

**CIRCULAR STATISTICAL ANALYSIS OF WIND  
DIRECTION AND WIND SPEED ON OZONE  
CONCENTRATION**

**NOORAIN ZINOL**

**UNIVERSITI SAINS MALAYSIA**

**2017**

**CIRCULAR STATISTICAL ANALYSIS OF WIND DIRECTION  
AND WIND SPEED ON OZONE CONCENTRATION**

**by**

**NOORAIN ZINOL**

**Thesis submitted in fulfillment of the  
requirements for the degree of  
Master of Science**

**August 2017**

## ACKNOWLEDGEMENTS

First of all, Alhamdulillah. I am very grateful to Allah s.w.t that had giving me His guidance, barakah, strength and determination for me to complete this project. Managing and writing of this dissertation has been one of the most significant academic challenges I have ever had to face. This dissertation would never be completed without the assistance of several peoples and organizations. I would like to take this opportunity to present my appreciation for their contributions.

I must express my deepest appreciation to my supervisor, Professor Ahmad Shukri Yahaya for his guidance, patience, and cooperation throughout this study. His valuable suggestions and encouragement enabled me to handle this project with confidence. This thesis would not have been possible without his vision and direction. I am very fortunate for having the opportunity to work with him and I will cherish these memories for my entire life. This dissertation would not have been possible without the constant support and encouragement of my family and friends especially during the difficult times faced during this research.

Deepest thanks to Jabatan Alam Sekitar Malaysia for the data used to carry out this research. Thank you all and I pray that The Almighty will grant each of you the happiness here and hereafter. My grateful thanks go to my mother, Puan Fatimah who paid my tuition fees so that I could pursue on my studies. Although she hardly knows how writing thesis is and understand what I research on, but she willing to support any decision I make. Her love in supporting me and keeping me in their prayers is truly appreciated. Last but not least thanks to my beloved fiancé Hasrul, always urging me to finish this writing as soon as possible so that we can get marry end of this year.

## TABLE OF CONTENTS

	<b>Page</b>
<b>ACKNOWLEDGEMENTS</b>	ii
<b>TABLE OF CONTENTS</b>	iii
<b>LIST OF TABLES</b>	viii
<b>LIST OF FIGURES</b>	x
<b>LIST OF SYMBOLS</b>	xii
<b>LIST OF ABBREVIATIONS</b>	xiii
<b>ABSTRAK</b>	xvi
<b>ABSTRACT</b>	xviii
<b>CHAPTER ONE – INTRODUCTION</b>	
1.1 General	1
1.2 Environmental Air Pollution In Malaysia	3
1.3 Circular Statistics	7
1.4 Problem Statement	9
1.5 Research Objectives	11
1.6 Scope of Study	11
1.7 Brief Methodology	12
1.8 Thesis Layout	13
<b>CHAPTER TWO – LITERATURE REVIEW</b>	
2.1 Background	14
2.2 Ozone as a Secondary Pollutant	15
2.2.1 Effect of Ozone	16

2.2.2	Ozone precursors	18
2.3	The Influence of Meteorological Parameters	20
2.3.1	Influence of Wind Direction and Wind Speed	20
2.4	Circular Distributions	23
2.5	Circular Regression and Correlation	25
2.6	Multivariate Analysis	27
2.6.1	Multiple Linear Regression (MLR)	28
2.7	Summary	29

### **CHAPTER THREE – METHODOLOGY**

3.1	General	30
3.2	Site Description and Measurement Techniques	32
3.2.1	Study Area	32
3.2.2	Weather Condition of Study Area	35
3.2.3	Measurement Techniques	36
3.2.3 (a)	Ground Level Ozone	36
3.2.3 (b)	Meteorological Parameters	38
3.2.4	Missing Records	39
3.3	Linear Data - Ozone Analysis	39
3.3.1	Box Plot	39
3.3.2	The Descriptive Analysis	40
3.3.2 (a)	Mean	41
3.3.2 (b)	Standard Deviation	41
3.3.2 (c)	Minimum and Maximum	42
3.3.2 (d)	Median	42

3.4	Circular – Wind Direction Analysis	43
3.4.1	Circular	43
3.4.1 (a)	Raw Circular Data Plot	44
3.4.1 (b)	Wind Rose	45
3.4.1 (c)	Kernel Density Estimate	46
3.4.2	The Descriptive Analysis	47
3.4.2 (a)	Mean	48
3.4.2 (b)	Mean Resultant Length	49
3.4.2 (c)	Circular Dispersion	49
3.4.2 (d)	Concentration, $\kappa$ using (MLE)	50
3.5	Fitting Distributions	51
3.5.1	Von Mises	51
3.5.2	Wrapped Normal	52
3.5.3	Wrapped Cauchy	53
3.5.4	Cardioid	54
3.5.5	Data Uniformity Test	54
3.5.5 (a)	Rayleigh	55
3.5.5 (b)	Kuiper's	55
3.5.5 (c)	Rao's Spacing	56
3.5.5 (d)	Range	57
3.5.6	Performance Indicator	58
3.5.6 (a)	Mean Chord Length	58
3.6	Correlation	60
3.6.1	Linear – Linear Correlation	60
3.6.2	Circular – Linear Correlation	61

3.7	Regression	62
3.7.1	Regression for Linear Response and a Circular Regressor	63
3.7.1 (a)	Local Linear Estimator	63
3.7.1 (b)	The Nadaraya - Watson estimator	64
3.7.2	Multivariate Regression with Circular Regressor	65
3.7.3	Performance Indicator	66
3.7.3 (a)	The Coefficient of Determination $R^2$	66
3.7.3 (b)	Index of Agreement IA	67
3.7.3 (c)	Prediction Accuracy PA	67
3.7.3 (d)	The Normalized Absolute Error NAE	68
3.7.3 (e)	The Root Mean Square Error RMSE	68
3.8	Summary	69

## **CHAPTER FOUR – RESULTS AND DISCUSSIONS**

4.1	Introduction	70
4.2	Box Plot and Descriptive Statistics of Ozone	70
4.3	Circular - Wind Directions	77
4.3.1	Circular Density and Wind Rose	78
4.3.2	Descriptive Statistics of Wind Directions	89
4.4	Probability Distributions	93
4.4.1	Uniformity Test for Circular Wind Directions	93
4.4.2	Circular Distribution	96
4.4.3	Best Distribution	106
4.5	Correlation & Regression	108
4.5.1	Johnson-Wherly-Mardia Correlation Coefficient	108

4.5.2	Multivariate Circular – Linear Regression	114
4.5.3	Multiple Linear Regression	124
4.5.4	Performance Indicator	127
4.6	Summary	130

## **CHAPTER FIVE - CONCLUSIONS AND RECOMMENDATIONS**

5.1	Conclusions	131
5.2	Recommendations	133

<b>REFERENCES</b>	<b>135</b>
-------------------	------------

## **APPENDICES**

Appendix A :	Summary of MLR models for O <sub>3</sub> concentration prediction in Shah Alam
Appendix B :	Summary of MLR models for O <sub>3</sub> concentration prediction in Prai
Appendix C :	Summary of MLR models for O <sub>3</sub> concentration prediction in Pasir Gudang
Appendix D :	Summary of MLR models for O <sub>3</sub> concentration prediction in Jerantut



## LIST OF TABLES

	<b>Page</b>	
Table 1.1	New Malaysia Ambient Air Quality Standard	6
Table 2.1	Different types of circular regression model	26
Table 2.2	Review prediction models using circular – linear and MLR	28
Table 3.1	Description of the selected monitoring stations	34
Table 3.2	Upper Quantiles of $V_n^*$	56
Table 3.3	Strength of correlation	62
Table 3.4	Errors and Accuracy Measure	66
Table 4.1	Circular descriptive statistics of wind direction for Shah Alam	89
Table 4.2	Circular descriptive statistics of wind direction for Prai	90
Table 4.3	Circular descriptive statistics of wind direction for Pasir Gudang	91
Table 4.4	Circular descriptive statistics of wind direction for Jerantut	91
Table 4.5	Result of uniformity test for DT	94
Table 4.6	Result of uniformity test for NT	95
Table 4.7	The Mean chord length for distribution	107
Table 4.8	Correlation in Shah Alam	110
Table 4.9	Correlation in Prai	111
Table 4.10	Correlation in Pasir Gudang	112
Table 4.11	Correlation in Jerantut	113

Table 4.12	The best MLR model for DT and NT model at all locations	125
Table 4.13	Error and accuracy measures for MLR model	129

## LIST OF FIGURES

		<b>Page</b>
Figure 1.1	Number of registered vehicles in Malaysia.	6
Figure 3.1	Flow of research methodologies	31
Figure 3.2	Selected location of monitoring stations across Peninsular Malaysia	33
Figure 3.3	010C wind speed sensor (left) and 020C wind direction sensor (right)	38
Figure 3.4	Element of Box and Whisker Plot	40
Figure 3.5	Linear plot (left) and raw circular data plot (right) for the 310 wind directions.	44
Figure 3.6	Raw circular data plot in combination with a rose diagram in the center for 310 wind directions data.	45
Figure 3.7	Raw circular data plots and rose diagrams of the 310 wind directions together with (left) three kernel density estimates with bandwidths of 10 (dotted), 40 (solid) and 75 (dashed).	46
Figure 3.8	Mean Direction	48
Figure 3.9	Chord Length	59
Figure 4.1	Box plot and descriptive statistics for Shah Alam	73
Figure 4.2	Box plot and descriptive statistics for Prai	74
Figure 4.3	Box plot and descriptive statistics for Pasir Gudang	75
Figure 4.4	Box plot and descriptive statistics for Jerantut	76
Figure 4.5	Rose diagrams with wind direction density (dotted green line) for Shah Alam daytime (left panel) and night time (right panel)	81
Figure 4.6	Rose diagrams with wind direction density (dotted green line) for Prai DT (left panel) and NT (right panel)	83
Figure 4.7	Rose diagrams with wind direction density (dotted green line) for Pasir Gudang DT (left panel) and NT (right panel)	85

Figure 4.8	Rose diagrams with wind direction density (dotted green line) for Jerantut DT (left panel) and NT (right panel)	87
Figure 4.9	Fitting distribution for Shah Alam daytime (left panel) and night time (right panel). Specifically wrapped normal (red), von Mises (dark blue), wrapped Cauchy (blue), and cardioid (yellow)	98
Figure 4.10	Fitting distribution for Prai DT (left panel) and NT (right panel). Specifically wrapped normal (red), von Mises (dark blue), wrapped Cauchy (blue), and cardioid (yellow)	100
Figure 4.11	Fitting distribution for Pasir Gudang DT (left panel) and NT (right panel). Specifically wrapped normal (red), von Mises (dark blue), wrapped Cauchy (blue), and cardioid (yellow)	102
Figure 4.12	Fitting distribution for Jerantut DT (left panel) and NT (right panel). Specifically wrapped normal (red), von Mises (dark blue), wrapped Cauchy (blue), and cardioid (yellow)	104
Figure 4.13	Shah Alam : circular representations of the Nadaraya-Watson estimator (red line) and local linear estimator (green line) with cross-validation smoothing parameter for ozone concentration (ppm) with respect to wind direction	116
Figure 4.14	Prai: Circular representations of the Nadaraya-Watson estimator (red line) and local linear estimator (green line) with cross-validation smoothing parameter for ozone concentration (ppm) with respect to wind direction.	118
Figure 4.15	Pasir Gudang: Circular representations of the Nadaraya-Watson estimator (red line) and local linear estimator (green line) with cross-validation smoothing parameter for ozone concentration (ppm) with respect to wind direction.	120
Figure 4.16	Jerantut: Circular representations of the Nadaraya-Watson estimator (red line) and local linear estimator (green line) with cross-validation smoothing parameter for ozone concentration (ppm) with respect to wind direction.	122

## LIST OF SYMBOLS

$\lambda$	Sunlight
$\theta$	Angle
$R^2$	Coefficient of determination
$\bar{x}$	Mean
$S$	Standard Deviation
$m$	Median
$\mu$	Mean Angle
$\rho$	Mean Resultant Length
$V$	Circular Variance
$v$	Circular Standard Deviation
$\delta$	Circular Dispersion
$\kappa$	Concentration
$Z$	Rayleigh Test
$V_n^*$	Kuiper's Test
$U$	Rao's Spacing Test
$w$	Range Statistics Test
$r$	Linear - Linear Correlation
$r_{x\theta}$	Circular – Linear
$bw$	bandwidth

## LIST OF ABBREVIATIONS

ANN	Artificial neural network
API	Air pollutant Index
CD	Cloud Direction
CO	Carbon Monoxide
CO <sub>2</sub>	Carbon Dioxide
DoE	Department of Environment (Malaysia)
DT	Daytime
DU	Dobson Unit
E	East
HO <sub>2</sub>	Hydroperoxy
h $\nu$	sunlight
IA	Index of Agreement
LL	Local Linear estimator
MAAQG	Malaysia Ambient Air Quality Guideline
MLR	Multiple Linear Regression
N	North
NAE	Normalized absolute error
NE	North East
NEM	Northeast Monsoon
NO	Nitrogen Oxide
NO <sub>2</sub>	Nitrogen Dioxide
NT	Nighttime
NW	Nadaraya-Watson estimator

NW	North West
O <sub>3</sub>	Ground Level Ozone
OH	Oxygen Hydroxide
P	Pressure
PA	Prediction Accuracy
PM	Particulate Matter
PM <sub>10</sub>	10 Micron
PM <sub>2.5</sub>	2.5 Micron
ppb	parts per billion
ppm	parts per million
R	Software Environment for Statistical Computing and Graphics
R <sup>2</sup>	Coefficient of Determination
RF	Rainfall
RH	Relative Humidity
RMSE	Root mean square error
RO <sub>2</sub>	Peroxy radicals
RSPM	Respiratory Suspended Particulate Matter
S	South
SE	South East
SO <sub>2</sub>	Sulphur Dioxide
SPSS	Statistical Package for the Social Science
SR	Solar Radiation
SW	South West
SWM	Southwest Monsoon
Temp	Temperature

UV	Ultraviolet
VOCs	Volatile Organic Compounds
W	West
WD	Wind Speed
WS	Wind Speed



# **ANALISIS BERSTATISTIK MEMBULAT BAGI ARAH DAN KELAJUAN ANGIN TERHADAP KEPEKATAN OZON**

## **ABSTRAK**

Perkembangan pembangunan bersama dengan peningkatan penduduk dan peningkatan penggunaan pengangkutan telah menyumbang dalam menurunkan kualiti udara di Malaysia. Tujuan kajian ini adalah untuk meramal dan memodelkan hubungan diantara pembolehubah linear dan pembolehubah membulat dengan menggunakan fungsi ketumpatan kebarangkalian dan analisis regresi. Pemerhatian dilakukan di stesen Shah Alam, Prai, Pasir Gudang dan Jerantut dalam tempoh sepuluh tahun dari tahun 2004 hingga 2013. Nilai min bagi kepekatan  $O_3$  untuk semua lokasi tidak melebihi had MAAQG. Ketumpatan membulat dan rajah angin telah digunakan untuk menunjukkan ciri arah angin termasuk arah min, min panjang paduan dan penumpuan. Kajian ini mencadangkan kehadiran angin yang menonjol telah bertiup dari utara dan arah selatan Semenanjung Malaysia disebabkan oleh monsun timur laut dan monsun barat daya. Empat fungsi taburan disuaikan dengan data arah angin dan telah mendapati bahawa taburan Cauchy membungkus menunjukkan prestasi yang baik berdasarkan analisis purata panjang perentas. Analisis diteruskan dengan korelasi linear-membulat, regresi linear-membulat dan regresi linear berganda. Kebanyakan korelasi positif ditemui antara arah angin dan kepekatan  $O_3$  di semua stesen pada waktu siang dan malam. Dari regresi linear-membulat, kepekatan  $O_3$  telah diramalkan dari arah angin dan keputusan menunjukkan bahawa beberapa bacaan di stesen Shah Alam melebihi had yang telah dibenarkan MAAQG (0.06 ppm). Model ramalan diteruskan dengan regresi linear berganda dengan memasukkan sinus dan kosinus arah angin sebagai pembolehubah

tak bersandar berserta dengan kelajuan angin. Didapati bahawa model ramalan sebelah malam memberikan ketepatan yang tinggi berbanding dengan waktu siang.

# **CIRCULAR STATISTICAL ANALYSIS OF WIND DIRECTION AND WIND SPEED ON OZONE CONCENTRATION**

## **ABSTRACT**

The rapid development along with increased population and transportation usage has deteriorated air quality status in Malaysia. The purpose of this research is to predict and modelling the casual relationship between circular and linear variables using probability density functions and regression analysis. The observations were made at Shah Alam, Prai, Pasir Gudang and Jerantut station over ten years period from 2004 until 2013 respectively. Mean values for O<sub>3</sub> concentrations at all locations did not exceed MAAQG level limit. Circular density and wind rose plot were used to describe the characteristics of wind direction including mean direction, mean resultant length and concentration. This study suggested the presence of prominent wind that were blowing from north and south direction of Peninsular Malaysia due to southwest monsoon and northeast monsoon. Four distributions function were fitted to the wind direction data and it was found that wrapped Cauchy fit the data very well based on mean chord length. The analysis was further conducted with circular-linear correlation, circular-linear regression and multi linear regression model. Most of positive correlations were found between wind direction and O<sub>3</sub> concentration in all stations during daytime and nighttime. From the circular – linear regression, O<sub>3</sub> concentration were predicted from wind direction and the results showed that few values in Shah Alam station exceeded the permissible limit by MAAQG (0.06 ppm). Prediction model were continued with multi linear regression by including the sine and cosine of wind direction as independent variables as well as wind speed variable. It was found that nighttime models provide better accuracy than daytime.

## CHAPTER ONE

### INTRODUCTION

#### 1.1 General

Air pollution is a term to explain the contamination of the air with harmful or poisonous substances that may harm animals, vegetation, humans, and materials (DEP, 2017). Emissions of unwanted chemicals gases or substances which exceeds the capacity of natural processes to convert or disperse them, can result in the degradation of air quality (Schwela and Haq, 2008). Pollutants emissions resulted from direct air emissions or through the production of secondary pollutants as a result of chemical reactions will take place in the air (Mohammed et al., 2013).

The goal of achieving industrial country status by the year 2020 and the associated rapid economic growth have started to impose costs in terms of industrial pollution and the degradation of urban environment (Begum et al., 2013). Depletion of fisheries, air and water pollution, and contamination by industrial wastes have become more serious in Malaysia in recent years. Among them, air pollution is the major issue that has been affecting human health, agricultural crops, forest species, and ecosystems.

Concentrations of air pollutants in Asia are higher than those are in Europe and North America, and levels are expected to worsen over time (Akimoto, 2003). As more air pollutants are emitted, ozone concentrations in East Asia have correspondingly increased. Created substantially by secondary production, PM<sub>2.5</sub> levels have reached a critical level, with levels frequently exceeding environmental standards (Shim, 2017).

According to data released by the UN Agency World Health Organization, Malaysia has become one of the least air polluted urban environments in Asia compared to Beijing, New Delhi and Jakarta (Rohde and Muller, 2015). Nevertheless, air pollutant might travelled miles from the origin driven by wind affected the neighborhood country. Ramanathan and Feng (2009) claimed that each part of the world is connected through fast atmospheric transport which open possibility for pollutant to travel at longer distance.

Fuel combustion is a key to air pollution source in Asian cities which tends to increase with population size along with economic activities. Major source of air pollution contributing to at least 70%–75% of the total air pollution is emission from mobile source such as motor vehicles. Another 20%-25% of the air pollution were contributed from stationary source such as power plant, factories and etc. However forest fire (open burning) contributed 3%-5% of total air pollution in Malaysia for the past five years (Afroz et al., 2003; Yahaya et al., 2006 and Azid et al., 2015).

Under those circumstances, Malaysia and Singapore has been affected by ‘haze’ caused by open burning of forests from neighborhood country, Indonesia. Repeated haze episodes in Southeast Asia recently has become an annual phenomenon during the drier months. On 24 September 2015, API reading had crossed into the ‘Very Unhealthy’ range on its way to a ‘Hazardous’ rating with value between 264-321; the highest levels in 2015 (Ewing, 2015). Most recent haze in 2016, more than half of the country’s 52 monitoring stations registered as “unhealthy” resulting in schools being closed. According to The Malaymail Online, (2016), API reading exceeding 223 at ‘very unhealthy’ stage has caused the closing of 83 schools in Sabah. People are

advised to avoid outdoor activities in order to prevent widespread illnesses particularly among the elderly and those with breathing difficulties.

Swathes of rainforest in Sumatra are periodically burnt off by farmers and settlers to open up land for oil palm plantations and agricultural for commercial purposes. The smoke were drifted over with the wind southwest monsoon to the Peninsular of Malaysia. East Malaysia (Sarawak and Sabah) also suffer from similar burning in Indonesia's Kalimantan province (Subekti et al., 2017). Apparently, Malaysia has its own hotspots resulting from local industrial and vehicular pollution which contribute to affects the situation to become becoming worst.

Unlike factory and vehicle emissions, wildfire smoke is not filtered by catalytic converters. The toxic components of the smoke vary according to the type of material that is burning. Sumatra's peat fires for example has produced smoke which is more harmful than ordinary forest fires (Qadri, 2001). Peat fires also last longer as they smoulder underground and are harder to put out. As the results result, poor air quality in Malaysia usually will continue continues until a change of in weather conditions and heavy rainstorms clear clearing the air.

## **1.2 Environmental Air Pollution In Malaysia**

Malaysia is a coastal nation located in the centre of Southeast Asia, surrounded by Thailand in the northwestern, Indonesia in the southern and Brunei in the eastern part (Malaysia, 2006). Malaysia consists of a Peninsula Malaysia and approximately one third of the island of Borneo East Malaysia. These two regions are separated by some

640 miles of the South China Sea. The country covers a total area of 330,252 square kilometers (km<sup>2</sup>), where the total land area is 328,550 km<sup>2</sup> and the water area, 1,200 km<sup>2</sup> (Malaysia, 2006). The country experiences relatively uniform temperature throughout the year, with the mean temperature in the lowlands ranging between 26 °C and 28°C (Malaysian Meteorological Department, 2006). It has a tropical climate with annual southwest monsoon from April to October and northeast monsoons from October to February (Hussin et al., 2006).

According to report on Urban Air Quality Management (2006), Malaysia is a middle-income country that has transformed itself in the late 1990's from a producer of raw materials into an emerging multi sector economy. The industries in Peninsular Malaysia including rubber and palm oil processing and manufacturing, light manufacturing industry, electronics, tin mining and smelting, and logging and processing timber. Sabah and Sarawak has logging and agriculture processing as well as petroleum production and refining. Urbanization process in population growth, industries and lifestyle leads to higher traffic density, increases energy demand and contributes to air pollution problems.

Significant changes in air pollution concentration for Malaysia are closely monitored by the DoE using 52 Continuous Ambient Air Monitoring Stations (CAQM) which have been installed across Malaysia. The station were strategically placed at urban, sub urban, industrial and background areas and the air quality were reported to citizens using Air Pollution Index (API) system. According to DoE (2015), New Ambient Air Quality Standard was established in order to replace the older Malaysia Ambient Air Quality Guideline (MAAQQ) that has been used since 1989.

Deputy Natural Resources and Environment Minister Datuk Hamim Samuri said that additional parameter PM<sub>2.5</sub> was outlined in 11<sup>th</sup> Malaysia Plan 2016-2020 to include the measurement in all 52 monitoring stations in Malaysia. It would take longer for Malaysia to fully implement it however, as for now 12 air quality monitoring stations have been set up in Kuala Lumpur, Putrajaya, Melaka, Kuala Terengganu, Tanah Merah, Langkawi, Ipoh, Kuching and Penang to measure PM<sub>2.5</sub>. (The Malaysian Times, 2015).

The New Ambient Air Quality Standard adopts 6 air pollutants criteria that include 5 existing air pollutants which are carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>), particulate matter with the size of less than 10 micron (PM<sub>10</sub>), sulphur dioxide (SO<sub>2</sub>), and ground level ozone (O<sub>3</sub>) and one additional parameter which is particulate matter with the size of less than 2.5 micron (PM<sub>2.5</sub>). The air pollutants concentration limit will be strengthen in stages until year 2020.

Table 1.1 shows the New Malaysia Ambient Air Quality Standard which form the basis for calculating the Air Pollution Index (API). There are 3 interim targets set which include interim target 1 (IT-1) in 2015, interim target 2 (IT-2) in 2018 and the full implementation of the standard in 2020. As depicted in Table 1.1, ground level O<sub>3</sub> were known as secondary pollutant contributed in API reading. Vehicle emissions accounted for largest share in formation of O<sub>3</sub> concentration in Malaysia beside industrial activities (DoE, 2015).