STATISTICAL AIR QUALITY MONITORING OF KLANG VALLEY DURING THE COVID-19 MOVEMENT CONTROL ORDER IN MALAYSIA

SITI AISHAH BINTI MOHD ZAHIDI

RESEARCH REPORT SUBMITTED TO THE FACULTY OF BUSINESS AND ECONOMICS UNIVERSITY MALAYA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED STATISTICS

2022

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: *Siti Aishah binti Mohd Zahidi*

Registration/Matric Number: *17122851*

Name of Degree: *Master of Applied Statistics*

Title of Project Paper / Research Report / Dissertation / Thesis: *Statistical Air Quality Monitoring of Klang Valley During The COVID-19 Movement Control Order In Malaysia*

Field of Study: *Quality Control*

I do solemnly and sincerely declare that:

(1) I am sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exist was done by way of fair dealing and for permitted purposes and any excerpt or extract from or references to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship has been acknowledged in this work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any action as may be determined by UM.

Candidate's Signature: Date: 17/06/2022

Subscribed and solemnly declared before,

Witness's Signature : Date: 17/06/2022
Name : MUZALWANA ABDUL TALIB
Designation : SENIOR LECTURER

ABSTRACT

Air quality is a very important indicator as it affects people and the environment. The pandemic of COVID-19 has caused Movement Control Orders (MCO) to be implemented to curb the number of cases in Malaysia. Hence, all economic activities have been halted except for essential services and working from home (WFH) has been implemented during the pandemic. Thus, this study aims to investigate significant differences in air quality before and after lockdown. Based on the non-parametric test, all air quality levels show significant differences except for SO_2 levels for Cheras and Shah Alam. By hierarchical clustering, the Klang area has the highest levels of particulate matter and SO_2. Meanwhile, Petaling Jaya has the highest levels of NO_2 and CO as in the same cluster for both periods while Batu Muda faced deteriorating CO levels after MCO. In addition, O_3 levels have improved for Cheras, Putrajaya, Banting and Shah Alam. However, ozone levels declined in Petaling Jaya after MCO. The air quality is further monitored by using a multivariate control chart approach in which it is found out that the air quality is out of control before and after the implementation of MCO. It can be concluded that in the Klang Valley area, some businesses are still operating during MCO and after MCO uplifted, almost all businesses are operating as usual to uplift the economy and result in deteriorated air quality compared to MCO periods.

ABSTRAK

Kualiti udara adalah aras penunjuk yang sangat penting kerana ia mempunyai kesan kepada manusia dan alam sekitar. Pandemik COVID-19 telah menyebabkan Perintah Kawalan Pergerakan (PKP) dilaksanakan bagi membendung bilangan kes di Malaysia. Oleh itu, semua aktiviti ekonomi telah dihentikan kecuali perkhidmatan penting dan bekerja dari rumah (WFH) telah dilaksanakan semasa pandemik. Oleh itu, kajian ini bertujuan untuk menyiasat perbezaan kualiti udara sebelum dan selepas perintah berkurung. Berdasarkan ujian tidak berparametrik, semua kualiti udara menunjukkan perbezaan signifikan kecuali tahap SO_2 di Cheras dan Shah Alam. Dengan pengelompokan hierarki, Klang mempunyai tahap zarah halus dan SO_2 yang paling tinggi. Sementara itu, Petaling Jaya mempunyai tahap tertinggi NO_2, dan CO kerana berada dalam kelompok yang sama untuk kedua-dua tempoh manakala tahap CO di Batu Muda merosot selepas PKP. Tambahan pula, tahap O_3 bertambah baik untuk Cheras, Putrajaya, Banting dan Shah Alam. Tetapi, tahap ozon Petaling Jaya menurun selepas PKP. Kualiti udara dipantau selanjutnya dengan menggunakan pendekatan carta kawalan multivariat dimana mendapati kualiti udara masih di luar kawalan selepas pelaksanaan PKP. Ini boleh disimpulkan bahawa di kawasan Lembah Klang, beberapa aktiviti perniagaan masih beroperasi semasa tempoh PKP dan selepas PKP dimansuhkan, hampir semua aktiviti perniagaan beroperasi seperti biasa untuk meningkatkan tahap ekonomi dan mengakibatkan kualiti udara merosot berbanding tempoh PKP.

ACKNOWLEDGEMENT

Thanks goes to my supervisor, Dr Muzalwana for her advice and support. I also feel thankful to all the lecturers for the knowledge. I also wish to express my gratitude towards my mother and sister for their encouragement and moral support. In addition, I feel extremely thankful towards the friends that I have met through my postgraduate studies. I also would like to express my gratitude to the Department of Environment for the air quality data that enabled me to conduct the research project. Finally, I would like to thank everyone who has helped me grow throughout my postgraduate experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

API	Air Pollution Index
MEWMA	Multivariate Exponentially Weighted Moving Average
MCO	Movement Control Order
RMCO	Recovery Movement Control Order
CMCO	Conditional Movement Control Order
UCL	Upper Control Limit
CL	Control Limit
LCL	Lower Control Limit
PM_{10}	Particulate Matter with Diameter < 10 Microns
$PM_{2.5}$	Particulate Matter with Diameter < 2.5 Microns
SO_2	Sulphur Dioxide
NO_2	Nitrogen Dioxide
O_3	Ground Level Ozone
CO	Carbon Monoxide

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

In this chapter, background of study, problem statement, objectives and scope of study are discussed. The background of the study is elaborated in terms of air quality and the statistical approach to monitoring air quality in Malaysia. The problem statement discusses the main objectives that led to this study. The objectives are specified that are required to be achieved throughout this study.

1.1 Background

Increases in the economic activities in Malaysia especially in the manufacturing, automotive, agricultural and construction industries have positively contributed to the economy. The Malaysian economic plan has contributed to the country's development and Malaysia is one of the leading exporters in the world. However, despite the flourishing Malaysian economy, there are negative consequences resulting from economic activities which is pollution that is harmful to the people, ecosystem and environment. One of the pollutions that has been a problem nowadays is air pollution. Air pollution is a global environmental problem that has influenced the urban population's health (Kanchan et al., 2015). With the rapid evolution of people's lives, the air quality has deteriorated and is impacting on the environment as well as people's health. An example of occurrences due to bad pollution levels is global warming and acid rain. Rapid weather changes are commonly caused by pollution such as air pollution, water pollution, plastic pollution, thermal pollution and radioactive contamination.

The outbreak of pandemic COVID-19 that started in late 2019 has affected all aspects worldwide and there is no exception to anyone. Thus, Malaysia is also one of the countries that cannot escape the impact of the pandemic. The first confirmed COVID-19 case in Malaysia was reported on 25th January 2020 involving three tourists from China that came in from Singapore through Johor on 23rd January 2020 (Berita Harian, 2020). The number of cases has kept increasing since March 2020 with several clusters with high positive cases (Rahman et al, 2021). With a high infection rate, especially with the emergence of Sri Petaling's cluster (Berita Harian,2020), the government decided to impose a Movement Control Order (MCO) to curb infection rates especially in crowded areas and industrial areas such as Klang, Kuala Lumpur and Johor.

Malaysia has three phases of MCO of which the first phase of MCO is from 18th March 2020 until 3rd May 2020. Succeeded by MCO, the conditional movement control order (CMCO) was implemented on 4th May 2020 until 9th June 2020. The Recovery Movement Control Order (RMCO) was implemented on 10th June 2020 until 31st March 2021. Since then, the cases have decreased and if there is any surge of COVID-19 cases reported, MCO is implemented depending on the number of cases in each state. The MCO phases in Malaysia are summarized in Table 1.1.

Table 1.1: Summary of MCO implementation in Malaysia

Stage of MCO	Phase of MCO	Duration
Movement Control Order (MCO)	Phase 1	18th March 2020-31st March 2020
	Phase 2	1st April 2020-14th April 2020
	Phase 3	15th April 2020 – 28th April 2020
	Phase 4	29th April 2020-3rd May 2020
Conditional Movement Control Order (CMCO)	Phase 1	4th May 2020-12th May 2020
	Phase 2	13th May 2020-9th June 2020
Recovery Movement Control Order (RMCO)	Phase 1	10th June 2020-31st August 2020
	Phase 2	1st September 2020-31st December 2020
	Phase 3	1st January 2021-31st March 2021

During the implementation of MCO, several restrictions have been imposed such as the prohibition of mass gatherings which include social, sports, religious and cultural activities. Furthermore, all schools and higher education institutes are closed. All industries are also closed except for essential industries such as finance, health and food supply. Most employees also work from home (WFH) with a restriction on the number of employees that can come to the office (Berita Harian, 2020). Along with the implementation of MCO, the pollution level in Malaysia such as water level and air quality in Malaysia has been changed drastically. Therefore, the environment has improved as a result of the cessation of human and economic activities.

1.2 Air Quality in Malaysia

Air quality has become a very important issue in Malaysia because it has worsened especially in areas where major industrial and cramped residential areas are located. Since air pollution comes from lots of sources, it has become a concerning issue since it has contributed to bad impacts on health such as asthma, cancer, bronchitis and cardiovascular diseases. The sources of air pollution come from industries, construction activities, motor vehicles, power generation, land clearing, open burning and forest fires (Department of Statistics, 2016).

Air quality in Malaysia is measured by the air pollution index (API). API is an indicator for air quality status (Department of Environment, 1997). API is calculated by six major air pollutants which are Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Particulate Matter with Diameter < 10 Microns (PM_{10}) , Particulate Matter with Diameter < 2.5 Microns ($PM_{2.5}$) and Ground Level Ozone (O_3). Malaysia has 68 stations for monitoring air quality nationwide. Air pollution is categorized by six different levels referring to the severity of air pollution which is summarized in Table 1.2.

Table 1.2: Level of air pollution index

Indicator	Status	API	Health Effect
	Good	0-50	Low pollution level without any bad effect on human health particularly.
	Moderate	51-100	Moderate pollution level that affecting on human health.
	Unhealthy	101-200	Worsen the health conditions of high-risk people especially persons with heart and lung implication.
	Very unhealthy	201-300	Very high risk for people's health.
	Hazardous	300-500	Dangerous level for people.
	Emergency	Above 500	Severe aggravations of symptoms and endangers public human health.

(Source: Department of Environment, 2000)

There are only several occasions where Malaysia experiences hazardous and emergency levels except for the impact of haze transboundary from neighboring countries and phenomena such as El Nino. The recent haze transboundary that impacted Malaysia occurred in September 2019 due to open forest burning in Sumatera and Kalimantan. 2469 schools were closed in Selangor, Putrajaya, Kuala Lumpur, Penang and Sarawak due to a surging number of asthma and conjunctivitis cases (Utusan Borneo, 2019). In 2015, Malaysia experienced the worst haze and was declared an emergency when the pollution levels were beyond safe levels for people. The schools were closed, and some main national events were cancelled (Utusan Borneo, 2019) for example the Kuala Lumpur Marathon Event and the 2015 FINA Swimming World Cup in Singapore were cancelled due to high API levels (Utusan Borneo,2019).

1.3 Monitoring air quality in Malaysia

Air pollution requires continuous monitoring as the monitoring is a measure to keep track of the ambient air pollution levels within an area. If the air pollution is at safe levels, it means that the air pollution is within a safe level for people. However, if the level of air pollution is exceeding the safe level, it means the air pollution level is hazardous and causes a bad impact on both the environment and people. The air pollution in Malaysia is monitored by the Air Pollution Index Management System (APIMS) and warned if there are areas that have unhealthy or very unhealthy air pollution within any station. Figure 1.1 is an example of air quality monitoring which was taken on 3rd April 2022 at 2pm. 36 stations have good pollution level whereas 28 stations have moderate pollution levels. However, there are four stations where the API is not available.

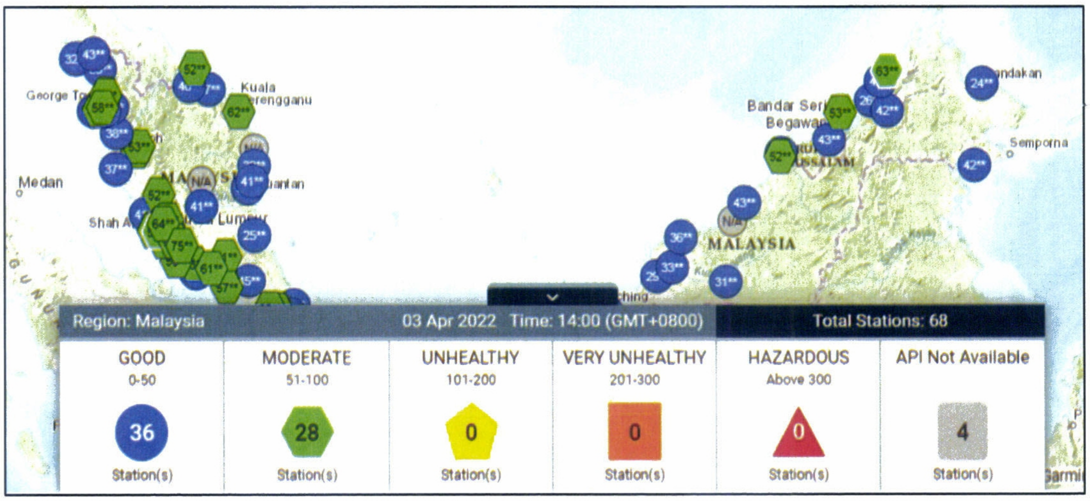


Figure 1.1: API monitoring on 3^{rd} April 2022 at 2pm
(Sources: Department of Environment, 2022)

Statistical approaches can be implemented for monitoring the air pollution levels in Malaysia. Time series analysis, modelling, clustering, control charts and multivariate analysis can be applied to monitor pollution levels in Malaysia. Descriptive statistics can provide insight information into the minimum, maximum and average of air quality level. Furthermore, descriptive statistics can provide knowledge of the air quality variation. Besides that, descriptive statistics can highlight the potential relationship between air quality. Clustering is a method that groups datasets that have similar characteristics and is frequently implemented in monitoring air quality. By the clustering technique, a comparison can be made between each group and how the dataset is grouped together.

Control charts are one of the statistical approaches that are used to monitor quality. The control chart consists of an upper control limit (UCL), central line for average (CL) and a lower control limit (LCL) to monitor quality. If any of the data falls beyond UCL and lower than LCL, the process is said to be out of control. The out of control process will be taken for further study to see whether it is occurring due to common causes or special causes. For datasets with more than one factor influencing air quality, univariate control charts can be extended to multivariate control charts.

1.4 Problem Statement

Air pollution in Malaysia has become a major issue as it affects people's health and the environment. Globally, air quality has also deteriorated over time. The health problems caused by bad air quality such as asthma, breathing difficulties and headache become common with bad air quality especially in industrial areas and urban areas where massive factories and transportation become sources of pollution. Shah Alam, Petaling Jaya and Klang are identified as areas with worse air quality due to rapid industrial activities. As a result, monitoring air quality in Klang Valley is necessary in order to determine the severity of air quality levels.

This study can alert and identify which air quality concentrations that can be harmful to peoples that stay and work in the Klang Valley. There are not many studies carried out, especially, on the use of statistical quality control and clustering techniques to identify the areas with high readings for all air pollutants. In order to analyze the air quality data in Klang Valley, it is essential to handle the data carefully to be able to extract the knowledge within the dataset.

The pandemic of COVID-19 has hit the entire nation and impacted all aspects of people's lives. The lockdown has been imposed to prevent the virus spreading hence all human activities have been stopped. The halt of human activities had a positive impact on the environment and people's health. Thus, by conducting hypothesis testing on the data, this study will determine whether there is a decrease in air quality in each area of Klang Valley.

The air quality in each area will be identified by using a clustering method to group the air quality levels of each area. The control chart technique is implemented for data that is divided before and after the implementation of MCO to investigate whether air quality is statistically in control by employing the Multivariate Exponentially Weighted Moving Average (MEWMA) control charting techniques. The study will be able to find out whether the air quality in Klang Valley has significantly reduced after the implementation of MCO. Many areas in the Klang Valley are known to have large number of infectious COVID-19 cases

. 1.5 Research Objective

The study aims to design a multivariate exponentially weighted moving average (MEWMA) control chart to monitor the air quality. The objectives of this study are;

i. To determine the differences of air quality level before and after pandemic COVID-19 for all air pollutants.
ii. To group the air monitoring stations accordingly based on the severity of air quality level.
iii. To construct a multivariate control chart in assessing the air quality level in Klang Valley
iv. To compare air quality of different areas in Klang Valley before and after the implementation of MCO.

1.6 Research questions

The research question of this study are;

i. Is there any difference in the air quality level in Klang Valley before and after implementation of the MCO?
ii. Which air monitoring stations in Klang Valley have similar severity of air quality level before and after the implementation of MCO?
iii. How to construct a multivariate control chart be to assess the air quality level in Klang Valley?
iv. Was there any difference in the air quality level before and after the implementation of MCO in Klang Valley?

1.7 Significance of study

There are a few studies conducted to investigate the air quality in Malaysia. This study will explore knowledge and information about air quality in the Klang Valley. The findings throughout this study can be basic references for further study of the air quality of Malaysia in many analysis fields such as time series, modelling and quality control fields. There are not many studies conducted to analyze data that is segregated into different periods before the implementation of MCO and after the implementation of MCO. This study will investigate whether the air quality improved after the implementation of MCO.

1.8 Research design

The scope of the study is divided into two sections which are the data on air quality and the methods used in this study. The air pollution data is secondary data which was requested from Department of Environment for Klang Valley. Six air quality levels of PM_{10}, $PM_{2.5}$,

CO, O_3, SO_2 and NO_2 of eight stations in Klang Valley data are given for this study. Furthermore, statistical quality control will be used to monitor and investigate the air quality before and prior to the pandemic COVID-19. The data is provided by hourly from 1st January 2019 until 31st December 2020 for two years which is divided into two different periods. The first periods start from 1st January 2019 until 17th March 2020 whereas the second periods start from 18th March 2022 until 31st December 2020 which is after the implementation of MCO. The missing data is treated using the mean imputations method.

In this study, the data was sorted using Microsoft Excel before further analysis. The eight stations chosen are monitored by using a non-parametric test on the significance between two periods. The data is further analyzed to group the air quality levels based on their severity groups by the hierarchical clustering method. The data is monitored by using a control chart technique of MEWMA by developing the first phase and further applying it to monitor the air quality level during the second phase of the control chart to detect the shift in the air quality levels.

1.9 Summary

In a nutshell, this chapter can be divided into five chapters. The first chapter is to provide a brief overview of air quality in Malaysia and discuss the purpose of this study such as problem statement, research objective, research questions, significance of study and research design. Then, in chapter two, which will discuss the literature review on both methods and dataset. Furthermore, chapter 3 will discuss the methodology for this study and the dataset selected. Chapter 4 will explore the results and discussion of the study conducted. The last chapter will be discussing the summary and conclusion based on the findings of this study.

CHAPTER 2

LITERATURE REVIEW

In chapter 1, air quality is briefly discussed in general along with the sources of air pollution in Malaysia. Throughout chapter 2, new knowledge of air quality is further discussed with the methodology used in this study. The brief overviews help to identify and grasp the idea of analyzing the data by using the statistical monitoring method.

2.1 Air Quality in Malaysia

Air pollution index (API) is an indicator to alert the severity level of air pollution to public in specific region and time occurrences (Kanchan et al., 2015). API is determined based on six pollution parameters of PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO and O_3 (Mohamed et al, 2016). The highest concentration of air pollutants will be determinant of API (Department of Environment, 1997). Normally, the highest air quality is $PM_{2.5}$ and become a determinant of API (Department of Environment, 1997). The reporting system of the API is very crucial as a tool of communication to people (Wong & Wai, 2012).

Since Malaysia's main goal is to achieve higher industrial economic growth, urban planning and large scale economic activity have impacted the environment through industrial pollution and the degradation urban environments (Amran et al., 2015). The emission of air pollutants and air quality influenced by natural factors such as climate and topography (Baek & Ban, 2020). Moreover, air pollutant also come from industry and vehicles (Baek & Ban, 2020).

Socioeconomic has become another factor that has impacted air quality where cramped residential areas and numerous industrial areas have worsened air quality. Several severe haze events which are caused by widespread biomass and open forest burning have also become one of the reasons for deteriorating air quality (Lee et al., 2018). Air pollution is one of the environmental health risks and particulate matter is the main air quality indicator that are most harmful pollutants (Beloconi & Vounatsou, 2021). Thus, it can be concluded that the higher the air quality parameters, the higher the air pollution index along together with consequences for humans and the environment.

Particle pollution which is also referred as particulate matter is a tiny particle of mixture of solid and liquid particles (Adams et al., 2015). Coarse particulate matter with a diameter 10 micrometers (Li et al., 2017) can contributed to higher risk to cardiovascular complication (Palacio et al., 2021) meanwhile fine particulate matter with a diameter of 2.5 micrometer (Li et al., 2017) can penetrate deeper into lungs (Xing et al., 2016). $PM_{2.5}$ are more hazardous since able to go to deep parts of the lungs and damage lung function (Xing et al., 2016). Particulate matter can cause health concerns such as respiratory disease, central nervous system dysfunction and cancer which can penetrate the respiratory system through inhalation (Manisalidis et al., 2020). Particulate matter can come from primary or secondary sources such as forest fires, factories, emissions from vehicles, industrial activities and power plants (Adams et al., 2015). Besides that, Malaysia is also often affected by uncontrolled open burning activities from neighboring countries (Afroz et al, 2003). During haze incident in 1997, PM_{10} is the predominant air pollutant hence the API values reported were primarily based on PM_{10} sub-index (Department of Environment, 1997).

Sulphur dioxide (SO_2) is an atmospheric pollutant that is produced from volcanoes explosions and human activity such as coal and petroleum combustion (Cullis & Hirschler, 1980). Sulfur deposition is harmful to ecosystems and human health which can cause radiative force and climate change (Carmichael et al., 2002). When SO_2 merged with water molecules, it forms sulfuric acid which is the main component of acid rain, a known contributor to deforestation (Yadav et al., 2022). In terms of health concern, SO_2 can cause bronchoconstriction, weaken the cardiovascular system (Tunnicliffe et al., 2001) lung damage, eye irritation and worsen asthma (Sensuła, 2016).

Nitrogen dioxide (NO_2) is a reddish-brown in color, a strong oxidant and soluble in water (WHO, 2000) pollutants. NO_2 are one of indicators for measuring air pollution that comes from vehicles and high temperature combustion such as volatile organic compounds or polycyclic aromatic hydrocarbons (Brook et al., 2007). Nitrogen oxide (NO) can be converted to NO_2 under many chemical reactions such as oxidation by ozone (Kimbrough et al., 2017). NO_2 can damage human respiratory tract (Chen et al., 2007) and increase human vulnerability to respiratory infections and asthma especially to children (Shima & Adachi, 2000). Moreover, high concentrations of ambient NO_2 can cause acute pulmonary edema and death with long-term exposure to NO_2 (Horvath, 1980).

Ozone is formed from the reaction dioxygen and oxygen in the presence of a third molecule body which is able to absorb the heat of the reaction (Zhang et al., 2019). Ozone exposure is linked to many negative health effects such as premature mortality (Anenberg et al., 2009), respiratory diseases and deteriorated lung function (Zhang et al., 2019). Besides that, ozone air pollution is a threat to global crop production that can threaten the world's food supply chain due to its warming impacts on crops (Tai & Val, 2017).

Carbon monoxide (CO) is a colorless, tasteless and odorless gas that is produced by incomplete combustion of carbon compounds that come from fire, faulty furnaces and incomplete combustion (Rose et al., 2017). CO poisoning can be a factor in deliberate or accidental death (Balzan et al., 1996). Since CO can binds to red blood cells with 200-fold greater than oxygen (Kester et al., 2012), CO can cause acute and chronic central nervous system by difficulty for oxygen transport (Balzan et al., 1996). CO poisoning can also cause headaches, dizziness, chest pain and confusion (Miller, 2011).

2.2 Non-Parametric Test of Wilcoxon Rank Sum Test

A parametric test requires strong assumptions for the data distribution whereby the data needs to follow a normal distribution (Scheff, 2016). Tests on data normality are very crucial and should be performed before proceeding with other tests for data analysis (Mishra et al., 2019). Checking the normality assumptions is very crucial to decide whether parametric or non-parametric techniques need to be used on the dataset (Orcan, 2020). Parametric tests have assumptions that the dataset must meet in order for test to be applied (Harris et al., 2008). Therefore, if the data does not meet the normal distribution assumptions, non-parametric test can be approached as an alternative to a parametric test. Non-parametric is used are when the assumptions of parametric test are violated, analyze nominal or ordinal data and data come with small sample size (Harris et al., 2008).

Kolmogorov-Smirnov test is a goodness of fit which is an alternative of non-parametric test that are used to decide if a data set are follows a specific distribution (Aslam, 2020). Two-sample Kolmogorov-Smirnov test are used to check on the distribution of two datasets (Teegavarapu, 2019). The data are always assumed to follow a statistical distribution however in real time application, it not always necessary for the data to follow normal distribution (Aslam, 2020).

If the p-value obtained by the Kolmogorov-Smirnov test is more than α, then null hypothesis is not rejected at level of significance, α (Razali & Yap, 2011). The normality of the data can be tested by using Shapiro-Wilk which is appropriate for a small sample size of < 50 meanwhile Kolmogorov-Smirnov test is more appropriate for data with $n \geq 50$ (Mishra et al., 2019).

The advantage of non-parametric test is can be applied in all data that does not obey normal distribution. Furthermore, non-parametric data can be applied to qualitative data including nominal and ordinal data (Scheff, 2016). Although the parametric and semiparametric approaches are preferred for simplicity and easy interpretation (Cheda et al., 2020), however, non-parametric tests are always chosen on data that doesn't follow the normality assumptions.

The Wilcoxon signed rank sum test is a non-parametric approach to test the median of a distribution that acts as an alternative to the one sample t-test and the paired t-test (Oyeka & Ebuh, 2012). These groups are assumed to be independent from each other (Shan et al., 2014). The Wilcoxon signed-rank test is a non-parametric technique that is applied to determine if the response variable sample has the same distribution (Refugio, 2018).

The Wilcoxon rank sum test is applied to compare data where data should be at least ordinal data (Kim, 2014). By the Wilcoxon rank sum test, the magnitude of differences are determine by test result where if the p-value is less than the level of significance, null hypothesis are rejected (Imam et al, 2014). The null hypothesis tested is that there are no differences between the data determined by the test conducted (Kim, 2014). Most of tests select a fixed significance level of 5% or 0.05 for hypothesis testing (Imam et al, 2014).

2.3 Hierarchical Agglomerative Clustering Analysis

Clustering is a method that is performed on big data to get insight on a dataset (Nerurkar et al., 2018). There are many methods of agglomerative hierarchical clustering that can be used on different types of data (Erman et al., 2015). Cluster analysis is a technique used to classify objects or datasets into homogeneous groups that are called as clusters (Maholtra, 2019). A cluster is defined as a set of objects that have a higher degree of similarity to each other compared to other objects that belong in the same cluster (Nerurkar et al., 2018). The clustering technique will group or segmenting the dataset (Erman et al., 2015) into subsets or clusters that are meaningful and useful (Jatain et al., 2013).

Clustering can allows to discover patterns in the dataset (Jatain et al., 2013). Clustering is also one of the data mining approaches that groups the data based on similarities (Karthikeyan et al, 2020). There are two types of hierarchical clustering which are agglomerative where adopts bottom up approach by each observation start its own cluster and cluster pairs are merged as one moves up the hierarchy (Sasirekha & Baby, 2013). Another one type of hierarchical clustering is divisive where using top down approach where all observation are start in one clusters and split as one moves down the hierarchy (Sasirekha & Baby, 2013).

A dendrogram is a visual tree graph that are visualize the cluster and very useful to determine when the clustering should be stopped until the data are in one large cluster (Yim

& Ramdeen, 2015). Classification are also visualize by hierarchical trees were also called as dendrogram (Saraçli et al., 2013). Dendrogram is an illustration of data with the same elements with the most relative position based on two points (Ridwan & Retnawati, 2021). Hence, the further the data, the less similar and higher differences can be concluded between the data (Ridwan & Retnawati, 2021). Hierarchical clustering has several advantages such as do not require number of clusters as initial parameters, can be applied to various type of data and capable to handle distance measure and similarity on the data applied (Kousiga, & Vadivu, 2019). Despite that, hierarchical clustering is very sensitive to goof initialization and coincident cluster may be formed (Sasirekha & Baby, 2013).

2.4 Multivariate Exponential Weighted Moving Average (MEWMA) Control Chart

To monitor a process, appropriate statistical tools need to be used that are capable of showing quality behavior over time (Mbaye et al., 2021). Control chart is an important tool to detect an out of control process (Mahmoud & Maravelakis, 2010). Nowadays, there is a higher interest in multivariate control chart due to many of processes are related to several quality characteristics which are correlated with each other (Mahmoud & Maravelakis, 2010). A control chart can be implemented where the performances characteristics of the system can be improved and monitored hence the quality can be enhanced (Wu & Shamsuzzaman, 2005). Furthermore, control charts can be beneficial in terms interpretability in decision making and simplify the communication of the dataset (Inkelas et al., 2021).

The Multivariate Exponential Weighted Moving Average (MEWMA) control chart was introduced by Lowry et al. (1992) as an extension of the univariate Exponential Weighted Moving Average (EWMA) control chart. In real-life situations, simultaneous monitoring of quality is necessary where the sample may have multivariate variables that influence quality (Montgomery, 2013). The MEWMA control chart is a multivariate control chart that is used to detect autocorrelation and detect mean vector shift (Ramadhani et al., 2019).

The MEWMA control chart is one of the best control charts to detect small shifts in process parameters and monitor correlated quality characteristics among other multivariate control charts (Ershadi et al., 2014). The MEWMA control chart is preferred due minimum cost and high statistical properties in handling multivariate data (Ershadi et al., 2014). By using Hotelling T^2 to estimate the parameters in Phase 1, MEWMA control chart can be used to monitor process variability in Phase 2. By using both control charts, maintenance costs are able to be reduced and maximize equipment performance is able to be achieved as well with online condition monitoring (Lampreia et al., 2015).

2.5 Summary

This chapter has been discussing the literature review on air quality and the methodologies of non-parametric tests, hierarchical clustering and control chart. The next chapter discusses the methodologies used in this study which are the non-parametric test of Wilcoxon rank sum test, hierarchical clustering and multivariate control chart.

CHAPTER 3

METHODOLOGY

This chapter discussed the methods implemented in this study. Besides that, the procedures of each method are explained in detail and summarized in a flowchart in subchapter 3.3. This chapter is divided into two parts which are data selection and methodologies applied. This chapter concludes with the summary of statistical analysis to be employed based on the research objectives of this study.

3.1 Selection of air monitoring station data

In Malaysia, there are 68 continuous air quality monitoring stations. Only eight stations in Klang Valley, including Kuala Lumpur, Selangor, and Putrajaya were chosen as study areas for this study. Thus, data is requested from January 2019 until December 2020 for further studies. The air quality data is continuously monitored hourly by the Department of Environment every day. The data is requested to the Department of Environment from January 2019 until December 2020. The air quality data is requested for six major air pollutants which are Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Particulate Matter with Diameter < 10 Microns (PM_{10}), Particulate Matter with

Diameter < 2.5 Microns ($PM_{2.5}$) and Ground Level Ozone (O_3). Thus, only eight air monitoring stations are considered in this study as per Table 3.1.

Table 3.1: Location of the air monitoring stations

State	Locations of monitoring stations
Kuala Lumpur	Batu Muda
Kuala Lumpur	Cheras
Putrajaya	Putrajaya
Selangor	Petaling Jaya
Selangor	Kuala Selangor
Selangor	Shah Alam
Selangor	Klang
Selangor	Banting

3.2 Methodologies of study

The methodology used for air quality monitoring of Klang Valley during the COVID-19 movement control order in Malaysia is a non-parametric test of the Wilcoxon rank sum test. The Kolmogorov-Smirnov test will be conducted to determine the data normality before proceeding to the Wilcoxon rank sum test. The second method used in this study is hierarchical agglomerative clustering analysis followed by a multivariate exponential weighted moving average (MEWMA) control chart.

3.2.1 Wilcoxon Rank Sum Test

Kolmogorov-Smirnov test are conducted first before proceeded to Wilcoxon rank sum test non-parametric test to determine the distribution of the data whether the data are met normal distribution. Kolmogorov-Smirnov test compares air quality data with a known distribution. The Kolmogorov-Smirnov test is given by where X_i are independent and identically distribution (IID) observations of a continuous dataset. The empirical cumulative distribution function, F_n is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(X_i)$$

where $I_{(-\infty,x)}(X_i)$ is the indicator function which $I_{(-\infty,x)}(X_i) = 1$ if $X_i \leq x$; otherwise equal to 0

The Kolmogorov-Smirnov methods are chosen as the sample size for the data is large where $n \geq 50$. The advantage of using Kolmogorov-Smirnov test is that there is no restriction on sample size whereby can be applied to small sample sizes. Furthermore, the Kolmogorov-Smirnov test is a distribution free test and can be applied without knowing underlying population distribution.

The hypothesis testing for the Kolmogorov-Smirnov test are

H_o = The data meet normal distribution assumptions.

H_1 = The data does not meet normal distribution assumptions.

The null hypothesis is rejected if the p-value is less than the significance value, α = 0.05 of 5% alpha level. Meanwhile, the null hypothesis is accepted if the p-values are larger than $\alpha = 0.05$ indicated that the data are normally distributed data. Therefore, if $p < 0.05$, the air quality data are concluded not obeyed a normal distribution.

The Wilcoxon rank sum test is a non-parametric method that is used to analyze the air quality data as the normal probability distribution cannot be met which are conducted and proved by the Kolmogorov-Smirnov test. Wilcoxon rank sum test is an alternative to the two-sample t- test. Wilcoxon ran sum test compare distribution of air quality data by comparing the significance differences in air quality during MCO implementation. The Wilcoxon rank sum test are given by;

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

where

w_1 = sum of ranks corresponding to n_1 observations

w_2 = sum of ranks corresponding to n_2 observations

The hypothesis testing for the Wilcoxon rank sum test are

H_o = There is no significant difference of air quality in Klang Valley before and after implementation of MCO

H_1 = There is significant difference of air quality in Klang Valley before and after implementation of MCO

The null hypothesis will be rejected if the p-value is less than significance value, $\alpha = 0.05$ of 5% alpha level for Wilcoxon rank sum test as evidence that there are significance differences in air quality after the implementation of MCO. Meanwhile, the null hypothesis is accepted if the p-values are larger than $\alpha = 0.05$ which indicates there are no significant differences in air quality after implementation of the MCO.

3.2.2 Hierarchical Agglomerative Clustering Analysis

In this study, cluster analysis is used to classify the air quality data into homogeneous groups that are called as clusters. Clusters are formed by finding similar characteristics in the structure on the air quality data. Objects in the cluster tend to be similar to each other and dissimilar to objects in the other clusters (Maholtra,2010). There are two types of hierarchical clustering which are divisive and agglomerative clustering. Agglomerative clustering is chosen to perform this method on the dataset. Agglomerative clustering worked by following steps (Maholtra,2010);

i. Agglomerative clustering start with each object in a separate cluster.
ii. Cluster are formed by grouping object into larger and larger cluster.
iii. This process is continuing until all objects are members of single cluster.

The type of linkage used to perform the clustering are average linkage. Average linkage is applied by the distance between each pair of objects are added and divided by numbers of pairs to obtain inter-cluster distance. The average linkage is given by

$$d(x,y) = \frac{1}{kl}\sum_{i=1}^{k}\sum_{j=1}^{l} d(x_i, y_j)$$

where *d(x,y)* is the distance between subject observation vector x with observation vector y.

The classification of air monitoring stations can be illustrated by a dendrogram which makes it easy to understand the result, especially used on big data. The dendrogram is a graphical aid to show the clustering results in which the vertical lines represent cluster that are joined together (Maholtra, 2010). The measure of similarity method used in this study is the squared euclidean distance between point a and point b which is given by

$$\|a - b\|_2^2 = \sum_i (a_i - b_i)^2$$

3.2.3 Multivariate Exponential Weighted Moving Average (MEWMA) Control Chart

The exponentially weighted moving average (EMWA) control chart is an alternative to the Shewhart control chart in detecting small shifts in univariate cases (Montgomery, 2013). A control chart consists of an upper control limit (UCL), center line and lower control limit (LCL) to determine whether the air quality in statistical control or out of control. If the air quality data falls below LCL and more than UCL, the air quality is said to be out of control. The EWMA control chart is constructed by plotting Z_i versus the sample number i (Montgomery,2013). The control limit for EWMA control chart are given by

$$\mathrm{UCL} = \mu_0 + \mathrm{L}\sigma \sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i})}$$

$$\mathrm{CL} = \mu_0$$

$$\mathrm{LCL} = \mu_0 - \mathrm{L}\sigma \sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i})}$$

where

μ_0 = starting value of first sample at i=1
L = width of control limits
λ= constant where $0 < \lambda < 1$ and starting value of first sample at i=1

The model for univariate EWMA is defined by

$$Z_i = \lambda x_i + (1-\lambda)Z_{i-1}$$

Where

Z_i = ith EWMA
X_i = ith observation
Z_0 = average form historical data where $0 < \lambda \leq 1$ is a constant

As an extension of multivariate data, a multivariate exponentially weighted moving average (MEWMA) control chart can be used to analyze the shifts of air quality data. The data that is taken under consideration for MEWMA control chart is the data should be correlated and in a continuous manner. The data should also in time order for us to apply the MEWMA control chart. The model for MEWMA can be extended to

$$Z_i = \Lambda x_i + (1 - \Lambda)Z_{i-1}$$

where

Z_i = ith EWMA vector
X_i = ith observation vector
Z_0 = vector of variable values from the historical data
Λ = diag $(\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p)$ which is a diagonal matric with $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ on the main diagonal, with p is number of variables in each vector.

Where the covariance matrix is

$$\sum\nolimits_{Z_i} = \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]\Sigma$$

The first phase is used to establish the control limit and all out of control limit will be eliminated during phase 1 for construct the second phase of MEWMA. The control chart is using F-statistics at $\alpha = 0.05$. The phase I control limit is given by

$$\text{UCL} = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$\text{LCL} = 0$$

In the phase II, the control chart is used to monitoring air quality level where control limit are given as below

$$\text{UCL} = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$\text{LCL} = 0$$

where

p = number of variables
m = number of subgroups in the reference sample
n = size of subgroup
$F_{\alpha,p,mn-m-p+1}$ = upper 100a percentile of F distribution with (p, mn -m -p + 1) degrees of freedom

3.3 Flowchart of study

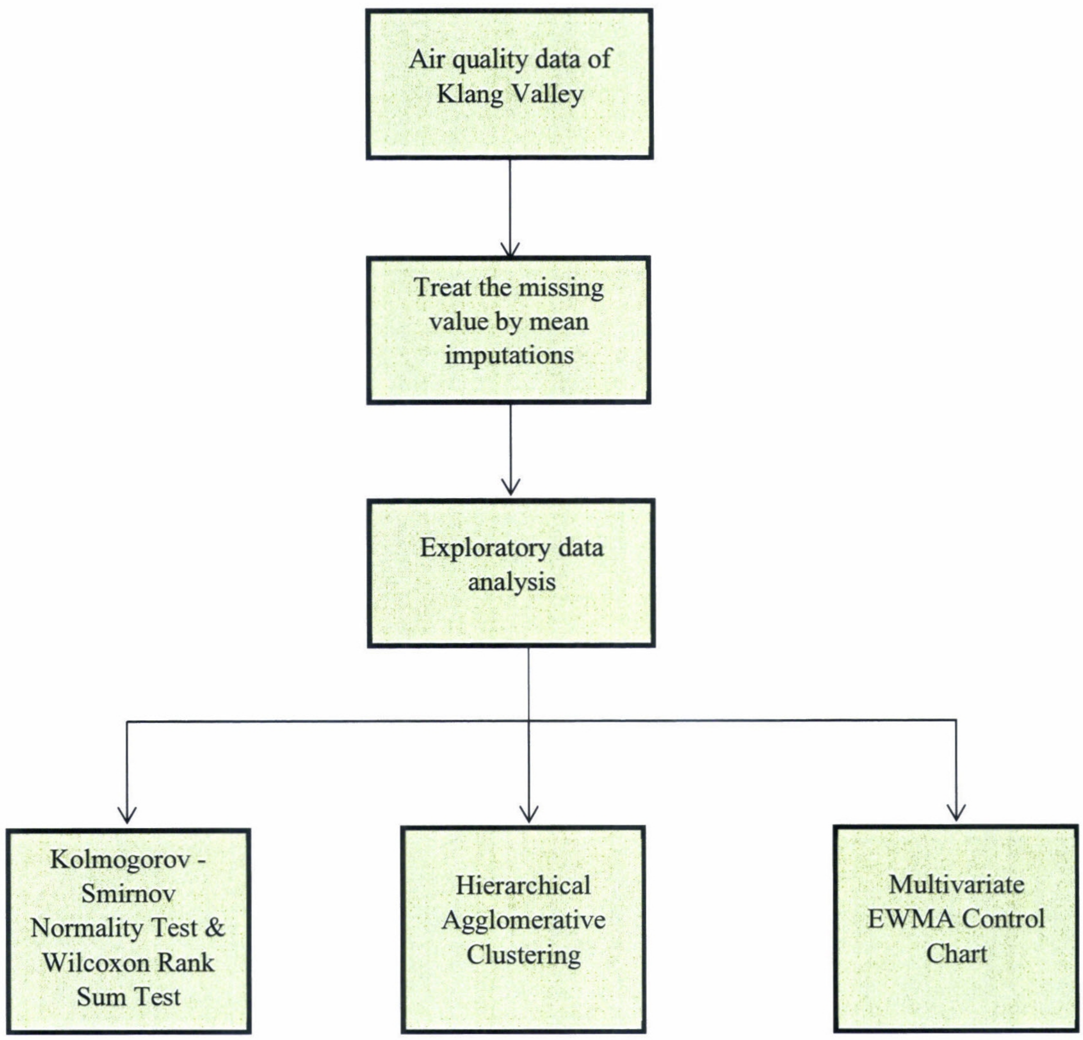


Figure 3.1: Flow of study methodology

3.4 Summary

In chapter 3, the methodologies used have been discussed and chapter 4 discusses results obtained through this study.

Objective	**Method of Analysis**
To check for any missing data	Mean imputation method
To run exploratory data analysis	Minimum, maximum and average values
Research Objective	
1. To determine the differences of air quality level before and after pandemic COVID-19 for all air pollutants.	Kolmogorov -Smirnov Normality Test & Wilcoxon Rank Sum Test
2. To group the air monitoring stations accordingly based on the severity of air quality level.	Hierarchical Agglomerative Clustering
3. To construct a multivariate control chart in assessing the air quality level in Klang Valley.	Multivariate Exponential Weighted Moving Average (MEWMA) Control Chart
4. To compare air quality of different areas in Klang Valley before and after the implementation of MCO	Multivariate Exponential Weighted Moving Average (MEWMA) Control Chart

CHAPTER 4

RESULT AND DISCUSSION

In this chapter, data collected is analyzed using Microsoft Excel, SPSS and JMP. All methodologies and steps discussed in chapter 3 are applied to the data in order to fulfil the objectives as stated in chapter 1. All results are explained in detail to obtain the information by using both numerical and graphical methods. Although numerical methods provide more accurate and precise knowledge, graphical methods are easier to understand.

4.1 EXPLORATORY DATA ANALYSIS

The first step of analyzing data is to conduct exploratory data analysis as it is an approach of data analysis to uncover underlying structure and insight to familiarize with data characteristics. By using exploratory data analysis, the data main characteristics can be summarized. Each data has its own unique underlying pattern. The daily average concentration before and after implementation of MCO is compared by finding out the minimum value, maximum value and average of each air quality level. Table 4.1 summarized air pollutants concentration of Klang Valley by minimum, maximum and average values.

Table 4.1: Summary of daily air pollutant concentrations

			PM_{10}	$PM_{2.5}$	SO_2	NO_2	O_3	CO
Banting	Before MCO	min	3.058	0.064	0.00	0.00	0.00	0.043
		max	303.953	279.573	0.021	0.054	0.101	3.114
		average	37.194	27.15	0.001	0.01	0.02	0.695
	After MCO	min	0.928	0.062	0.00	0.00	0.00	0.042
		max	169.666	150.209	0.020	0.0336	0.0788	2.307
		average	23.915	17.092	0.001	0.009	0.016	0.604
Batu Muda	Before MCO	min	0.00	0.00	0.00	0.00	0.00	0.04
		max	283.126	263.752	0.0171	0.0567	0.1	4.914
		average	33.713	26.653	0.001	0.016	0.017	0.963
	After MCO	min	1.633	0.075	0.00	0.00	0.00	0.086
		max	110.457	86.574	0.006	0.051	0.087	3.004
		average	29.33	17.21	0.001	0.013	0.015	0.954
Cheras	Before MCO	min	1.77	0.071	0.00	0.0002	0.00	0.079
		max	229.279	205.283	0.01	0.057	0.131	2.895
		average	34.583	26.405	0.001	0.016	0.023	0.813
	After MCO	min	2.217	0.098	0.00	0.00	0.00	0.248
		max	127.768	93.834	0.017	0.05	0.098	2.518
		average	21.465	15.221	0.001	0.012	0.018	0.728
Klang	Before MCO	min	1.177	0.071	0.00	0.00	0.00	0.046
		max	370.449	348.253	0.061	0.068	0.091	4.814
		average	41.064	31.8	0.001	0.018	0.014	1.027
	After MCO	min	2.16	0.062	0.00	0.00	0.00	0.00
		max	336.826	312.651	0.044	0.005	0.079	3.022
		average	32.282	22.525	0.001	0.014	0.017	0.745
Kuala Selangor	Before MCO	min	1.054	0.076				
		max	303.163	287.63				
		average	29.653	22.422				
	After MCO	min	0.11	0.074				
		max	148.303	131.183				
		average	21.247	15.448				
Petaling Jaya	Before `MCO	min	2.208	0.399	0.00	0.001	0.00	0.164
		max	374.34	219.342	0.018	0.097	0.094	4.538
		average	39.237	28.458	0.001	0.023	0.013	1.212
	After MCO	min	1.778	0.265	0.00	0.00	0.00	0.043
		max	169.267	158.768	0.019	0.111	0.056	6.735
		average	25.753	19.746	0.001	0.018	0.009	0.974

Putrajaya	Before MCO	min	1.407	0.147	0.00	0.00	0.00	0.04
		max	261.645	241.325	0.023	0.047	0.136	2.208
		average	36.527	27.526	0.001	0.008	0.026	0.637
	After MCO	min	0.00	0.049	0.00	0.00	0.00	0.04
		max	80.201	71.937	0.015	0.033	0.093	1.254
		average	23.656	16.57	0.001	0.007	0.02	0.524
Shah Alam	Before MCO	min	2.23	0.046	0.00	0.00	0.00	0.044
		max	264.642	252.305	0.023	0.065	0.156	3.202
		average	39.883	30.114	0.001	0.017	0.024	0.887
	After MCO	min	1.802	0.084	0.00	0.00	0.00	0.24
		max	110.409	94.175	0.014	0.059	0.099	2.265
		average	24.671	17.406	0.001	0.013	0.019	0.717

Table 4.1 summarized the exploratory data analysis conducted on minimum, maximum and average air pollutants readings in all locations. The minimum and maximum value are conducted to determine the highest and lowest air quality meanwhile, the average air quality is determined by mean values. By Table 4.1, all locations show most of air pollutants level are reduce before and after implementation of MCO. Overall, the air quality did reduce in terms of minimum, maximum and average concentration except for slight increase for minimum value of PM_{10}, $PM_{2.5}$ and CO for Batu Muda. PM_{10}, $PM_{2.5}$ and CO minimum reading also slightly increased with maximum value of SO_2 for Cheras. Klang experience a slight increase for minimum value for PM_{10} however it should notice that Klang did not have much air quality reading reduce due to Klang is an industrial area and maintain to operate during MCO. Klang also experience a slight increase for ozone readings. There is significance reduce on Kuala Selangor. From here, it can be noticed that only data PM_{10} and $PM_{2.5}$ that are available for Kuala Selangor. Besides that, SO_2, NO_2 and CO maximum air quality readings did slightly increase in Petaling Jaya. The minimum value of $PM_{2.5}$ and CO of Shah Alam also increased. However, despite the small increase in air quality, it is shown that air quality levels did reduce during the MCO implementation in Malaysia.

4.2 Wilcoxon Rank Sum Test

Kolmogorov-Smirnov are carried out to test whether normality distribution is met on the air quality dataset whether fulfil the normality assumption. Table 4.2 are summarized the results for the Kolmogorov-Smirnov test based on Appendix B. Based on Table 4.2, all datasets are compared for both periods for all air quality level. Based on the Kolmogorov-Smirnov value, all p-value are less than significance level of α=0.05. Thus, null hypothesis is rejected and no dataset met the normality assumptions. By Appendix A, the non-normal distribution of air quality dataset has been supported by boxplot where each distribution in the dataset are skewed and has outliers which supporting the Kolmogorov-Smirnov test result that the data does not met the normality assumption. Hence, the Wilcoxon rank sum test is chosen since they opted for non-parametric methods to find out whether there are differences in air quality after MCO implementation.

Table 4.2: Summary of Kolmogorov-Smirnov test

Location	Air Quality	Duration	Statistics	df	Sig.	Result
Kuala Selangor	PM_{10}	1	0.152	10608	0.00	Ho is rejected
		2	0.094	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.158	10608	0.00	Ho is rejected
		2	0.084	6936	0.00	Ho is rejected
Banting	PM_{10}	1	0.134	10608	0.00	Ho is rejected
		2	0.097	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.143	10608	0.00	Ho is rejected
		2	0.1	6936	0.00	Ho is rejected
	SO_2	1	0.208	10608	0.00	Ho is rejected
		2	0.254	6936	0.00	Ho is rejected
	NO_2	1	0.056	10608	0.00	Ho is rejected
		2	0.071	6936	0.00	Ho is rejected
	O_3	1	0.146	10608	0.00	Ho is rejected
		2	0.147	6936	0.00	Ho is rejected
	CO	1	0.085	10608	0.00	Ho is rejected
		2	0.067	6936	0.00	Ho is rejected

Batu Muda	PM_{10}	1	0.132	10608	0.00	Ho is rejected
		2	0.051	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.152	10608	0.00	Ho is rejected
		2	0.056	6936	0.00	Ho is rejected
	SO_2	1	0.138	10608	0.00	Ho is rejected
		2	0.116	6912	0.00	Ho is rejected
	NO_2	1	0.056	10608	0.00	Ho is rejected
		2	0.053	6936	0.00	Ho is rejected
	O_3	1	0.163	10608	0.00	Ho is rejected
		2	0.178	6936	0.00	Ho is rejected
	CO	1	0.074	10608	0.00	Ho is rejected
		2	0.077	6936	0.00	Ho is rejected
Cheras	PM_{10}	1	0.141	10608	0.00	Ho is rejected
		2	0.044	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.154	10608	0.00	Ho is rejected
		2	0.05	6936	0.00	Ho is rejected
	SO_2	1	0.17	10608	0.00	Ho is rejected
		2	0.173	6936	0.00	Ho is rejected
	NO_2	1	0.059	10608	0.00	Ho is rejected
		2	0.071	6936	0.00	Ho is rejected
	O_3	1	0.155	10608	0.00	Ho is rejected
		2	0.16	6936	0.00	Ho is rejected
	CO	1	0.084	10608	0.00	Ho is rejected
		2	0.107	6936	0.00	Ho is rejected
Klang	PM_{10}	1	0.127	10608	0.00	Ho is rejected
		2	0.094	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.134	10608	0.00	Ho is rejected
		2	0.095	6936	0.00	Ho is rejected
	SO_2	1	0.279	10608	0.00	Ho is rejected
		2	0.198	6936	0.00	Ho is rejected
	NO_2	1	0.053	10608	0.00	Ho is rejected
		2	0.063	6936	0.00	Ho is rejected
	O_3	1	0.16	10608	0.00	Ho is rejected
		2	0.129	6936	0.00	Ho is rejected
	CO	1	0.097	10608	0.00	Ho is rejected
		2	0.123	6936	0.00	Ho is rejected

Petaling Jaya	PM_{10}	1	0.152	10608	0.00	Ho is rejected
		2	0.104	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.165	10608	0.00	Ho is rejected
		2	0.116	6936	0.00	Ho is rejected
	SO_2	1	0.176	10608	0.00	Ho is rejected
		2	0.186	6936	0.00	Ho is rejected
	NO_2	1	0.063	10608	0.00	Ho is rejected
		2	0.079	6936	0.00	Ho is rejected
	O_3	1	0.189	10608	0.00	Ho is rejected
		2	0.185	6936	0.00	Ho is rejected
	CO	1	0.074	10608	0.00	Ho is rejected
		2	0.069	6936	0.00	Ho is rejected
Putrajaya	PM_{10}	1	0.15	10608	0.00	Ho is rejected
		2	0.064	6936	0.00	Ho is rejected
	$PM_{2.5}$	1	0.159	10608	0.00	Ho is rejected
		2	0.075	6936	0.00	Ho is rejected
	SO_2	1	0.209	10608	0.00	Ho is rejected
		2	0.19	6936	0.00	Ho is rejected
	NO_2	1	0.082	10608	0.00	Ho is rejected
		2	0.094	6936	0.00	Ho is rejected
	O_3	1	0.115	10608	0.00	Ho is rejected
		2	0.118	6936	0.00	Ho is rejected
	CO	1	0.086	10608	0.00	Ho is rejected
		2	0.06	6936	0.00	Ho is rejected
Shah Alam	PM_{10}	1	0.118	10608	0.00	Ho is rejected
		2	0.062	6912	0.00	Ho is rejected
	$PM_{2.5}$	1	0.136	10608	0.00	Ho is rejected
		2	0.065	6912	0.00	Ho is rejected
	SO_2	1	0.188	10608	0.00	Ho is rejected
		2	0.204	6936	0.00	Ho is rejected
	NO_2	1	0.048	10608	0.00	Ho is rejected
		2	0.05	6936	0.00	Ho is rejected
	O_3	1	0.164	10608	0.00	Ho is rejected
		2	0.15	6936	0.00	Ho is rejected
	CO	1	0.071	10608	0.00	Ho is rejected
		2	0.08	6936	0.00	Ho is rejected

The result of non-parametric of Wilcoxon rank sum test are summarized in Table 4.3 based on output in Appendix B. Based on Table 4.3, based on the p- value, all result shows significance differences by p-value less than α=0.05 except for SO_2 level for Cheras and Shah

Alam that are unable to reject the null hypothesis. Thus, it can be said that there are significance differences for air quality after implementation of MCO except for SO_2 level for Cheras and Shah Alam The p-value for SO_2 level for Cheras and Shah Alam are highlighted in Table 4.3.

Table 4.3: Summary of non-parametric test

Location		PM_{10}	$PM_{2.5}$	SO_2	NO_2	O_3	CO
Kuala Selangor	Z	-25.841	-24.959				
	Asymp. Sig. (2-tailed)	0.00	0.00				
Banting	Z	-44.943	-34.069	-11.327	-16.454	-7.122	-18.311
	Asymp. Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00
Batu Muda	Z	-5.884	-34.494	-4.428	-23.317	-6.651	-4.031
	Asymp. Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00
Cheras	Z	-49.448	-46.873	-0.854	-32.171	-11.821	-14.476
	Asymp. Sig. (2-tailed)	0.00	0.00	0.393	0.00	0.00	0.00
Klang	Z	-23.058	-32.035	-27.225	-32.219	-11.821	-45.404
	Asymp. Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00
Petaling Jaya	Z	-50.669	-36.72	-32.646	-35.199	-14.959	-32.522
	Asymp. Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00
Putrajaya	Z	-44.586	-42.366	-23.091	-14.545	-14.695	-34.088
	Asymp. Sig. (2-tailed)	0.00	0.00	0.00	0.00	0.00	0.00
Shah Alam	Z	-51.404	-50.005	-1.692	-23.736	-2.665	-28.744
	Asymp. Sig. (2-tailed)	0.00	0.00	0.091	0.00	0.00	0.00

4.3 Hierarchical Agglomerative Clustering Analysis (HACA)

The air quality is monitored by locations and classified based on levels of similarities and differences. Locations with a high level of similarities are grouped together. By clustering methods, there are formation of clusters in dendrograms as per Appendix C. The summary of clustering is summarized in Table 4.4 based on output in Appendix C. The Table 4.4 summarizes the output from clustering methods that has been perform on the air quality dataset. Based on dendrograms as per Appendix C, the data are partition to two groups which can be defined as low polluted area and high polluted area. Cluster 1 classified as low polluted area meanwhile cluster 2 is classified as high polluted area. Since the Kuala Selangor data is only available for particulate matter, the level of air quality Kuala Selangor can only be identified by particulate matter.

Table 4.4: Summary results of hierarchical clustering

Air Quality	Duration	Group	Location
PM_{10}	Before MCO	Cluster 1	Batu Muda, Cheras, Banting, Putrajaya, Petaling Jaya, Shah Alam, Kuala Selangor
		Cluster 2	Klang
	After MCO	Cluster 1	Batu Muda, Cheras, Banting, Putrajaya, Shah Alam, Kuala Selangor, Petaling Jaya
		Cluster 2	Klang
$PM_{2.5}$	Before MCO	Cluster 1	Batu Muda, Cheras, Banting, Putrajaya, Petaling Jaya, Shah Alam, Kuala Selangor
		Cluster 2	Klang
	After MCO	Cluster 1	Batu Muda, Cheras, Banting, Putrajaya, Shah Alam, Kuala Selangor, Petaling Jaya
		Cluster 2	Klang
SO_2	Before MCO	Cluster 1	Batu Muda, Cheras, Putrajaya, Petaling Jaya, Shah Alam, Banting
		Cluster 2	Klang
	After MCO	Cluster 1	Batu Muda, Cheras, Putrajaya, Shah Alam, Petaling Jaya, Banting
		Cluster 2	Klang
NO_2	Before MCO	Cluster 1	Banting, Putrajaya, Batu Muda, Cheras, Klang, Shah Alam

		Cluster 2	Petaling Jaya
	After MCO	Cluster 1	Banting, Putrajaya, Batu Muda, Cheras, Klang, Shah Alam
		Cluster 2	Petaling Jaya
O_3	Before MCO	Cluster 1	Batu Muda, Petaling Jaya, Klang
		Cluster 2	Cheras, Putrajaya, Banting, Shah Alam
	After MCO	Cluster 1	Batu Muda, Cheras, Putrajaya, Shah Alam, Banting, Klang
		Cluster 2	Petaling Jaya
CO	Before MCO	Cluster 1	Banting, Putrajaya, Cheras, Shah Alam, Batu Muda, Klang
		Cluster 2	Petaling Jaya
	After MCO	Cluster 1	Banting, Putrajaya, Cheras, Shah Alam, Klang
		Cluster 2	Batu Muda, Petaling Jaya

Cluster 1 - low level of air pollutant concentration
Cluster 2 - high level of air pollutant concentration

By clustering, Klang is highest location that contain fine particles of PM_{10} and $PM_{2.5}$ before and after implementation of MCO. Besides that, by clustering method, it is found that Klang is also area with high pollution level of SO_2 for both periods. Meanwhile, Petaling Jaya is has highest pollution level of NO_2 before and after lockdown imposed. Ozone level showing the biggest improvement after lockdown imposed where the high polluted level is only Petaling Jaya after MCO. Before MCO, Cheras, Putrajaya, Banting and Shah Alam are area with high ozone level however improving after lockdown. Despite that, the ozone level for Petaling Jaya worsen after MCO implementation. Furthermore, CO level is Batu Muda deteriorate after MCO joining Petaling Jaya with high CO level after MCO.

Therefore, it can be concluded that Klang has the highest level of particulate matter and SO_2. Meanwhile, Petaling Jaya has highest level of NO_2 and CO with declining O_3 levels after lockdown. In addition, the CO level is Batu Muda has worsened after MCO. This is due to Klang is an industrial area with massive factories and operating during lockdown. This can be concluded that the high level of NO_2, O_3 and CO which can be contributed by vehicle emission and combustion of fossil fuels in Petaling Jaya and Klang. Industrial activities in Klang and Petaling Jaya is also a factor has been contributed to bad air quality.

4.4 Multivariate Exponential Weighted Moving Average (MEWMA) Control Chart

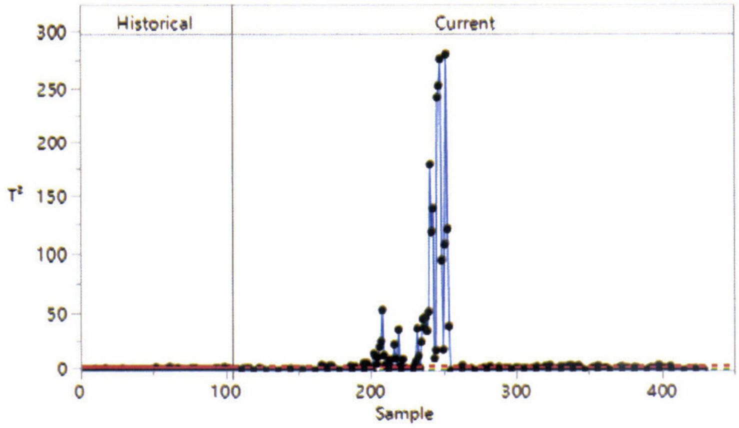
Since the air quality dataset are multivariate data, the control chart of MEWMA are applied. The data are given for two years hence the data are divided into two periods which are before and the implementation of MCO. Hence there are two phase I conducted for both periods. During phase I, all outliers are removed and control limits are calculated to be applied during phase II. The phase I for before implementation of phase I are 90 days and 44 days for periods after implementation of MCO. The phase I for period after lockdown imposed are shorter considering the volume of data are smaller compared to data before lockdown imposed. The out of control points are sequentially removed to get the control limit which are used during phase II. The alpha used in this control chart are $\alpha=0.05$.

Before proceeding to use the multivariate control chart, some fundamental checking is required to be done before constructing the MEWMA. The preliminary analysis that are done is check if the variables are correlated. After checking the correlation between air quality, the first phase develops by eliminating out of control data. The same mean and covariance matrix during phase I are used during monitoring phase II. Thus, after constructing control chart on all locations, based on all control charts, it is shown that all locations have out of control points for both periods. The multivariate control chart for all locations shown below from Figure 4.1 until Figure 4.16.

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 1 Principal Components



T^2 Limit Summaries

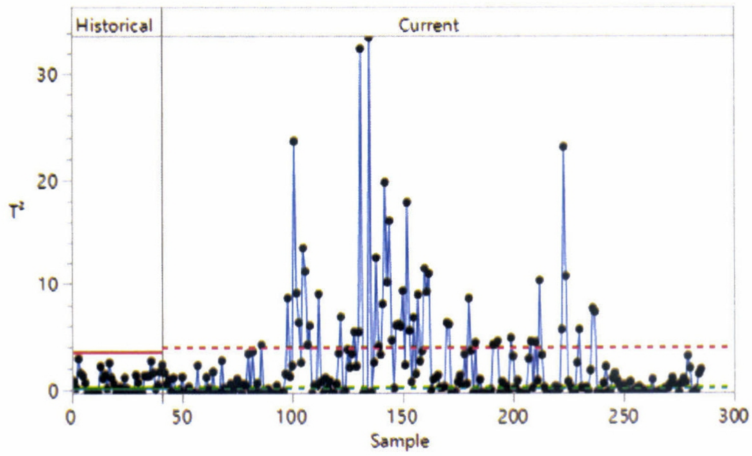
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	0.45619	3.7885	0	0.05
Current	0.46257	3.97116	63	

Figure 4.1: Control chart of Kuala Selangor before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 1 Principal Components



T^2 Limit Summaries

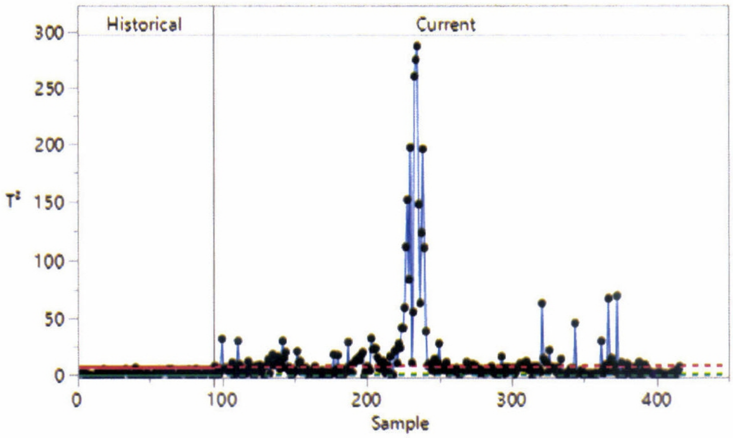
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	0.45847	3.70166	0	0.05
Current	0.47512	4.19356	50	

Figure 4.2: Control chart of Kuala Selangor after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 4 Principal Components



T^2 Limit Summaries

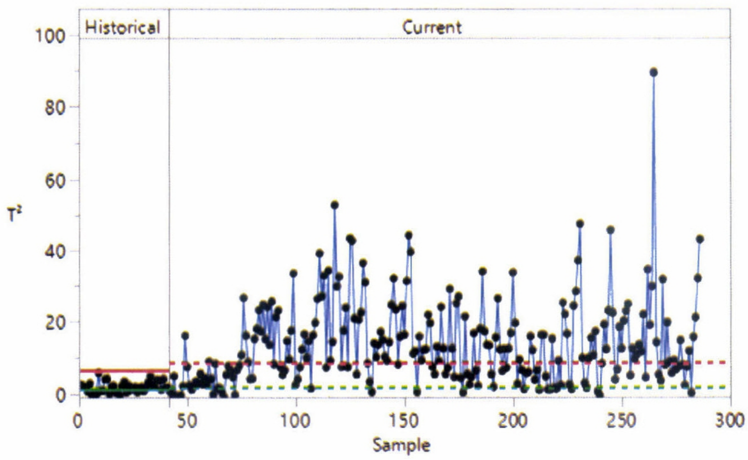
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	3.3688	9.20732	0	0.05
Current	3.53204	10.3302	105	

Figure 4.3: Control chart of Shah Alam before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

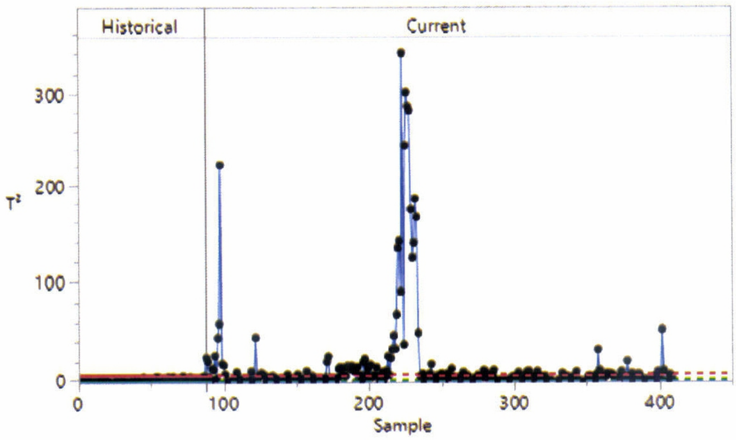
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38659	7.34346	0	0.05
Current	2.59765	9.22515	147	

Figure 4.4: Control chart of Shah Alam after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

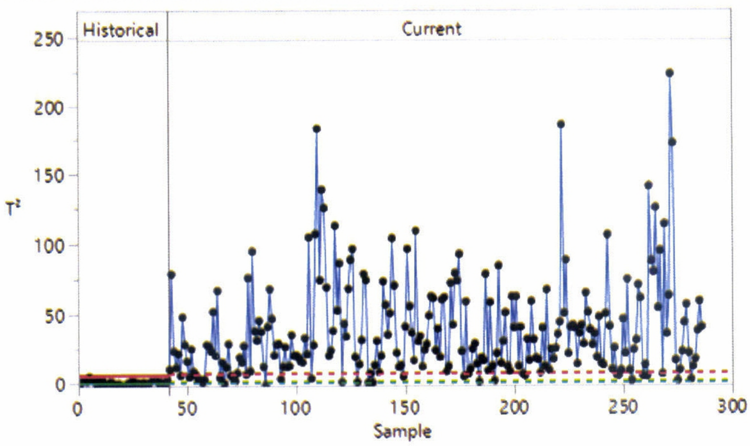
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.3751	7.59578	0	0.05
Current	2.47018	8.42927	101	

Figure 4.5: Control chart of Putrajaya before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

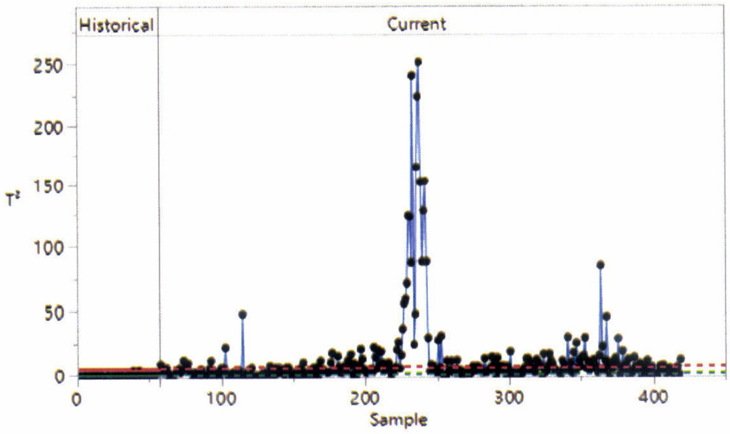
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38659	7.34346	0	0.05
Current	2.59765	9.22515	208	

Figure 4.6: Control chart of Putrajaya after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

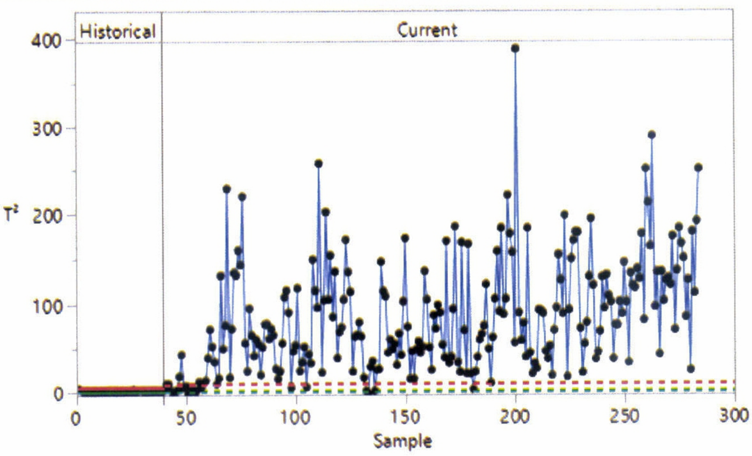
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38032	7.47832	0	0.05
Current	2.52849	8.78721	119	

Figure 4.7: Control chart of Petaling Jaya before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 4 Principal Components



T^2 Limit Summaries

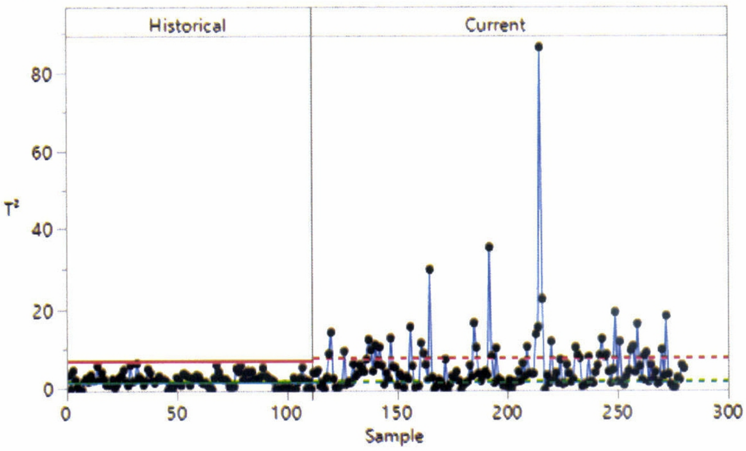
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	3.38806	8.79955	0	0.05
Current	3.81124	11.7656	226	

Figure 4.8: Control chart of Petaling Jaya after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

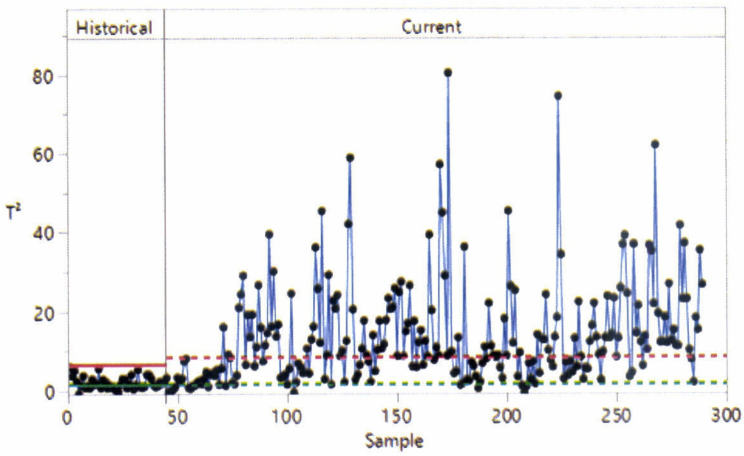
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.37304	7.64359	0	0.05
Current	2.44694	8.28946	40	

Figure 4.9: Control chart of Klang before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

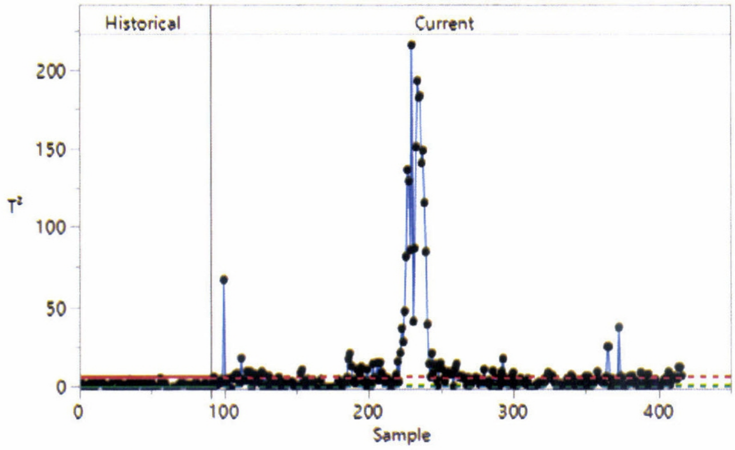
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38503	7.37641	0	0.05
Current	2.58053	9.11535	146	

Figure 4.10: Control chart of Klang after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

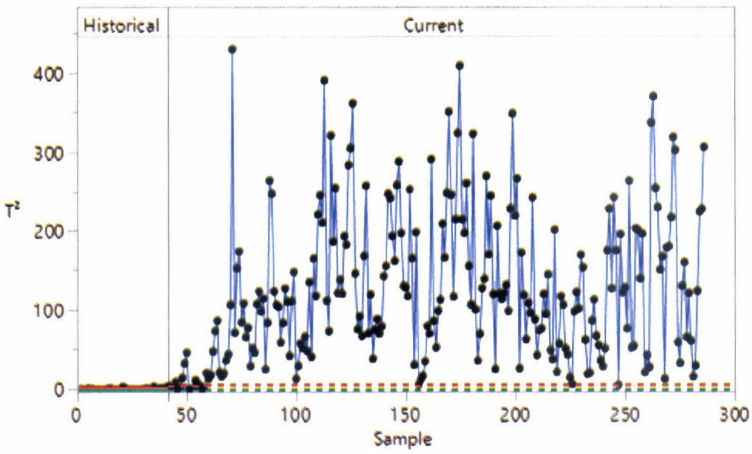
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.37478	7.60317	0	0.05
Current	2.46657	8.40745	88	

Figure 4.11: Control chart of Cheras before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

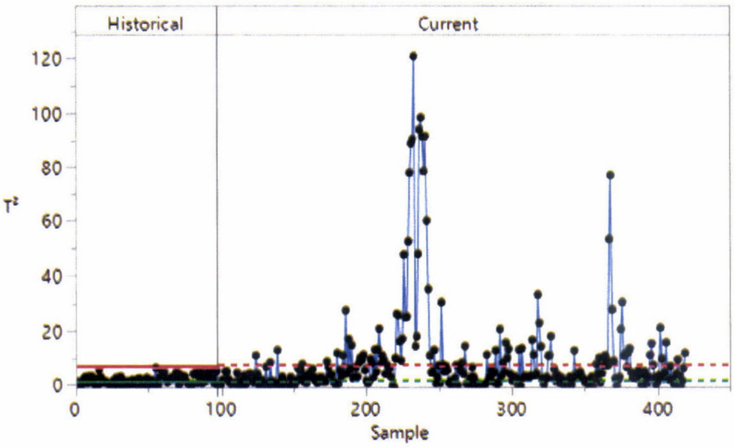
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38659	7.34346	0	0.05
Current	2.59765	9.22515	232	

Figure 4.12: Control chart of Cheras after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

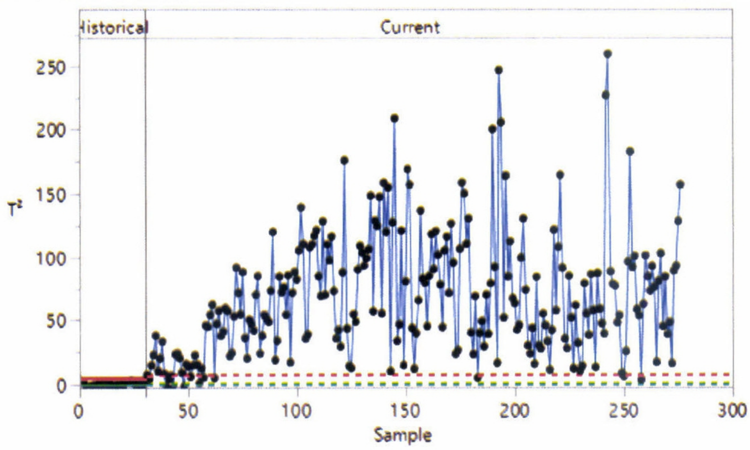
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.37411	7.61861	0	0.05
Current	2.45905	8.3621	96	

Figure 4.13: Control chart of Batu Muda before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

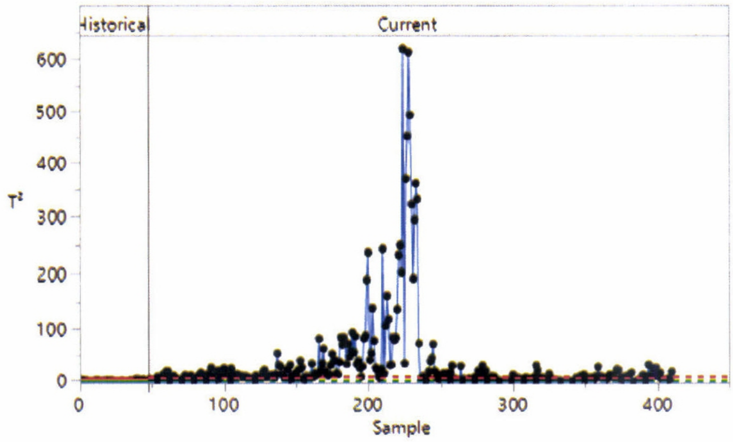
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.3954	7.16413	0	0.05
Current	2.69349	9.85687	233	

Figure 4.14: Control chart of Batu Muda after implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 4 Principal Components



T^2 Limit Summaries

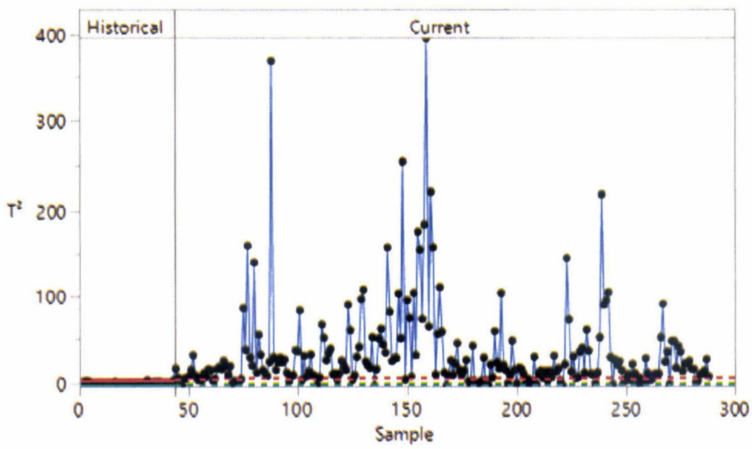
Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	3.38218	8.91971	0	0.05
Current	3.72575	11.3135	196	

Figure 4.15: Control chart of Banting before implementation of MCO

PCA Model Driven Multivariate Control Chart

Monitor the Process

T^2 for 3 Principal Components



T^2 Limit Summaries

Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	2.38553	7.36595	0	0.05
Current	2.58595	9.15	193	

Figure 4.16: Control chart of Banting after implementation of MCO

Table 4.5: Summary of air quality before the implementation of MCO

Locations	Median first phase	Median second phase	UCL first phase	UCL second phase	Out of control points
Kuala Selangor	0.4562	0.4626	3.7885	3.9712	63
Shah Alam	3.3688	3.5320	9.2073	10.3302	105
Putrajaya	2.3751	2.4702	7.5958	8.4293	101
Petaling Jaya	2.3803	2.5285	7.4783	8.7872	119
Klang	2.3730	2.4469	7.6436	8.2895	40
Cheras	2.3748	2.4666	7.6032	8.4075	88
Batu Muda	2.3741	2.4591	7.6186	8.3621	96
Banting	3.3822	3.7258	8.9197	11.3135	196

Table 4.5: Summary of air quality after the implementation of MCO

Locations	Median first phase	Median second phase	UCL first phase	UCL second phase	Out of control points
Kuala Selangor	0.4585	0.4751	3.7017	4.1936	50
Shah Alam	2.3866	2.5977	7.3435	9.2252	147
Putrajaya	2.3866	2.5977	7.3435	9.2252	208
Petaling Jaya	3.3881	3.8112	8.7996	11.7656	226
Klang	2.3850	2.5805	7.3764	9.1154	146
Cheras	2.3866	2.5977	7.3435	9.2252	232
Batu Muda	2.3954	2.6935	7.1641	9.8569	233
Banting	2.3855	2.5860	7.3660	9.15	193

Table 4.5 and Table 4.6 compare the air quality in terms of median, UCL and out of control points for all locations which summarized the results. When the median is statistically in control, it is indicating that the air quality is in control. By MEWMA control charts, shifts in air quality of Klang Valley areas are quickly detected. All the air qualities are contributing to air pollution index and MEWMA are one of the control charts that able to detect shifts of air quality. Based on the control charts plotted, only Kuala Selangor and Banting shows the reduction of out of control points. In contrast, Shah Alam, Putrajaya, Petaling Jaya, Klang,

Cheras and Batu Muda show the higher shifts after the implementation of MCO compared to normal periods.

The UCL for the first phase are decreasing for all locations except for Petaling Jaya. However, the control limits for monitoring phase of Kuala Selangor, Putrajaya, Petaling Jaya, Klang, Cheras and Batu Muda are higher compared to before the implementations of MCO. The occurrences of out of control after the implementation of MCO is due to the periods during MCO or total lockdown has been set as first phase and periods during CMCO and RMCO has been set as second phase. Thus, the periods of MCO between 18^{th} March 2020 until 3^{rd} May 2020 are having low air quality and causing many of out of controls points when air quality during CMCO and RMCO are compared to periods during MCO. This can be supported when total lockdown are ends, CMCO are implemented followed by RMCO where flexibility imposed to uplift the economy. Some flexibility is imposed such as opening of almost sectors with fulfil all the SOP. The public transport also fully re-operate with the inter-state travel restriction has been uplifted which is approaching the new norm after pandemic of COVID-19.

Therefore, based on MEWMA applied, it can be concluded that air quality in the Klang Valley are not statistically in control for periods before implementation of MCO and after the implementations of MCO. The causes of the poor air quality are emission from vehicles, industrial activities and open burning. From the control chart applied, it can be concluded that Kuala Selangor and Banting are improving after total lockdown. Meanwhile, other locations besides Kuala Selangor and Banting have deteriorating air quality after total lockdown uplift. Despite all locations having a higher number of shifts, most occurrences of out of control are occurring during CMCO and RMCO has been implemented compared to total lockdown periods.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

This chapter is to conclude the findings of this study and to provide a conclusion for each chapter. Besides that, this chapter relates the results and discussion to the research objectives. In this final chapter, some recommendations are suggested for further study and improvements that can be made in the same study field.

5.1 MAIN FINDINGS

Based on the analysis and results in Chapter 4, all the data has been analyzed to fulfil the objectives of this study. Through the exploratory data analysis, it is found that the average air quality has been decreased and proven that the air quality is better after the implementation of MCO except for SO_2 level for Cheras and Shah Alam. However, on overall average, daily air quality concentrations are decreasing, indicating that air pollution levels have improved since the lockdown was imposed.

The result of the Kolmogorov-Smirnov test has been concluded that the null hypothesis is rejected that the data does not meet the normality assumption and non-parametric approaches have been applied to find the significant differences before and after lockdown implementation.

Research Objectives 1: To determine the differences of air quality level before and after pandemic COVID-19 for all air pollutants.

By the non-parametric test of Wilcoxon rank sum test, all the tests conducted on all air quality are able to reject H_o based on p-value less than α=0.05. However, the SO_2 level for Cheras and Shah Alam accept the null hypothesis meaning that the SO_2 level in both locations do not show statistical significance before and after the implementation of MCO. With this, it can be concluded that air quality is better after implementation of MCO except for SO_2 level for Cheras and Shah Alam.

Research Objectives 2: To group the air monitoring stations accordingly based on the severity of air quality level.

With hierarchical clustering, it is found that the Klang area has the highest level of particulate matter and SO_2 before and prior to lockdown. Meanwhile, Petaling Jaya has the highest level of NO_2 and CO. In addition, CO level is Batu Muda has worsened after MCO was implemented. Nevertheless, there has been a major improvement on ozone levels in Cheras, Putrajaya, Banting and Shah Alam and ozone level for Petaling Jaya has declined since prior to lockdown. Hence, by clustering, it is discovered that Petaling Jaya has high levels of NO_2 and CO along with O_3 levels after lockdown.

Research Objectives 3: To construct a multivariate control chart in assessing the air quality level in Klang Valley.

By multivariate control chart of MEWMA, after checking the correlation between air quality, the control chart is constructed by develop first phase whereby out of control data is eliminated and using mean and covariance matrix during monitoring phase II.

Research Objectives 4: To compare air quality of different areas in Klang Valley before and after the implementation of MCO.

Due to economic activity and vehicle emissions, it is possible to conclude that Klang Valley experienced poor air quality levels during both periods. Therefore, the air quality is statistically out of control for periods before and after the implementation of MCO based on the MEWMA control charts.

Therefore, it can be concluded that all objectives are achieved. The Wilcoxon rank sum test, hierarchical clustering and multivariate control chart of MEWMA are successfully applied to dataset by using the Microsoft Excel, SPSS and SAS JMP software. Findings from hourly air quality data for two years has been successfully extracted from all methods employed.

5.2 RECOMMENDATION

In this study, the main constraints are limited to carrying out the study. Time constraints limit the methods that can be used to analyze datasets. There are numerous methods that can be used to analyze the data to obtain information from the dataset. Moreover, the different imputation methods for missing values can be utilized which could be more precise in estimating the true value of missing data. The data may yield different results and yield different findings and insights into the Klang Valley's air quality.

REFERENCES

Adams, K., Greenbaum, D. S., Shaikh, R., Van Erp, A. M., & Russell, A. G. (2015). Particulate matter components, sources, and health: Systematic approaches to testing effects. *Journal of the Air & Waste Management Association*, *65*(5), 544–558. https://doi.org/10.1080/10962247.2014.1001884

Amran, M. A., Azid, A., Juahir, H., Toriman, M. E., Mustafa, A. D., Hasnam, C. N. C., Azaman, F., Kamarudin, M. K. A., Saudi, A. S. M., & Yunus, K. (2015). Spatial Analysis of Thecertain Air Pollutants Using Environmetric Techniques. *Jurnal Teknologi*, *75*(1), 241–249. https://doi.org/10.11113/jt.v75.3977

Oyeka, I. C. A., & Ebuh, G. U. (2012). Modified Wilcoxon Signed-Rank Test. *Open Journal of Statistics*, *02*(02), 172–176. https://doi.org/10.4236/ojs.2012.22019

Anenberg, S. C., Jaffe, D. A., Prather, M. J., Bergmann, D., Cuvelier, K., Dentener, F. J., Duncan, B. N., Jonson, J. A. N. E., Lupu, A., Mackenzie, I. A. N. A., Marmer, E., Park, R. J., Sanderson, M. G., Schultz, M., Shindell, D. T., Szopa, S., Vivanco, M. G., Wild, O., & Zeng, G. (2009). *Intercontinental Impacts of Ozone Pollution on Human Mortality*. *43*(17), 6482–6487.

Aslam, M. (2020). Introducing Kolmogorov-Smirnov Tests under Uncertainty: An Application to Radioactive Data. *ACS Omega*, *5*(1), 914–917. https://doi.org/10.1021/acsomega.9b03940

Karthikeyan, B., George, D. J., Manikandan, G. & Thomas, T. (2020). A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering. *International Journal of Emerging Trends in Engineering Research*, *8*(5), 1600–1604. https://doi.org/10.30534/ijeter/2020/20852020

Baek, J. I., & Ban, Y. U. (2020). The impacts of urban air pollution emission density on air pollutant concentration based on a panel model. *Sustainability (Switzerland)*, *12*(20), 1–26. https://doi.org/10.3390/su12208401

Balzan, M. V., Agius, G., & Debono, A. G. (1996). Carbon monoxide poisoning: Easy to treat but difficult to recognise. *Postgraduate Medical Journal*, *72*(850), 470–473. https://doi.org/10.1136/pgmj.72.850.470

Beloconi, A., & Vounatsou, P. (2021). Substantial Reduction in Particulate Matter Air Pollution across Europe during 2006-2019: A Spatiotemporal Modeling Analysis. *Environmental Science and Technology*, *55*(22), 15505–15518. https://doi.org/10.1021/acs.est.1c03748

Berita Harian, (2020). Kronologi Covid-19 di Malaysia. Retrieved from https://www.bharian.com.my/berita/nasional/2020/03/666122/kronologi-covid-19-di-malaysia

Brook, J. R., Burnett, R. T., Dann, T. F., Cakmak, S., Goldberg, M. S., Fan, X., & Wheeler, A. J. (2007). Further interpretation of the acute effect of nitrogen dioxide observed in Canadian time-series studies. *Journal of Exposure Science & Environmental Epidemiology*, *17*(2), S36–S44. https://doi.org/10.1038/sj.jes.7500626

Carmichael, G. R., Streets, D. G., Calori, G., Amann, M., Jacobson, M. Z., Hansen, J., & Ueda, H. (2002). Changing trends in sulfur emissions in Asia: Implications for acid deposition, air pollution, and climate. *Environmental Science and Technology*, *36*(22), 4707–4713. https://doi.org/10.1021/es011509c

Cheda, A. L., Jacome, M. A., Cao, R. & Salazar, P. M. D. (2020). Estimating lengths-of-stay of hospitalized COVID-19 patients using a non-parametric model: a case study in Galicia (Spain). *The Preprint Server for Health Sciences*, 1–29.

Chen, T.-M., Kuschner, W. G., Gokhale, J., & Shofer, S. (2007). Outdoor Air Pollution: Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide Health Effects. *The American Journal of the Medical Sciences*, *333*(4), 249–256. https://doi.org/https://doi.org/10.1097/MAJ.0b013e31803b900f

Cullis, C. F., & Hirschler, M. M. (1980). Atmospheric sulphur: Natural and man-made sources. *Atmospheric Environment (1967)*, *14*(11), 1263–1278. https://doi.org/https://doi.org/10.1016/0004-6981(80)90228-0

Department of Environment (1997). A Guide to Air Pollutant Index (API) in Malaysia. *Ministry of Science, Technology and the Environment.*

Department of Environment & Ibarahim, H. R. (2000). A Guide to Air Pollutant Index (API) in Malaysia. *Department of Environment Malaysia*, *4*, 20.

Department of Statistics, (2016). Sources of Pollution, *Federal Government Administartive Centre.*

Erman, N., Korosec, A., & Suklan, J. (2015). Performance of Selected Agglomerative Hierarchical Clustering Methods. *Innovative Issues and Approaches in Social Sciences*, *8*(1), 180–204. https://doi.org/10.12959/issn.1855-0541.iiass-2015-no1-art11

Ershadi, M. J., Taghi, S., & Niaki, A. (2014). *Economic-Statistical Design of MEWMA Control Chart Using Genetic Algorithm*. *February 2009*.

Harris, J. E., Boushey, C., Bruemmer, B., & Archer, S. L. (2008). Publishing Nutrition Research: A Review of Nonparametric Methods, Part 3. *Journal of the American Dietetic Association*, *108*(9), 1488–1496. https://doi.org/10.1016/j.jada.2008.06.426

Horvath, S. M. (1980). Nitrogen dioxide, pulmonary function, and respiratory disease. *Bulletin of the New York Academy of Medicine*, *56*(9), 835–846. https://pubmed.ncbi.nlm.nih.gov/6161660

Imam, A., Mohammed, U., Abanyam & Chiawa, M. A. (2014). On Consistency and Limitation of paired t-test, Sign and Wilcoxon Sign Rank Test. *IOSR Journal of Mathematics*, *10*(1), 01–06. https://doi.org/10.9790/5728-10140106

Inkelas, M., Blair, C., Furukawa, D., Manuel, V. G., Malenfant, J. H., Martin, E., Emeruwa, I., Kuo, T., Arangua, L., Robles, B., & Provost, L. P. (2021). Using control charts to understand community variation in COVID-19. *PLoS ONE*, *16*(4 April), 1–11. https://doi.org/10.1371/journal.pone.0248500

Jatain, A., Nagpal, A., & Gaur, D. (2013). Agglomerative Hierarchical Approach for Clustering Components of Similar Reusability. *International Journal of Computer Applications*, *68*(2), 33–37. https://doi.org/10.5120/11553-6832

Kanchan, Gorai, A. K., & Goyal, P. (2015). A review on air quality indexing system. *Asian Journal of Atmospheric Environment*, *9*(2), 101–113. https://doi.org/10.5572/ajae.2015.9.2.101

Kester, M., Karpa, K. D., & Vrana, K. E. (2012). *3 - Toxicology* (M. Kester, K. D. Karpa, & K. E. B. T.-E. I. R. P. (Second E. (Second E. Vrana (eds.); pp. 29–39). W.B. Saunders. https://doi.org/https://doi.org/10.1016/B978-0-323-07445-2.00003-3

Kim, H.-Y. (2014). Statistical notes for clinical researchers: Nonparametric statistical methods: 1. Nonparametric methods for comparing two groups. *Restorative Dentistry & Endodontics*, *39*(3), 235–239. https://doi.org/10.5395/rde.2014.39.3.235

Kimbrough, S., Owen, R. C., Snyder, M., & Richmond-Bryant, J. (2017). NO to NO(2) Conversion Rate Analysis and Implications for Dispersion Model Chemistry Methods using Las Vegas, Nevada Near-Road Field Measurements. *Atmospheric Environment (Oxford, England : 1994)*, *165*, 23–24. https://doi.org/10.1016/j.atmosenv.2017.06.027

Kousiga, T. & Vadivu, R. S. (2019). Hierarchical Clustering Algorithms in Data Mining. *International Journal of Scientific Development and Research (IJSDR)*, *4*(9), 1–3. http://www.waset.org/publications/10002625

Lampreia, S. S., Requeijo, J. G., Dias, J. M., & Vairinhos, V. (2015). Implementation of MEWMA control chart in equipment condition monitoring. *Journal of Vibration Engineering and Technologies*, *3*(6), 667–677.

Lee, H. H., Iraqui, O., Gu, Y., Yim, S. H. L., Chulakadabba, A., Tonks, A. Y. M., Yang, Z., & Wang, C. (2018). Impacts of air pollutants from fire and non-fire emissions on the regional air quality in Southeast Asia. *Atmospheric Chemistry and Physics*, *18*(9), 6141–6156. https://doi.org/10.5194/acp-18-6141-2018

Li, X., Chen, X., Yuan, X., Zeng, G., León, T., Liang, J., Chen, G., & Yuan, X. (2017). Characteristics of particulate pollution (PM2.5 and PM10) and their spacescale-dependent relationships with meteorological elements in China. *Sustainability (Switzerland)*, *9*(12), 1–14. https://doi.org/10.3390/su9122330

Lowry, C. A., Woodwall, W. H., Champ, C. W. & Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics,* Vol. 34(1), pp 45-63.

Mahmoud, M. A., & Maravelakis, P. E. (2010). The performance of the MEWMA control chart when parameters are estimated. *Communications in Statistics: Simulation and Computation*, *39*(9), 1803–1817. https://doi.org/10.1080/03610918.2010.518269

Maholtra, N. K. (2019), Marketing Research: An Applied Orientation. ***Pearson.***

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, *8*(February), 1–13. https://doi.org/10.3389/fpubh.2020.00014

Mbaye, M. F., Sarr, N., & Ngom, B. (2021). Construction of Control Charts for Monitoring Various Parameters Related to the Management of the COVID-19 Pandemic. *Journal of Biosciences and Medicines*, *09*(03), 9–19. https://doi.org/10.4236/jbm.2021.93002

Miller, B. G. (2011). *4 - The Effect of Coal Usage on Human Health and the Environment* (B. G. B. T.-C. C. E. T. Miller (ed.); pp. 85–132). Butterworth-Heinemann. https://doi.org/https://doi.org/10.1016/B978-1-85617-710-8.00004-2

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Mohamed, R. M. S. R., Nizam, N. M. S., Al-Gheethi, A. A., Lajis, A. & Kassim, A. H. M. (2016). Particulate Matter Levels in Ambient Air Adjacent to Industrial Area. *Soft Soil Engineering International Conference*, *136*. https://doi.org/10.1088/1757-899X/136/1/012056

Montgomery, G. C. (2013). Introduction to Statistical Quality Control, *John Wiley & Sons.*

Razali, N. M., & Yap, B. W. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 13–14.

Nerurkar, P., Shirke, A., Chandane, M., & Bhirud, S. (2018). Empirical Analysis of Data Clustering Algorithms. *Procedia Computer Science*, *125*, 770–779. https://doi.org/10.1016/j.procs.2017.12.099

Orcan, F. (2020). Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison. *International Journal of Assessment Tools in Education*, *7*(2), 236–246. https://doi.org/10.21449/ijate.656077

Palacio, L. C., Pachajoa, D. C., Durango-Giraldo, G., Zapata-Hernandez, C., Ugarte, J. P., Saiz, J., Buitrago-Sierra, R., & Tobón, C. (2021). Atrial proarrhythmic effect of lead as one of the PM10 metal components of air pollution. An in-silico study. *PLOS ONE*, *16*(10), e0258313. https://doi.org/10.1371/journal.pone.0258313

Rahman. S. Z. A., Ahad. N. A. & Zinzendoff. O. F. (2021). Status Of Air Quality Before And During Reinforcement Of Mco Due To Covid-19 Outbreak In Central And Southern Regions Of Peninsular Malaysia Malaysia. *Journal of Quality Measurement and Analysis*, *17*(2), 65–78.

Ramadhani, E., Mawengkang, H., & Ramli, M. (2019). Controlling Industrial Processes Using Multivariate Exponential Weighted Moving Average (Mewma). *International Journal of Recent Technology and Engineering (IJRTE)*, *Volume-7*(Issue-6S5), 1448–1453.

Refugio, C. N. (2018). An Empirical Study on Wilcoxon Signed Rank Test An Empirical Study on Wilcoxon Signed Rank Test. *Journal Of*, *December*, 12. https://doi.org/10.13140/RG.2.2.13996.51840

Ridwan, M. R., & Retnawati, H. (2021). Application of Cluster Analysis Using Agglomerative Method. *Numerical: Jurnal Matematika Dan Pendidikan Matematika*, *5*, 33–48.

Rose, J. J., Wang, L., Xu, Q., McTiernan, C. F., Shiva, S., Tejero, J., & Gladwin, M. T. (2017). Carbon monoxide poisoning: Pathogenesis, management, and future directions of therapy. *American Journal of Respiratory and Critical Care Medicine*, *195*(5), 596–

606. https://doi.org/10.1164/rccm.201606-1275CI

Saraçli, S., Doğan, N., & Doğan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, *2013*, 1–8. https://doi.org/10.1186/1029-242X-2013-203

Sasirekha, K. & Baby, P. (2013). Agglomerative Hierarchical Clustering Algorithm- A Review. *International Journal of Scientific and Research Publications*, *3*(3), 2–4.

Sensuła, B. M. (2016). The Impact of Climate, Sulfur Dioxide, and Industrial Dust on δ(18)O and δ(13)C in Glucose from Pine Tree Rings Growing in an Industrialized Area in the Southern Part of Poland. *Water, Air, and Soil Pollution*, *227*, 106. https://doi.org/10.1007/s11270-016-2808-0

Shan, G., Young, D., & Kang, L. (2014). A New Powerful Nonparametric Rank Test for Ordered Alternative Problem. *PLOS ONE*, *9*(11), e112924. https://doi.org/10.1371/journal.pone.0112924

Shima, M., & Adachi, M. (2000). Effect of outdoor and indoor nitrogen dioxide on respiratory symptoms in schoolchildren. *International Journal of Epidemiology*, *29*(5), 862–870. https://doi.org/10.1093/ije/29.5.862

Tai, A. P. K., & Val, M. (2017). Impacts of ozone air pollution and temperature extremes on crop yields : Spatial variability , adaptation and implications for future food security. *Atmospheric Environment*, *169*, 11–21. https://doi.org/10.1016/j.atmosenv.2017.09.002

Teegavarapu, R. S. V. (2019). *Chapter 1 - Methods for Analysis of Trends and Changes in Hydroclimatological Time-Series* (R. B. T.-T. and C. in H. V. Teegavarapu (ed.); pp. 1–89). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-810985-4.00001-3

Tunnicliffe, W. S., Hilton, M. F., Harrison, R. M., & Ayres, J. G. (2001). The effect of sulphur dioxide exposure on indices of heart rate variability in normal and asthmatic adults. *European Respiratory Journal*, *17*(4), 604–608. https://doi.org/10.1183/09031936.01.17406040

Utusan Borneo, (2019). Kronologi Jerebu di Malaysia. Retrieved from https://www.utusanborneo.com.my/2019/09/22/kronologi-jerebu-di-malaysia

Wong, P., & Wai, T. (2012). *A Study of the Air Pollution Index Reporting System*. *June*, 1–51.

Wu, Z., & Shamsuzzaman, M. (2005). Design and application of integrated control charts for monitoring process mean and variance. *Journal of Manufacturing Systems*, *24*(4), 302–314. https://doi.org/https://doi.org/10.1016/S0278-6125(05)80015-9

Xing, Y.-F., Xu, Y.-H., Shi, M.-H., & Lian, Y.-X. (2016). The impact of PM2.5 on the human respiratory system. *Journal of Thoracic Disease*, *8*(1), E69–E74. https://doi.org/10.3978/j.issn.2072-1439.2016.01.19

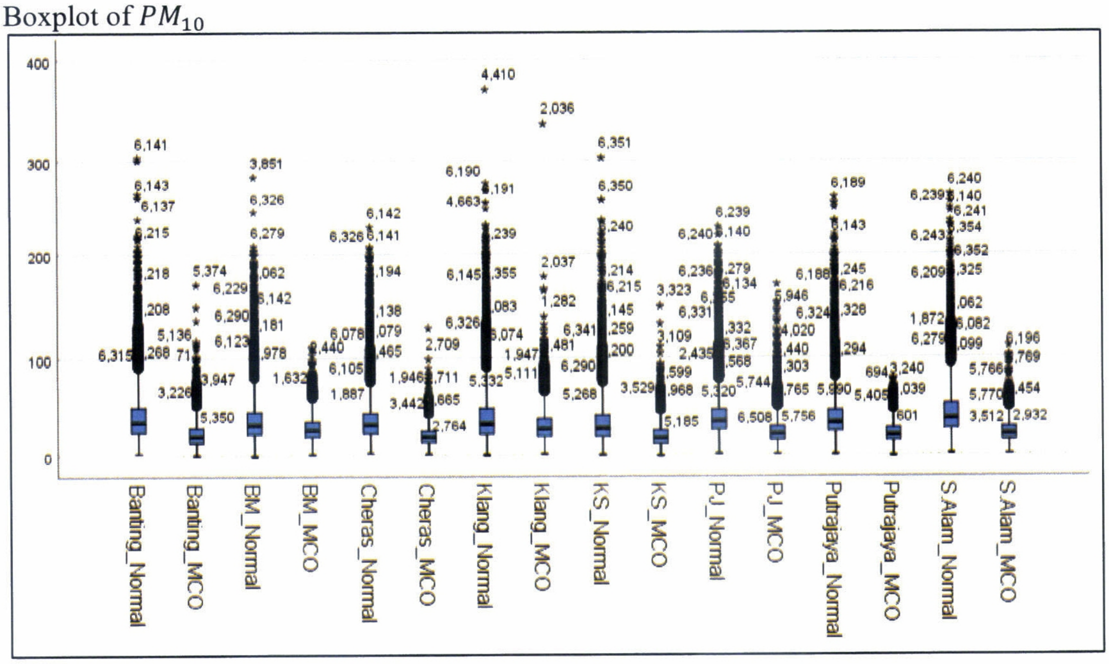
Yadav, P., Usha, K., & Singh, B. (2022). *Chapter 10 - Air pollution mitigation and global dimming: a challenge to agriculture under changing climate* (A. K. Shanker, C. Shanker, A. Anand, & M. B. T.-C. C. and C. S. Maheswari (eds.); pp. 271–298). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-816091-6.00015-8

Yim, O., & Ramdeen, K. T. (2015). Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *The Quantitative Methods for Psychology*, *11*(1), 8–21. https://doi.org/10.20982/tqmp.11.1.p008

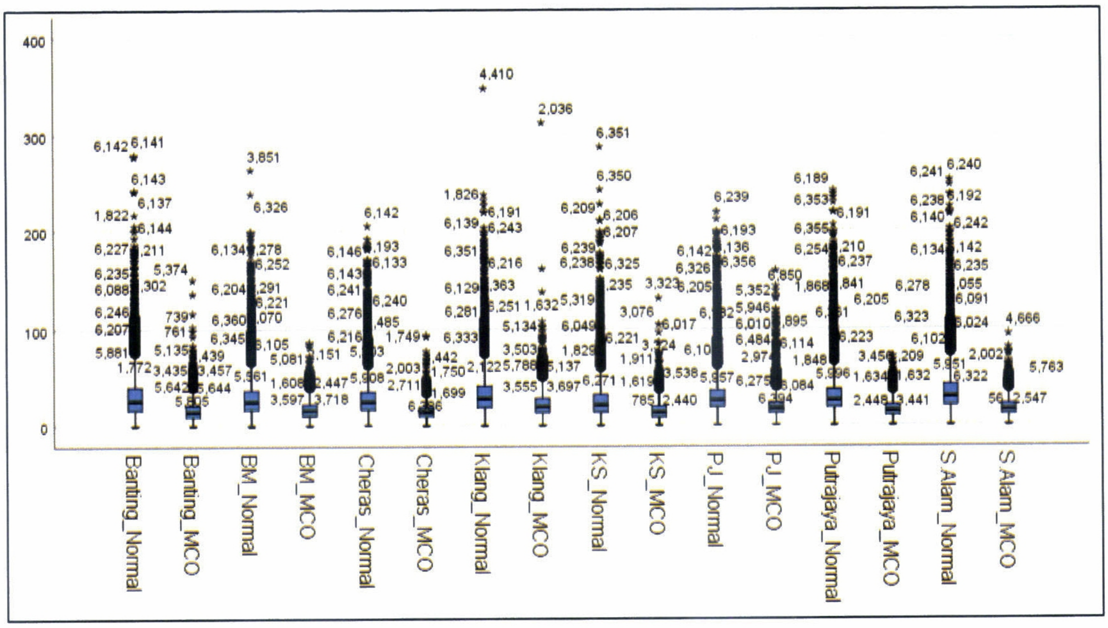
Zhang, J. J., Wei, Y., & Fang, Z. (2019). *Ozone Pollution : A Major Health Hazard Worldwide*. *10*(October), 1–10. https://doi.org/10.3389/fimmu.2019.02518

APPENDIX A: Boxplot of each air quality before and after implementation of MCO

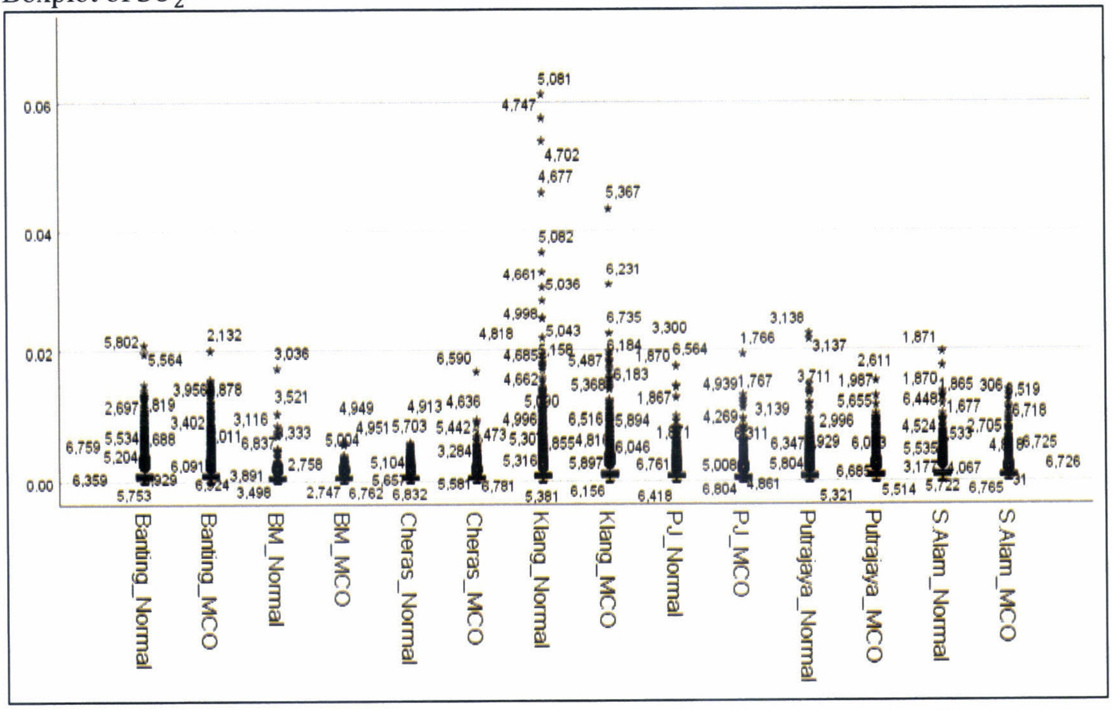
Boxplot of PM_{10}



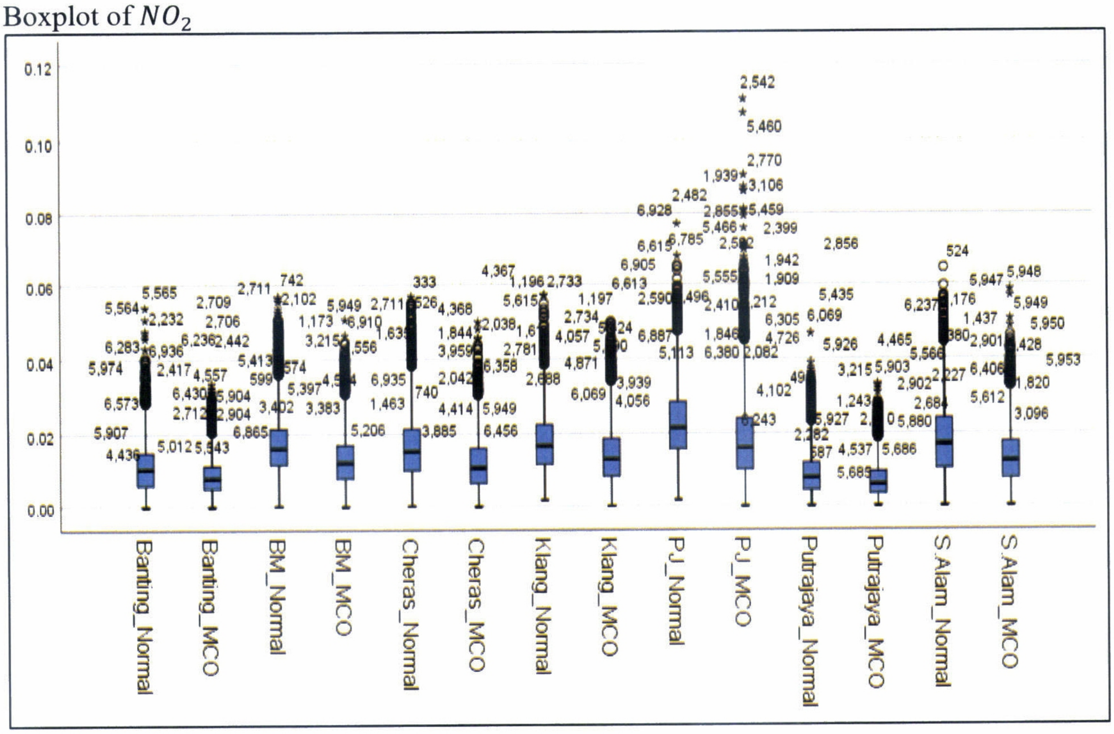
Boxplot of $PM_{2.5}$



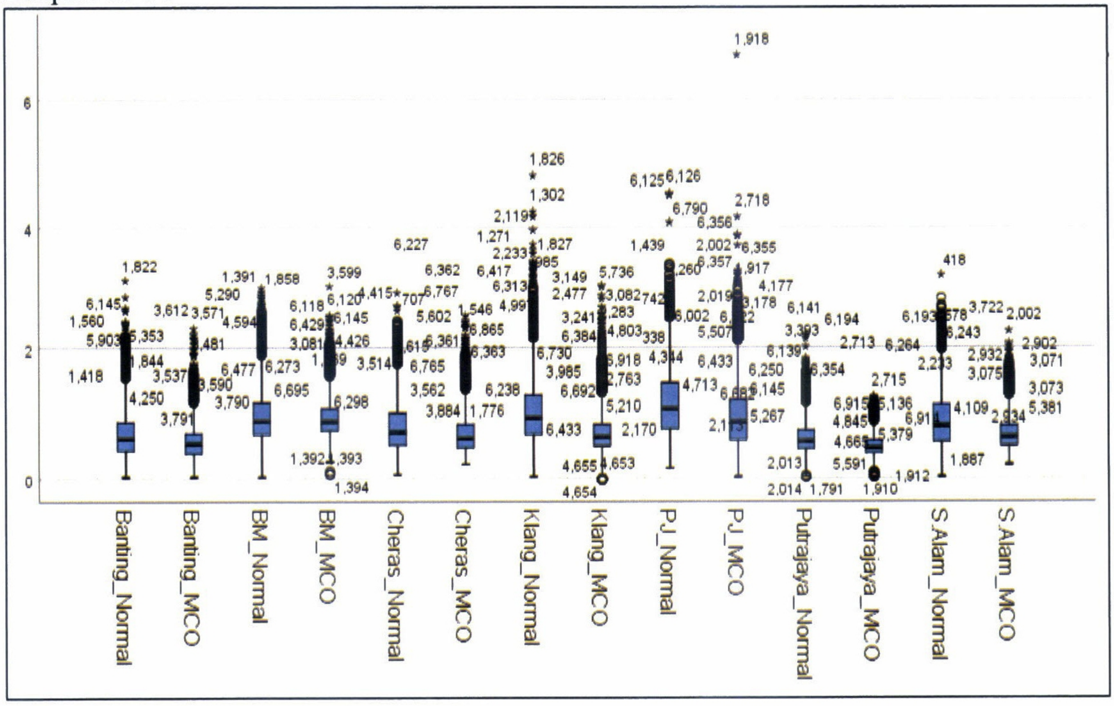
Boxplot of SO_2



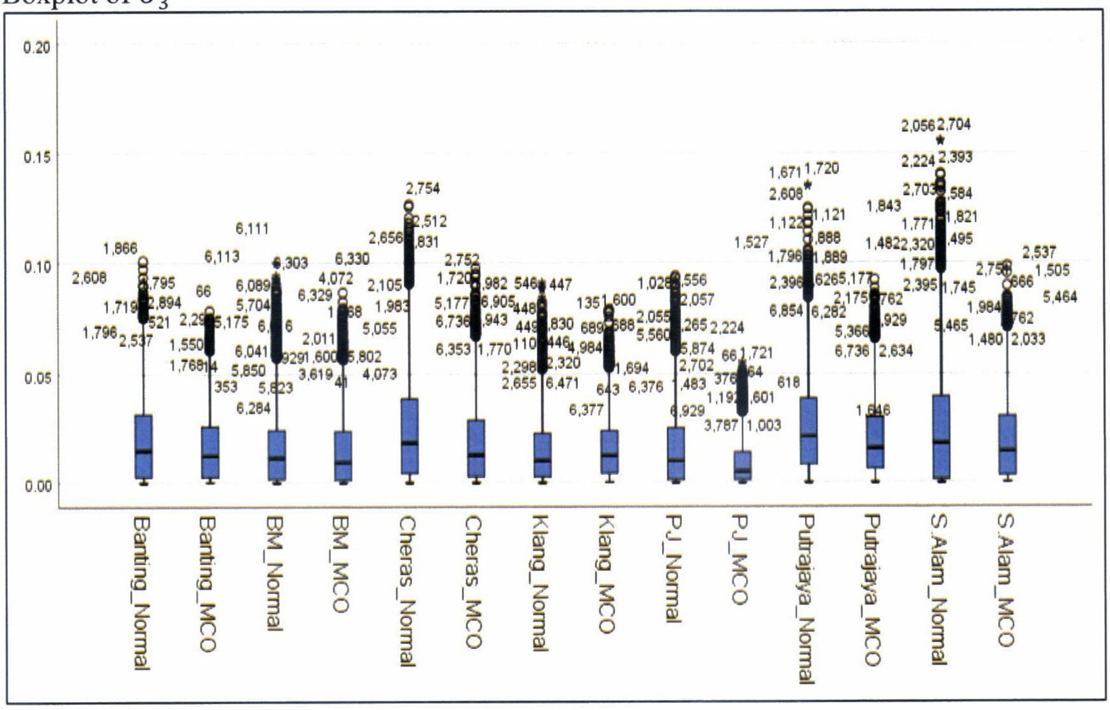
Boxplot of NO_2



Boxplot of CO



Boxplot of O_3



APPENDIX B: Results of Kolmogorov-Smirnov test and Wilcoxon Rank Sum test

Test Results of Kuala Selangor

Tests of Normality

		Kolmogorov-Smirnov[a]		
	duration	Statistic	df	Sig.
pm10	1.00	.152	10608	.000
	2.00	.094	6936	.000
pm2.5	1.00	.158	10608	.000
	2.00	.084	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a]

	pm10	pm2.5
Mann-Whitney U	28313027.500	28602334.000
Wilcoxon W	52370543.500	52659850.000
Z	-25.841	-24.959
Asymp. Sig. (2-tailed)	.000	.000

a. Grouping Variable: duration

Test Statistics[a,b]

	pm10	pm2.5
Kruskal-Wallis H	667.756	622.947
df	1	1
Asymp. Sig.	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Banting

Tests of Normality

	duration	Kolmogorov-Smirnov[a] Statistic	df	Sig.
pm10	1.00	.134	10608	.000
	2.00	.097	6936	.000
pm2.5	1.00	.143	10608	.000
	2.00	.100	6936	.000
so2	1.00	.208	10608	.000
	2.00	.254	6936	.000
no2	1.00	.056	10608	.000
	2.00	.071	6936	.000
o3	1.00	.146	10608	.000
	2.00	.147	6936	.000
co	1.00	.085	10608	.000
	2.00	.067	6936	.000

a. Lilliefors Significance Correction

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	22047698.000	25614185.000	33079424.000	31391854.500	34452596.000	30782912.500
Wilcoxon W	46105214.000	49671701.000	89349560.000	55449370.500	58510112.000	54840428.500
Z	-44.943	-34.069	-11.327	-16.454	-7.122	-18.311
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	2019.901	1160.728	128.295	270.739	50.727	335.277
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Batu Muda

Tests of Normality

	duration	Kolmogorov-Smirnov[a] Statistic	df	Sig.
pm10	1.00	.132	10608	.000
	2.00	.051	6936	.000
pm2.5	1.00	.152	10608	.000
	2.00	.056	6936	.000
so2	1.00	.138	10608	.000
	2.00	.116	6912	.000
no2	1.00	.056	10608	.000
	2.00	.053	6936	.000
o3	1.00	.163	10608	.000
	2.00	.178	6936	.000
co	1.00	.074	10608	.000
	2.00	.077	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	34858772.000	25475035.000	35215602.500	29141043.000	34607204.500	35466339.000
Wilcoxon W	58916288.000	49532551.000	91485738.500	53198559.000	58664720.500	91736475.000
Z	-5.884	-34.494	-4.428	-23.317	-6.651	-4.031
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

a. Grouping Variable: duration

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	34.618	1189.816	19.603	543.663	44.238	16.251
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Cheras

Tests of Normality

	duration	Kolmogorov-Smirnov[a] Statistic	df	Sig.
pm10	1.00	.141	10608	.000
	2.00	.044	6936	.000
pm2.5	1.00	.154	10608	.000
	2.00	.050	6936	.000
so2	1.00	.170	10608	.000
	2.00	.173	6936	.000
no2	1.00	.059	10608	.000
	2.00	.071	6936	.000
o3	1.00	.155	10608	.000
	2.00	.160	6936	.000
co	1.00	.084	10608	.000
	2.00	.107	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	20570214.000	21414911.000	36509516.500	26237029.500	32911534.000	32040599.000
Wilcoxon W	44627730.000	45472427.000	92779652.500	50294545.500	56969050.000	56098115.000
Z	-49.448	-46.873	-.854	-32.171	-11.821	-14.476
Asymp. Sig. (2-tailed)	.000	.000	.393	.000	.000	.000

a. Grouping Variable: duration

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	2445.105	2197.042	.729	1034.951	139.732	209.555
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.393	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Klang

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	29225680.000	26281575.500	27870978.000	26221254.500	32911444.000	21896615.000
Wilcoxon W	53283196.000	50339091.500	84141114.000	50278770.500	89181580.000	45954131.000
Z	-23.058	-32.035	-27.225	-32.219	-11.821	-45.404
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

a. Grouping Variable: duration

Tests of Normality

		Kolmogorov-Smirnov[a]		
	duration	Statistic	df	Sig.
pm10	1.00	.127	10608	.000
	2.00	.094	6936	.000
pm2.5	1.00	.134	10608	.000
	2.00	.095	6936	.000
so2	1.00	.279	10608	.000
	2.00	.198	6936	.000
no2	1.00	.053	10608	.000
	2.00	.063	6936	.000
o3	1.00	.160	10608	.000
	2.00	.129	6936	.000
co	1.00	.097	10608	.000
	2.00	.123	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	531.689	1026.219	741.222	1038.050	139.741	2061.520
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Petaling Jaya

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	20169682.0 00	24744886.5 00	26100261.5 00	25243775.0 00	31882607.0 00	26121854.0 00
Wilcoxon W	44227198.0 00	48802402.5 00	50157777.5 00	49301291.0 00	55940123.0 00	50179370.0 00
Z	-50.669	-36.720	-32.646	-35.199	-14.959	-32.522
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

a. Grouping Variable: duration

Tests of Normality

		Kolmogorov-Smirnov[a]		
	duration	Statistic	df	Sig.
pm10	1.00	.152	10608	.000
	2.00	.104	6936	.000
pm2.5	1.00	.165	10608	.000
	2.00	.116	6936	.000
so2	1.00	.176	10608	.000
	2.00	.186	6936	.000
no2	1.00	.063	10608	.000
	2.00	.079	6936	.000
o3	1.00	.189	10608	.000
	2.00	.185	6936	.000
co	1.00	.074	10608	.000
	2.00	.069	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	2567.366	1348.348	1065.744	1238.965	223.778	1057.657
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Putrajaya

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	22165044.500	22893064.500	29227247.500	32017949.000	31968936.000	25608235.000
Wilcoxon W	46222560.500	46950580.500	85497383.500	56075465.000	56026452.000	49665751.000
Z	-44.586	-42.366	-23.091	-14.545	-14.695	-34.088
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

a. Grouping Variable: duration

Tests of Normality

	duration	Kolmogorov-Smirnov[a] Statistic	df	Sig.
pm10	1.00	.150	10608	.000
	2.00	.064	6936	.000
pm2.5	1.00	.159	10608	.000
	2.00	.075	6936	.000
so2	1.00	.209	10608	.000
	2.00	.190	6936	.000
no2	1.00	.082	10608	.000
	2.00	.094	6936	.000
o3	1.00	.115	10608	.000
	2.00	.118	6936	.000
co	1.00	.086	10608	.000
	2.00	.060	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	1987.870	1794.867	533.203	211.566	215.929	1161.968
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000	.000

a. Kruskal Wallis Test

b. Grouping Variable: duration

Test Results of Shah Alam

Test Statistics[a]

	pm10	pm2.5	so2	no2	o3	co
Mann-Whitney U	19842029.000	20299892.000	36234789.000	29003462.500	35914520.000	27360835.000
Wilcoxon W	43733357.000	44191220.000	92504925.000	53060978.500	59972036.000	51418351.000
Z	-51.404	-50.005	-1.692	-23.736	-2.665	-28.744
Asymp. Sig. (2-tailed)	.000	.000	.091	.000	.008	.000

a. Grouping Variable: duration

Tests of Normality

		Kolmogorov-Smirnov[a]		
	duration	Statistic	df	Sig.
pm10	1.00	.118	10608	.000
	2.00	.062	6912	.000
pm2.5	1.00	.136	10608	.000
	2.00	.065	6912	.000
so2	1.00	.188	10608	.000
	2.00	.204	6936	.000
no2	1.00	.048	10608	.000
	2.00	.050	6936	.000
o3	1.00	.164	10608	.000
	2.00	.150	6936	.000
co	1.00	.071	10608	.000
	2.00	.080	6936	.000

a. Lilliefors Significance Correction

Test Statistics[a,b]

	pm10	pm2.5	so2	no2	o3	co
Kruskal-Wallis H	2642.389	2500.482	2.861	563.399	7.102	826.225
df	1	1	1	1	1	1
Asymp. Sig.	.000	.000	.091	.000	.008	.000

a. Kruskal Wallis Test

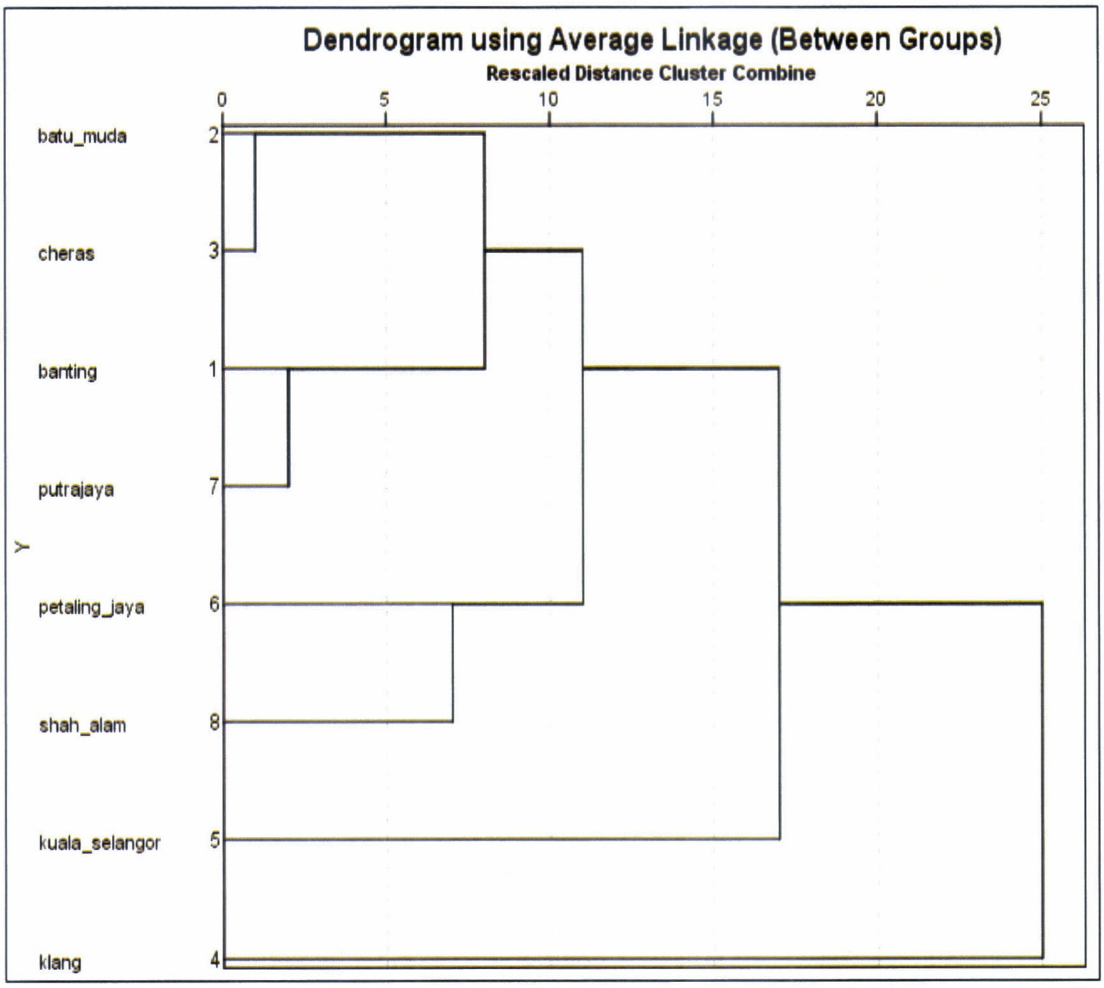
b. Grouping Variable: duration

APPENDIX C: Results of Hierarchical Agglomerative Clustering Analysis (HACA)

Result of clustering of PM_{10} before MCO

Agglomeration Schedule

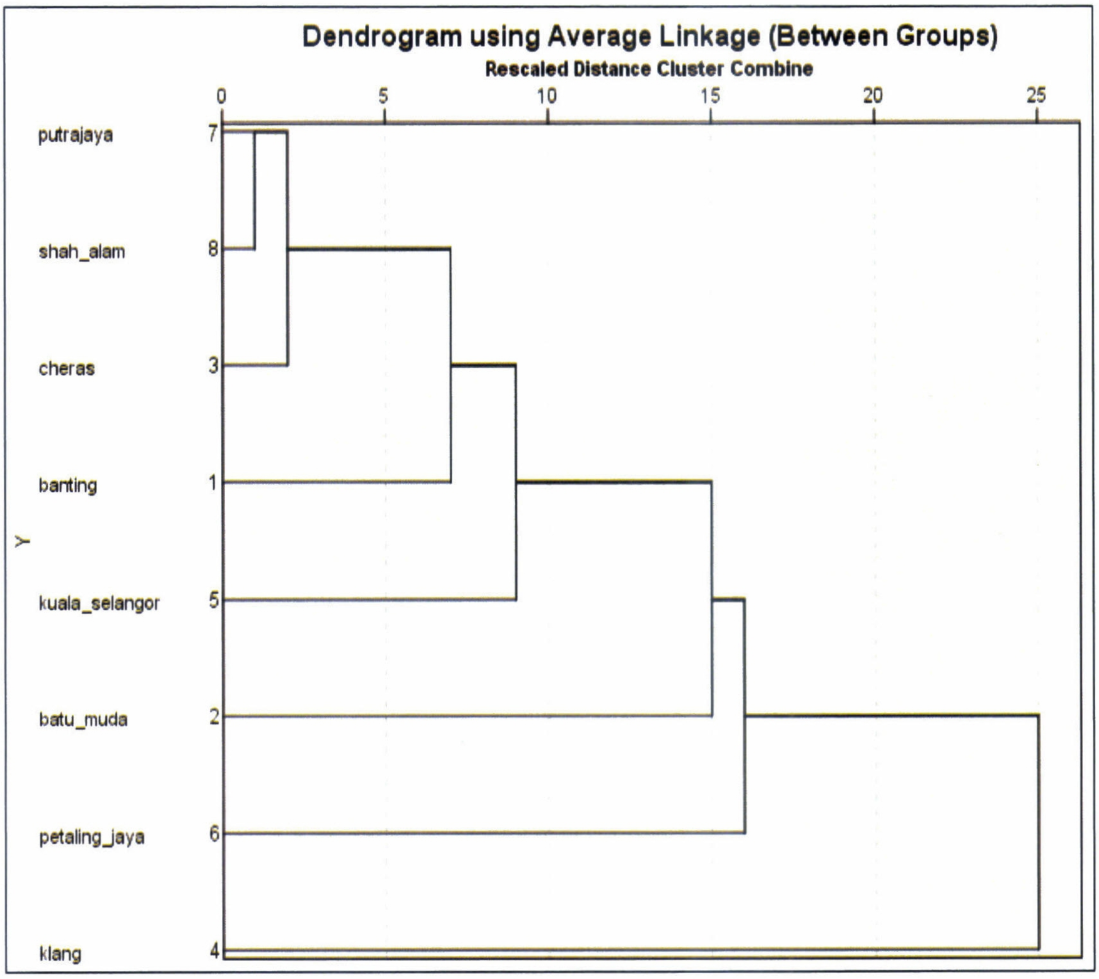
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	3	1759963.987	0	0	4
2	1	7	2021124.894	0	0	4
3	6	8	2693232.742	0	0	5
4	1	2	2783649.831	2	1	5
5	1	6	3170168.276	4	3	6
6	1	5	3967916.021	5	0	7
7	1	4	5127007.510	6	0	0



Result of clustering of PM_{10} after MCO

Agglomeration Schedule

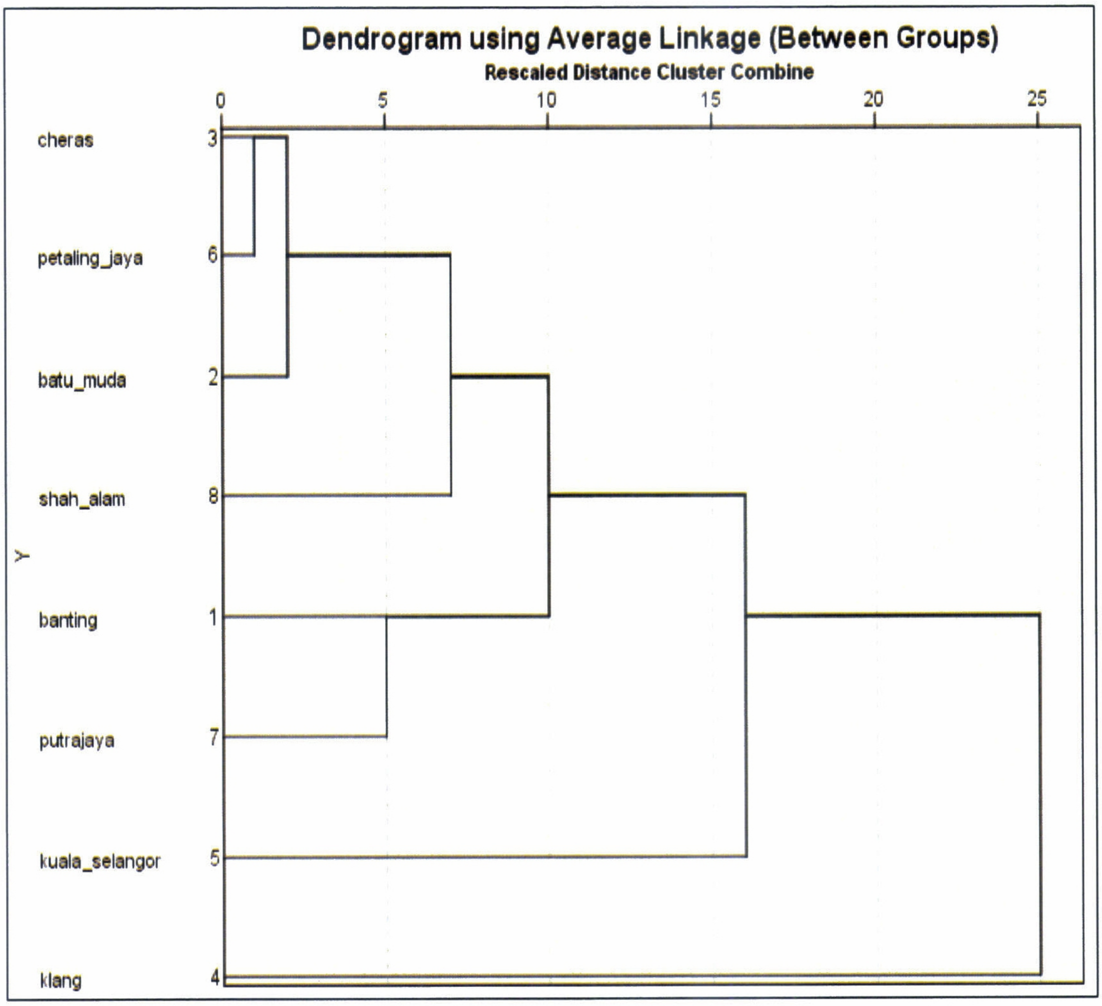
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	7	8	739144.977	0	0	2
2	3	7	807742.951	0	1	3
3	1	3	1159226.148	0	2	4
4	1	5	1336670.900	3	0	5
5	1	2	1698315.839	4	0	6
6	1	6	1767009.736	5	0	7
7	1	4	2424896.647	6	0	0



Result of clustering of $PM_{2.5}$ before MCO

Agglomeration Schedule

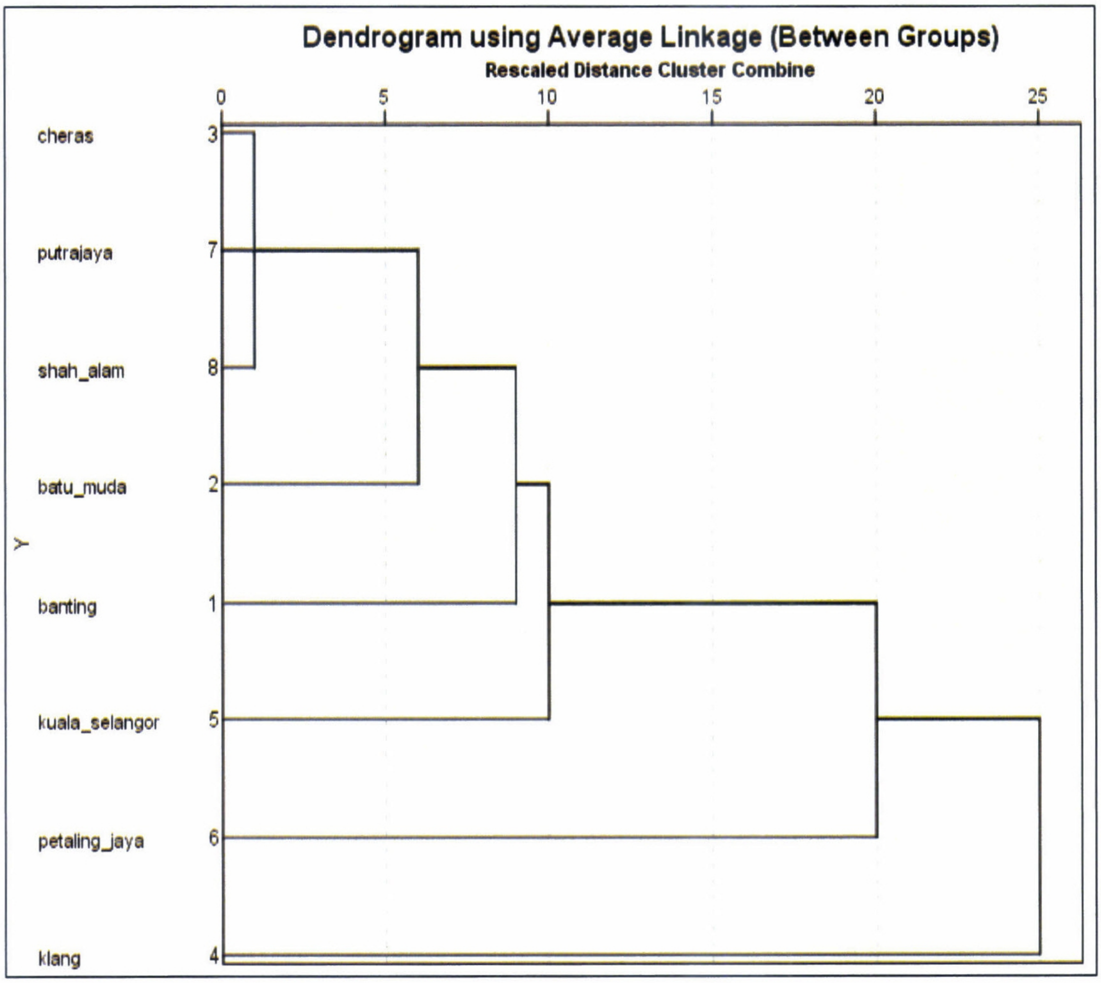
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	3	6	1284015.369	0	0	2
2	2	3	1468427.877	0	1	4
3	1	7	1666819.227	0	0	5
4	2	8	1881420.812	2	0	5
5	1	2	2211537.567	3	4	6
6	1	5	2723257.964	5	0	7
7	1	4	3652040.755	6	0	0



Result of clustering of $PM_{2.5}$ after MCO

Agglomeration Schedule

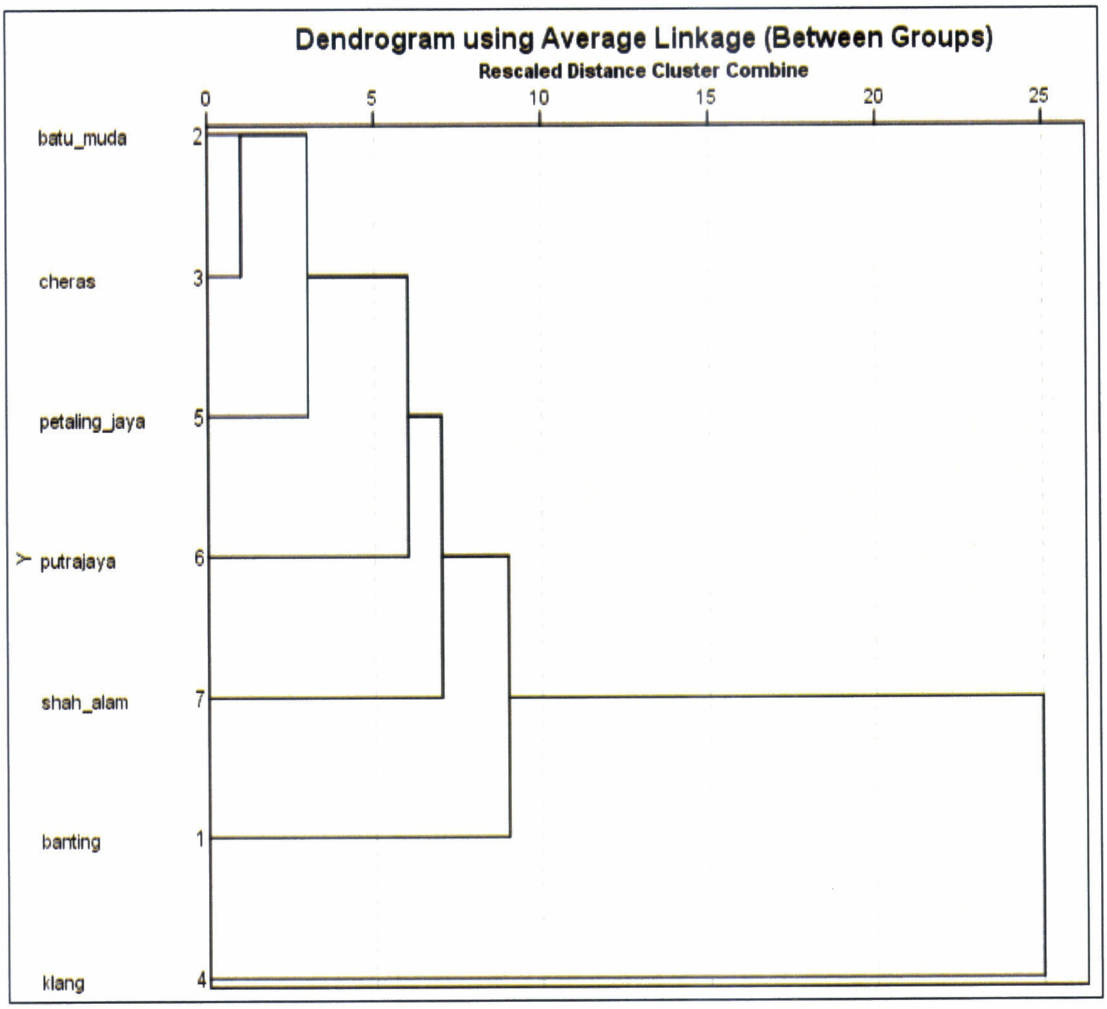
Stage	Cluster Combined: Cluster 1	Cluster Combined: Cluster 2	Coefficients	Stage Cluster First Appears: Cluster 1	Stage Cluster First Appears: Cluster 2	Next Stage
1	3	7	528807.332	0	0	2
2	3	8	545227.198	1	0	3
3	2	3	747996.042	0	2	4
4	1	2	870732.456	0	3	5
5	1	5	927612.642	4	0	6
6	1	6	1329852.566	5	0	7
7	1	4	1579873.148	6	0	0



Result of clustering of SO_2 before MCO

Agglomeration Schedule

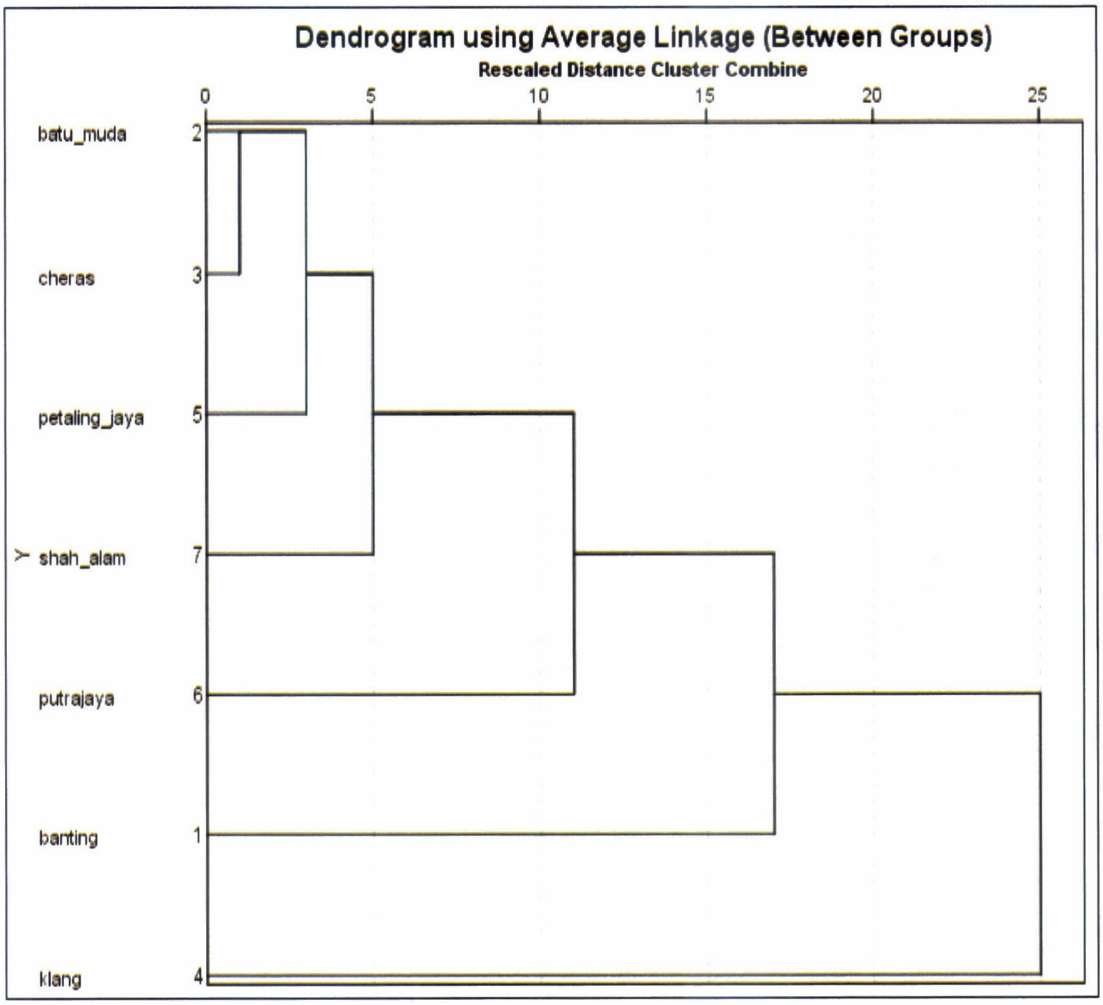
Stage	Cluster Combined: Cluster 1	Cluster Combined: Cluster 2	Coefficients	Stage Cluster First Appears: Cluster 1	Stage Cluster First Appears: Cluster 2	Next Stage
1	2	3	.006	0	0	2
2	2	5	.010	1	0	3
3	2	6	.014	2	0	4
4	2	7	.015	3	0	5
5	1	2	.018	0	4	6
6	1	4	.043	5	0	0



Result of clustering of SO_2 after MCO

Agglomeration Schedule

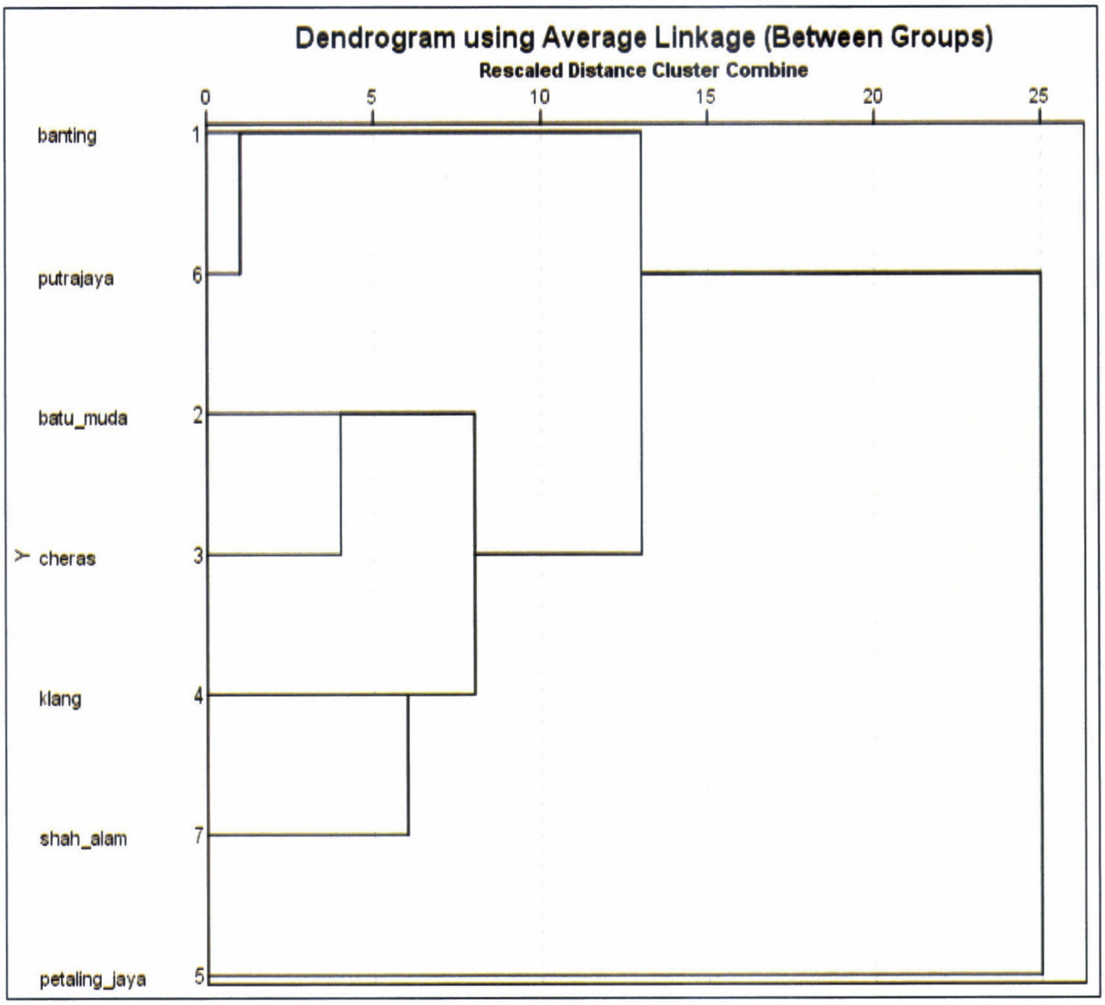
Stage	Cluster Combined: Cluster 1	Cluster Combined: Cluster 2	Coefficients	Stage Cluster First Appears: Cluster 1	Stage Cluster First Appears: Cluster 2	Next Stage
1	2	3	.003	0	0	2
2	2	5	.004	1	0	3
3	2	7	.006	2	0	4
4	2	6	.010	3	0	5
5	1	2	.014	0	4	6
6	1	4	.019	5	0	0



Result of clustering of NO_2 before MCO

Agglomeration Schedule

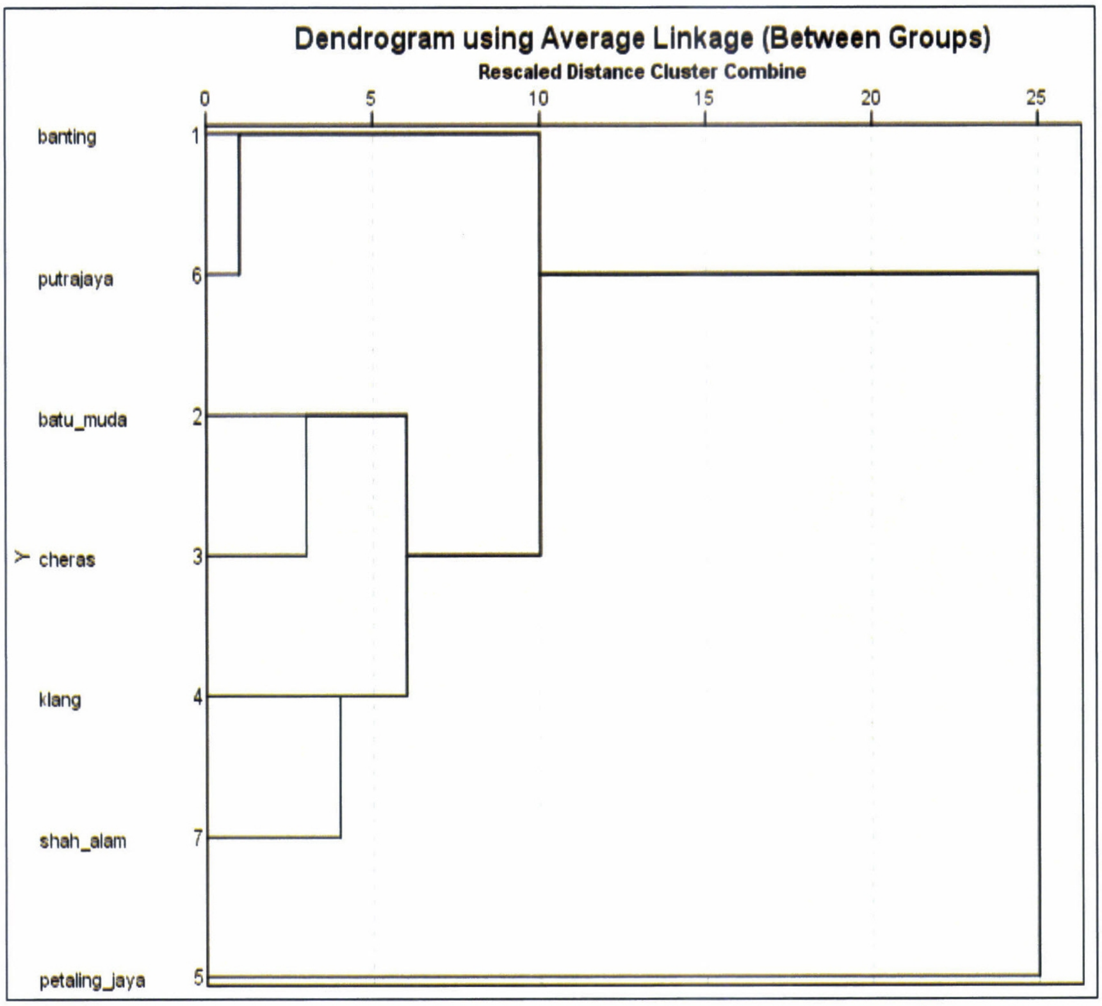
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	6	.398	0	0	5
2	2	3	.649	0	0	4
3	4	7	.784	0	0	4
4	2	4	.885	2	3	5
5	1	2	1.261	1	4	6
6	1	5	2.081	5	0	0



Result of clustering of NO_2 after MCO

Agglomeration Schedule

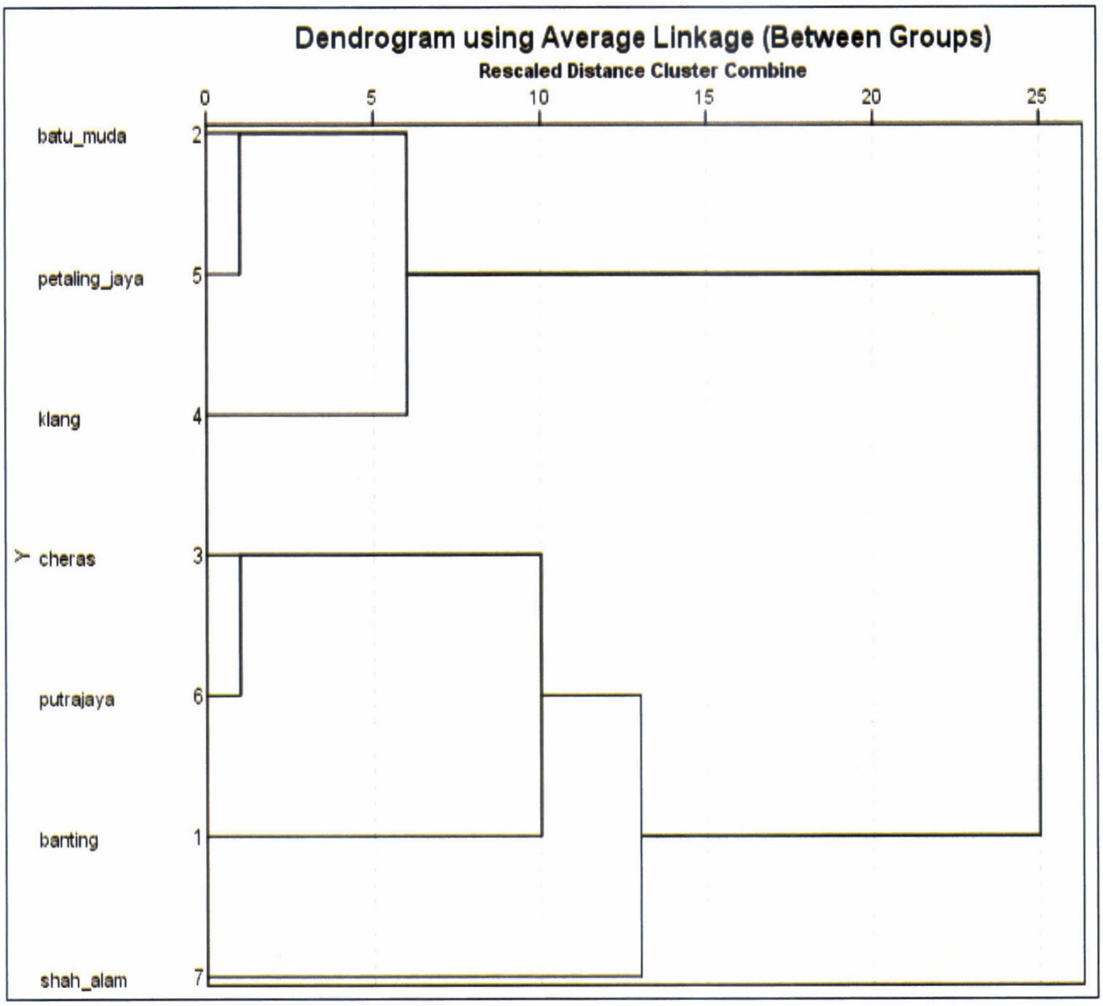
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	1	6	.169	0	0	5
2	2	3	.240	0	0	4
3	4	7	.296	0	0	4
4	2	4	.358	2	3	5
5	1	2	.507	1	4	6
6	1	5	1.042	5	0	0



Result of clustering of O_3 before MCO

Agglomeration Schedule

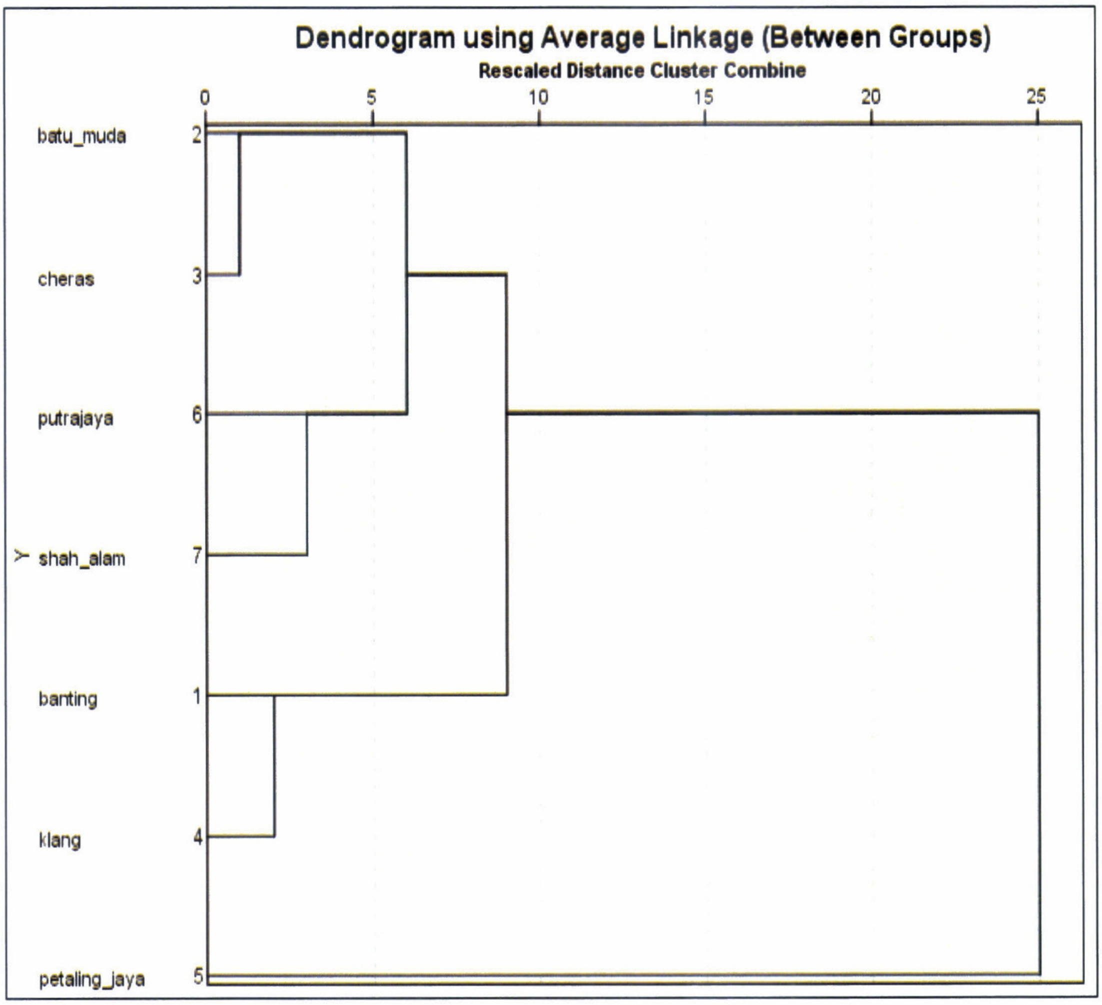
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	5	1.333	0	0	3
2	3	6	1.367	0	0	4
3	2	4	1.669	1	0	6
4	1	3	1.909	0	2	5
5	1	7	2.110	4	0	6
6	1	2	2.921	5	3	0



Result of clustering of O_3 after MCO

Agglomeration Schedule

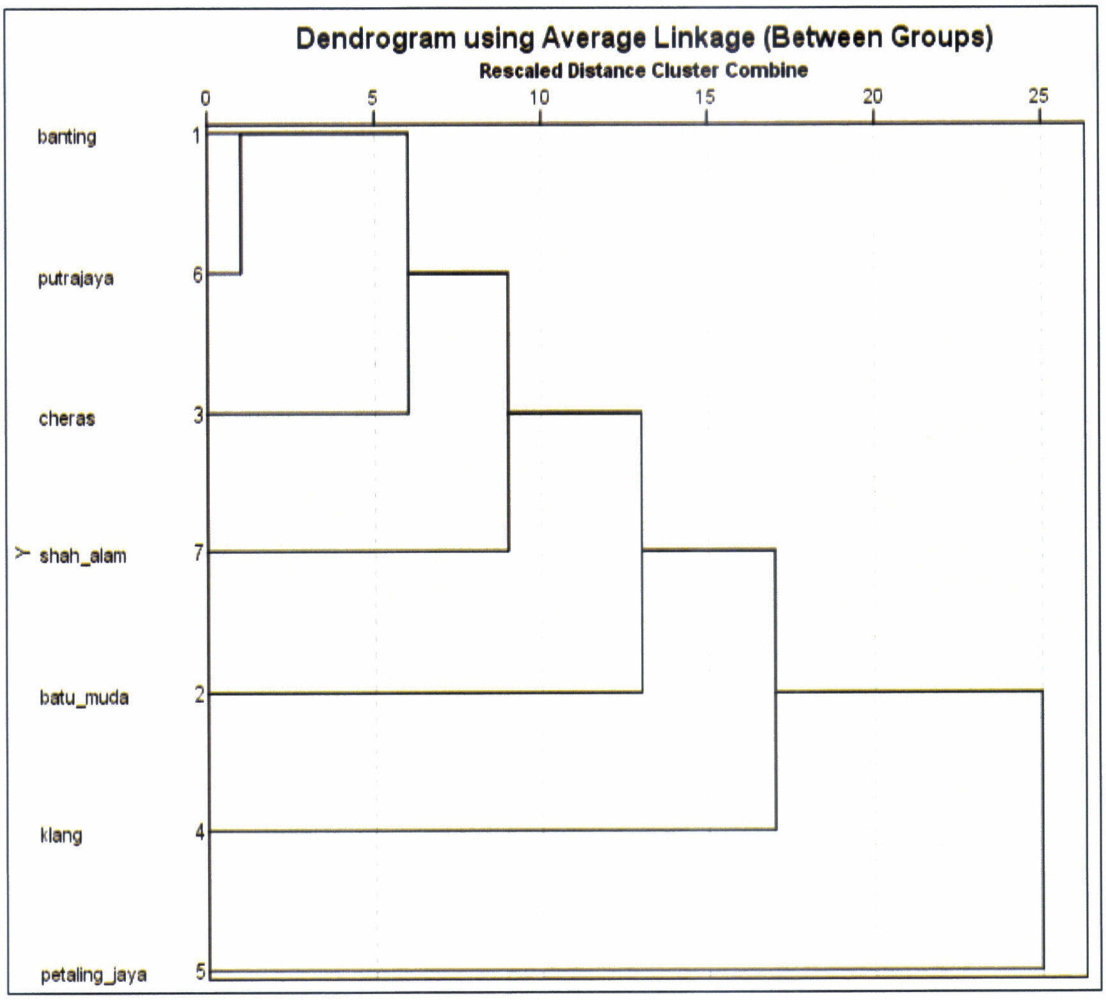
Stage	Cluster Combined: Cluster 1	Cluster Combined: Cluster 2	Coefficients	Stage Cluster First Appears: Cluster 1	Stage Cluster First Appears: Cluster 2	Next Stage
1	2	3	.540	0	0	4
2	1	4	.594	0	0	5
3	6	7	.611	0	0	4
4	2	6	.729	1	3	5
5	1	2	.817	2	4	6
6	1	5	1.329	5	0	0



Result of clustering of *CO* before MCO

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	1	6	645.329	0	0	2
2	1	3	1329.478	1	0	3
3	1	7	1718.551	2	0	4
4	1	2	2229.144	3	0	5
5	1	4	2807.457	4	0	6
6	1	5	3852.329	5	0	0



Result of clustering of *CO* after MCO

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	1	6	410.709	0	0	4
2	3	7	440.838	0	0	3
3	3	4	528.295	2	0	4
4	1	3	764.243	1	3	6
5	2	5	1182.434	0	0	6
6	1	2	1603.198	4	5	0