2020
(Source: (Department of Environment, 2020))

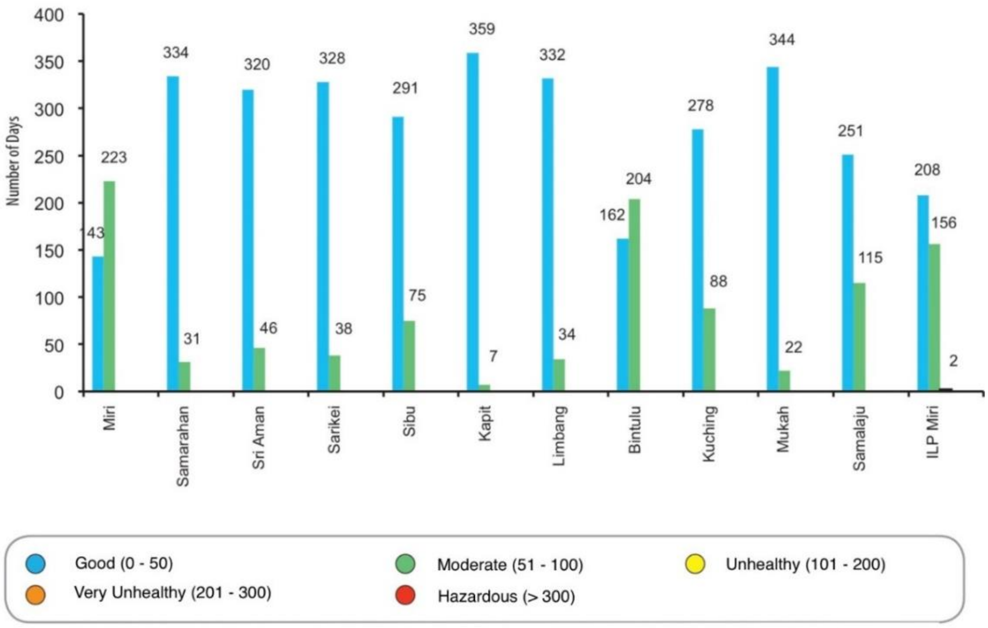


Figure 1.9 Air Quality Status, Sarawak, 2020
(Source: (Department of Environment, 2020))

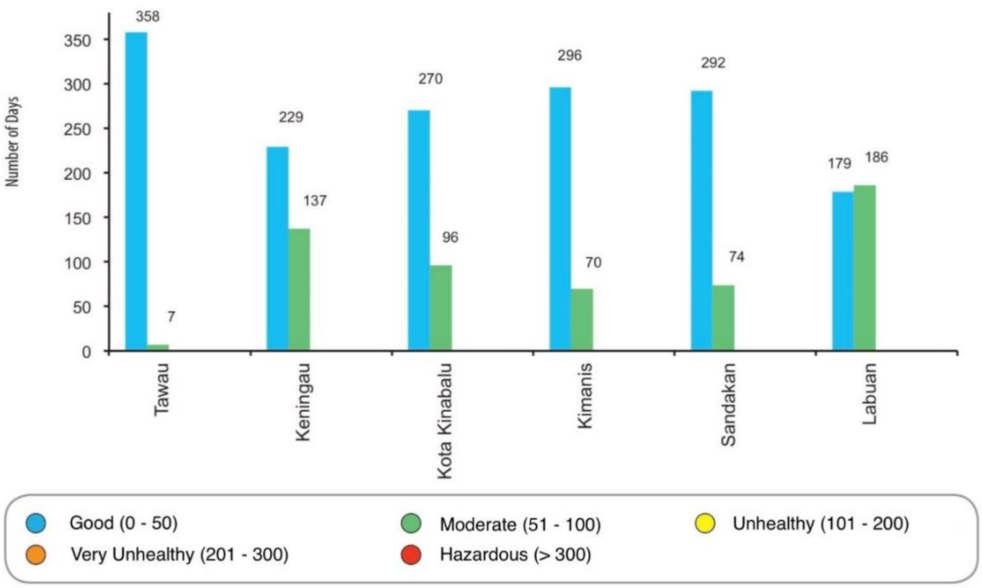


Figure 1.10 Air Quality Status, Sabah and Labuan, 2020
(Source: (Department of Environment, 2020))

Malaysia Ambient Air Quality Standard (MAAQS) is used as the guideline for each pollutant. It is calculated based on the averaging time for each pollutant. To improve air quality, DOE has introduced a new guideline which drew 2-interim targets.

The first interim target was in 2015, the second interim target was in 2018 and the new standard would be in 2020. But for the year 2020, the MAAQS interim target 2018 was used. Table 1.2 shows the details of MAAQS (Department of Environment, 2020).

Table 1.2 Malaysia Ambient Air Quality Standard
(Source: (Department of Environment, 2020))

PARAMETER	AVERAGING TIME	UNIT	EXISTING GUIDELINES	MALAYSIAN AMBIENT AIR QUALITY STANDARD		
				IT-1 (2015)	IT-2 (2018)	STAN-DARD (2020)
PM_{10}	1 Year	µg/m^3	50	50	45	40
	24 Hours	µg/m^3	150	150	120	100
$PM_{2.5}$	1 Year	µg/m^3	-	35	25	15
	24 Hours	µg/m^3	-	75	50	35
SO_2	1 Hour	µg/m^3	350	350	300	250
		ppm	0.135	0.135	0.115	0.095
	24 Hours	µg/m^3	105	105	90	80
		ppm	0.040	0.040	0.035	0.030
*CO	1 Hour	mg/m^3	35	35	35	30
		ppm	30.6	30.6	30.6	26.2
	8 Hours	mg/m^3	10	10	10	10
		ppm	8.75	8.75	8.75	8.75
NO_2	1 Hour	µg/m^3	320	320	300	280
		ppm	0.170	0.170	0.160	0.150
	24 Hours	µg/m^3	75	75	75	70
		ppm	0.040	0.040	0.040	0.037
O_3	1 Hour	µg/m^3	200	200	200	180
		ppm	0.100	0.100	0.100	0.090
	8 Hours	µg/m^3	120	120	120	100
		ppm	0.060	0.060	0.060	0.050

Note: *mg/m^3 IT-Interim Target (year)

In Table 1.2, the ‘Existing Guidelines” refer to the previous guidelines used before 2015, also known as the Malaysia Ambien Air Quality Guidelines (MAAQG) (Department of Environment, 2015). The previous guidelines for PM_{10} were 50 µg/m^3 for the yearly mean and 150 µg/m^3 for the daily mean. After the introduction of MAAQS, the new standard is used (Department of Environment, 2020). For the first interim target, the yearly and daily limits of PM_{10} were the same as the previous guidelines. While for $PM_{2.5}$, the yearly and daily limits were 35 µg/m^3 and 75 µg/m^3.

For the second interim target, the yearly and daily limits for PM_{10} were reduced to 45 μg/m^3 and 120 μg/m^3, and the yearly and daily limits for $PM_{2.5}$ were reduced to 25 μg/m^3 and 50 μg/m^3 respectively. Then, the new standard for the yearly and daily limits of PM_{10} was 40 μg/m^3 and 100 μg/m^3, while 15 μg/m^3 and 35 μg/m^3 for $PM_{2.5}$.

1.2 Particulate Matter

Particulate matter (PM) is a common physical classification of particles found in the air, such as dust, dirt, soot, smoke, and liquid droplets. Unlike other pollutants, PM is not a specific chemical entity but a mixture of particles from different sources and sizes, compositions, and properties. Particles that are sufficiently small and low in mass can be suspended in the air for long periods of time. Some particles can be found in soot and smoke while others can only be seen using special tools (Vallero, 2014).

Particulate matter is categorized by its total mass per unit volume (μg/m^3). There are three categories of particulate matter. The first category is called coarse particles. Its size in aerodynamic diameter ranges from 2.5 μm to 10 μm, also known as PM_{10}. The second category is called fine particles, also known as $PM_{2.5}$, having particle size of 2.5 μm or less in aerodynamics diameter. The third category is called ultrafine particles, having the size of an aerodynamic diameter of 0.1 μm or less, also known as $PM_{0.1}$ (Frederickson et al., 2021; Tran et al., 2020).

In indoor settings, PM sources came from indoor activities including cooking, smoking, and residential hobbies (Tran et al., 2020). While in outdoor settings, PM can be generated from natural sources and anthropogenic sources. Some natural sources are wildfire, sea spray and resuspension of organic matter, while anthropogenic sources mainly come from combustion activities including motor vehicle emissions and wood-smoke emissions (National Research Council (U.S.), 2010).

Since PM is airborne, it can penetrate human organs during inhalation and cause many adverse effects on human health. PM exposure towards the heart and lungs can cause cardiovascular system diseases like acute coronary syndrome, myocardial infarction and ischemic heart disease, and respiratory diseases like asthma, lung cancer and lung dysfunction. PM exposure can also cause adverse effects on other organs like the brain, liver, spleen, kidneys, pancreas and gastrointestinal tract, thus causing oxidative stress and inflammation (Zhang J. et al., 2024).

In Malaysia, PM_{10} and $PM_{2.5}$ are among the major pollutants and are included in the calculation of API. According to DOE, PM is still a major concern that contributes to the API. The main sources that produce PM are power plants, followed by industries and motor vehicles (Department of Environment, 2021). Figure 1.11 shows the yearly trend of PM_{10} mean concentration from 2010 to 2021. The line graph shows that the yearly mean of PM_{10} has improved over 12 years. Figure 1.12 shows the yearly trend of $PM_{2.5}$ mean concentration from 2018 to 2021. Since $PM_{2.5}$ was included in the API readings, the $PM_{2.5}$ yearly mean concentration did not exceed the MAAQS IT-2 limit (Department of Environment, 2021; Malek et al., 2021).

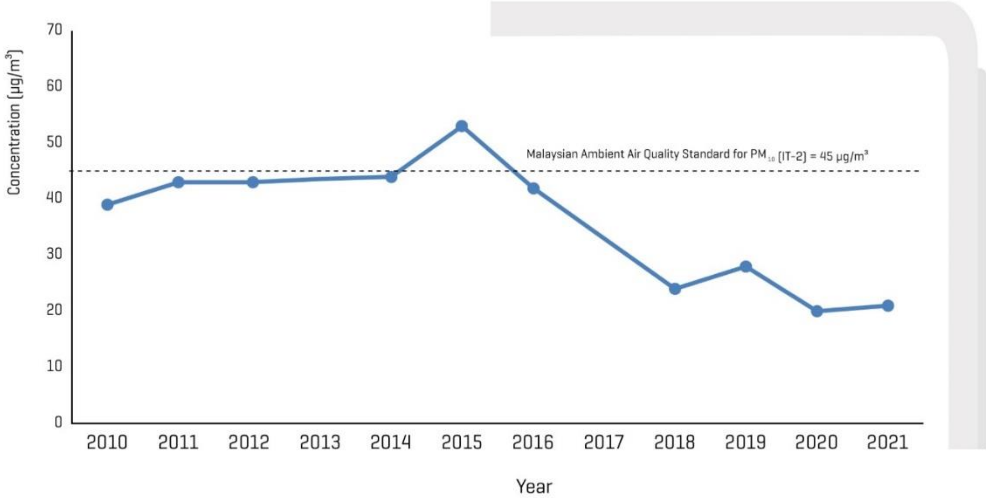


Figure 1.11 Yearly Mean Concentration of PM_{10}, 2010-2021
(Source: (Department of Environment, 2021))

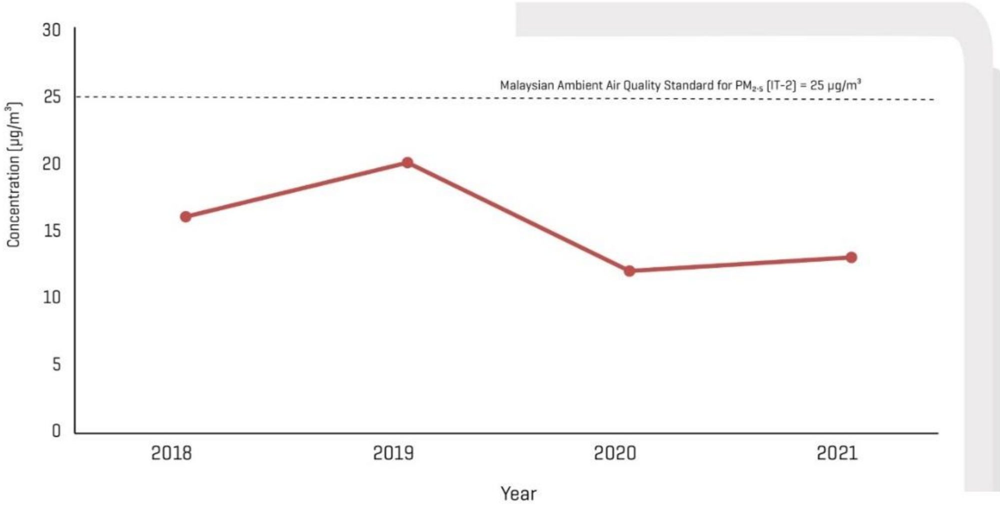


Figure 1.12 Yearly Mean Concentration of $PM_{2.5}$, 2018-2021
(Source: (Department of Environment, 2021))

1.3 Statistical Distribution

To better understand the particulate matter, statistical models are employed. A statistical distribution is a mathematical function that defines the likelihood of different outcomes for a given experiment (Forbes et al., 2010). It indicates how the values of a random variable are dispersed or distributed. Distributions can be visualized through graphs like histograms or probability density functions (PDFs), which display the frequency or probability of various outcomes.

Statistical distributions can be categorized into two categories. The first category is discrete distributions like Bernoulli, Binomial, and Poisson distributions. The second category is continuous distributions like normal, lognormal and exponential distributions. Each statistical distribution has distinct characteristics that determine its shape and behaviour. These include the mean, variance, standard deviation, kurtosis, and skewness (Evans & Rosenthal, 2010).

Furthermore, statistical distribution is crucial for understanding and analyzing particulate matter concentrations. It offers a strong method for evaluating pollution, spotting patterns, and helping to make policy decisions. Some statistical distributions that are widely used in air pollution are Weibull, gamma, lognormal, exponential, Gumbel Rayleigh and Frechet distributions (M. Ismail et al., 2018; Jiang et al., 2019; Kalimeri et al., 2019; Norazian Mohamed et al., 2011; Peter et al., 2023; Razak et al., 2014). These distributions help explain the random nature of particulate matter concentrations, making it easier to estimate important factors like mean, variations, and the chances of exceeding certain limits.

According to Masseran et al. (2021), a statistical distribution can offer insights into the probabilistic behaviour of the magnitude of an air pollution event. For example, the distribution of general air pollution index data can be represented using various unimodal distributions, including Gamma, Lognormal, and Weibull (Al-Dhurafi et al., 2016).

1.4 Problem Statement

As Malaysia progresses towards a developed country, many activities and developments like industrialisation, urbanization, mining, and quarrying take place. These contribute to the emission of air pollutants that affect health and the environment. Of all air pollutants monitored in Malaysia, particulate matter turned out to be the main pollutant (Department of Environment, 2021; Malek et al., 2021).

Particulate matter has caused severe adverse effects on human health (Zhang J. et al., 2024). To understand the PM better, statistical modelling is used. In Malaysia, several distribution models have been used to fit PM, namely Weibull, gamma, lognormal, Gumbel Rayleigh and Frechet distributions (Jaffar et al., 2018; Norazian Mohamed et al., 2011; Razak et al., 2014; Sansuddin et al., 2011). It turns out that the best-fit distribution to fit PM_{10} and $PM_{2.5}$ in Malaysia is the lognormal distribution (LN) (Hamid et al., 2013; Yunus & Hasan, 2017).

Lognormal fitting requires two parameters to be estimated. These two parameters, namely location and scale parameters, will affect the accuracy and errors of the fitting. Previous studies showed that commonly used estimators are maximum likelihood estimators (MLE) (Kan & Chen, 2004; Perišić et al., 2015; Razak et al., 2014; Wang et al., 2013) and method of moments (MoM) (Hamid et al., 2013; Lu, 2002; Md Yusof et al., 2011; Mijić et al., 2009). Other estimators that are not commonly used are probability weighted moments (PWM) (Hamid et al., 2013; Tosunoğlu, 2018; Wan Deraman et al., 2017) and uniformly minimum variance unbiased estimator (UMVUE) (Ginos et al., 2009; Shen, 1998).

Currently, the selection of estimators in a study is based solely on the researcher's preference without any guide or proper justification of the estimators' selection (Gulia et al., 2017; Maciejewska et al., 2015; Masseran et al., 2021; Perišić et al., 2015; Peter et al., 2023; Zhao et al., 2020). Most previous research used MoM and MLE without any specific justification, though there were researches that showed PWM (Hamid et al., 2013) and UMVUE (Ginos et al., 2009) can perform better.

So, this study will further investigate which estimators are best to estimate LN to fit PM_{10} and $PM_{2.5}$. This study will then help authorities and other relevant parties to take action to help reduce air pollution or any control planning, and better human health.

1.5 Research Objectives

To ensure the best estimator to fit lognormal distribution is found, and as to why such an estimator is selected, the study will focus on these three objectives:

1. To describe the trend of PM_{10} and $PM_{2.5}$ concentrations.

2. To determine the most appropriate estimator for lognormal distribution in predicting PM_{10} and $PM_{2.5}$ concentrations based on the secondary data.

3. To determine the most appropriate estimator for lognormal distribution in predicting PM_{10} and $PM_{2.5}$ concentrations based on various sample sizes using simulation data.

1.6 Scope of Research

This study focused on four air monitoring stations with three of the stations located in the northern region of Peninsular Malaysia and one as a background station. Each monitoring station was selected based on the category of suburban, urban and industrial as categorised by DOE. The stations were Jerantut (background), Sungai Petani (suburban), Alor Setar (urban) and Perai (industrial).

Secondary data on PM_{10} and $PM_{2.5}$ concentrations were obtained from the Department of Environment, Malaysia. The PM_{10} and $PM_{2.5}$ concentrations which were analysed consisted of four years of data, spanning from 2017 to 2020.

The statistical model that was employed in this study was limited to only the lognormal distribution since it was the best distribution to fit PM_{10} and $PM_{2.5}$ in Malaysia (Hamid et al., 2013; Yunus & Hasan, 2017). As for parameter estimation, four estimators were used to estimate the lognormal distribution, namely the method of moments (MoM), method of likelihood estimator (MLE), probability weighted moments (PWM) and uniformly minimum variance unbiased estimator (UMVUE). Each estimator was accessed to determine the best estimator that fits the lognormal distribution.

1.7 Thesis Structure

This thesis consists of five chapters. Chapter 1 will cover the background of the study, problem statement, research objectives, scope of the research and thesis structure.

Chapter 2 will discuss the literature review of previous research on PM_{10} and $PM_{2.5}$ concentrations using lognormal distribution. The trends of PM_{10} and $PM_{2.5}$ from previous studies will be discussed. Then, the results based on different existing estimators from previous studies will also be discussed.

Chapter 3 will describe the methodology used in this study. This includes how the data is obtained, the description of the location of the monitoring stations, how the data is processed and analysed using software like Microsoft Excel, Matlab and SPSS, how the data will be displayed and portraited, and how the preliminary results from the obtained data are used for further analysis using simulations using different sample sizes.

Chapter 4 will present and discuss the results. Graphs and tables will be illustrated to visually describe the results. It begins with the discussion of descriptive statistics of PM_{10} and $PM_{2.5}$ for each station for each year. Then it continues with the discussion of the trend of PM_{10} and $PM_{2.5}$ for each station for each year. The chapter continues with a discussion of the performance of the estimators at each station which results in the best estimator. After that, the results of the best estimator for simulation using different sample sizes, $N = 50, \dots, 10000$, will be discussed.

Lastly, chapter 5 will give the conclusion of the study. Recommendations for further research will also be given.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter will discuss the adverse effects of PM_{10} and $PM_{2.5}$ followed by the trends of particulate matter. The discussion continues with lognormal distribution and parameter estimation to fit the PM_{10} and $PM_{2.5}$ concentrations data.

2.2 PM_{10} and $PM_{2.5}$

Particulate matter (PM) is one of the main pollutants in Malaysia. PM is defined as tiny particles suspended in the air which range in size from a few hundred micrometers to nanometers (Tiwari & Mishra, 2019). There are two categories of PM included in the calculation of API in Malaysia, i.e., particles with an aerodynamic diameter less than 10 µm (PM_{10}) and particles with an aerodynamic diameter less than 2.5 µm ($PM_{2.5}$) (Department of Environment, 2021).

Since PM is airborne and its size is so tiny, it can enter the human body through inhalation. PM can be absorbed by cells tissues and organs especially fine particles. Not only it can penetrate the lungs and heart, but also other organs like the brain, liver, kidneys, spleen, pancreas, gastrointestinal tract, joints and reproductive organs. Thus, it can cause adverse effects on human health, especially cardiovascular disease, respiratory disease, reproductive system diseases, motor system diseases, cancer and death (Zhang J. et al., 2024).

2.3 The Effects of Particulate Matter

Studies have shown that particulate matter can harm people's health, both over long periods and short periods (Peter et al., 2023). Even on days when the particulate matter levels are low, vulnerable groups like children and women can still be affected. Particulate matter can lead to a variety of health problems, including breathing and heart issues, cancer, high blood pressure, and problems with thinking and memory. Thus, some of the effects will be discussed in the next sub-subchapter.

2.3.1 The Effects on Children

Children are among those who are affected by the adverse effects. A study by Abdul Rahman et al. (2017) in Klang Valley hospitals from 2006 to 2010 showed that more than 50,000 cases of acute respiratory infections (ARI) in children were related to air pollutants, especially PM_{10}. PM_{10} increased the risk of children getting acute bronchiolitis, acute upper respiratory infection of multiple and unspecified sites, and unspecified acute lower respiratory infection. More boys than girls were affected and most of the affected were under five years old. In another study near natural gas factories in Bintulu, higher levels of $PM_{2.5}$ were linked to hospital admissions for respiratory issues among children especially those aged between 0-4 years old (Ibrahim et al., 2022). Even when PM_{10} and $PM_{2.5}$ were within WHO guideline limits, children below five years old who were exposed towards PM_{10} and $PM_{2.5}$ were prone to respiratory diseases, developmental disorders, allergies, ear infections, cancer, obesity, and more (Spencer-Hwang et al., 2023).

Furthermore, a place where children are supposed to be safe is not exempted from PM_{10} and $PM_{2.5}$ exposure. It was reported that daycare centers (DCCs) in

Seremban which were located in urban areas and near major roads had poor indoor air quality which caused the children sent there to experience respiratory symptoms like wheezing (Khamal et al., 2019). Other contributing factors that contributed to poor indoor air quality are indoor air ventilation, floor area per child, temperature and types of ventilation. DCCs which used air-conditioning and mechanical ventilation increased the risk of children getting a persistent cough and catching a cold (Malik et al., 2021).

2.3.2 The Effects on Pregnant Mothers

Pregnant mothers who are exposed to PMs can experience adverse effects not only on themselves but also on their babies. Higher exposure of $PM_{2.5}$ to pregnant mothers in Jakarta, Indonesia, caused their babies to be slightly shorter when they were born (Soesanti et al., 2023). This is supported by another study in Durban, South Africa, that found the babies to be smaller than expected for their age (Mitku et al., 2023). The babies are also at risk of having preterm birth. A review study by Song et al. (2023) found that when pregnant women were exposed to higher levels of $PM_{2.5}$, there was a higher chance that their babies had certain problems. These problems included having babies with low birth weight, being born too early, stillbirth, being smaller than expected for their age, and having birth defects. The pregnant mothers also were at a higher risk of having health issues like high blood pressure during pregnancy, diabetes during pregnancy, and a serious condition called preeclampsia.

To understand more about the adverse effects on pregnant mothers, a study was done on mice. The study looked at how $PM_{2.5}$ exposure during pregnancy could affect the brain and behaviour of baby mice. Male baby mice had trouble with learning and memory while female baby mice did not have the same problems. The brains of the male baby mice showed changes in genes related to brain development and their brain cells did not grow properly (Hou et al., 2023).

2.3.3 The Effects on Workers

Those who work outside and near PM sources are at higher risk of getting adverse effects. A study on Malaysian traffic policemen in Kuala Lumpur and Johor Bahru showed that they experienced coughing issues and deterioration in their lung functions (Jamil et al., 2019; Putri Anis Syahira et al., 2020). A new study on outdoor workers in Kuala Lumpur, Malaysia, found that exposure to higher $PM_{2.5}$ changed the appetite and increased the calorie intake. This could increase the risk of having metabolic syndromes (Sundram et al., 2022).

2.3.4 The Effects on Diabetic Patients

People with diabetic problems are at higher risk of getting adverse effects. From 2006 to 2015, it was reported that the increment in pollutant levels in Malaysia increased the rate of diabetes by 9.8% (Wong et al., 2020). This is further proven by a study on mice. Mice with type 1 diabetes (T1D) and diet-induced obesity (DIO) were exposed to $PM_{2.5}$ at a mean concentration of 95.77 $\mu g/m^3$ for four weeks in Shijiazhuang City, China. Both groups of mice showed signs of lung inflammation and damage. T1D mice were more severely affected in their lung with changes in markers related to inflammation, oxidative stress and tissue repair (Chen S. et al., 2023).

2.3.5 The Effects on Lung and Heart

Short-term and long-term exposure to PM has different adverse effects. A review study by Wan Mahiyuddin et al. (2023) in countries with higher income like Eastern Asia, Europe and North America found that $PM_{2.5}$ was more strongly linked to respiratory problems than heart problems, especially in studies in less-than-seven-year period. For studies with more than seven years period, heart problems became more of a concern. This is further proven by a study by Sun et al. (2023) in China. Over a million people aged above 35 years old from 13 different check-up centres that were long-term exposed to PM_{10} and $PM_{2.5}$ were linked to a heart condition called atrial fibrillation (AF). AF is a type of irregular heartbeat that can lead to heart problems. The connection was stronger in women, younger people, and those with health conditions like high blood pressure and diabetes. A substance in $PM_{2.5}$ called sulfate had the biggest impact on increasing the risk of AF.

2.3.6 The Effects on Brain and Others

Long exposure to PM can affect the brain and cause cancer. A study by Ma et al. (2023) on thirty thousand individuals in China who had a ten-year follow-up period showed an increase in brain impairment. High $PM_{2.5}$ exposure was also linked to higher levels of amyloid proteins in the cerebrospinal fluid, which are associated with Alzheimer's disease. Moreover, the decline in cognitive function related to $PM_{2.5}$ exposure appeared to be partly explained by the accumulation of amyloid proteins in the brain. Another study by Lequy et al. (2023) showed that there was an increased risk of cancer in 12,000 individuals in France who were exposed to $PM_{2.5}$ for twenty-two period. About 2,000 cases of cancer like bladder, lung, breast and prostate cancers were found.

Exposure to PM also affects reproductive systems. A study by Y. Zhang et al. (2024) on 200 Chinese adult men found that their semen motility was reduced. The exposure caused alterations in D-aspartate levels. It was revealed that D-aspartate might affect both total and progressive motility during the spermatogenesis period.

All these adverse effects are from the exposure to PM_{10} and $PM_{2.5}$. To understand more about particulate matter, the next subchapter will discuss the trends of particulate matter.

2.4 Trends of Particulate Matter

The concentrations of particulate matter fluctuated over a year. The trend of the PM's concentrations is affected by the change of seasons. In the winter season, PM_{10} and $PM_{2.5}$ concentrations were higher than in the summer. A study by Lv et al. (2019) showed that all three urban-category places: Beijing, China; Suning, China; and Islamabad, Pakistan had higher PM_{10} and $PM_{2.5}$ concentrations during winter than summer. In Urumqi, China, particulate matter concentrations increased significantly in winter, with $PM_{2.5}$ being much more abundant than PM_{10} (Meng et al., 2019). In spring, the concentration of particulate in the air decreases because of air coming in from the desert, making the particles coarser. In summer, the flow of clean air, driven by wind speed, affected the concentration of these particles. In autumn, the PM concentration was greater compared to summer.

Another study by Chauhan et al. (2022) in Varanasi, India, which was categorised as semi-urban, showed that fine particles dominated the air in winter, while in summer, larger particles were more common. The ratio between different-sized particles ($PM_{2.5}/PM_{10}$) was lowest in summer and highest in winter. Winter had higher particle concentrations, likely due to stable atmospheric conditions and more biomass burning. When the atmosphere is stable, it tends to trap pollutants, making their concentrations higher and the air quality worse (Yavuz, 2023).

Furthermore, the PM_{10} concentrations at a particular place increase during the dry season (Redzuan et al., 2023). In, Malaysia, the dry season which took place between June and September, combined with the slow wind conditions during the southwest monsoon contributed to the elevation of PM_{10} concentrations. This caused the three industrial-category monitoring stations—Petaling Jaya, Bukit Rambai, and Pasir Gudang—to experience an increase in PM_{10} concentrations during this period.

In addition, other categories of monitoring stations are also affected by different periods of seasons. Naidin et al. (2024) reported that in a study of PM_{10} concentrations from 13 monitoring stations across Malaysian Borneo—comprising 1 industrial, 3 urban, 8 suburban, and 1 rural station—conducted between 2006 and 2016, higher particulate matter concentrations were observed during the southwest monsoon (June to September), while lower concentrations were recorded during the northeast monsoon (December to March). The southwest monsoon is known as a dry season while the northeast monsoon is known as a wet season due to the high rainfall events.

The same trend was observed in Peninsular Malaysia. A study by Zheng et al. (2023) on PM_{10} from 25 monitoring stations, covering all categories—industrial, urban, suburban, and background—over 20 years (2000–2019), showed that the first peak in PM_{10} concentrations occurred in February or March. The PM_{10} concentration reached its highest in September during the southwest monsoon where the wind speed was relatively low. During intermonsoon (April and October), the wind was mild causing