Assessment of Surface Ozone Concentration in Northern Peninsular Malaysia

Nur Izzah Mohd Hashim¹, Nur Alia Izzati Mohd Yusoff¹, Norazian Mohamed Noor¹, Ahmad Zia Ul-Saufie²

¹School of Environmental Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia
²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Pulau Pinang, Malaysia

Email: norazian@unimap.edu.my

Abstract. This study investigate the surface ozone variations in northern Peninsular Malaysia which is Penang and Ipoh. A database with hourly air quality data and meteorological parameters covering the period of 2013 – 2016 obtained from Air Quality Division, the Department of Environment (DOE) were analysed. The association between the O₃ concentration and meteorological parameter as well as the other air pollutants were identified by using Pearson Correlation Analysis and Principle Component Analysis. In order to predict the next three hour of O₃ concentration, two models were developed which is MLR and PCA-MLR. The results indicate that Ipoh has higher ozone concentration with maximum concentration of 0.124 ppm. The diurnal pattern of ozone displayed a peak during afternoon and low concentration during night-time until morning. On the other hand, the NO and NO₂ concentration shows opposite pattern from ozone concentration. Maximum concentration of O₃ reached the peak at late afternoon at approximately 2 p.m. in Penang and Ipoh while the lowest reading was recorded was at 8 a.m. A significant correlation was observed between O₃ concentration with temperature and wind speed in both urban areas. MLR was chosen as best predictive and more accurate model in predicting O₃ concentration during the peak hour compared to PCA-MLR.

1. Introduction

Ozone is a combination of three oxygen atoms which majority found in lower stratosphere, to maintain the thermal structure of the atmosphere and absorption of ultraviolet rays (UV) [1]. Due to rapid increasing in industries and anthropogenic activities in the ground-level atmosphere lead to higher concentration of the nitrogen oxides (NOx) and volatile organic compounds (VOC) which is the ozone precursors. When the nitrogen oxides (NOx) interact with the VOCs under the action of UV radiation at the earth surface, it will result to the form of secondary pollutant, which is ozone, O₃ [2].

Surface ozone is one of the major causes of air pollution in urban areas around the world. Breathing ozone can trigger to respiratory problems, lung diseases such as asthma and also can give harm on vegetation and environment [3,4].

There are several factors that can affect human health including the exposure duration, concentration of surface ozone in atmosphere, length of time between short-term exposures and average amount of air inhaled per minute (ventilation rate) [5-7].
Hence, surface ozone should be firstly identify and analyses. The understanding of the conditions that contribute to its formation [8] and the influence of meteorological parameters such as temperature and humidity will help to thoroughly understand the photochemical reactions and the relationship of $O_3$ concentration with the weather parameters. Prediction of ground-level ozone concentration can be done with the good understanding of the pollutants. Thus, this paper evaluate the relationship of $O_3$ with its precursors and weather parameters in the northern areas of Peninsular Malaysia. Then, 2 models were developed and the performances were compared.

2. Methodology

2.1 Study Area
The chosen areas in this study are Seberang Perai, Penang and Ipoh, Perak. These two areas are known as urbanized area and was chosen due to lack of studies were conducted in these two selected areas. Seberang Perai and Ipoh had a tropical rainforest climate also borders on tropical monsoon climate [9]. Seberang Perai is located in the mainland of Penang. The data for air quality and meteorological parameters used in this study were collected from the monitoring station at Sekolah Seberang Perai II located at latitude 5.3975, 100.4048. Seberang Perai is served by a major highways, which experience heavy traffic in the morning and peak hours. Besides that, this city also was surrounded with a heavy industry area with high pollution. Meanwhile, the monitoring station at Sekolah Menengah Kebangsaan Jalan Tasek, Ipoh (4.6305, 101.1156) were used to collected the air quality and meteorological parameters data which will used in this study. This city consist a lot of industrial areas located at nearby the urban area which may lead to formation of pollution, especially ozone.

2.2 Statistically Analysis
Descriptive statistics is used to present the data in simple and easier way to interpret. It include the descriptive statistic table and boxplot generated by using Statistical Package for Social Science (SPSS) software version 23. Process of collecting, organizing, summarizing and presenting the data also part of the descriptive statistic.

2.3 Diurnal Plot
It is use to identify the behaviour, plotting and average of hourly value on chosen time of 24-h scale. It can show the relationship and effect of ozone precursors towards the surface ozone concentration, and show the variations were in 24-h scale.

2.4 Contour Plot
Contour plot is used to show three dimensional surface on two dimensional plane. The line are drawn for connecting the x and y axis while the contour value represents as z. In this study, the contour plot is used to show the monthly variation of surface $O_3$ concentration by using hourly data. The contour was plotted by using open software, Plot.ly.

2.5 Pearson Correlation Analysis
Pearson correlation analysis use to measure the strength of the relationship between two continuous variable X and Y. In Pearson Correlation, the letter r stand for a value from +1 and -1 where 1 is the positive correlation while 0 means no correlation and -1 is negative correlation. It can show the strength of linear relationship between surface ozone concentration and meteorological parameter.

2.6 Principle Component Analysis (PCA)
PCA was used to find a small sets of linear combinations of covariates that not correlated with one another. This will avoid multicollinearity problems. Additionally, it can also ensure that the selected linear combinations have the maximum variance [10]. Principal Component Analysis is generally expressed as follows:
Where:

$PC_i = l_1 X_1 + l_2 X_2 + \ldots + l_n X_n$  \hspace{1cm} (1)

$PC_i$ = $i^{th}$ principal component

$l_{ji}$ = loading of observed variable $X_j$

2.7 **Multiple Linear Regression (MLR)**

MLR have capabilities to predict the contribution of selected variables to the variations of $O_3$ concentration. This was an attempts to model $O_3$ concentration the relationship between two or more explanatory variables such as the traces gases and a response variable surface ozone concentration with meteorological factors by fitting a linear equation to predict the next hour of $O_3$ concentration. 80 percent of the data will be used to develop the model while 20 percent will be used to test the performance of the model developed using performance indicator [11].

2.8 **PCA-MLR**

The original variables (dataset of NO$_2$, NOx, PM$_{10}$ and other) and PCA (PC1, PC2, PC3 and PC4) results were used as the input in MLR analysis. Mostly PCA is used to reduce the various dimensions associated with MLR which then creates a new variable called the principle component (PC). The PCs has been extracted so that the first principle component (PC1) accounts for the largest number of variations in the data set, while the following components contribute to variations that have not been considered on PC1.

2.9 **Model Performance Measure**

Performance measure are used to evaluate the predictive model between Multiple Linear Regression (MLR) and Principle Component Analysis (PCA). This will ensure that the model developed for prediction of ozone concentration in the studied areas were comparable and had reliable performances. The performance measures used in this study were Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Normalised Absolute Error (NAE), Index of Agreement (IA) and Prediction Accuracy (PA) and coefficient of determination ($R^2$). The equations for all the performance measures applied in this study is shown in Table 1 [12].

| No | Performance Indicator       | Formula                                                                 | Better Predictability if |
|----|-----------------------------|-------------------------------------------------------------------------|--------------------------|
| 1  | Mean Absolute Error         | $MAE = \frac{1}{n-1} \sum_{t=1}^{n} (P_t - O_t)$                        | Closer to 0              |
| 2  | Normalized Absolute Error   | $NAE = \frac{\sum_{t=1}^{n} |P_t - O_t|}{\sum_{t=1}^{n} O_t}$         | Closer to 0              |
| 3  | Root Mean Square Error      | $RMSE = \frac{1}{n-1} \sum_{t=1}^{n} (P_t - O_t)^2$                    | Closer to 0              |
| 4  | Prediction Accuracy         | $PA = \frac{\sum_{t=1}^{n} (P_t - O_t)^2}{\sum_{t=1}^{n} (O_t - \bar{O})^2}$ | Closer to 1              |
3. Results and Discussion

3.1 Boxplot of Ozone Concentration

Figure 1 represents the box plot of hourly ozone concentration in both studied area, Penang and Ipoh from 2013-2016. Based on the plot, $O_3$ concentration in Ipoh was higher than Penang in 2013 and 2014, and the median in Ipoh is the most highest compared to other year for Penang and Ipoh. Other than that, the dataset in Ipoh was more varies than Penang since the boxplot for Ipoh on 2013 and 2014 have the longer distribution ozone concentration due to high value of standard deviation. Then, the boxplot for ozone concentration in both Penang and Ipoh for all years are skewed to the right. Hence, it shows that the $O_3$ concentration is quite high. However, the distributions in Penang on 2015 and 2016 were more varies because the concentration of $O_3$ in Ipoh was droped from 2014 to 2016. This is due to the decreasing in $O_3$ precursor concentration, NO$_2$ in that years.

![Boxplot of Surface Ozone Concentration](image)

Figure 1. Boxplot of Surface Ozone Concentration

3.2 Boxplot of Other Pollutant and Weather Parameter

The boxplot of traces gases (NOx, NO$_2$, CO, SO$_2$, PM$_{10}$) and meteorological parameter (temperature, humidity, wind speed) in Penang and Ipoh were displayed in Figure 2. The boxplots for NOx concentration in Ipoh has longer distribution than Penang because monitoring station in Ipoh surrounded with high volume of vehicles. The variation of dataset for NO$_2$ concentration and temperature for both studies area are just similar. Furthermore, boxplot for SO$_2$ and PM$_{10}$ concentration show very small distribution in both area, even though the concentration in Penang for
both parameter was higher than Ipoh except for PM\textsubscript{10} concentration in Ipoh for 2014. This is due to the location of monitoring station in Penang which surrounded by heavy industrial areas.

![Boxplot of other pollutant and weather parameter](image)

**Figure 2.** Boxplot of other pollutant and weather parameter

### 3.3 Variation of Surface Ozone Concentration

The surface ozone concentration was analysed diurnally and the result was presented in in the Figure 3. The ozone peak was observed in both sites and higher ozone concentration was observed in Ipoh compared to Penang. This is because the location of the monitoring station in Ipoh situated nearby
schools and surrounded with high traffic flow. The minimum ozone concentration was observed during the night-time and early morning hours. Urbanization, industrialization, high-speed propulsion, and atmospheric emissions of a large numbers of active substances such as carbon dioxide and aerosol are some of the major sources that contribute the increase of ozone concentration. The lowest ozone concentration consistently measured at 0800 and reached the peak around 1300-1500 during midday and appear to reduce in the night at the both site. The value is reached at approximately 2 p.m. in Penang and Ipoh.

![Figure 3. Diurnal concentration of surface ozone in Penang and Ipoh](image)

### 3.4 Monthly Variation of Ozone Concentration

The surface ozone concentration was analysed by using average hourly data for every month to identify the monthly variation of ozone concentration. Figure 4 presented the contour plot of surface \( O_3 \) concentration for Penang and Ipoh. Based on the plot, the \( O_3 \) concentration in Penang was higher in March and April, meanwhile in Ipoh, the highest \( O_3 \) concentration was in February. This is because, Peninsular Malaysia especially northern area were experiencing the effects of the El Nino phenomenon and northeastern monsoon winds before the monsoon wind shifts to wet weather and thunderstorms from April to middle May [10].

a) Penang

![Figure 4. Contour plot of monthly \( O_3 \) concentration](image)

b) Ipoh
This phenomenon causes the current temperature to rise and enhance the production rate of surface ozone concentration. The concentration of surface ozone are lower after July until end of December in both studies area especially Ipoh. This regarding to heavy rain and thunderstorm between the months which causes low temperature and wind speed to increase [14]. This will affect the photochemical reaction between O₃ precursors.

3.5 Diurnal Fluctuation of Ozone and Its Precursor

Figure 5 (a-b) compares the diurnal variation of surface ozone and its precursor which are NO₂, NOx and CO. The diurnal plots of O₃ concentration for Penang (a) and Ipoh (b) were plotted according to the average hourly data set from 2013 to 2016. The dependency of the O₃ formation to the precursor concentration is associated with a clear hourly variation.

The diurnal variation for all urbanized area around the world shows similar pattern which displayed peak concentration during afternoon and low concentration during night-time until morning. While the NOx and NO₂ concentration shows the opposite pattern from ozone concentration. Increasing ozone concentration correspond to the decreasing of its precursors concentration. However, CO concentration shows only slight changes for both areas.

NOx played an important role in forming and destroying the surface ozone. When NO concentrations is high, it will effectively titrate O₃ in urban areas. During rush hour between 8.00 until 10.00, the NO₂ concentration was recorded at the highest levels. O₃ precursors such as NOx and CO are originated from the sources like motor vehicles and industrial activities [15]. In addition, high concentrations of NO₂ will produce especially in areas with high number of vehicles since high emission of NO levels. Other than that, VOCs also a precursor to O₃, which is released from anthropogenic activity and natural resources as well [16].

The concentration of surface O₃ was highest during midday and reached the peak around 13.00-15.00 and appear to reduce in the night. This pattern happen when the intensity of sunlight was at the maximal level and it will promote the formation of O radical, which then responsible for the formation of O₃ from its precursors [11,17]. Meanwhile, in the absence of sunlight during evening until early morning, the surface O₃ concentration reached minimum value since the UV intensity was very small, hence the photochemical reaction is stopped.

The concentrations of the NO₂ peaked again in the evening when people travel back home from work. The concentrations of O₃ precursors showed a less steep drop during the night due to the higher atmospheric stability and the absence of sunlight, which reducing the scavenging reaction of photochemical oxidants [16].

![Figure 5. Diurnal plot of O₃ and its precursors in (a) Penang, (b) Ipoh](image)

3.6 Pearson Correlation Analysis

Figure 6 show linear relationship between the two variables in this two urbanized areas which are Penang and Ipoh. In this analysis, the surface ozone concentration were correlated between its precursors and the meteorological parameter. Based on this result, ozone concentration in Penang and Ipoh were strongly correlated with temperature, humidity and wind speed.

In Penang, there is a strong positive relationship between O₃ concentration with temperature with the r value of 0.970 and negative correlated with the wind speed with r = -0.962 respectively.
Meanwhile, in Ipoh, the most correlated meteorological parameter were temperature and wind speed with the r value of 0.959 and r = -0.850. This result shows that formation of surface ozone are affected by temperature, as well as the wind speed, which both affect the rate of chemical reaction of O₃ production [18].

(a) Penang

![Graphs showing relationship of O₃ concentration with temperature and wind speed in Penang.](a.png)

(b) Ipoh

![Graphs showing relationship of O₃ concentration with temperature and wind speed in Ipoh.](b.png)

**Figure 6:** Relationship of O₃ concentration with temperature and wind speed in (a) Penang (b) Ipoh.

3.7 Principle Component Analysis (PCA)

There are three principal component (PCs) in Ipoh where the total variance is 31.35% which mainly comprised of O₃, temperature and relative humidity same as in Penang for the first component, thus was classified as the meteorological factor. The second component (PC2) comprises of NO₂ and NOx (16.96% of total variance) since the monitoring station is surrounded by densely residential areas and schools which, contributed to high number of vehicle.

Besides that, the third component (PC3) which represented by 14.25% of the total variance shows the positive loading of CO (0.872) and PM₁₀ (0.686) as the main source from diesel vehicles such as lorries and bus [17]. While the fourth component (PC4) as the meteorological parameter which represented by wind speed with total variance of 11.19%. In short, first principle component (PC) group in Penang and Ipoh could be renamed as meteorological parameters and the second component could be appointed as the O₃ precursors which was similar to the finding from [10] that conducted PCA for the same 9 parameters in Shah Alam.
Table 2. Rotated Component Matrix

| Air pollutant | Principle Component Penang | Principle Component Ipo | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 | PC4 |
|---------------|---------------------------|------------------------|-----|-----|-----|-----|-----|-----|-----|
| O₃            | 0.671                     | 0.725                  |     |     |     |     |     |     |     |
| NO₂           | 0.847                     | 0.861                  |     |     |     |     |     |     |     |
| NOₓ           | 0.775                     | 0.872                  |     |     |     |     |     |     |     |
| SO₂           | 0.642                     | -0.475                 | 0.358|     |     |     |     |     |     |
| CO            | 0.599                     | 0.863                  |     |     |     |     |     |     |     |
| PM₁₀          | 0.312                     | 0.730                  | 0.442|     |     | 0.686|     |     |     |
| Temperature   | 0.865                     | 0.839                  |     |     |     |     |     |     |     |
| Humidity      | -0.835                    | -0.840                 |     |     |     |     |     |     |     |
| Wind speed    | -0.723                    |                       |     |     |     |     |     |     | -0.875|

3.8 Model Development

The original air quality data in Penang and Ipoh was further analysed by development of model by using Multiple Linear Regression for prediction of O₃ concentration for the next hour, next two hour and next three hour during the peak hours (1300-1500). Then the hybrid model was also developed by using Principle Component (PCs) as the input in order to enhance the predictability of MLR model. The data for three peak hours which are from 13.00 – 15.00 were selected for both model. The model for O₃ prediction for both method was presented in Table 3. The input for MLR model include O₃, NO₂, NOₓ, CO, SO₂, PM₁₀, temperature, humidity and wind speed. While for PCA-MLR, the significant variable which was greater than 0.5 was used as input.

Table 3 : Summary model for O₃ concentration predictions

| Method | Prediction Day | Model                                                                 |
|--------|----------------|----------------------------------------------------------------------|
|        | Next hour      | O₃ₜ+₁,PNG = 0.067 – 0.001T – 0.141NOₓ + 0.494SO₂ + 0.307NO₂ + 0.289CO + 0.002CO |
|        | Next two hour  | O₃ₜ+₁,PNG = 0.058 + 2.123*10⁻⁶WS – 0.001T + 0.095NOₓ – 0.249SO₂ – 0.33NO₂ + 0.375O₃ – 0.002CO |
|        | Next three hour| O₃ₜ+₁,PNG = 0.058 – 9.837*10⁻⁵T + 0.018NOₓ + 0.087NO₂ – 1.461SO₂ – 0.296NO₂ + 0.287O₃ – 0.001CO + 9.366*10⁻⁵PM₁₀ |
| PCA-MLR| Next hour      | O₃ₜ+₁,PC1 = 0.026 + 0.001PC1 + 0.001PC2 – 9.837*10⁻⁵PC3                |
|        | Next two hour  | O₃ₜ+₁,PC1 = 0.043 – 6.926*10⁻⁵PC1 - 0.248PC2 + 8.346*10⁻⁵PC3 - 4.24*10⁻⁵PC4 |
|        | Next three hour| O₃ₜ+₁,PC1 = 0.039 + 5.195*10⁻⁵PC1 - 0.153PC2 + 1.064*10⁻⁵PC3 - 4.809*10⁻⁵PC4 |

3.9 Performance Index

The developed MLR and PCA-MLR model were assessed by using a few performance indicators. The calculated performance indicators for both urbanized areas were summarized in Table 3. The MLR
was chosen as the most reliable method compared to PCA-MLR with smaller error (MAE=0.0191: NAE=0.0562: RMSE=0.0236) and higher accuracy (PA=0.4675: R²=0.4351: IA=0.5092) for the next hour in Penang. In general, both models give low R² values indicating that the performance of the models in prediction of O₃ concentration (next hour, next two hour and next three hour) are not very good. However, MLR shows better performance than PCA-MLR. This is because MLR models is better in prediction of O₃ levels for time periods of greater than 1 hour. It has also successfully used in predictions of O₃ concentrations one day ahead as proved by previous researches [19].

In this study, Multiple Linear Regression give the best performance of O₃ concentration prediction for the next hour, next two hour and next three hour in both area. Most of the models studied on ozone concentration have shown that PCA-MLR has higher accuracy than MLR except for the model studied by [20] on the variation of O₃ concentration in Dilovasi, Turkey which had similar results with this study. It is due to the dataset used in this study was minimal which the three peak hours (1300 - 1500) of dataset were used. Thus, the usage of PCs as input in MLR could not improve the performance of the model to study the variations of O₃ concentrations.

### Table 4. Performance Comparison of O₃ Prediction Models

| Study Area | Method   | MAE   | NAE   | RMSE  | PA    | R²   | IA   |
|------------|----------|-------|-------|-------|-------|------|------|
| Next hour  | MLR      | 0.0191| 0.0562| 0.0236| 0.4675| 0.4351| 0.5092|
|            | PCA-MLR  | 0.0254| 0.0643| 0.0467| 0.4361| 0.3164| 0.4672|
| Next two hour | MLR | 0.0221 | 0.0648 | 0.0264 | 0.4675 | 0.4351 | 0.4701 |
|            | PCA-MLR  | 0.0276 | 0.0704 | 0.0436 | 0.4105 | 0.3132 | 0.4361 |
| Next three hour | MLR | 0.0253 | 0.0708 | 0.0316 | 0.4443 | 0.4177 | 0.4033 |
|            | PCA-MLR  | 0.0297 | 0.0738 | 0.0395 | 0.3848 | 0.2961 | 0.3942 |
| Next hour  | MLR      | 0.0184 | 0.0463 | 0.0209 | 0.3812 | 0.2447 | 0.3827 |
|            | PCA-MLR  | 0.0239 | 0.0510 | 0.0259 | 0.3170 | 0.2212 | 0.3594 |
| Next two hour | MLR | 0.0220 | 0.0542 | 0.0269 | 0.3497 | 0.2224 | 0.3784 |
|            | PCA-MLR  | 0.0294 | 0.0548 | 0.0297 | 0.3043 | 0.2203 | 0.3236 |
| Next three hour | MLR | 0.0228 | 0.0610 | 0.0276 | 0.3029 | 0.2186 | 0.3526 |
|            | PCA-MLR  | 0.0301 | 0.0654 | 0.0306 | 0.2834 | 0.2135 | 0.3192 |

The relationship between the prediction and actual for the next hour O₃ concentration for both studies area is presented by using scatter plot as shown in Figure 7. The figure shows that all plots have positive trend and showed weak association between the predicted and the actual values. The MLR model for both Penang and Ipoh has higher R² values than the PCA-MLR model with the value of 0.4351 and 0.2447.

(a) Penang
4. Conclusions
In this study, MLR model was developed and modified MLR model was also develop using Principle Component (PCs) as the input to enhance the predictability of Multiple Linear Regression model. Then, the performance indicator were used to find the goodness of the models. The result shows that the MLR model was more accurate model in predicting $O_3$ concentration prediction during the peak hour compared to PCA-MLR in both study areas.

By conducting this study, a sufficient database of ozone concentration, can be used as reference for future development since the weather parameter at the study area has been achieved. Other than that, a prediction of $O_3$ by using MLR is capable and a really useful tool in providing early information to the public in an urban area.

Acknowledgements
Author would like to thank Department of Environment, Malaysia for the air pollutant dataset and Ministry of Education Malaysia for the Fundamental Reserach Grant Scheme (FRGS 9003-00508).

References
[1] Burghardt T E et al 2016 14 714.
[2] Abdul-Wahab S A et al 2002 Environmental Modelling & Software 17 219.
[3] Alkasassbeh M et al 2013 International Journal of Advanced Science and Technology 55 1.
[4] Noor N M et al 2015 Revista de Chimie 66 1443.
[5] Noor N M et al 2014 Key Engineering Materials 594-595 902.
[6] Sara Y Y et al 2013 Advanced Materials Research 620 502.
[7] Karlsson P E et al 2017 Science of total Environment 576 22.
[8] Kumar N et al 2017 Urban Climate 20 148.
[9] Department os Statistics Malaysia 2010 https://www.dosm.gov.my/v1 (Retrieved on 15 Jan 2018).
[10] Awang N R et al 2015 Atmospheric Pollution Research 6 726.
[11] Zhao W et al 2016 Atmospheric Environment 144 182.
[12] Noor N M et al Australian Journal of Basic and Applied Sciences 5 2796.
[13] Utusan Online http://www.utusan.com.my/berita/nasional/negeri-di-utara-berhadapan-cuaca-panas-lebih-tinggi-sehingga-mac-15- (Retrieve on 5 May 2018).
[14] Department of Environment 2015 Environmental Quality Report. Department of Environment: Putrajaya.
[15] Vingarzan R 2004 Atmospherics Environment 38 3431.
[16] Latif M T et al 2016 Science of the Total Environment 573 494.
[17] Xie M et al 2016 Science Total Environment 551.
[18] Alghamdi M et al 2014 Atmospherics Environment 94 205.
[19] Sousa S I V et al 2006 Chemosphere 64 1141.
[20] Ozbay et al 2011 Atmospheric Research 102 57.