Development of Air Polluter Model for the Carbon
Monoxide (CO) Element Based on Mixed
Geographically Temporal Weighted Regression
(MGTWR) Kriging

Kukuh Winarso
Post Graduate Student of MIPA, Universitas Airlangga and Department of Industrial Engineering, Faculty of Engineering, Universitas Trunojoyo Madura, JL. Raya Telang P.O. Box 2 Bangkalan, Madura, Indonesia

Hari Basuki Notobroto
Department of Biostatistics and Demography, Faculty of Public Health, Airlangga University, JL. Mulyorejo, Kampus C, Unair Surabaya, Indonesia

Fatmawati
Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, JL. Mulyorejo, Kampus C, Unair Surabaya

Copyright © 2014 Kukuh Winarso, Hari Basuki Notobroto, Fatmawati. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract
Motor vehicle exhaust is the main source of Carbon Monoxide (CO). In urban areas, where motor vehicles are easily found, CO is commonly identified as one of the air polluter sources. CO is hazardous, and human being will suffer from serious health problems if they are exposed to this polluting agent in a prolonged period. This study aims to develop a model that will aid the effort to minimize the negative effects of CO. The model is built in which it will be able to show the cause-factors and the preventing-factors of CO generation. The Mixed Geographically Temporally Weighted Regression (MGTWR) approach is used as
the basis of the model. MTGWR is a spatial-temporal regression model, which takes into account the geographical and temporal aspects of the pollution. However, the MGTWR model cannot be used to predict the effect if it is used outside the sample location of the research, unless we predict the associated regression coefficients in the respected area beforehand. In this case, we use the estimated predictor parameter based on the Kriging method to predict the regression parameters outside the research location. As the result, the Kriging Predictor-based MTGWR model can be used to estimate the pollution level caused by CO outside the sample location of the research.

**Keywords:** Air Polluter, Carbon Monoxide, Spatio-Temporal, MGTWR, Kriging-predictor

1. **INTRODUCTION**

Previous study shows that there is a parallel relationship between the number of polluted air-generated disease sufferers and the intensity of industrialization and urbanization in countries whose level of air pollution are severely high [1]. Carbon Monoxide (CO) is one of the air polluters, which is generated from the imperfect fuel combustion of motor vehicles and industrial machines. Motor vehicle exhaust is the main generator of Carbon Monoxide and it is common to be found in big cities. Based on data, about 60% of air pollution in major urban areas is related with public transportation [2]. Within the context of human health, Carbon Monoxide is toxic. It causes lower fetal weight, higher rate of infant death, and leads to human brain damage. The air quality standard that is related with Carbon Monoxide is 10.000 ug/Nm3 [3]

Demographically, the potential of the effect and causes of air pollution will be different one to another in different regions. Also, it is inappropriate to analyze the effect and causes of air pollution based on a global approach, as we will be unable to examine the influence of local variations if we are to use global approach [4-5]. The common spatial regression method known is the Geographically Weighted Regression (GWR). GWR is a weighted-regression technique that is based on the simple regression approach. It is a common statistical technique used for spatial heterogeneity. The term heterogeneity in this case is the measurement relationship of different variables among various locations [6]. Spatial heterogeneity occurs when an independent variable gives different responses to different locations within the same research location [7]. The crux of GWR approach is determining the regression model for each of the location nodes so that the regression models obtained are unique, in which a model for a node will be different to the others [8]. There is another model aside from GWR, which is Mixed Geographically Temporal Weighted regression (MGTWR). MTGWR is a combined modeling approach between global regression and local regression (GWR) [9]. This paper takes into consideration the element of location, as well as the time into the local variation based on MGTWR approach. The
Development of air polluter model

usefulness of global regression is that it can be used to predict in each location. However, in the spatial regression models such as GWR, MGWR (Mixed Geographically Weighted Regression), and MGTWR, the global regression cannot be used for predicting outside the research sample location, unless we predict the respected location’s regression coefficients beforehand [10]. In order to eliminate this obstacle, we estimate the predictor parameter using the Kriging method to predict the regression parameter outside the research location.

2. RESEARCH OBJECTIVE

This research aims to develop the air polluter model for the Carbon Monoxide (CO) element based on the Mixed Geographically Temporal Weighted Regression (MGTWR) approach, in which the developed model can be used to predict outside the research location using the Kriging predictor estimation. This way, the developed model can be used to predict the regression parameter outside the research location.

3. METHODOLOGY

3.1 Mixed Geographically Temporal Weighted Regression (MGTWR)

MGTWR is the advancement of the MGWR model [11] with the incorporation of the temporal element into the model. The temporal factor is aimed to predict the observation time, and to complement the location (location coordinate) factor. Mathematically, the temporal model can be stated as the following:

\[ y_i = \sum_{j=1}^{q} \beta_i x_{ij} + \beta_j \left( u_i, v_i, t_i \right) x_{ij} + e_i \]  

(1)

Where: \( i = 1, 2, 3 \ldots n \)
- \( \beta_i \) is global variable \( x_i \)
- \( \beta_j \) is local variable \( x_j \)
- \( y_i \) is Response of variables
- \( x_i \) is predictors of variables
- \( u_i \) is longitude at the time-\( i \)
- \( v_i \) is latitude at the time-\( i \)
- \( t_i \) is the time length at the time-\( i \)

3.2. Kriging Spatial Predictor Model

Kriging method is used to predict the regression parameter outside the research location [12]. The assumption that is associated with the predictor is:

\[ p(b; s_0) = \sum_{i=1}^{n} \lambda_i b(s_i), \quad \sum_{i=1}^{n} \lambda_i = 1 \]  

(2)

Where:
- \( \lambda_i \) = the weight \( b(s_i) \) for the estimation of location \( s \), with the value of \( b(s_i) \) has different weighted-coefficients for the estimation in different locations
- \( s_i \) = location vector
- \( s_0 \) = Location of the data to be estimated
- \( n \) = number of the sample data
The Kriging spatial predictor matrix notation can be stated as the following [12]:

\[
\begin{bmatrix}
\gamma(s_i - s_j) & f_t(s_i) \\
0 & f_t(s_j)
\end{bmatrix}
\begin{bmatrix}
\lambda_j \\
m_t
\end{bmatrix}
= \begin{bmatrix}
\gamma(s_i - s_j) \\
f_t(s_0)
\end{bmatrix}, \text{ for } i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, p
\]

Based on the previous matrix, we can compute the weight matrix for the Universal Kriging, which is:

\[
\begin{bmatrix}
\lambda_j \\
m_t
\end{bmatrix}
= \begin{bmatrix}
\gamma(s_i - s_j) & f_t(s_i) \\
f_t(s_j) & 0
\end{bmatrix}^{-1}\begin{bmatrix}
\gamma(s_i - s_j) \\
f_t(s_0)
\end{bmatrix}
\]

(3)

Notations:
\(\gamma(s_i - s_j)\) = Semivariogram between sampled nodes
\(\gamma(s_0 - s_i)\) = Semivariogram between sampled node and estimated node
\(f_t(s_i), f_t(s_j)\) = Location coordinate of the sampled data
\(\lambda_j\) = Weighting value to be computed
\(m_t\) = Value of lagrange parameter
\((s_i, s_j)\) = Location of the sampled data
\(s_0\) = Location of the data to be estimated
\(p\) = Number of order in the trend equation

Semivariogram is used to determine the distance where the observation data values are not dependent one to another or have no correlation among them [12]. The equation in the Kriging model is used to model the trend of the low order polynomial, which is the first or the second order [13]. If the obtained trend has the first order at \(R^3\), then the equation is:

\[
m(s) = m(x,y,z) = a_0 + a_1x + a_2y + a_3z
\]

where \(x,y,z\) are the sampled location coordinates. Thus, the equation (3) becomes as the following:

\[
\sum_{j=1}^{n} \lambda_j \gamma(s_i - s_j) + m_0 + m_1x_i + m_2y_i + m_3z_i = \gamma(s_0 - s_i), \text{ for } i = 1, 2, \ldots, n
\]

With assumption:

\[
\sum_{j=1}^{n} \lambda_j x_j = x, \quad \sum_{j=1}^{n} \lambda_j y_j = y, \quad \sum_{j=1}^{n} \lambda_j z_j = z, \quad \sum_{j=1}^{n} \lambda_j = 1
\]

If we want to transform the equation above into the Kriging matrix in the
Development of air polluter model

The following first order Kriging trend (KT):

\[
K_{KT} = \begin{bmatrix}
\gamma(s_1 - s_1) \cdots \gamma(s_1 - s_n) & 1 & x_1 & y_1 & z_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\gamma(s_n - s_1) \cdots \gamma(s_n - s_n) & 1 & x_n & y_n & z_n \\
1 & \cdots & 1 & 0 & 0 & 0 \\
x_1 & \cdots & x_n & 0 & 0 & 0 \\
y_1 & \cdots & y_n & 0 & 0 & 0 \\
z_1 & \cdots & z_n & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_n \\
m_0 \\
m_1 \\
m_2 \\
m_3
\end{bmatrix}
\]

\[
k_{KT} = \begin{bmatrix}
\gamma(s_0 - s_1) \\
\vdots \\
\gamma(s_0 - s_n) \\
1 \\
x \\
y \\
z
\end{bmatrix}
\]

where: \( K_{KT} \lambda_{KT} = k_{KT} \)

then the weighting matrices can be stated as the followings [13]:

\[
\lambda_{KT} = K_{KT}^{-1} k_{KT}
\]

The case study in this research uses the Carbon Monoxide (CO) element as the response variable \( Y \), while the predictor variables are the air temperature (\( X_1 \)), the wind velocity (\( X_2 \)), the air humidity (\( X_3 \)), the traffic velocity (\( X_4 \)), the area size of the urban forest (\( X_5 \)), the population density (\( X_6 \)), and the business center aspect (\( X_7 \)) [14]

4. RESULTS AND DISCUSSION

The first MGTWR model test is the model fitness, in which the hypothesis can be stated as the followings:
H₀ : \( \beta_k \left( u_i, v_i, t_i \right) = \beta_k \left( u_i, v_i \right) \) \( k = 0, 1, 2, \ldots \)
(MGTWR model is not different from the MGWR model)

H₁ : At least one \( \beta_k \left( u_i, v_i, t_i \right) \neq \beta_k \left( u_i, v_i \right) \)
(MGTWR model is significantly different from the MGWR model)

The test results are summarized in Table 1:

**Table 1** Fitness Test of MGTWR Model for the CO Response

<table>
<thead>
<tr>
<th>Source of Error</th>
<th>Sum Square</th>
<th>Degree of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>7,7679</td>
<td>15,9260</td>
<td>0,4877</td>
<td>4,2433</td>
<td>0,0000</td>
</tr>
<tr>
<td>MGTWR</td>
<td>46,1362</td>
<td>401,3706</td>
<td>0,1149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGWR</td>
<td>53,9041</td>
<td>417,2966</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows that the F test statistical value is 4,2433, and the p-value of 0,0000. Using the significance value \( \alpha \) of 5\%, we must reject \( H_0 \), and conclude that the MGTWR model is significantly different from the MGWR model. Therefore, we can further conclude that the MGTWR model is more proper to model the Air Polluter Standard Index (APSI) for Carbon Monoxide (CO). This means that the time element is influential in the APSI modeling for CO so that not only the location factor is considered, but the observation time is taken as the influential factor to the APSI model for CO as well.

Next step is the global parameter test that is conducted to identify which global predictor variables significantly influence the response. In this test, we perform the partial global parameter test using the hypothesis below:

H₀ : \( \beta_k = 0 \) (global variable \( x_k \) is not significant)

H₁ : \( \beta_k \neq 0 \) (global variable \( x_k \) is significant)

Table 2 shows that by using the level of significance \( \alpha \) 5\%, we can conclude that the global predictor variable that significantly influence the APSI for CO are the area size of the urban forest (\( X_5 \)) and the business center aspect (\( X_7 \)), because these variables have the p-value that are less than 0.05.
Table 2  Parsial Test of Global Variable of MGTWR Model

<table>
<thead>
<tr>
<th>Global Parameter</th>
<th>Variable</th>
<th>Beta</th>
<th>T</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X_5</td>
<td>-0.0445</td>
<td>-0.4171</td>
<td>0.0080*</td>
</tr>
<tr>
<td></td>
<td>X_6</td>
<td>0.0088</td>
<td>0.3556</td>
<td>0.3613</td>
</tr>
<tr>
<td></td>
<td>X_7</td>
<td>-0.0456</td>
<td>2.0045</td>
<td>0.0228*</td>
</tr>
</tbody>
</table>

Note: *) significant at α = 5%

By using α = 5%, the location and the observation time of APSI for CO in Surabaya city can be classified based on the significant variables for APSI values for CO. The detailed significant variable list in each location and in each observation time is presented in Table 3:

Table 3 MGTWR Model for CO at Five Observation Location Nodes

<table>
<thead>
<tr>
<th>No</th>
<th>Observation Location</th>
<th>Observation Time</th>
<th>Significant Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SUF 1 (Location 1)</td>
<td>Morning</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>X_1, X_2 and X_4</td>
</tr>
<tr>
<td>2</td>
<td>SUF 3 (Location 2)</td>
<td>Morning</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>X_1, X_2, X_3, X_4</td>
</tr>
<tr>
<td>3</td>
<td>SUF 4 (Location 3)</td>
<td>Morning</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>X_1, X_2, X_3, X_4</td>
</tr>
<tr>
<td>4</td>
<td>SUF 5 (Location 4)</td>
<td>Morning</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>X_1, X_2, X_3, X_4</td>
</tr>
<tr>
<td>5</td>
<td>SUF 6 (Location 5)</td>
<td>Morning</td>
<td>X_3, X_4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>X_2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>X_2, X_3, and X_4</td>
</tr>
</tbody>
</table>

The MGTWR model obtained in each of the location and observation time will be different one to another, depending on the parameter value of MGTWR and on which predictor variables that significantly influence the response variable of APSI (CO) with the R^2 value of 36.60%.
To predict the MGTWR model at a certain location node, we use the Kriging approach that is presented in the following Table 4:

Table 4 Kriging-based MGTWR Model with Local Parameter Lokal for CO Element

<table>
<thead>
<tr>
<th>No</th>
<th>Sub-district</th>
<th>Time</th>
<th>Constant</th>
<th>X_1</th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Location z</td>
<td>Morning</td>
<td>4.5498</td>
<td>-0.0541</td>
<td>0.0690</td>
<td>-0.1437</td>
<td>0.0614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>4.4902</td>
<td>0.0164</td>
<td>0.0180</td>
<td>0.0932</td>
<td>0.2797</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>4.5415</td>
<td>0.1745</td>
<td>0.0870</td>
<td>0.1284</td>
<td>0.2571</td>
</tr>
<tr>
<td>2</td>
<td>Location y</td>
<td>Morning</td>
<td>4.5523</td>
<td>-0.0425</td>
<td>0.0667</td>
<td>-0.1441</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>4.4903</td>
<td>0.0145</td>
<td>0.0192</td>
<td>0.0833</td>
<td>0.2773</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>4.5366</td>
<td>0.1706</td>
<td>0.0840</td>
<td>0.1233</td>
<td>0.2567</td>
</tr>
<tr>
<td>3</td>
<td>Location x</td>
<td>Morning</td>
<td>4.5457</td>
<td>-0.0618</td>
<td>0.0665</td>
<td>-0.1335</td>
<td>0.0693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noon</td>
<td>4.4897</td>
<td>0.0202</td>
<td>0.0206</td>
<td>0.0923</td>
<td>0.2775</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evening</td>
<td>4.5441</td>
<td>0.1686</td>
<td>0.0874</td>
<td>0.1262</td>
<td>0.2612</td>
</tr>
</tbody>
</table>

In Table 4, if we are to estimate for location x, the Kriging-MGTWR with the CO response for evening is:

$$\ln (\text{CO}) = 4.544 + 0.0874X_2 + 0.2612X_4 + 0.0445X_5 - 0.0456X_7$$
The classification map in Surabaya based on CO polluter can be seen in the following:

![Classification map of sub-districts in Surabaya based on CO polluter](image)

**Figure 1. Classification of Sub-districts in Surabaya Based on the Air Polluter CO**

There are eleven sub-districts that are classified as unhealthy (yellow-colored) and twenty sub-districts that are classified as fair (blue-colored).

5. CONCLUSION

Aside from the global influence, the MGTWR model of the air pollution is also influenced significantly by the location factor (geographical factor). Therefore, in this study, the mixed model is suitable to model the air polluter. The local influence of MTGWR for CO element for observation time morning and noon show that the two significantly influencing predictor variables are the air humidity ($x_3$) and the traffic velocity ($x_4$). This indicates that in the morning and noon, the influence of motor-vehicle exhaust and air humidity are very dominant in polluting the air with CO. For the night observation, almost all of the predictor variables influence the air pollution level, with the global influence factors of it are the area size of the urban forest ($x_5$) and the business center ($x_7$). To predict an
MGTWR model outside the observation nodes, we can use the Kriging estimation approach.

REFERENCES


Received: May 1, 2014