ACHIEVING SUSTAINABLE ENVIRONMENT THROUGH PREDICTION OF AIR POLLUTANTS IN YOGYAKARTA USING ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

MUHAMMAD R. A. PURNOMO$^{1,*}$, ADHE R. ANUGERAH$^{2}$

$^{1}$Department of Industrial Engineering, Universitas Islam Indonesia, Jalan Kaliurang KM 14.5, 55584, Sleman, Yogyakarta, Indonesia
$^{2}$Institute of Tropical Forestry and Forest Products (INTROP), Universiti Putra Malaysia, Jalan Universiti, 43400, Serdang, Selangor DE, Malaysia
*Corresponding Author: ridwan_ie@uii.ac.id

Abstract
Air pollution has been proven as one of the major problems of metropolitan areas in the world, especially in Yogyakarta, Indonesia, where it listed as top 3 asthma prevalence in Indonesia. Formalising the air pollutant concentration prediction model in Yogyakarta becomes essential to lessen the number of victims and to develop more comprehensive sustainable policies by government agencies. This research aims to predict six air pollutants in the province of Yogyakarta which include sulphur dioxide (SO$_2$), nitrogen dioxide (NO$_2$), ozone (O$_3$), hydrocarbon (HC), lead (Pb), and particulate matter 10 micrometres (PM10). The model was developed using Adaptive neuro-fuzzy inference system (ANFIS). The model performance was fair, with the best error rate of 27.22%. The unsatisfactory prediction performance was caused by the unfavourable data input. This province was collected and analysed the air pollutant data manually that resulted to human error, inconsistency, and did not present an up to date data. The theory of constraint (TOC) was utilised to improve the performance of monitoring system by analysing the root-caused error. Changing the current monitoring system from manual to automatic air quality monitoring system was expected to improve the input data quality, thus improving the prediction performance model. Furthermore, HC and PM10 concentration levels in this Province were recoded beyond the national standard and several strategies to decrease the concentration have been discussed in this paper. This research is the first to develop prediction model of air pollutants and analysed the current air quality in Yogyakarta.

Keywords: ANFIS, Air pollutant, Prediction, Sustainable environment, TOC.
1. Introduction

Air pollution is an air contamination in the atmosphere caused by gas, liquid, or solid waste that affects the human’s health and biosphere, decreases visibility, and damages the materials. Air pollution is a developing country’s problem which is believed may raise more victims more than AIDS, malaria disease, breast cancer, and tuberculosis [1]. There are several main air pollutants that may harm human’s health, i.e. particulate matter (PM), Sulphur Dioxide (SO\textsubscript{2}), Nitrogen Oxide (NO\textsubscript{x}), and Carbon Monoxide (CO).

Air pollution concentration has closely linked to mortality in all respiratory diseases and ischemic heart disease (IHD) [2]. There are several studies have demonstrated the relationship between air pollutants and human health, such as particulate matter [3], carbon monoxide [4], nitrogen dioxide [5], ozone [6], hydrocarbons [7], and lead [8]. Besides, study proves that air pollutants cause 1.2% of the total mortality rate in the world and causes more than two million early deaths each year [9], among deaths caused by air pollutants, nearly half of which occurred in developing countries [10].

Indonesia is included in the list of countries with large victims caused by air pollutants, sits on the 4th rank after India, China, and Russia. In the year of 2012, there were 61,792 deaths caused by air pollutants, which means that in 100,000 people, there are 25 deaths caused by air pollutants in Indonesia. The increase in population will affect the vehicle demand and industrial growth, which will impact to the chemical composition of the atmosphere, especially particulate matter concentration [11, 12]. Yogyakarta is one of the provinces in Indonesia that has relatively high population growth and density. In the year of 2015, the number of populations in Yogyakarta reached 3,679,176 people or increased by 126,714 people in 3 years. In addition, there was an increase in industrial production in the large and medium scale by 5.44%. In 2017, the number of vehicles has highly increased to the number of 279,359. The increasing number of vehicles is known as one of the increasing factors of PM concentration [13].

Respiratory disease (asthma, bronchitis, pneumonia) is one of the ten communicable diseases that are most occurred in few last years in Yogyakarta. In the year of 2013, Yogyakarta was ranked in the 3rd for asthma prevalence in Indonesia [14]. Some of these respiratory problems are caused by allergens outside the room, weather changes, and pollution outside the room.

The above statements show that there is a close relationship between air pollutants and human health; hence, measuring the pollutant concentration in a specific time period, which is obtained from air quality monitoring systems is important [15]. It is necessary to give an early warning to society, to do the preventive action and the settlement of government regulations regarding air pollution, therefore the goal of environmental sustainability can be achieved [16].

The effect of air pollution does not solely affect human’s health, but also impacts climate and ecosystems, including agriculture. The United Nations Environment Assembly (UNEA) in 2014 stated that clean air is essential to protect human health and simultaneously benefit climate and ecosystems, including food security. In order to solve the abovementioned problems, predicting several harmful air pollutants in Yogyakarta is important to achieving a sustainable environment.
This research aims to construct prediction models for several harmful air pollutants using adaptive neuro fuzzy inference system (ANFIS). ANFIS is one of the popular models in artificial intelligence that takes the strength of both the neural network and fuzzy model [17]. ANFIS is an effective method for the prediction process because the error rate is smaller than the ANN methods [18]. ANFIS can produce better performance because it has adaptation ability and data fitting to describe the physical process that is not linear [19, 20].

The constructed air pollutant models can be used to predict future air pollutant concentrations; hence government can take early action if some pollutant concentrations exceeding the standard. However, the air quality monitoring system conducted by Yogyakarta was done manually. In the manual data collection, the air quality is collected by a schedule, then be written as an environment or pollution data report, and finally archived. The entire process may take up to 2-7 days to get the air pollutant information. The manual processes were not only done in the data collection but also in the data analysis, which may account for human error. Therefore, improvement in the current air quality monitoring system is needed, while to achieve this goal, the theory of constraint (TOC) is utilised in this research.

2. Methodology

2.1. Study area

This research was conducted in the Province of Yogyakarta. Yogyakarta is located in the centre of Java island, with a total area of 3133 km2, listed as the smallest province in Indonesia after Jakarta according to the land area. Yogyakarta was also a favourite destination for higher education students because of its education quality and quantity. Moreover, according to the domestic tourism statistics in 2018, Yogyakarta was listed as the top 5 domestic tourist numbers. The high number of arrivals in Yogyakarta, either for tourism or education purposes, led to an increase in vehicle ownership. As a result, pollution caused by vehicle emissions has become more serious [21].

In 2019, there are more than 5 million vehicles in Yogyakarta and listed as the second-most dense vehicles in Indonesia after Jakarta. In another words, there was 1600 vehicles per km2 followed by Bali (936 vehicles/ km2), Banten (350 vehicles/km2), etc. This means, that the vehicle density in Yogyakarta was significantly higher compare to other provinces in Indonesia. The characteristics of vehicle ownership in Yogyakarta are identical to other provinces in Indonesia where motorcycles were dominant. As many as 86% of the total vehicles were motorcycles, followed by cars (9%), buses (1%), and trucks (4%).

The exhaust gas emissions from vehicles release harmful pollutants to the air. Yogyakarta is not an industrial area; therefore, the outdoor air pollutants were highly contributed by vehicles. Moreover, this province has one active volcano located in Sleman and often causes an eruption; this condition would aggravate the air quality in Yogyakarta. When an eruption occurred, SO$_2$ was released and reacted with sunlight, aerosol, and greenhouse gases and transformed into fine particles [22].

Six air pollutants will be predicted in this research: NO$_2$, SO$_2$, O$_3$, Pb, HC, and PM10 using three predictors: air temperature, humidity, and wind speed. The data were given by the Environmental Agency of Yogyakarta, covers 5 years from 2009 to 2015, where data in year 2010 and 2012 were not available. This research covers
all cities in the Province of Yogyakarta: Sleman, Bantul, Gunung Kidul, Kota Yogyakarta, and Kulon Progo. Figure 1 shows the study area of this study.

![Map of Yogyakarta province and its cities.](image)

The data were obtained from the Environmental Agency of Yogyakarta, the only agency that has the right to do air quality monitoring in Yogyakarta. The monitoring was conducted twice annually with uncertain time, sometimes conducted in March and August, sometimes in different months. The locations were scattered in 30 locations, which are 8 locations in Sleman Regency, 5 locations in Yogyakarta City, 8 locations in Bantul Regency, 4 locations in Kulon Progo Regency, and 5 locations in Gunung Kidul Regency. The total data that could be processed were 186 records with 1674 instances. The data later will be classified as training data to establish the model and testing data to check the accuracy.

### 2.2. Input and output variables

There are number of publications have found that meteorological aspects can be used to predict air pollution concentrations. Meteorological aspects can affect pollutant dispersion and the chemical transformation of pollutants [23]. Several meteorological aspects that may affect the pollution concentration are the mean temperature, wind speed, relative humidity, atmospheric pressure, and sunshine duration [24]. In this study, only three meteorological aspects were used to predict the air pollution concentration due to the availability of data from the environmental agency: wind speed, temperature, and humidity. The use of meteorological aspects to predict air pollutant concentrations is explained in Table 1.
Table 1. Researchers using meteorological aspects to predict air pollutants.

<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Input Variable</th>
<th>Output Variable</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>[25]</td>
<td>a. Humidity</td>
<td>a. NO$_2$</td>
<td>O$_3$ can be predicted favourably by using meteorological variables, while another three air pollutants could have better prediction by adding variation of human activities and irregularity of the processes, as such: traffic, dust storms, resuspension etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Pressure</td>
<td>b. O$_3$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Temperature</td>
<td>c. CO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. Wind speed, and</td>
<td>d. PM10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>e. Wind components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>[24]</td>
<td>a. NO</td>
<td>a. PM10</td>
<td>The best predictors were NO$_x$, PM10 and wind speed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. NO$_2$</td>
<td>b. NO$_x$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. NO$_x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. CO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>e. O$_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>f. PM2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>g. PM10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>[26]</td>
<td>a. NO$_2$, NO, PM10</td>
<td>a. PM10</td>
<td>The most impactful variables in predicting PM$_{10}$ were road traffic emissions (NO$_2$, NO, PM10) and temperature while humidity has small impacts.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Temperature</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Wind direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. Wind intensity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>e. Humidity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>[27]</td>
<td>a. Wind speed</td>
<td>a. NO$_x$</td>
<td>PM was highly correlated with NO$_x$ and SO$_2$ and wind speed was a great predictor for NO$_x$, SO$_2$ and PM.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Temperature</td>
<td>b. SO$_2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. PM</td>
<td>c. PM</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>[28]</td>
<td>a. Wind direction</td>
<td>a. O$_3$</td>
<td>The meteorological aspects as input variables were effective to predict the concentration of all air pollutants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Wind speed</td>
<td>b. PM10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Temperature</td>
<td>c. NO$_2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. Humidity</td>
<td>d. CO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>e. PM2.5</td>
<td>e. PM2.5</td>
<td></td>
</tr>
</tbody>
</table>
2.3. Data processing

2.3.1. Adaptive neuro fuzzy inference systems (ANFIS)

In order to form an air pollutant concentration prediction model in Yogyakarta, there are several steps that need to be done as shown in Fig. 2.

Here are the explanations for every process:

- Step 1: Data that contains errors, for example misspelling, incomplete, or other invalid data should be removed. Clean data could increase data quality.
- Step 2: Dividing the data into two big groups: training and checking. The training set is used to build networks and adjust the connected weights of the constructed data, while the checking set is used to evaluate the model performance. There is no specific procedure for allocating data for training and checking sets. In order to choose which mechanism resulted in smaller error, this research compared the ratio of 2:3 and 1:2 for NO, and O3 as a sample by using 70, 80, and 90 epochs.
- Step 3: Initiation the fuzzy inference system is done from the problems, including the membership function type and membership function number. In the number of membership functions, the fuzzy value for every variable was determined. Besides, the membership function type used in this research is a generalized bell because it has a better rate of detection [29].
- Step 4: Deciding the membership function for each input by defining its fuzzy value that comes from the opinion of the expert. The expert gives opinions on defining the number of classes and ranges for input data (meteorology aspects: temperature, humidity, and wind speed) through an interview process. Temperature is defined into three different fuzzy values, namely low (24°C - 30°C), medium (28°C-32°C) and high (31°C -37°C); humidity classified as dry (14-46%), moderate (44-70%), and humid (68-90%); while wind speed was classified as low (0.3-1.7 m/s) and high (1.5-4 m/s)
- Step 5: Conducting the ANFIS training by determining the error tolerance, which is the stop limit to training the Fuzzy Inference System (FIS) and determining the total epoch, which was the total iteration. In order to minimise the researcher’s subjectivity, the error tolerance was set to 0.
- Step 6: Calculate the mean absolute deviation as the performance evaluation on training and checking data. It was intended to examine the previous training process, whether the ANFIS could learn the pattern properly.
- Step 7: Post-examining was designated to repeat FIS training if MAD was still considered as large and had a high gap between training and checking by eliminating the data that caused deviation or error. After following the steps above, the ANFIS structure is constructed as shown in Fig. 3.
This model has three inputs, where temperature and humidity have three fuzzy values, while the wind speed has only two. Therefore, 27 rules were constructed to predict air pollutant concentrations for each pollutant or 162 rules in total. Below are some examples of rules in this research:

1) If (Temperature is Low) and (Humidity is Dry) and (Wind Speed is Low) then (NO₂ is out1mf1)
2) If (Temperature is Low) and (Humidity is Dry) and (Wind Speed is High) then (NO₂ is out1mf2)
3) If (Temperature is Low) and (Humidity is Moderate) and (Wind Speed is Low) then (NO₂ is out1mf3)
4) If (Temperature is High) and (Humidity is Dry) and (Wind Speed is Low) then (SO₂ is out1mf4)
5) If (Temperature is High) and (Humidity is Dry) and (Wind Speed is High) then (SO₂ is out1mf5)
6) If (Temperature is High) and (Humidity is Moderate) and (Wind Speed is High) then (SO₂ is out1mf6)
7) If (Temperature is High) and (Humidity is Moderate) and (Wind Speed is Low) then (SO₂ is out1mf6)

The prediction model was evaluated using Mean Absolute Percentage Error (MAPE). Performance metrics or error measures are key components of the evaluation framework. In the prediction model, this measure is used to compare the training set with the actual data from checking set data. If the error is too large, the error-caused were removed, and the model was re-constructed.

2.3.2. Theory of constraint in improving air quality monitoring systems

The problem related to the sustainable environment in Indonesia, especially in the Province of Yogyakarta will be analysed. The theory of constraint (TOC) is one of the management approaches based on the process of continuous improvement to identify constraints in the overall process and overcome them [30]. Two TOC tools will be used to understand the problems in achieving a sustainable environment: a
current reality tree and a future reality tree. These two tools are able to answer two basic questions in TOC: a) what to change and b) what to change to [31].

In the initial stage, root cause or core conflict underlying a significant majority of the sustainable environment in Yogyakarta’s problem is revealed. CRT is used to establish a stream of cause and effect logical relationship that links the core conflict with the various undesirable effects. In the next stage, FRT will be established to determine what changes need to be implemented to successfully eliminate the core problem.

3. Results and Discussion
3.1. Current condition of air quality in Yogyakarta

Six air pollutant concentration data from five cities in Yogyakarta were calculated and visualized. The distribution of air pollutants was not centralized in one city, although Kota Yogyakarta is the capital city of the province of Yogyakarta, and we expect to be high in any air pollutant concentration but only HC, NO$_2$, and SO$_2$ showed the highest average value in this city. Besides, although Sleman is the most populous city, where one-third of the province population is living there, there is no any significant air pollutant concentration in this city. This means, that there were no any significant differences in air pollutant concentrations in all cities in this province. Figure 4 shows six air pollutant concentration distributions in Yogyakarta.

The average values for the air pollution indicators, namely HC, NO$_2$, O$_3$, Pb, SO$_2$ and PM10 were 73.94 µg/m$^3$ (range: 3.7 µg/m$^3$ - 199.5 µg/m$^3$), 28.21 µg/m$^3$ (range: 7.6 µg/m$^3$ - 78 µg/m$^3$), 13.25 µg/m$^3$ (range: 3 µg/m$^3$ - 36.5 µg/m$^3$), 0.34 µg/m$^3$ (range: 0 µg/m$^3$ - 1.2 µg/m$^3$), 52.11 µg/m$^3$ (range: 14.9 µg/m$^3$ - 166 µg/m$^3$), and 33.94 µg/m$^3$ (range: 10.7 µg/m$^3$ - 88.2 µg/m$^3$) respectively. Yogyakarta City has the highest HC, NO$_2$, and SO$_2$ concentrations, while the highest PM10 and O$_3$ concentrations were located in Kulon Progo. In addition, the highest Pb concentration was located in Bantul.

However, there were only two pollutants that exceeded the national air quality standard: HC and PM10. All cities except Gunung Kidul had a concentration above the standard historically where Bantul is the only city with PM$_{10}$ shown above the standard. The standard of air quality in Indonesia refers to the Republic of Indonesia Regulation number 41 in 1999 regarding air pollution control. The air quality standard was differentiated according to the parameters and measurement duration. Different countries have different standards and are affected by several factors, for instance: the existing modern technologies and scientist capability. This means that the values that considered safe in Indonesia might be considered dangerous in several developed countries.

Four out of five cities in Yogyakarta had a high-level concentration of HC. This was due to two main reasons: a) age of the vehicles, and b) intersection with traffic lights where the vehicles were completely stopped or running at low speed. The age of vehicles was one of the main reasons for the increasing HC concentration, which is in line with research conducted by [32], who found that among other pollutants, HC was the most sensitive to vehicle age. In Indonesia, there are no regulations regarding the age limitations of vehicles. Moreover, the increasing HC concentration was affected by the speed of the vehicles. The lower speed of motor vehicles increases HC concentration. Regarding to our data, the exceeded HC
concentrations were located at intersections with traffic lights. In this area, the drivers were accelerating and decelerating the speed at which produce higher pollutant concentrations than the steady-speed driving modes [33].

Fig. 4. Distribution of air pollutant concentrations (µg/m³) that consisting of a) SO₂; b) NO₂; c) O₃; d) HC; e) Pb; f) PM10
3.2. Current condition of air quality in Yogyakarta

3.2.1. Training and checking data selection

In the training data selection, the requirement was that the data should be representative, which means that the data should contain the lowest and the highest values from each variable, while the rest would be chosen randomly. This was aimed to let ANFIS to learn a wide range of data, thereby improving the prediction accuracy. For the rest, the training and checking data members were collected randomly [34].

In addition, another important aspect was the determination of the training and checking ratio. In order to determine the comparison ratio, the two testing mechanisms were performed, which were dividing 1/2 and 2/3 data for training and checking set, and then the results of those two mechanisms were evaluated. NO$_2$ and O$_3$ were selected as samples in determine the ratio between the training and checking sets. Table 2 shows the comparison result of two different ways in determining the number of training and testing sets.

<table>
<thead>
<tr>
<th>Air Pollutant</th>
<th>Testing Mechanisms</th>
<th>Epoch</th>
<th>Error</th>
<th>Average Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO$_2$</td>
<td>2/3 training set</td>
<td>70</td>
<td>8.1</td>
<td>8.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>8.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
<td>8.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/2 training set</td>
<td>70</td>
<td>4.98</td>
<td>4.88*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>4.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
<td>4.79</td>
<td></td>
</tr>
<tr>
<td>O$_3$</td>
<td>2/3 training set</td>
<td>70</td>
<td>4.14</td>
<td>4.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
<td>4.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/2 training set</td>
<td>70</td>
<td>3.85</td>
<td>3.79*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
<td>3.72</td>
<td></td>
</tr>
</tbody>
</table>

According to Table 2, it can be concluded that dividing the data into 1/2 for the training set and another 1/2 for the checking set resulted in a lower error. In NO$_2$, the performance of the prediction model was increased by 65% and increased by nearly 8% in O$_3$. There is no rule of thumb in determining the ratio between the training and the testing sets because every case is unique. However, in most published research, the ideal proportion for the training set appeared to be between 40% and 80% [35]. Therefore, computing the splitting performance should be done to choose the ratio with the smallest error.

3.2.2. The performance of the prediction model

ANFIS is a part of adaptive networks, which is a combination of the neural network principle and fuzzy logic. Neural networks are supervised learning using several data obtained from the past to determine the value for the future. While the fuzzy logic model is unable to learn the data pattern, fuzzy logic is easier to understand because it uses linguistic language rather than numbers in if-then rules structure.
ANFIS can predict SO\textsubscript{2} better than other pollutants with a mean absolute percentage error of 27.22%. However, several air pollutants have higher errors, such as NO\textsubscript{2} (36.26%), and PM10 (43.43%). The high error was mostly contributed by the checking set data. ANFIS was able to predict the air pollutant concentration better in the training set but relatively poor in checking set data because several predicted points were extremely far from the observation. Figure 5 presents the ANFIS prediction performance of six air pollutants.

The error difference between the training and the checking sets were huge. This means that ANFIS successfully learnt the historical data but unable to predict the air pollutant concentration. The large difference is caused by the low quality of historical data, for instance the presence of extreme data, and the less precise of training and checking data selection. Besides, most of the supervised learning algorithms are only focused on error minimisation without supervising the
complexity of the models, which may cause adverse situations, such as overfit or even underfit, which on both cases they may lead to large variant of the data [36].

In this case, the data is overfitting caused by the over-trained data by ANFIS. Overfitting is the situation when the model performs well on the training data but poorly on the test or validation data [37]. Each data set trained by ANFIS has the maximum limit of epoch determination before the overfitting cases happened. Thus, several mechanisms must be done to test several epoch with the smallest error values. In this research, the mechanism of maximum epochs total determination was not done, but the data quality was improved by eliminating the data that caused the error. The data that caused the big error were eliminated from the checking data, and the performance was improved significantly.

In this province, the data were collected twice in a year in 30 locations spreading in 5 cities. The measurement units of some pollutants in different years were usually different and contained several blank data. Besides, between 2009 and 2015, data in 2011 and 2012 were collected in different areas, this resulting to discontinuity of data. The performance of the prediction model can be improved if the input is clean and continuous. Therefore, in the next section, we suggest several ways to improve the air quality monitoring system in Yogyakarta; thus, environmental sustainability can be achieved.

3.3. Air quality monitoring system improvement and sustainable environment

The current reality tree is utilised to structurally visualized the cause-and-effect relationship of problems related to air pollution management in Yogyakarta. The use of a manual monitoring system is believed as one of the major sources of uncontrollable air pollution concentration and the increase of victims due to air pollution in the province of Yogyakarta. Manual active sampling is the most common monitoring air quality method in Indonesia, especially in the province of Yogyakarta. This province does not have an automatic air quality monitoring system, which can provide real-time air quality, but currently only available in 14 provinces in Indonesia. This become the reason for the low precise air pollution prediction model because manual active sampling is done twice a year. Most of the published research collected the data in a daily basis for several range of years, which can clearly detect any change in air pollution concentration. Figure 6 visualizes the current reality tree and future reality tree of air quality monitoring system in the province of Yogyakarta.

The use of a manual monitoring system will lead to three preconditions: incomplete data history, unrepresentative data, and misunderstanding of the current environment situation. The environment agency takes samples of air pollution at 30 points, spreading in 5 regions, twice a year. This condition would affect the incomplete data history, and it does not fully represent the current environment condition. By changing the manual monitoring system to an automatic air monitoring system, the above-mentioned problems can be managed. An automatic air quality monitoring system could produce larger and more complete data, which can be more valuable assets for prediction models [38].

If the prediction model has higher accuracy, the policy will be well designed, supported by the understanding of the current environment situation. The related government agencies need to make long-term policies according to the forecasting
results and ensure that the concentration of the substances is below the maximum permitted level to achieve environmental sustainability goals [39]. We suggest several solutions: first implementing an automatic air quality monitoring system, second developing convenient public transportation, and third developing regulations to control the age of vehicles.

First, the real-time air pollution monitoring system and predicting air pollution concentration is one of the ways to minimise the negative impacts of air pollution in humans and ecology; thus, environmental sustainability could be achieved. By predicting air pollution, the related government able to determine the short- and long-term planning for their societies; therefore, mortality and disease rates can be pressed. Therefore, providing a real-time air quality monitoring system is among the first steps the government should take to achieve a sustainable environment.

Installing new automatic air quality monitoring systems is also a dilemma because of the relatively high cost, ranging between € 5000 and € 30,000 per device [40]. Besides it requires routine maintenance and calibration of the instrument, which adds another operational cost. Therefore, the determination of air quality monitoring station should be done wisely. Previously, the environmental agency in Yogyakarta did monitoring of air quality in 30 points spreading in 5 cities, which will result in high costs. K-Means clustering was conducted to divide the locations according to the pollutant levels (NO₂, SO₂, CO, O₃, HC, and PM10). Two-, three, and four- clusters procedures were utilised to find which cluster can provide groups with significantly different profiles. Table 3 shows the results comparison of k-means clustering.

Fig. 6. Current reality tree and future reality of air pollution problem in Yogyakarta.
Table 3. K-means clustering result comparison.

<table>
<thead>
<tr>
<th>NO₂</th>
<th>1</th>
<th>2</th>
<th>NO₂</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.63</td>
<td>36.71</td>
<td></td>
<td>26.78</td>
<td>32.44</td>
<td>36.17</td>
<td></td>
</tr>
<tr>
<td>SO₂</td>
<td>35.06</td>
<td>36.54</td>
<td></td>
<td>50.22</td>
<td>34.60</td>
<td>35.04</td>
</tr>
<tr>
<td>CO</td>
<td>577.54</td>
<td>806.27</td>
<td></td>
<td>387.09</td>
<td>644.28</td>
<td>889.20</td>
</tr>
<tr>
<td>HC</td>
<td>80.61</td>
<td>69.82</td>
<td></td>
<td>102.63</td>
<td>77.54</td>
<td>61.97</td>
</tr>
<tr>
<td>PM₁₀</td>
<td>65.00</td>
<td>61.66</td>
<td></td>
<td>80.21</td>
<td>65.49</td>
<td>52.49</td>
</tr>
</tbody>
</table>

Cluster 1 is a group where air pollutants were relatively higher compared to other groups, while cluster 4 was a group with the highest HC and PM₁₀ concentrations. For this reason, the air monitoring stations are advised to be installed in cluster 1 with 6 locations and cluster 4 with 2 locations. Figure 7 explains the advised location for automatic air quality monitoring device installation.

Members of the cluster is the prospect locations to install the automatic air quality monitoring system and will be chosen according to the highest average concentration level. This is because a cluster with a higher value indicates the severity of pollutants.

We have done several mechanisms by dividing the location into two, three, and four locations; however, dividing the data into two and three groups, resulting in non-specific cluster profile. For instance, in 3-means clustering, cluster two and cluster three were similar. Therefore, we divided the locations into four clusters, in which cluster 1 is a group where air pollutants were relatively higher compared to other groups, while cluster 4 was a group with the highest HC and PM₁₀ concentrations.

![Fig. 7. Recommended location for applying automatic air quality monitoring system.](image-url)
Second, providing convenient public transportation because the large volume of vehicles on the road was the major cause of air pollutants in Yogyakarta. In this province, public transportation was not well developed and inconvenient. Yogyakarta has one bus system called transjogja; however, not all areas were covered by the bus route and it was lack of comfort (e.g., waiting room, schedule, etc.); therefore, people tended to use their own vehicles because it offers more flexibility and comfort.

Third, several previous studies have proven the correlation between the ability to maintain the vehicles and the production of emission concentration. Research in China suggested that air pollutants were more sensitive to maintenance quality than the model year of the vehicles [32]. In Yogyakarta, the vehicle emission test was conducted three days annually with only targeted 2,000 vehicles, which means nearly 0.04% of the total vehicles were tested. Furthermore, there were no any regulations or sanctions applied if the vehicles exceeded the standard. However, in Jakarta, the regulations have been made in 2005 but have not implemented at all. The regulation was written in the Regional Province Regulation No. 2 Year 2005 concerning air pollution control. Several ways to control air pollutants are a) motor vehicles must undergo an emissions test at least every six months, b) the emission test results are used as a tax payment requirement, etc. In this case, Yogyakarta can adopt similar regulations, socialisation to vehicle users and implement it.

4. Conclusion

Air pollution is the main issue of most developing countries, including Indonesia. The high mortality rate caused by air pollution has been increasing annually due to insufficient policy and regulations formulated by the government. Forecasting by using an expert system is one of the ways to minimise the effect of air pollution because of its accuracy; therefore, policy can be better designed. This research has implemented ANFIS to predict six harmful air pollutants.

The performance of the prediction model was fair, with the highest performance of 72.78% in predicting SO\textsubscript{2}. However, several air pollutants were forecasted unsatisfactory. In this model, the error rate in the training set was significantly better compared to the testing set. This difference was caused by several factors, such as data quality containing noise, too focusing on error minimisation, which causes overfitting, and the less precise data selection in building the prediction model.

This research has limited data since Yogyakarta is implementing manual active sampling and collected the data twice, this makes the changes of meteorological aspects and air pollution concentration cannot be recorded in detail. Indonesia through its sustainable development goals are concerned to have sustainable climate action and managing disasters by putting low carbon development on the national development agenda.

Yogyakarta, as the top 3 number of respiratory disease victims currently do not have automatic air quality monitoring and have limited air pollution historical data which is important to achieving sustainable development goals. The real time air pollutant data would increase the performance of the prediction model; hence policy would be better designed. The practical implication is that it helps government-related agency to have better data collection, and this research has pointed out the important area for installation. In addition, due to the high HC and
PM10 in several cities, we recommend the government to prepare comfortable and convenient public transportation, thus people have interested to use public transportation. Other than that, the government should ensure to check the vehicle emissions periodically.

References


