RESEARCH ARTICLE

Advanced Science Letters
Vol. 23, 6545–6547, 2017

Modeling of Air Polluter NO_2 and SO_2 Using Geographically Weighted Multivariate Regressions (GWMR)

Hasbi Yasin*, Dwi Ispriyanti, and Abdul Hoyyi

Department of Statistics, Diponegoro University, Indonesia

Air pollution is one of the most concerned problems on earth today. It is closely related with and mostly generated from the transportation and industrialization sectors, as well as from the environmentally degrading effect of the urban physical development. Air pollution promotes the lower level of air quality, which in turn promotes the greater risk on health, especially that of the human being. This research aims to aid the government in the policy making process related to air pollution mitigation by developing a standard index model for air polluter (Air Polluter Standard Index—APSI) based on the Geographically Weighted Multivariate Regression (GWMR) approach. This GWMR-based APSI model is aimed as the indicator of the prevailing air pollution. Two of the five polluting elements in the APSI are Nitrogen Dioxide (NO_2) and Sulfur Dioxide (SO_2). The GWMR approach used in modeling the APSI in this research is a spatial multivariate regression model, which is expected to be able to show the effect of the air polluters to the level of air pollution with regard to the geographical aspects of the prevailing event. In this case, the result concluded that there are no significance differences between Multivariate Regression and GWMR in this case. But GWMR is the best model to modeling NO_2 and SO_2 because it has smallest AIC and MSE.

Keywords: Air Polluter, APSI, Multivariate Regressions, GWMR.

1. INTRODUCTION

Patients with diseases caused by air pollution with increased industrialization and urbanization in developing countries showed a parallel relationship.[1] One type of pollutant gases emitted by motor vehicles is NO_2 and SO_2.[2] When high concentrations of SO_2 and NO_2 will cause acid rain that can endanger the health of the respiratory tract. Gas Sulphur dioxide (SO_2) is one element of air pollutants generated from the combustion of lubricant and incomplete combustion from motor vehicles and industrial machinery. Fumes are a major source for SO_2 in various cities. There is 60 percent of air pollution in large cities contributed by public transport.[3] Sulphur dioxide is a poison that causes shortness of breath, nervous disorders. At levels above the threshold limit, can cause death. Victims of sulphur dioxide are not only humans, but also buildings and plants. The existence of this gas in the air can cause acid rain which damaged building materials and impede growth plants. Standard quality that is allowed is ≤ 365 mg/Nm^3.[4] An appropriate model for this case is to look at the elements of the location (geography), because of the potential impact and cause air pollution will vary between locations.[5,6]

Regression model by looking at the spatial element is *Geographically Weighted Regression* (GWR). GWR is a statistical method used to analyze Spatial heterogenity referred to is a state measurement of the relationship between variables vary from one location to another.[7] Spatial heterogeneity occurs when the same independent variables are not the same response in different locations within the study area.[8] GWR models can be used when the response variable consists of only one variable. If the response variables were analyzed over one and correlated, then the model can be developed into a regression model spatial multivariate, or known as Geographically Weighted Multivariate Regression (GWMR).[9,10] This is consistent with the phenomenon of air pollutant elements allegedly correlated with one another. Therefore, in this study will be carried out modeling of pollutant elements using GWMR.

2. RESEARCH DETAILS

This research aims to developt the Geographically Weighted Multivariate Regression (GWMR) model, which is intended to see the influence of the predictor variables to the Air Polluter Standard Index (APSI) locally. Espeespecially for NO_2 and SO_2 pollutant elements.

*Author to whom correspondence should be addressed.

 1936-6612/2017/23/6545/003 doi:10.1166/asl.2017.9678

 Adv. Sci. Lett. 23, 6545–6547, 2017

2.1. Data

- Location:
  1. Centre Surabaya (Jalan Ketabang kali//SUF1),
  2. South Surabaya (Jalan Masjid Al Akbar Gayungan/SUF4),
  3. West Surabaya (Jalan Simomulyo/SUF3),
  4. East Surabaya (Jalan Arief Rahman Hakim/SUF5),
  5. Wonorejo Surabaya (Jalan Kendal Sari 117/SUF6).
- Respon Variables: the APSI data (NO_2 and SO_2)
- Predictor Variable:
  1. The traffic velocity,
  2. The population density,
  3. The business center aspect,
  4. The air humidity,
  5. The wind velocity,
  6. The air temperature,
  7. The size of the urban forest.

3. METHODOLOGY

3.1. Multivariate Regression

A multivariate regression model regression models were constructed of several response variables which are correlated based on some of the predictor variables.[11] For q response variable and p predictor variables, a multivariate regression model can be written as follows:

$$\mathbf{Y}_{n\times q} = \mathbf{X}_{n\times(p+1)}\mathbf{B}_{(p+1)\times q} + \boldsymbol{\varepsilon}_{n\times q} \tag{1}$$

$\mathbf{Y}$ is the matrix of the response variable sized $n\times q$, $\mathbf{X}$ is the matrix of variable size $n\times(p+1)$, $\mathbf{B}$ is a matrix of regression coefficient measuring $(p+1)\times q$, and $\boldsymbol{\varepsilon}$ matrix of residual sized $n\times q$ are assumed to be identical, independent and normally multivariate distributed with zero mean and constant variance.

Multivariate regression parameter estimation is done using Ordinary Least Squares (OLS) as well as on the linear regression model. In order to obtain estimates of the following: Multivariate regression model is a regression model with more than one response variable. Simultaneous multivariate regression model consisting of q models can be expressed as:

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

3.2. Geographically Weighted Regression (GWR)

GWR Model is a development of the regression model where the basic idea is taken from the non-parametric regression.[12] Model parameters are local to each observation location. GWR models can be written as follows:[8]

$$y_i = \beta_0(i) + \sum_{k=1}^{p}\beta_k(i)x_{ik} + \varepsilon_i \tag{2}$$

Parameter estimation using Weighted Least Squares (WLS) is to give a different weighting for each observation location. So that the parameter model for each location are:

$$\hat{\boldsymbol{\beta}}(i) = (\mathbf{X}^T\mathbf{W}(i)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}$$

$$\mathbf{W}(i) = \mathrm{diag}(w_1(i), w_2(i), \ldots, w_n(i))$$

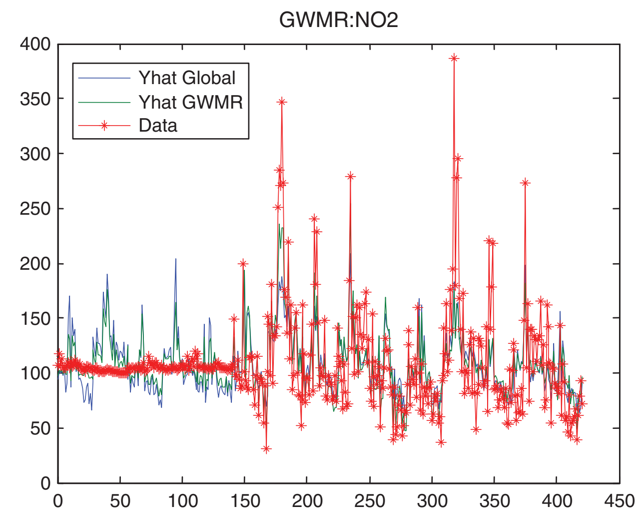

Fig. 1. Predicted and actual value of NO_2 element.

3.3. Geographically Weighted Multivariate Regression (GWMR)

GWMR is the development of a multivariate regression model with parameters locally to each observation location:[9,10]

$$\mathbf{Y}_{n\times q} = \mathbf{X}_{(n\times(p+1))}\mathbf{B}(u_i, v_i)_{((p+1)\times q)} + \boldsymbol{\varepsilon}_{n\times q} \tag{3}$$

Model parameters estimated by the WLS:

$$\begin{aligned}\hat{\mathbf{B}}(u_i, v_i) &= \begin{bmatrix} \hat{\underset{\sim}{\beta}}_1(u_i, v_i) \\ \hat{\underset{\sim}{\beta}}_2(u_i, v_i) \\ \vdots \\ \hat{\underset{\sim}{\beta}}_q(u_i, v_i) \end{bmatrix}^T = \begin{bmatrix} [\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\underset{\sim}{\hat{y}_1} \\ [\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\underset{\sim}{\hat{y}_2} \\ \vdots \\ [\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\underset{\sim}{\hat{y}_q} \end{bmatrix}^T \\ &= [\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(u_i, v_i)\mathbf{Y}\end{aligned}$$

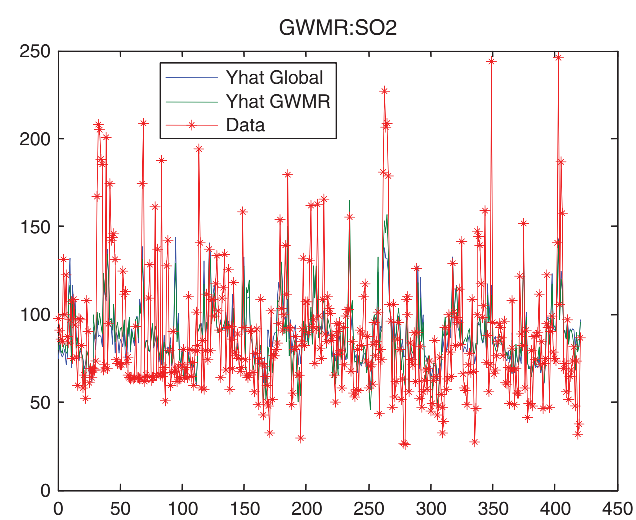

Fig. 2. Predicted and actual value of SO_2 element.

3.4. Steps of Data Analysis

1. Prepare the APSI data (NO_2 and SO_2) and the predictors variable from 5 location
2. Determine the euclidean distance
3. Find the optimum bandwith with CV
4. Determine the weighting matrix by gaussian kernel function
5. Estimate the parameters model in each location
6. Test the significance model and parameters
7. Interprate the result.

4. RESULTS AND DISCUSSION

The first step taken in modeling is to determine the location of each sample to be used is the geographic location of statsiun observation of air pollution in the city of Surabaya. Then search for the optimum bandwidth based on the coordinates of the location of observation with cross validation procedure to draw a spatial weighting matrix. Based on the weighting matrix, then do GWMR model estimation. Plot the predicted value and the actual value shown in Figures 1 and 2.

Based on Table I shows that there are no significant differences between the multivariate regression model with GWMR. This indicates that the spatial effect does not provide a real change to the modeling of air pollution for NO_2 and SO_2 elements. However, in this case more appropriate MGWR models when compared with multivariate regression model because the model GWMR has the smallest AIC and MSE as shown in Table II.

Table I. Goodness of fit test of GWMR model.

Variable	Source	*SS*	*df*	*MS*	*F*
NO_2	Global errors	459,510.12	412.00		
	GWR improvement	115,224.27	77.03	1,495.92	
	GWR errors	344,285.75	368.78	933.59	1.602
SO_2	Global errors	427,360.94	412.00		
	GWR improvement	73,945.04	77.03	960.01	
	GWR errors	353,415.90	368.78	958.35	1.002

Table II. Comparison of models.

Model	AIC	MSE
Multivariate regression	44,4605	2,755
GWMR	44,4603	0,454

5. CONCLUSION

Based on the results and discussion can be concluded that there is no significant difference between the multivariate regression model with GWMR. However, in this case more appropriate MGWR models when compared with multivariate regression model because the model GWMR has the smallest AIC and MSE.

Acknowledgments: We would like to give thank to Directorate of Research and Public Services, The Ministry of Research, Technology and Higher Education Republic of Indonesia for their support. This research was funded by "Fundamental" Research Grant 2016.

References and Notes

1. Ebtekar, *Allergy Asthma Immunol.* 5, 47 **(2006)**.
2. K. Winarso, H. B. Notobroto, and Fatmawati, *International Journal of Academic Research* 6, 193 **(2014)**.
3. BAPEDAL, Indeks Standar Pencemar Udara, Menteri Lingkungan Hidup, Jakarta **(1997)**.
4. M. Fahimi, B. Dharma, D. Fetararayani, A. Baskoro, G. Sugiarto, and C. Effendi, *Jurnal Penyakit Dalam.* 1 **(2012)**.
5. D. Robinson and J. M. Lloyd, *International Journal of Applied Earth Observation and Geoinformation* 21, 374 **(2013)**.
6. A. Gilbert and J. Chakraborty, *Social Science Research* 40, 273 **(2011)**.
7. J. Mennis, *The Cartographic* 43, 171 **(2006)**.
8. A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression, Jhon Wiley & Sons, Chichester, UK **(2002)**.
9. S. Harini, Purhadi, M. Mashuri, and S. Sunaryo, *International Journal of Applied Mathematics and Statistics* 29, 110 **(2012)**.
10. M. N. Hayati and Purhadi, *Statistika* 2 **(2014)**.
11. A. C. Rencher, Linear Models in Statistics, John Wiley & Sons, New York **(2002)**.
12. C. Mei, S. He, and K. Fang, *Journal of Regional Science* 2, 143 **(2004)**.

Received: 5 September 2016. Accepted: 13 December 2016.