MODELING OF AIR POLLUTANTS SO2 ELEMENTS USING GEOGRAPHICALLY WEIGHTED REGRESSION (GWR), GEOGRAPHICALLY TEMPORAL WEIGHTED REGRESSION (GTWR) AND MIXED GEOGRAPHICALLY TEMPORALWEIGHTED REGRESSION (MGTWR)

Kukuh Winarso[1] and HasbiYasin[2]
[1]Department of Industrial Engineering, Faculty of Engineering, Trunojoyo Madura University, Indonesia
[2]Department of Statistics, Faculty of Sciencee and Mathematics, Diponegoro University, Indonesia
JL. Raya Telang Kamal, Bangkalan, Madura, Indonesia
E-Mail: kukuhutm@gmail.com

ABSTRACT

Sulphur dioxide gas (SO2) is derived from the combustion of fuels containing sulphur. Aside from fuel, sulphur is also contained in the lubricant. Sulphur dioxide gas is difficult to detect because it is colourless gas. Sulphur dioxide can cause respiratory disorders, indigestion, headache, chest pain, and nerve. A necessary preventive measures to reduce the impact of air pollutants SO2 particular elements, one of them by making the modeling that can bring the causes and factors resistor element of air pollutants SO2. The modeling is Geographically Weighted Regression (GWR), Temporal Geographically Weighted Regression (GTWR) and Mixed Geographically Weighted Temporal Regression (MGTWR). All three models are regression models spatial, temporal and spatial temporal spatial- combined, which models the effects of air pollutants SO2 element with a direct view of geography and time of occurrence of air pollution. The third model is then compared to obtain the best model in the modeling of air pollutants SO2 elements.

Keywords: air polluter, sulphur dioxide, GWR, GTWR, MGTWR.

1. INTRODUCTION

Patients with diseases caused by air pollution with increased industrialization and urbanization in developing countries showed a parallel relationship [1]. Gas Sulphur dioxide (SO2) is one elements of air pollutants generated from the combustion of lubricant and incomplete combustion from motor vehicles and industrial machinery. Fumes are a major source for sulphur dioxide (SO2) in various cities. There is 60 percent of air pollution in large cities contributed by public transport [2]. 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 / Nm3 [3]. 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 [4-5]. Regression model by looking at the spatial element is *Geographically Weighted Regression* (GWR). GWR is a statistical method used to analyse Spatial *heterogeneity* referred to is a state measurement of the relationship (*Measurement of relationship*) Between variables vary from one location to another location [6]. Spatial heterogeneity occurs when the same independent variables are not the same response in different locations within the study area [7]. Namely the Model GWR regression model to the point that differs from the other dots [8]. Besides GWR, spatial regression model that also involves the element of time is Temporal Geographically Weighted Regression (GTWR) as well as the model mix of GWR, GTWR and OLS regression. Mixed Geographically Temporal Weighted Regression (MGTWR) is a global regression modeling combined with local regression (GWR) [9]. Later in this paper a compare of the three models, namely GWR, GTWR and MGWTR to see the best model.

2. RESEARCH OBJECTIVE

This study aims to model air pollutant elements of SO2 with GWR approach, GTWR and MGTWR and comparing them.

3. METODOLOGY

3.1Geographically weighted regression (GWR)

Geographically Weighted Regression (GWR) is the development of a regression model where each parameter is calculated at each observation location, so that each location has a value parameter regression observation Different. GWR model can be written as follows [6-7]:

$$y_i=\beta_0\left(u_i,v_i\right)+\sum_{k=1}^{p}\beta_k\left(u_i,v_i\right)x_{ik}+\varepsilon_i \tag{1}$$

Where

$\left(u_i,v_i\right)$ = Stating geographical location coordinates (*Longitude*, *latitude*) From the location of the observation – *i*

$\beta_k\left(u_i,v_i\right)$ = The regression coefficient predictor variables -*k* andlocation of the observation-*i*

VOL. 11, NO. 13, JULY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

3.2 Geographically temporal weighted regression (GTWR)

GTWR not only accommodate the spatial heterogeneity, but also pay attention to heterogenitas temporally [10].STWR for variable predictor of the dependent variables yi coordinate {(ui, vi, ti)} for each observation is as follows:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i \qquad (2)$$

Where
i = 1,2,3….n
u_i=longitude at the time-i
v_i= latitude at the time-i
t_i = time length at the time-i

3.3 Mixed geographically temporal weighted regression (MGTWR)

MGTWR is the development of a model MGWR [11], with the added element of time that is temporal. Temporal factors developed is intended to predict the observed time serves to complement the location factors (location coordinates). Mathematically, temporal models that can be expressed as follows:

$$y_i = \sum_{j=1}^{q} \beta_j x_{ij} + \beta_j\left(u_i, v_i, t_i\right) x_{ij} + \varepsilon_i \qquad (3)$$

Where: i = 1,2,3….n
u_i=longitude at the time-i
v_i= latitude at the time-i
t_i = time length at the time-i

In the case study of this research used the response variable Y is an element of Sulphur Dioxide (SO2) and predictor variables, the water temperature (X1), the wind velocity (X2), the air humidity (X3), the traffic velocity (X4), the area size of the urban forest (X5), the population density (X6), the business center aspect (X7) [12].

4. RESULTS AND DISCUSSIONS

4.1 Statistical test model GWR, GTWR and MGTWR

Test Model GWR, GTWR and MGTWR the first is testing the suitability of the model.

1. The first GWR model test is the model fitness, in which the hypothesis can be stated as the followings:

H_0 : $\beta_k\left(u_i, v_i, t_i\right) = \beta_k$ k=0,1,2,⋯,4, dan i=1,2,⋯,5
(GWR model is no different from the global Regression)

H_1 : At least one $\beta_k\left(u_i, v_i, t_i\right) \neq \beta_k$
(GWR model is significantly different from the global Regression Model)

The test results are summarized in Table 1:

Table-1. Fitness test of GWR model for the SO_2 Response.

Source Error	Sum Square	Degree of Fredom	Mean Square	F	p-value
Improvement	14.235	78.991	0,1789	6,264	0,0093
GWR	56,1587	4.103.512	0,1369		
Regression	57,5722	4.100.000			

Table-1 indicates that the value of the F test statistic of 6.2641 with a p-value of 0.0093. Out by using a level of significance (α) 5% then Reject H0 and concluded that differ significantly from the GWR models Global regression, so the model is more feasible GWR to describe the data modeling elements SO2. Very influential in the location element modeling elements SO2.

GWR statistical summary of the parameters are as follows:

Table-2. Summary statistics local parameter GWR model with weighted Exponential.

Parameter Statistics Summary GWR					
Variable	Min	Max	Mean	Range	StDv
B_0	45.331	45.526	45.404	0.0195	0.0073
B_1	0.0202	0.0246	0.0218	0.0044	0.0016
B_2	-0.0568	-0.0373	-0.0460	0.0194	0.0066
B_3	0.0099	0.0178	0.0144	0.0079	0.0029
B_4	0.1365	0.1475	0.1415	0.0110	0.0042
B_5	0.0845	0.1001	0.0933	0.0157	0.0060
B_6	-0.1218	-0.0907	-0.1061	0.0310	0.0106
B_7	0.0614	0.0753	0.0687	0.0140	0.0054

To see where the predictor variables that influence different at each location of the observation, it can be used to test the partial influence of geographical factors for each predictor variable. Table-3 shows that using a significance level (α) 5%it can be concluded that the variables Influence locally, the wind velocity (X2), the traffic velocity (X4), the area size of the urban forest (X5), the population density (X6), the business center aspect (X7).

Table-3. Test of influence of geographic factors predictorsof GWR with weighting variables Exponential

Variable	F	P-Value
Constant	238.094	0.0000*
Z_1	0.1311	0.8641
Z_2	53.318	0.0094*
Z_3	0.4647	0.6149
Z_4	31.722	0.0467*
Z_5	229.289	0.0000*
Z_6	173.740	0.0000*
Z_7	186.414	0.0000*

Note: *) significant at α = 5%

2. The fitness Test of GTWR model, in which the hypothesis can be stated as the followings:

H_0 : $\beta_k\left(u_i,v_i,t_i\right)=\beta_k$ k=0,1,2,⋯,4, dan i=1,2,⋯,5
(GTWR model is not different from the global regression model)

H_1 : At Least One $\beta_k\left(u_i,v_i,t_i\right)\neq\beta_k$ (GTWR model is significantly different from the global regression model)

The test results are summarized in Table-4:

Table-4. Fitness Test of GTWR Model for the SO_2 Response.

Source of Error	Sum Square	Degree of Fredom	Mean Square	F	p-value
Improvement	113.219	249.389	0,454	4,667	0,0001
GTWR	46,2503	3.914.643	0,1181		
Regression	57,5722	4.120.000			

Table-4 shows that the F test statistical value is 4,6665, and the p-value of 0,000. Using the significance value(α)of 5%, we must reject H_0, and conclude that the GTWR model is significantly different from the Global Regression model. Therefore, we can further conclude that the GTWR model is more proper to model the Air Polluter Standard Index (APSI) for SO_2. This means that the time element is influential in the APSI modeling for SO_2 so that not only the location factor is considered, but the observation time is taken as the influential factor to the APSI model for SO_2 as well.

GTWR statistical summary of the parameters are as follows:

Table-5. Summary statistics local parameter model GTWR with weighted exponential.

Parameter Statistics Summary GTWR					
Variable	Min	Max	Mean	Range	StDv
B_0	43.949	44.958	44.505	0.1010	0.0400
B_1	-0.0065	0.0629	0.0220	0.0695	0.0263
B_2	-0.0950	0.0312	-0.0491	0.1262	0.0496
B_3	-0.0545	0.2128	0.0487	0.2672	0.1058
B_4	-0.0294	0.2330	0.1388	0.2624	0.1141
B_5	0.0579	0.1362	0.0927	0.0783	0.0276
B_6	-0.1383	-0.0271	-0.0850	0.1111	0.0359
B_7	0.0294	0.0607	0.0464	0.0313	0.0088

To see where the predictor variables that influence different at each location of the observation, it can be used to test the partial influence of geographical factors for each predictor variable. Table 6 shows that the level of significance (α) 5% it can be concluded that the variables that influence local, the traffic velocity (X4).

Table-6. Test variables influent geographic factor predictor GTWR with weighted exponential.

Variable	F	P-Value
Constant	0.2984	0.7185
Z_1	0.1007	0.8984
Z_2	19.120	0.1494
Z_3	44.467	0.0127
Z_4	204.349	0.0000*
Z_5	10.869	0.3389
Z_6	0.9692	0.3812
Z_7	0.0702	0.9324

Note: *) significant at α = 5%

3. The fitness Test of MGTWR Model, in which the hypothesis can be stated as the followings:

H_0 : $\beta_k\left(u_i,v_i,t_i\right)=\beta_k\left(u_i,v_i\right)$ k=0,1,2,⋯,4, dan i=1,2,⋯,5
(MGTWR model is not different from the MGWR model)

H_1 : 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)

The test results are summarized in Table-7:

Table-7. Fitness test of MGTWR model for the SO_2 response.

Source of Error	Sum Square	Degree of Fredom	Mean Square	F	p-value
Improvement	86.937	98.418	0,8833	9,2852	0,0000
MGTWR	481.719	4.055.713	0,1188		
MGWR	568.656	4.118.569			

Table-7 shows that the F test statistical value is 9,2852, and the p-value of 0,000. Using the significance value (α) of 5%, we must reject H_0, and conclude that the MGTWR model is significantly different from the MGWR. Therefore, we can further conclude that the MGTWR model is more proper to model the Air Polluter Standard Index (APSI) for SO_2. This means that the time element is influential in the APSI modeling for SO_2 so that not only the location factor is considered, but the observation time is taken as the influential factor to the APSI model for SO_2 as well.

MGTWR statistical summary of the parameters are as follows:

Table-8. Summary statistics local parameter model MGTWR with weighted exponential.

Parameter Statistics Summary MTGWR					
Variable	Min	Max	Mean	Range	StDv
B_1	44.552	44.910	44.804	0.0358	0.0121
B_3	-0.0709	0.2317	0.0534	0.3026	0.1170
B_4	-0.0293	0.2389	0.1386	0.2682	0.137

4.2 Comparison model GWR, GTWR and MGTWR

The next grouping of observation location by using three models, GWR, GTWR and MGTWR

Table-9. GWRmodel for SO_2 at five observation location nodes.

No.	Observation Location	Significant variables
1	SUF 1	X4, X5, X6 and X7
2	SUF 3	X2, X4, X5, X6 and X7
3	SUF 4	X2, X4, X5, X6 and X7
4	SUF 5	X2, X4, X5, X6 and X7
5	SUF 6	X2, X4, X5, X6 and X7

Table-10. GTWRmodel for SO_2 at five observation location nodes.

No.	Observation Location	Observation Time	Significant variables
1	SUF 1	Morning	X3, X5 and X6
		Noon	X4dan X5
		Evening	X2, X4, X5 and X6
2	SUF 3	Morning	X3, X5, and X6
		Noon	X4dan X5
		Evening	X2, X4, X5 and X6
3	SUF 4	Morning	X3, X5 and X6
		Noon	X4dan X5
		Evening	X2, X4, X5 and X6
4	SUF 5	Morning	X3, X5 and X6
		Noon	X4dan X5
		Evening	X2, X4, X5 and X6
5	SUF 6	Morning	X3, X5 and X6
		Noon	X4dan X5
		Evening	X2, X4, X5 and X6

Table-11. MGTWRmodel for SO_2 at five observation location nodes.

No	Observation Location	Observation Time	Local Significant Variables	Global Significant Variables
1	SUF 1	Morning	X_3	X_2, X_5, X_6 dan X_7
		Noon	X_4	
		Evening	X_4	
2	SUF 3	Morning	X_3	
		Noon	X_3 dan X_4	
		Evening	X_4	
3	SUF 4	Morning	X_3	
		Noon	X_3 dan X_4	
		Evening	X_4	
4	SUF 5	Morning	X_3	
		Noon	X_3 dan X_4	
		Evening	X_4	
5	SUF 6	Morning	X_3	
		Noon	X_3 dan X_4	
		Evening	X_4	

To see which models are most appropriate to describe the pollutant, then used R2, AIC and MSE, as shown in Table-12 below:

Table-12. Comparison of estimated GWR, GTWR and MGTWR.

Analysis	GWR	GTWR	MGTWR
Weighted	Exponential	Exponential	Exponential
R^2	0.7157	0.7541	0.8273
MSE	0.1369	0.1181	0.1104
AIC	367.122	324.1381	271.6596

Criteria for the good of the model used are to compare the value of Mean Square Error (MSE), the coefficient of determination (R2), Akaika Information Criterion (AIC) of the MGTWR models. The smallest MSE, R2 largest and smallest AIC. Shown in Table-12 that the most appropriate model to describe the air pollutant is MGTWR model.

5. CONCLUSIONS

Comparison of estimates of air pollutant models most suitable is MGTWR, which is influenced by factors location (geographically) showed a significant effect as well as a global influence. So the model *Mixed* is suitable for modeling of air pollutants in this study. The influence of local models MGTWR elements of SO2 for observation time morning and afternoon shows two predictor variables that influence that the air humidity (X3), the traffic velocity (X4), it indicates that in the morning and during the day the effects of smoke from the burning vehicle motor and the humidity is very dominant influence of air pollution SO2 elements. for night time observation almost all predictor variables affect air pollution. whereas the global influence is the wind velocity (X2), the area size of the urban forest (X5), the population density (X6), the business center aspect (X7).

REFERENCES

[1] Ebtekar 2006. Air Pollution induced asthma and aleterations in cytokine pattern. Allergy Asthma Immunol. 5(2):47-56. ISSN 1735-1502.

[2] BAPEDAL. 1997. Indeks Standar Pencemar Udara. Jakarta: Menteri Lingkungan Hidup.

[3] Fahimi M., Dharma B., Fetararayani D., Baskoro 2012. Asosiasi antara polusi udara dengan IgE total serum dan tes faal paru pada polisi lalul intas. Jurnal Penyakit Dalam. pp. 1-9.

[4] Robinson D. and Lloyd J. M. 2011. Increasing the accuracy of nitrogen dioxide (NO2) pollution mapping using geographically weighted regression and geostatistic . International Journal of Applied Earth Observation and Geoinformation, In Press 1-10. ISSNs: 0303-2434

[5] Gilbert A. and Chakraborty J. 2011. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. Social Science Research, 40: 273-286. DOI:10.1016/j.ssresearch.2010.08.006.

[6] Mennis J. 2006. Mapping the results of geographically weighted. The Cartographic. 43(2): 171-179. 171-179.DOI: 10.1179/000870406X114658.

[7] Fotheringham A. S., Brunsdon C., Chartlon. 2002. Geographically Weighted Regression. The analysis of spatially varying relationship. John Wiley and Sons, LTD. ISBN 0-471-49616-2.

[8] Brunsdon C., Fotheringham A. and Charlton. 1999. Some notes on parametric significance tests for Geographically Weighted Regression. Journal of Regional Science. 39(3): 497-524. DOI: 10.1111/0022-4146.00146.

[9] Purhadi and Hasbi Y. 2012. Mixed geographically weighted regression model (Case study: The percentage of poor households in Mojokerto 2008). European Journal of Scientific Research. 69(2): 188-196. ISSN: 1450202X.

[10] Yu P.H. and LayJ.G. 2011. Exploring Non-Stationarity of Local Mechanism of Crime Events with Spatial-Temporal Weighted Regression. Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on. pp. 7-12.

[11] Leung Y., Mei Chang L., and ZhangWen X. 2000a.Statistical test for spatial nonstationarity based on the geographically weighted regression model. Environment and Planning, 32 (5): 871-890. DOI: 10.1068/a32117.

[12] Winarso. K, NotobrotoH.B and Fatmawati. 2014. Development of Air Pollutant Standard Index Model Based On Mixed Geographically Temporal Weighted Regression Approach. Applied Mathematical Science. 8 (118): 5863-5873.DOI 10.12988/ams.