PM_{10} and NO_2 Concentration Variations in Indonesia using Geospatial Technologies: A Case Study of Surabaya

Liadira Kusuma Widya[1,2], Chih-Da Wu[2,3], Ya-Ping Hsiao[1], Lalu Muhamad Jaelani[2],

[1] Department of Geomatics, National Cheng Kung University, Tainan, 70101, Taiwan
[2] Department of Geomatics Engineering, Sepuluh Nopember Institute of Technology, Surabaya 60111, Indonesia
[3] National Health Research Inst., No. 35, Keyan Rd, Zhunan, Miaoli County 35053, Taiwan

Email: liadira13@mhs.geodesy.its.ac.id; chidawu@mail.ncku.edu.tw; ypiu1111@gmail.com; lmjaelani@gmail.com;

KEYWORDS: Geospatial Technologies, Particulate Matter (PM_{10}), Nitrogen Dioxide (NO_2), GIS, Remote Sensing

ABSTRACT:
Air pollution has emerged as a major health, environmental, economic and social problem all over the world. In this study, geospatial technologies coupled with a Land Use Regression (LUR), a Geographically Weighted Regression (GWR), and a Geographically and Temporal Weighted Regression (GTWR) approach applied to assess the spatial-temporal distribution of two types of air pollutants, particulate matter (PM_{10}) and nitrogen dioxide (NO_2) in Surabaya, Indonesia. In-situ observations of air pollutants from seven monitoring stations during 2010 to 2018 were used as dependent variables, while the land-use/land cover allocations surrounding the monitoring stations from 250 to 5000 m buffer ranges were collected as spatial predictors from GIS and remote sensing databases. A supervised stepwise linear regression approach was employed to develop the LUR models, and a 10-fold cross validation was employed to test the model robustness. According to the obtained model R^2, the developed models from LUR, GWR and GTWR explained 49%, 50%, and 51% of PM_{10} variations and 46%, 47%, and 48% of NO_2 variations, respectively. In the PM10 model, public facility with radius 5000 m, industry and warehousing with radius 500 m, paddy field with radius 2500 m, and NDVI with radius 250 m were selected as the top four important predictors variables for PM_{10}. On the other hand, paddy field with radius 4250 m, a residential area with radius 4000 m, rainfall, and temperature played the most important roles in explaining NO_2 variations while their partial. The result of cross-validated R^2 was 0.6 for PM_{10} and NO_2, confirming the model robustness.

1. Introduction

Air pollution is a major issue for the global that requires serious attention due to effected on human health and environmental. It has been related to increased levels of mortality and morbidity in megacities, and is leading factor to global disease burden (Crouse et al. 2019; Gurjar et al. 2010; Khaniabadi et al. 2017; Lazaridis 2011; West et al. 2016). The highest percentage sources of particulate matter were from residential and industrial area. The sources were identified, such as lead industry mixed road dust, diesel vehicles, oil and coal fired, power plant, road dust, and biomass burning mixed with road dust (Ryou, Heo, and Kim 2018; Santoso et al. 2011). Air pollution in cities were mostly contributed through land use and land cover categories (Weng and

Yang 2006). Meteorological conditions have the most affect to increase dispersion of pollutants. The comparison of seasons between summer and post monsoon season, the higher air pollution has been found in summer season (Verma and Desai 2016). Normalized Difference Vegetation Index (NDVI) is a greenness index that estimate and monitors vegetation density based on satellite (Crouse et al. 2019; Y. L. Guo et al. 2017) .

Land Use Regression (LUR) is an statistical method that developed a multiple linear regression to measure pollutant concentrations as dependent variable with spatial parameters as independent variables to estimate concentrations for non-measurement locations (Novotny *et al.*, 2011; Guo *et al.*, 2017b) Shairsingh *et al.*, 2019). Previous studies that utilized Land Use Regression (LUR) models to estimate air pollution, such as PM_{10} and NO_2 in urban area (Habermann, 2018).

This research developed a Land Use Regression (LUR), a Geographically Weighted Regression (GWR), a Geographically and temporally Weighted Regression (GTWR) model that can calculate for spatial and temporal variability in the correlation between ambient air pollution and several predictors, such as land use variables, meteorological fields, and greenness from Moderate Resolution Imaging Spectroradiometer (MODIS). Developed a GTWR model to determine with the spatial and temporal and simultaneously through integrating temporal effects into GWR model. GTWR models shows much better than GWR (Huang, 2010; Guo *et al.*, 2017a).

The objective of this study was to develop a LUR, a GWR and a GTWR model for PM_{10} and NO_2 in Surabaya, East Java. It was considering multi-temporal meteorology condition and greenness (NDVI) variability from MODIS and the contribution of land use variability.

2. Materials and Methods

Surabaya City is located at 7º21º South Latitudes and 112º54º East Latitudes. The area of Surabaya City is about 326,36 Km and devide into 31 districts and 154 villages. The north and east parts were bounded by Madura Bay, South part is bounded by Sidoarjo County, West part is bounded by Gresik County. The population densities of Surabaya city is 8,463 people/km^2 (Statistics Bureau of Surabaya City 2014). Fig. 1 shows the land use of Surabaya City was dominated by residential area in 2014 (City Development Planning Bureau of Surabaya 2014). Land use development is followed by establishment of road network in the region causes increase people activity mobilization (Rahayu, 2016).

2.1. Air pollution monitoring database

PM_{10} and NO_2 concentrations data in Surabaya was obtained by Environmental Bureau of Surabaya City. Ground monitoring measurements of PM_{10} and NO_2 mass concentration were obtained from 7 automatic monitoring stations. The stations distributed within the study area. Daily concentration observations from 2010 to 2018 were aggregated into annual averages for the model analysis.

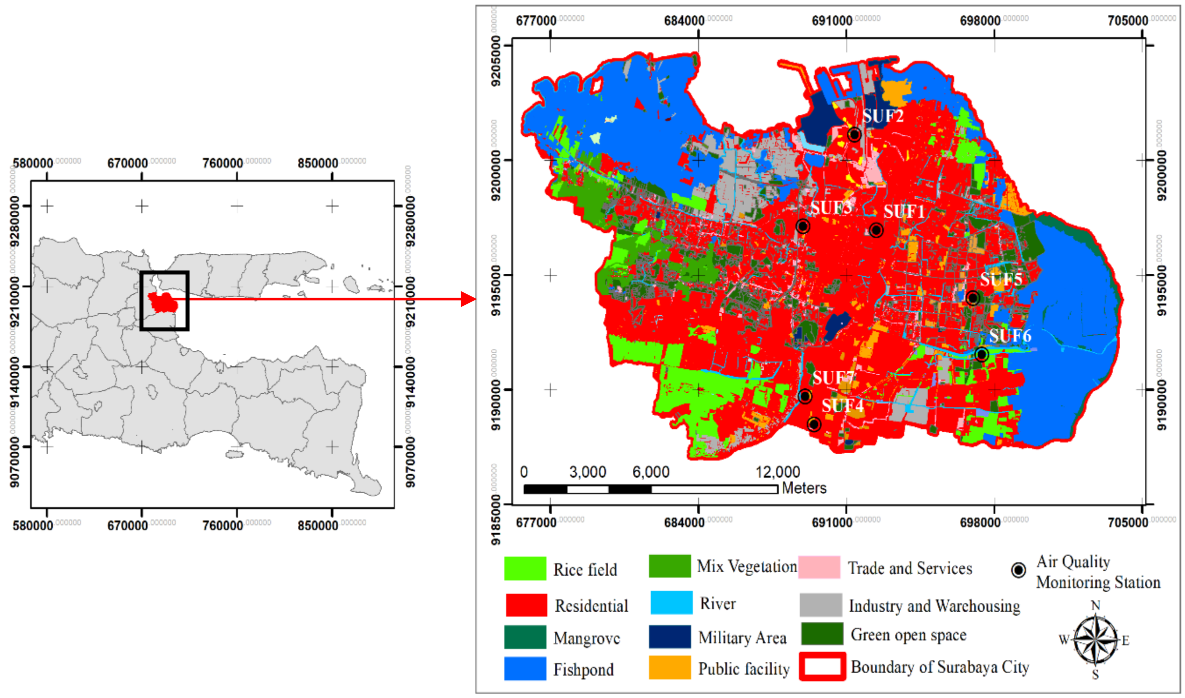

Figure 1. A summary map of the study area

2.2. Spatial databases

Land Use data is formed by raster with calculations of focal statistics, 250 - 5000 meters radius using ArcGIS and Python, for each type of land use. Daily meteorological data is collected by Meteorological, Climatological, and Gephysical Bureau (BMKG) as Government Bureau. The monitoring stations in Java Island in Indonesia, including temperature, wind direction, wind speed, relative humidity, solar radiation, and rainfall, were obtained from BMKG database center (Data Online-BMKG Database Center : http://dataonline.bmkg.go.id/home) from Januari 1st, 2010 to December 31st, 2018. Vegetation Indices (MOD13Q1) version 6 data were produced each 16 days at 250 meter spatial resolution as a level 3 product by Terra Moderate Resolution Imaging Spectroradiometer (MODIS). The algorithm chosen is based on the best pixel value from all acquisition each 16-day period, which have the criteria, such as low clouds, low view angle, and the highest NDVI/EVI value (Didan 2015). Meteorological data from each station point is obtained using the Inverse Distance Weighting method. The next step after the data becomes raster, data collection for each variable is carried out at each point of the PM10 and NO2 monitoring station.

Raster maps (50 x 50 m) to develop PM_{10}, and NO_2 models were create for each element based on the predictors, such as land use types, meteorology conditions, and NDVI. The focal statistics function in ArcMap was used to describe the predictor data for each types of land use and NDVI. Circle neighborhoods were used to generate the map from radius 250 to 5000 meters, respectively. ArcGIS 10.5 was used for the spatial analysis.

2.3. Correlation analysis and LUR modelling

Association between spatial predictors and air pollutions were assessed using the Spearman correlation coefficient. Variables with intuitive direction of correlations were remained for the model analysis. A multivariate linear regression with stepwise variable selection procedures were applied to identify the important predictions for the LUR model development. The statistical criterion used for variable selections was $\rho < 0.1$ and VIF <3 (Wang et al., 2018). The equation of the developed LUR model is defined as follows (1) :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad \ldots 1$$

Where Y is PM_{10} or NO_2 concentration; β_0 is constant intercept; β_1 to β_n are regression coefficients; $X_1 \ldots X_n$ are potential predictors.

In the second step, all of the selected variables were further entered into the GWR and GTWR process for developing the prediction models. GWR is an applicability approach to solve the model based on the spatial. The equation of the GWR model is defined as follows (2) :

$$Y_i = \beta_{(U_i, v_i)} + \sum_k \beta_{k(U_i, v_i)} X_{ik} \qquad \ldots 2$$

Where (u_i, v_i) denotes the coordinates of the point in location; Y_i is PM10 or NO2 concentartion; $\beta_{(U_i, v_i)}$ represents the intercept value; $\beta_{k(U_i, v_i)}$ is a set of values of parameters at point i; and X_{ik} are potential predictors.

A GTWR model is an applicability approach to solve the spatial and temporal nonstationarity simultaneously. The equation of the GTWR model is defined as follows (3) (Huang, 2010).

$$Y_i = \beta_{(U_i, v_i, t_i)} + \sum_k \beta_{k(U_i, v_i, t_i)} X_{ik} \qquad \ldots 3$$

Where (u_i, v_i, t_i) denotes the coordinates of the point in space-time; Y_i is PM10 or NO2 concentartion; $\beta_{(U_i, v_i, t_i)}$ represents the intercept value; $\beta_{k(U_i, v_i, t_i)}$ is a set of values of parameters at point i; and X_{ik} are potential predictors. The bandwith value was selected by utilizing corrected Akaike Information Criterion (AICc). The choice attempts to analyze an ideal fixed distance (Hu 2009).

Model comparisons were based on the R^2, adjusted R^2, and AICc. All of the statistical analyses were conducted using SPSS version 20 packages and R statistical packages 3.333 .

3. Results

3.1. PM_{10} and NO_2 concentrations of Surabaya

Figure 2 (a) shows demonstrates the time series trend of PM_{10} and NO_2 concentrations of 7 stations from 2010 to 2018. The annual average of PM_{10} and NO_2. Based on the WHO guidelines indicate that the PM_{10} in Surabaya already on the edge because the standard of annual average PM10 from WHO was 20 µg/m^3.

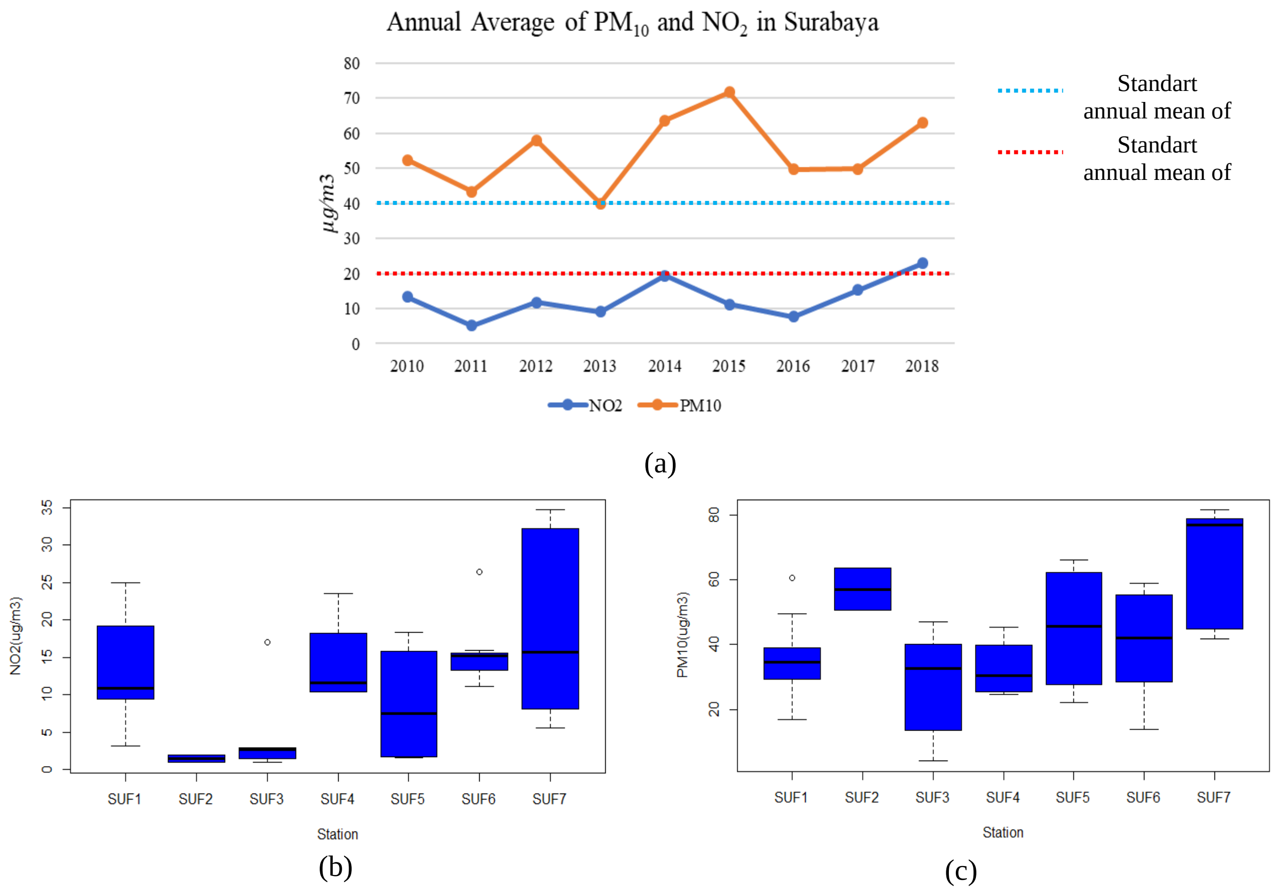

Figure 2. (a) Time series trend of PM_{10} and NO_2 concentrations of the 7 monitoring stations from 2010 to 2018 (b) box plots of NO_2 (c) box plots of PM_{10}

3.2. Correlation analysis

The result indicates that PM_{10} has positive correlation with public facility, industrial and warehousing, paddy field, and negative correlation with NDVI. NDVI construct to be correlated with lower pollution and reduce air pollution around it (Macnaughton et al. 2017). Paddy field, residential area, temperature have positive correlation with NO_2, and negative correlation to rainfall. However, rainfall has effect of reducing air pollution as well (Kwak et al. 2017). The correlation between industry and warehousing, paddy field to PM_{10}, NO_2 achived significant level at 0.05, respectively.

3.3. LUR model development

The final LUR model for PM_{10} and NO_2 is shown in Table 1. The distribution of VIF values was less than 3 which means the predictors indicate no multicollinearity. The R^2 of each model for PM_{10} and NO_2 were 0.49 and 0.48.

The predictor, such as public facility with radius 5000 meters, industry and warehousing with radius 500 meters, paddy field with radius 2500 meters, and NDVI with radius 250 meters were selected in LUR model for PM_{10}, with partial R^2 0.095, 0.109, 0.116, and 0.169, respectively. The variables were selected in NO_2 model, such as paddy field with radius 4250 meters, residential area with radius 4000 meters, rainfall, and average temperature, with partial R^2 0.164, 0.153, 0.083, 0.056, respectively.

Table 1. Coefficient estimates of the developed LUR model

PM10					NO2				
Parameter	B	p-value	VIF	Partial R2	Parameter	B	p-value	VIF	Partial R2
Intercept	36.278	0.01826			Intercept	-374.13	0.0778		
Public Facility$_{5000m}$	.562	0.00017	1.387	.095	Paddy Field$_{4250m}$	.146	3.4E-05	2.695	0.164
Industry and Warehousing$_{500m}$	.027	0.17349	1.328	.109	Residential Area$_{4000m}$	.013	0.00769	1.896	0.153
Paddy Field$_{2500m}$	.185	0.00024	2.220	.116	Rainfall	-3.028	0.00833	1.706	0.083
NDVI$_{250m}$	-191.001	0.00308	2.888	.169	Temperature	13.212	0.08346	2.747	0.056
Diagnostic information									
R2		0.49					0.457		
Adj R2		0.424					0.387		
AICc		310					252		

3.4. Model comparisons

GWR and GTWR models show can improvements LUR model in terms R^2 and AICc measures. However, It was still necessary to investigate whether the GTWR models have better performance than the GWR models from a statistical description in Table 2 and 3. From the statistics coefficient of the model, public facility, industry and warehousing, paddy field has a positive correlation and a negative correlation for NDVI (greenness) with dependent variables PM_{10}, constantly. Paddy field, residential area, rainfall, and temperature influenced the development of the NO_2 model. The coefficients were stable with a positive correlation to paddy field, residential area, and temperature. Inversely, negative correlation between rainfall. The map prediction shows in Figure 3 and 4. It is figure shows spatial temporal from GTWR models from 2010 to 2018. From the statistical analysis, the PM_{10} increases according to industry and warehousing, public facilities, and paddy fields. For NO_2, increases according to a residential area and paddy field.

Table 2. Coefficient statistics of the developed PM_{10} model using GWR and GTWR

Parameter	GWR (bandwith = 1.989)					GTWR (bandwith = 1.414)				
	Min	LQ	Med	UQ	Max	Min	LQ	Med	UQ	Max
Intercept	36.2	36.2	36.3	36.4	36.5	35.9	35.9	36.1	36.4	37.5
Public Facility$_{5000m}$	0.559	0.559	0.56	0.56	0.56	0.544	0.554	0.558	0.562	0.563
Industry and Warehousing$_{500m}$	0.027	0.027	0.0271	0.0271	0.0272	0.0251	0.0264	0.0275	0.028	0.0288
Paddy Field$_{2500m}$	0.184	0.184	0.184	0.185	0.185	0.18	0.182	0.184	0.185	0.186
NDVI$_{250m}$	-191	-191	-190	-190	-190	-193	-191	-189	-188	-186
Diagnostic information										
R^2			0.504					0.511		
Adj R^2			0.441					0.448		
AICc			305.142					305.027		

Table 3. Coefficient statistics of the developed NO_2 model using GWR and GTWR

Parameter	GWR (bandwith = 1.987)					GTWR (bandwith = 1.985)				
	Min	LQ	Med	UQ	Max	Min	LQ	Med	UQ	Max
Intercept	-377	-377	-374	-372	-367	-377	-377	-376	-374	-366
Paddy $Field_{4250m}$	0.145	0.146	0.146	0.146	0.146	0.144	0.146	0.146	0.146	0.146
Residential $Area_{4000m}$	0.0129	0.013	0.013	0.0131	0.0131	0.0127	0.0129	0.013	0.0131	0.0131
Rainfall	-3.07	-3.06	-3.03	-2.990	-2.99	-3.06	-3.06	-3.03	-3.000	-2.95
Temperature	12.9	13.2	13.2	13.3	13.3	12.9	13.2	13.3	13.3	13.3
Diagnostic information										
R2			0.473					0.480		
Adj R^2			0.405					0.410		
AICc			252.079					251.81		

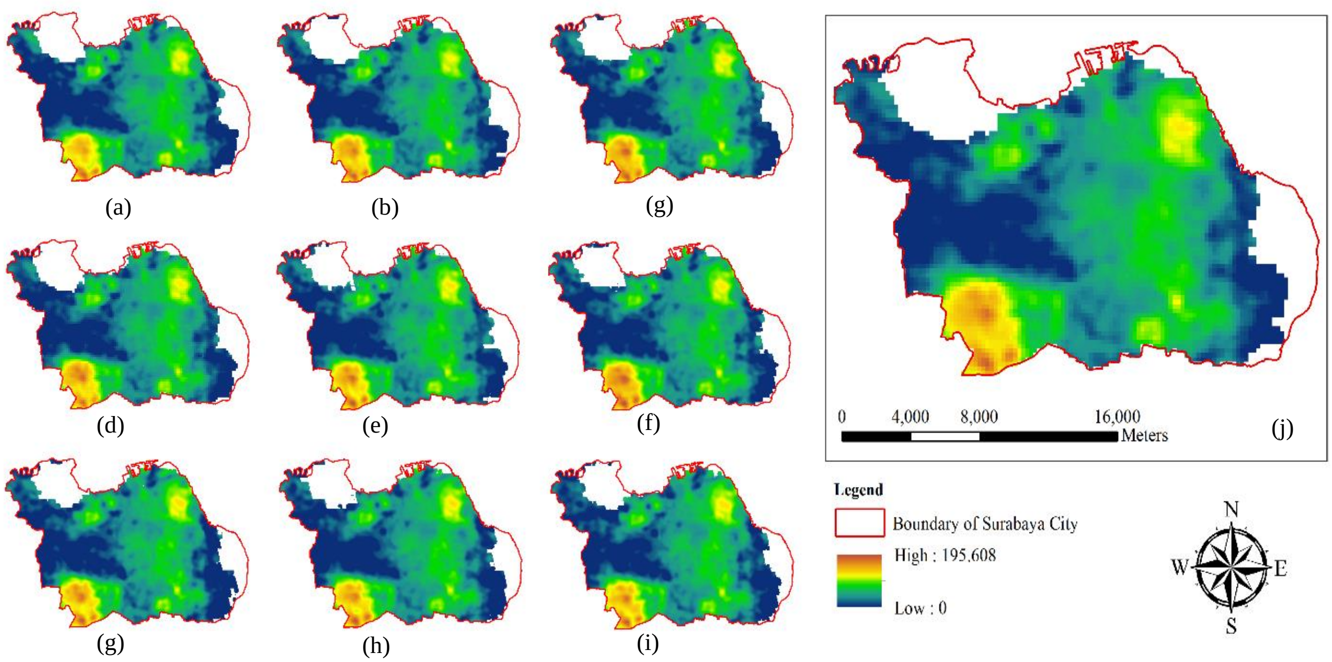

Figure 3. Prediction maps of spatial-temporal of PM10 concentration using GTWR model (a) 2010 (b) 2011 (c) 2012 (d) 2013 (e) 2014 (f) 2015 (g) 2016 (h) 2017 (i) 2018 (j) 2010 to 2018 average

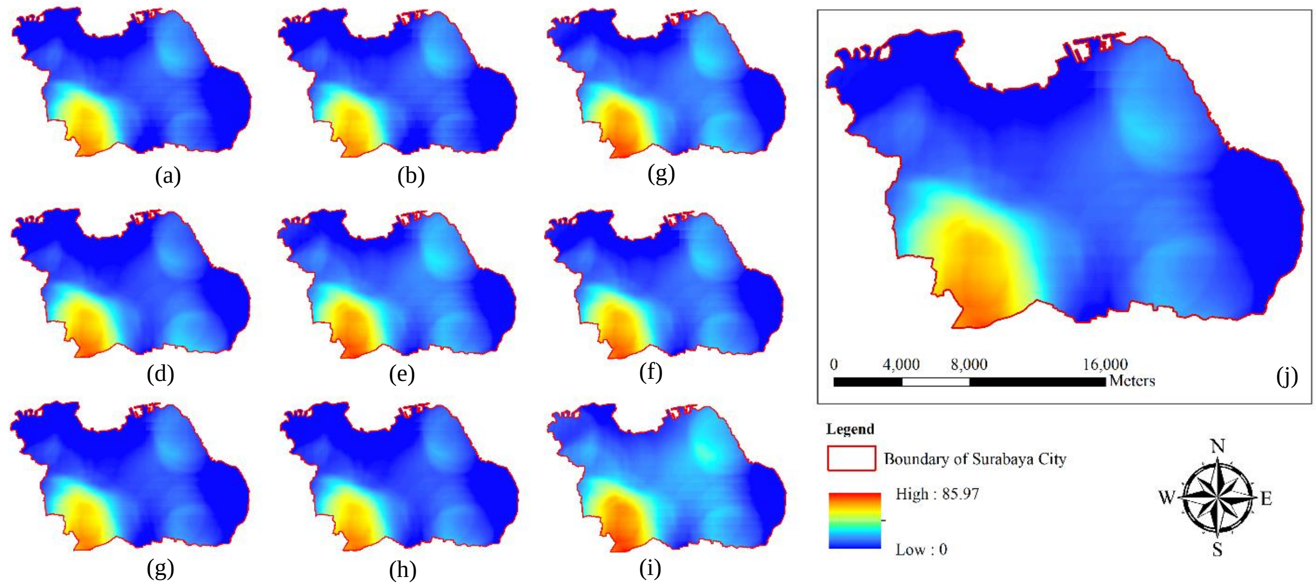

Figure 4. Prediction maps of spatial-temporal of NO_2 concentration using GTWR model (a) 2010 (b) 2011 (c) 2012 (d) 2013 (e) 2014 (f) 2015 (g) 2016 (h) 2017 (i) 2018 (j) 2010 to 2018 average

3.5.Estimation of spatial temporal variability of PM_{10} and NO_2

Map estimation was applied by each model. Figure 5 shows the different of map estimation from LUR, GWR, and GTWR. The map from developing a model that can identify the area which have the high air pollution concentrations.

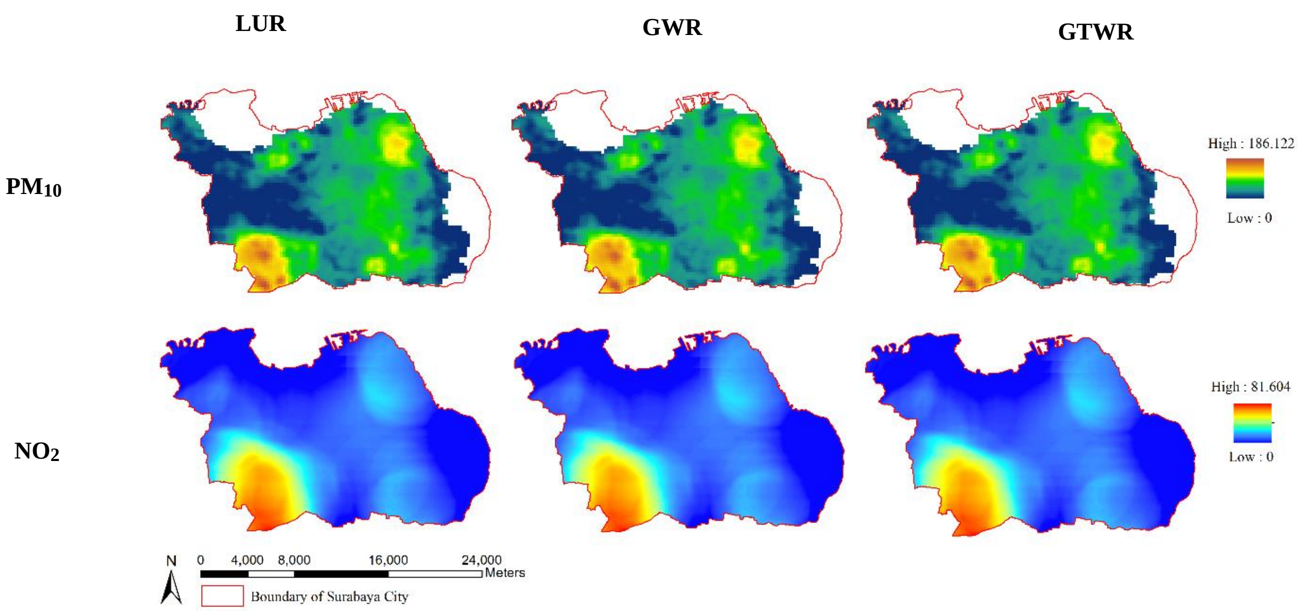

Figure 5. Comparison of map prediction from LUR, GWR, and GTWR models

4. Conclusion

GTWR was an effective method to interpret the spatial non-stationer in the association between land use types, meteorological conditions, and NDVI between PM_{10} and NO_2. It improved the prediction variability of the PM_{10} and NO_2 LUR models. The parameters have different effects on PM_{10} and NO_2 in the case study. Compared with the LUR, GWR and GTWR models, the R^2 were 0.46, 0.47, and 0.48 for the NO_2 model, respectively. For the PM_{10} model using LUR, GWR, and GTWR, the R^2 were 0.49, 0.50, and 0.51. The limitation from this model was NDVI for PM_{10} model uncovered all the area of Surabaya because of not covered map estimation for PM_{10}. And the total of station from 2014 has a different, because of in 2014 there was a new station from SUF 7 in Surabaya. The prediction of PM_{10} and NO_2 can be used to estimate the health exposure risk and to hold urban air quality management. And recommendation for government of Environmental Bureau in Surabaya should add a new station in West Surabaya to monitor area around an industrial area and inter-city.

References :

City Development Planning Bureau of Surabaya. 2014. “Land Use Map of Suarabaya in 2014.”

Crouse, Dan L. et al. 2019. “Complex Relationships between Greenness, Air Pollution, and Mortality in a Population-Based Canadian Cohort.” *Environment International* 128(April): 292–300. https://doi.org/10.1016/j.envint.2019.04.047.

Didan, Kamel. 2015. “MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006.” *NASA EOSDIS Land Processes DAAC*. https://lpdaac.usgs.gov/products/mod13q1v006/.

Guo, Yuanxi, Qiuhong Tang, Dao Yi Gong, and Ziyin Zhang. 2017a. “Estimating Ground-Level PM2.5 Concentrations in Beijing Using a Satellite-Based Geographically and Temporally Weighted Regression Model.” *Remote Sensing of Environment* 198: 140–49. http://dx.doi.org/10.1016/j.rse.2017.06.001.

Guo, Yue Leon et al. 2017b. “Land-Use Regression with Long-Term Satellite-Based Greenness Index and Culture-Specific Sources to Model PM2.5 Spatial-Temporal Variability.” *Environmental Pollution* 224: 148–57. http://dx.doi.org/10.1016/j.envpol.2017.01.074.

Gurjar, B. R. et al. 2010. “Human Health Risks in Megacities Due to Air Pollution.” *Atmospheric Environment* 44(36): 4606–13. http://dx.doi.org/10.1016/j.atmosenv.2010.08.011.

Habermann, Mateus, Monica Billger, and Marie Haeger Eugensson. 2018. “Land Use Regression Approach to Model NO2–Concentrations in a Swedish Medium-City.” *Environmental Pollution and Protection* 3(3): 71–89.

Hu, Zhiyong. 2009. “Spatial Analysis of MODIS Aerosol Optical Depth, PM2.5, and Chronic Coronary Heart Disease.” *International Journal of Health Geographics* 8(1): 1–10.

Huang, Bo, Bo Wu, and Michael Barry. 2010. “Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices.” *International Journal of Geographical Information Science* 24(3): 383–401.

Khaniabadi, Yusef Omidi et al. 2017. “Exposure to PM10, NO2, and O3and Impacts on Human Health.” *Environmental Science and Pollution Research* 24(3): 2781–89. http://dx.doi.org/10.1007/s11356-016-8038-6.

Kwak, Ha Young, Joonho Ko, Seungho Lee, and Chang Hyeon Joh. 2017. “Identifying the Correlation between Rainfall, Traffic Flow Performance and Air Pollution Concentration in Seoul Using a Path Analysis.” *Transportation Research Procedia* 25: 3552–63. http://dx.doi.org/10.1016/j.trpro.2017.05.288.

Lazaridis, Mihalis. 2011. *First Principles of Meteorology and Air Pollution*. Environmen. eds. Brain

J. Alloway and Jack T. Trevors. New York: Springer.
Macnaughton, Piers et al. 2017. “Impact of Particular Matter Exposure and Surrounding ‘Greenness’ on Chronic Absenteeism in Massachusetts Public Schools.” *International Journal of Environmental Research and Public Health* 14(2).
Novotny, Eric V., Matthew J. Bechle, Dylan B. Millet, and Julian D. Marshall. 2011. “National Satellite-Based Land-Use Regression: NO2 in the United States.” *Environmental Science and Technology* 45(10): 4407–14.
Rahayu, Yoanita Eka, Ervina Ahyudanari, and Nursakti Adhi Pratomoadmojo. 2016. “Land Use Development and Its Impact on Airport Access Road.” *Procedia - Social and Behavioral Sciences* 227(November 2015): 31–37.
Ryou, Hyoung gon, Jongbae Heo, and Sun Young Kim. 2018. “Source Apportionment of PM10 and PM2.5 Air Pollution, and Possible Impacts of Study Characteristics in South Korea.” *Environmental Pollution* 240: 963–72. https://doi.org/10.1016/j.envpol.2018.03.066.
Santoso, Muhayatun et al. 2011. “Preliminary Study of the Sources of Ambient Air Pollution in Serpong, Indonesia.” *Atmospheric Pollution Research* 2(2): 190–96. http://dx.doi.org/10.5094/APR.2011.024.
Shairsingh, Kerolyn K. et al. 2019. “Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas?” *Atmospheric Environment* 196(May 2018): 143–51. https://doi.org/10.1016/j.atmosenv.2018.10.013.
Statistics Bureau of Surabaya City. 2014. *Surabaya in Figures*. ed. Statistics Bureau of Surabaya. Surabaya: Statistics Bureau of Surabaya.
Verma, Sonal S, and Birva Desai. 2016. “Effect of Meteorological Conditions on Air Pollution of Surat City.” *Journal of International Environmental Application and Science* 8(5): 358–67.
Wang, Jian-bing et al. 2018. “National Scale Spatiotemporal Land-Use Regression Model for PM2 . 5 , PM10 and NO2 Concentration in China National Scale Spatiotemporal Land-Use Regression Model for PM 2 . 5 , PM 10 and NO 2 Concentration in China.” *Atmospheric Environment* 192(August): 48–54. https://doi.org/10.1016/j.atmosenv.2018.08.046.
Weng, Qihao, and Shihong Yang. 2006. “Urban Air Pollution Patterns, Land Use, and Thermal Landscape: An Examination of the Linkage Using GIS.” *Environmental Monitoring and Assessment* 117(1–3): 463–89.
West, J Jason et al. 2016. “‘ What We Breathe Impacts Our Health: Improving Understanding of the Link between Air Pollution and Health .’”