Article

Comparison of Spatial Modelling Approaches on PM$_{10}$ and NO$_2$ Concentration Variations: A Case Study in Surabaya City, Indonesia

Liadira Kusuma Widya [1,2]**, Chin-Yu Hsu** [3]**, Hsiao-Yun Lee** [4]**, Lalu Muhamad Jaelani** [2]**, Shih-Chun Candice Lung** [5,6,7]**, Huey-Jen Su** [8] **and Chih-Da Wu** [1,9,*]

[1] Department of Geomatics, National Cheng Kung University, Tainan City 70101, Taiwan; liadira13@mhs.geodesy.its.ac.id

[2] Department of Geomatics Engineering, Institut Teknologi Sepuluh Nopember, Surabaya City 60111, Indonesia; lmjaelani@geodesy.its.ac.id

[3] Department of Safety, Health, and Environmental Engineering, Ming Chih University of Technology, New Taipei City 24301, Taiwan; gracecyhsu@mail.mcut.edu.tw

[4] Department of Leisure Industry and Health Promotion, National Taipei University of Nursing and Health Sciences, Taipei City 112303, Taiwan; hsiaoyun07@ntunhs.edu.tw

[5] Research Center for Environmental Changes, Academia Sinica, Taipei City 11529, Taiwan; sclung@rcec.sinica.edu.tw

[6] Department of Atmospheric Sciences, National Taiwan University, Taipei City 10617, Taiwan

[7] Institute of Environmental Health, National Taiwan University, Taipei City 100025, Taiwan

[8] Department of Environmental and Occupational Health, National Cheng Kung University, Tainan City 70101, Taiwan; hjsu@mail.ncku.edu.tw

[9] National Institute of Environmental Health Sciences, National Health Research Institutes, Miaoli County 35053, Taiwan

[*] Correspondence: chidawu@mail.ncku.edu.tw

check for updates

Abstract: Because of fast-paced industrialization, urbanization, and population growth in Indonesia, there are serious health issues in the country resulting from air pollution. This study uses geospatial modelling technologies, namely land-use regression (LUR), geographically weighted regression (GWR), and geographic and temporal weighted regression (GTWR) models, to assess variations in particulate matter (PM$_{10}$) and nitrogen dioxide (NO$_2$) concentrations in Surabaya City, Indonesia. This is the first study to implement spatiotemporal variability of air pollution concentrations in Surabaya City, Indonesia. To develop the prediction models, air pollution data collected from seven monitoring stations from 2010 to 2018 were used as dependent variables, while land-use/land cover allocations within a 250 m to 5000 m circular buffer range surrounding the monitoring stations were collected as independent variables. A supervised stepwise variable selection procedure was applied to identify the important predictor variables for developing the LUR, GWR, and GTWR models. The developed models of LUR, GWR, and GTWR accounted for 49%, 50%, and 51% of PM$_{10}$ variations and 46%, 47%, and 48% of NO$_2$ variations, respectively. The GTWR model performed better ($R^2 = 0.51$ for PM$_{10}$ and 0.48 for NO$_2$) than the other two models ($R^2 = 0.49$–0.50 for PM$_{10}$ and 0.46–0.47 for NO$_2$), LUR and GWR. In the PM$_{10}$ model four predictor variables, public facility, industry and warehousing, paddy field, and normalized difference vegetation index (NDVI), were selected during the variable selection procedure. Meanwhile, paddy field, residential area, rainfall, and temperature played important roles in explaining NO$_2$ variations. Because of biomass burning issues in South Asia, the paddy field, which has a positive correlation with PM$_{10}$ and NO$_2$, was selected as a predictor. By using long-term monitoring data to establish prediction models, this model may better depict PM$_{10}$ and NO$_2$ concentration variations within areas across Asia.

1. Introduction

It is well documented that exposure to ambient air pollution can lead to increased mortality and morbidity and a shortened life expectancy [1,2]. Among a variety of particles in the atmosphere, pollutants such as PM_{10} (coarse particulate matter with an aerodynamic diameter smaller than 10 μm) and nitrogen dioxide (NO_2) are commonly used as indicators of ambient air pollution [3–5]. This is because these pollutants are believed to be associated with several acute and chronic health conditions [6]. Moreover, they have demonstrated associations with increased levels of mortality and morbidity in megacities [7–10]. As a developing country with several densely-populated industrial cities, Indonesia faces a significant challenge relating to worsening environmental quality resulting from increased air pollution [11]. A study conducted by Greenstone and Fan demonstrated that, in the last two decades, Indonesia has experienced dramatic changes in air quality [12]. The study confirmed that high levels of air pollution are now undermining Indonesians' health and may continually reduce the average life expectancy of citizens. Furthermore, previous studies have confirmed that in Indonesia exposure to air pollution can worsen health conditions, including respiratory diseases and lung cancer, and even general health burdens. [13–15].

Exposure to air pollution is a major issue for residents of developing countries, and it arises due to rapid industrialization and urbanization [16,17]. Increased levels of air pollution can be attributed to population growth, which brings with it increased human activities and removal of green space [18,19]. According to the literature, fast-growing cities in Southeast Asia, the Middle East, and the Western Pacific Region are suffering from severe air pollution levels that are five to 10 times greater than the levels recommended by the World Health Organization (WHO), implying potential negative health impacts on local residents [2]. Take Indonesia as an example: Indonesia is undergoing a rapid transformation with a great percentage of the population moving from rural towns to urban cities in search of more lucrative jobs. Nonetheless, residents may lose several years of life if high levels of air pollution are present throughout their lifetime [20]. Public awareness regarding the negative impacts of air pollution in Indonesia remains low due to a lack of evidence by in-situ observation. Therefore, accurate methods to assess within-city variability of air pollution are necessary [21].

In connection with the air pollution estimation methods, several spatio-temporal modeling approaches have been developed in previous studies, including land-use regression (LUR), geographically weighted regression (GWR), and geographically and temporally weighted regression (GTWR) [22–25]. In LUR, a multiple linear regression model is developed, which then links the air pollution concentrations observed in the network to the most predictive environmental characteristics (e.g., traffic, industry) [24]. Previous studies have applied this method to estimate the spatial variations in exposure to pollution [26,27]. In recent years, researchers have developed various statistical approaches in order to deal with spatial issues in modeling air pollution. GWR was introduced as an extended traditional model fitted by least squares regression and can effectively deal with spatial heterogeneity and autocorrelation problems [28,29]. This method refers to local modeling techniques and fits a regression model at each geographic area based on neighbors within a specific bandwidth [30]. Furthermore, researchers have extended GWR to a temporal dimension for spatio-temporal modeling, which is named GTWR. GTWR can address spatial and temporal non-stationarities simultaneously by developing a weight matrix based on spatiotemporal distance [30]. This method expands the boundary of local modeling techniques and has been applied in various disciplines [31,32].

Given the paucity of air pollution information and the low levels of awareness in local residents in Indonesia, this study has two objectives: (1) to estimate spatio-temporal concentration of PM_{10} and NO_2 by in situ observation and application of advanced methods, (2) to explore the determinants

influencing the concentrations of air pollution. It is noted this study further applied GWR and GTWR to examine and verify the estimation in order to replicate accurate results in addition to LUR. To our knowledge, this is the first study estimating air pollution concentration in Indonesia by comparing these three advanced methods. The study results can be used as references for establishing policies or regulations focusing on air quality control.

2. Materials and Methods

2.1. Study Area

Surabaya City (7°21′ S, 112°54′ E), the second largest city in Indonesia, was chosen as the study area. The size of Surabaya City is about 326.36 km^2, which is divided into 31 districts and 154 villages. The northern and eastern portions of the city are surrounded by Madura Bay, while neighboring Sidoarjo County is in the southern portion and Gresik County is in the western portion. The population density is 8463 people/km^2 [33]. Figure 1 illustrates most of Surabaya City was residential areas in 2014 [34]. Surabaya fixed station (SUF) 1, 4, and 7 are in rural areas, SUF 2 and 3 are in proximity to industrial areas, and SUF 5 and 6 are in urban areas. The development and establishment of road networks has helped residents in the region to more easily travel from one portion to another [35]. The population in Surabaya has increased each year, and air pollution due to transportation and industrial activity, too, have increased. The transportation sector contributes nearly 60% of pollutants, and, specifically, are responsible for 25% of the carbon dioxide (CO_2), 90% of the carbon monoxide (CO), and 50% of the nitrogen oxide (NOx) present. In 2016 there were 2,244,317 units of gasoline-fueled vehicles and 150,413 units of diesel-fueled vehicles [36]. This number of vehicles will act to increase NO_2 and PM_{10} levels due to motor vehicle fumes and dust agitated from the road [37].
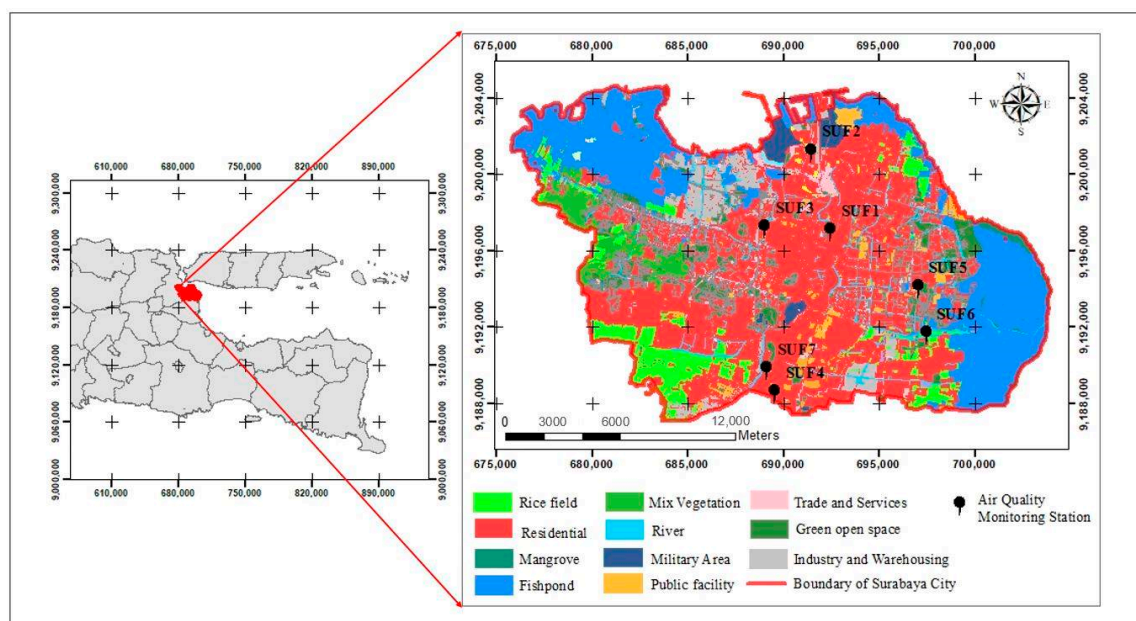

Figure 1. Location of Surabaya City and its land-use allocations.

2.2. Air-Pollution Monitoring Database

The Environmental Bureau of Surabaya City monitors PM_{10} and NO_2 concentrations in Surabaya City via seven automatic monitoring stations distributed within the study area, as shown in Figure 1. For the model analysis, we aggregated hourly concentration observations from 2010 to 2018 into annual averages for model analysis.

2.3. Spatial Databases

Daily climate data were collected by Meteorological, Climatological, and Geophysical Bureau (BMKG) in the East Java Province (http://dataonline.bmkg.go.id/home). The monitoring stations collect information, including temperature, wind direction, wind speed, relative humidity, solar radiation, and rainfall. It is noted that the meteorological monitoring stations and air quality monitoring stations are in different locations. Therefore, the inverse distance weighting (IDW) interpolation method was employed to estimate meteorological conditions at air quality monitoring stations.

Land-use data for Surabaya City were provided by City Development Planning Bureau of Surabaya City (BAPPEKO) with its most recent update in 2014. As shown in Figure 1, the land use in Surabaya City was dominated by 47.85% residential area (red color), followed by 20.4% fishpond, 7.95% green open space, 6.04% paddy field, 5.35% industry and warehousing, and 2.79% public facility [34]. The land-use map used WGS 84 UTM Zone 49S as its coordinate reference system.

A greenness inventory can be obtained by satellite imagery, allowing easy access to multi-temporal greenness [21]. The normalized difference vegetation index (NDVI) was incorporated to represent the surrounding greenness during the study period. The Terra Moderate Resolution Imaging Spectroradiometer (MODIS) (MOD13Q1) version 6 was applied to adopt NDVI for a spatial resolution of 250 m × 250 m (https://ladsweb.modaps.eosdis.nasa.gov/search/order/1). Moreover, the best pixel value from all satellite images was chosen in the 16-day period based on the criteria of low clouds, low view angle, and the highest NDVI/enhanced vegetation index (EVI) value, as set by Didan [38]. There were two greenness values of NDVI for each month. We picked the one which was closest to the middle of the month (date, 15th) and then aggregated those for the annual average of each year. The greenness and all land use variables were abstracted from 250 m to 5000 m, with 250 m interval circular buffer ranges surrounding each PM_{10} and NO_2 measurement site, representing the land-use/land cover allocations in the neighborhoods. Table S1 lists the potential predictor variables used in this study.

2.4. Model Developments Using Three Approaches and Validation

Numerous studies have presented LUR, GWR, and GTWR as methods suitable for investigating air pollution exposure, comparing the performance of the three methods in their analyses. For example, by using data from 56 monitoring stations, a study in Heilongjiang, China, applied several global and geographically-temporally weighted regression models to investigate $PM_{2.5}$ related to O_2, NO_2, PM_{10}, CO, and O_3 [30]. The results demonstrated that temporally weighted regression (TWR) and GTWR yielded slightly better model performance than least square regression and GWR, indicating there are impacts on spatio-temporal variation in air quality. A study conducted in Eastern China also confirmed GTWR obtained the highest model performance in estimating ground-level $PM_{2.5}$ concentrations compared to least square regression and GWR [23]. Furthermore, a recent study by Zeng and colleagues performed GTWR coupled with kriging-based hybrid models by applying three geospatio-temporal modeling approaches to analyze air quality data from specific industrial monitoring stations in Taiwan. It concluded GWTR had the best performance for predicting PM_{10} and O_3 concentrations compared to LUR and GWR methods [39].

In this study, the land-use regression (LUR) models were built based on methodologies developed and verified in previous studies [16,25,26]. The first steps to develop the LUR model determined all parameters with the calculated pollutant as a dependent variable according to the strength of each association. Then, statistical analyses using Spearman correlation were used to look at the bivariate correlation between ambient air pollution and the three types of variable, such as land use, meteorological, and greenness variables. After that, variables that have absolute correlation to a high factor are kept in each sub-category (Spearman's r ≥ 0.4). Then, all selected variables are entered into a stepwise linear regression. A supervised stepwise procedure was applied to increase the percentage of justified variability. For all potential predictor variables, we chose an a priori direction of effect to PM_{10} and NO_2 concentrations (e.g., positive for road length and residential area, negative for NDVI and

green spaces) [21,22,25]. The initial model included variables with the highest explained variance in a univariate analysis and the regression slope with the expected direction. After that, all other variables were added to the model separately, so as to assess whether or not the p-value was less than 0.1 and whether or not the variance inflation factor (VIF) was less than three. This procedure was repeated until none of variables met the criteria listed. Finally, R^2, adjusted R^2, and root mean square error (RMSE) were applied to evaluate the model performance.

The equation of the developed LUR model is defined as follows (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \ldots \tag{1}$$

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

In the second step, each of the selected variables was further examined by GWR and GTWR for developing the prediction models. GWR is an approach applicable for solving a model based on the spatial variation of parameters by region [40]. The GWR was created to extend the traditional global model fitted by ordinary least squares and can efficiently address spatial heterogeneity and autocorrelation issues [28,41]. The equation of the GWR model is defined as follows (2): an extension of the linear regression model, which allows researchers to bring data from surrounding samples to each region and establish individual regression models for which parameters vary by region,

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

where $(U_i, \, v_i)$ denotes the coordinates of the point in location; Y_i is concentration of PM_{10} or NO_2; $\beta_{(U_i, v_i)}$ represents the intercept; $\beta_{k(U_i, v_i)}$ is a set of values of parameters at point i; and X_{ik} are potential predictors.

The GTWR model is an approach suitable for simultaneously solving non-stationarity spatial and temporal data. The equation of the GTWR model is defined as follows [42] (3):

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

where $(U_i, \, v_i, \, t_i)$ denotes the coordinates of the points in space-time; Y_i is concentration of PM_{10} or NO_2; $\beta_{(U_i, \, v_i, \, t_i)}$ represents the intercept; $\beta_{k(u_i, \, v_i, \, t_i)}$ is a set of values of parameters at point i; and X_{ik} are potential predictors. The bandwidth value was selected by utilizing a corrected Akaike information criterion (AIC) which used to analyze the model performance and accuracy [43]. The corrected AIC was used since this information criterion method is one of the most commonly applied goodness-of-fit criteria for model comparisons [30,43,44]. Moreover, as noted in Fotheringham's study, AIC offers the advantage of being more general in application and, thus, can be used to assess, by considering the degree of freedom, whether a regression model can provide a better fit than a global model [28].

This study used LUR to identify important prediction variables. The variables selected by LUR were then used for the GWR and GTWR. The developed models from the three approaches were validated for verification accuracy. For purposes of cross-validation, 90% of the air pollution measurements were randomly selected for training the model and the other 10% of data were used to validate the model predictions. This procedure was repeated 10 times to ensure each measurement had been used as out-of-sample data for the model evaluation. All of the statistical analyses were conducted using SPSS version 20 packages and R statistical packages x64 3.5.2 software. The spatial analyses were performed using ArcGIS 10.5.

3. Results

3.1. Particulate Matter (PM$_{10}$) and Nitrogen Dioxide (NO$_2$) Concentrations of Surabaya

Figure 2 shows the annual trend of PM$_{10}$ and NO$_2$ concentrations calculated from information collected in all stations from 2010 to 2018 in Surabaya. Regarding PM$_{10}$, the annual mean concentration of PM$_{10}$ in Surabaya was 41.31 μg/m^3, well beyond the 20 μg/m^3 limit recommended by WHO [29]. The highest level was observed in 2015 (60.46 μg/m^3) and the lowest level was recorded in 2013 (30.85 μg/m^3). Meanwhile, the annual mean concentration of NO$_2$ during the study period was 12.86 μg/m^3. The highest concentration of NO$_2$ was observed in 2018 (22.91 μg/m^3) and the lowest concentration was detected in 2011 (5.2 μg/m^3) (Figure 2). Overall, the annual level of NO$_2$ did not change significantly during the 9-year period and did not exceed the 40 μg/m^3 limit recommended by WHO or the 100 μg/m^3 limit recommended by the Indonesian government. Figure 3 illustrates the boxplots of PM$_{10}$ and NO$_2$ concentrations at the seven monitoring stations. The highest concentrations of PM$_{10}$ and NO$_2$ were both observed in SUF7 (64.77 ± 19.72 μg/m^3 for PM$_{10}$ and 19.24 ± 13.49 μg/m^3 for NO$_2$) and the lowest concentration was recorded in SUF3 (27.502 ± 18.095 μg/m^3 for PM$_{10}$ and 5.021 ± 6.782 μg/m^3 for NO$_2$).
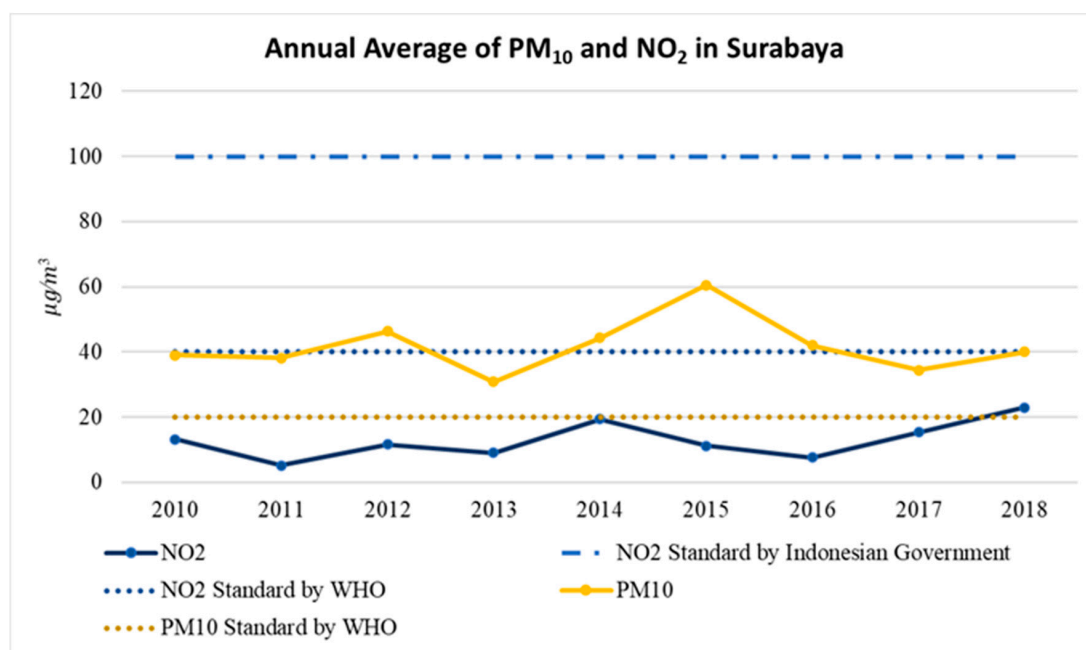

Figure 2. Time-series trend of air pollutants (particulate matter (PM$_{10}$) and nitrogen dioxide (NO$_2$) concentrations) in Surabaya from 2010 to 2018.
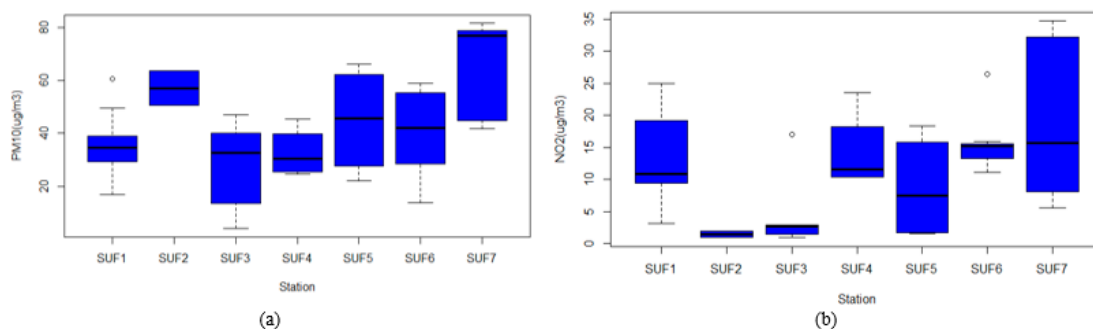

Figure 3. Box plots of (**a**) PM$_{10}$ concentrations (**b**) NO$_2$ concentrations of the 7 monitoring stations.

Int. J. Environ. Res. Public Health **2020**, *17*, 8883

7 of 15

3.2. Model Developments and Validation

Table 1 lists the coefficients of selected variables with their respective *p* values, variance inflation factor (VIF), and partial R^2 of LUR models developed in this study. For PM_{10} the final LUR model includes predictors such as public facility with a 5000 m radius, industry and warehousing within a 500 m radius, paddy field within a 2500 m radius, and NDVI within a 250 m radius, and their values, which were 0.10, 0.11, 0.12, and 0.17 partial R^2, respectively. The results demonstrate PM_{10} to be positively related to proximity to a public facility, industrial and warehousing area, and paddy field. In contrast, PM_{10} had a negative relationship with NDVI. Moreover, NDVI accounted for 17% of PM_{10} variation and, thus, is the dominant variable for the developed model.

Regarding NO_2, the predictors of paddy field within a 4250 m radius, residential area within a 4000 m radius, rainfall, and temperature were selected in the LUR model, yielding values of 0.16, 0.15, 0.08, and 0.06 partial R^2, respectively. The concentration of NO_2 was positively correlated with proximity to paddy field, proximity to residential area, and temperature, but it was negatively correlated with rainfall. The paddy field variable accounted for 16.4% of NO_2 variation and, thus, is the dominant variable in the model.

Table 1. The developed land-use regression (LUR) model for PM_{10} and NO_2.

PM_{10}				
Variable	**Coefficients**	**p-Value**	**VIF**	**Partial R^2**
Intercept	36.28	<0.01	-	-
[a] Public Facility$_{5000m}$	0.562	<0.01	1.39	0.10
[b] Industry and Warehousing$_{500m}$	0.027	0.01	1.33	0.11
[c] Paddy Field$_{2500m}$	0.185	<0.01	2.22	0.12
[d] NDVI$_{250m}$	−191	<0.01	2.89	0.17
NO_2				
Variable	**Coefficients**	**p-Value**	**VIF**	**Partial R^2**
Intercept	−374.13	0.08	-	-
[e] Paddy Field$_{4250m}$	0.146	<0.01	2.70	0.16
[f] Residential Area$_{4000m}$	0.013	<0.01	1.90	0.15
Rainfall	−3.028	<0.01	1.71	0.08
Temperature	13.212	0.08	2.75	0.06

[a] Public facility within a radius of 5000 m; [b] Industry and Warehousing within a radius of 500 m; [c] Paddy field within a radius of 2500 m; [d] Average normalized difference vegetation index (NDVI) within a radius of 250 m; [e] Paddy field within a radius of 4250 m; [f] Residential area within a radius of 4000 m.

Tables 2 and 3 both display comparison results of the three different methods for PM_{10} and NO_2, respectively, based on R^2, adjusted R^2, and Akaike information criterion (AIC). It is noted every selected variable was correlated with PM_{10} and NO_2 in the same direction in the GWR and GTWR models as those in LUR. In Table 2, GWR model and GTWR model have higher values of R^2 and lower values of AIC, indicating their model performances were better than the LUR model. Similarly, GWR model and GTWR model also show higher values of R^2 and lower values of AIC than the LUR model when comparing models for NO_2 (Table 3), indicating better performances for GWR and GTWR. Among the three approaches, GTWR-based models had the best prediction performance and a moderate explanatory power. As for model validation, the 10-fold cross validation R^2 were 0.52 and 0.53 for the GTWR-based models for PM_{10} and NO_2, respectively. The 10-cross validation confirms the robustness of the developed prediction models.

Int. J. Environ. Res. Public Health **2020**, *17*, 8883

8 of 15

Table 2. Comparison of Final Model for PM_{10} using LUR, geographically weighted regression (GWR) and geographic and temporal weighted regression (GTWR).

	LUR	GWR (Bandwidth: 1.989)	GTWR (Bandwidth: 1.414)
Intercept	36.28 [a]	36.20–36.50 [b]	35.90–37.50 [b]
Public Facility$_{5000m}$ [c]	0.562	0.559–0.560	0.544–0.563
Industry and Warehousing$_{500m}$ [d]	0.027	0.0271–0.0272	0.025–0.029
Paddy Field$_{2500m}$ [e]	0.185	0.184–0.185	0.180–0.186
NDVI$_{250m}$ [f]	−191	−191.00~−190.00	−193~−186
R^2	0.49	0.50	0.51
adjusted-R^2	0.42	0.44	0.45
AIC	310.00	305.14	305.03

[a] coefficient estimates; [b] minimum and maximum of the coefficient estimates; [c] Public facility within a radius of 5000 m; [d] Industry and Warehousing within a radius of 500 m; [e] Paddy field within a radius of 2500 m; [f] Average NDVI within a radius of 250 m.

Table 3. Comparison of final model for NO_2 using LUR, GWR and GTWR.

	LUR	GWR (Bandwidth: 1.987)	GTWR (Bandwidth: 1.985)
Intercept	−374.13 [a]	−377~−367 [b]	−377~−366 [b]
Paddy Field$_{4250m}$ [c]	0.146	0.145~0.146	0.144~0.146
Residential Area$_{4000m}$ [d]	0.013	0.0129~0.0131	0.0127~0.0131
Rainfall	−3.028	−3.07~−2.99	−3.06~−2.95
Temperature	13.212	12.9~13.3	12.90~13.30
R^2	0.46	0.47	0.48
adj-R^2	0.39	0.41	0.41
AIC	252.00	252.08	251.81

[a] coefficient estimates; [b] minimum and maximum of the coefficient estimates; [c] Paddy field within a radius of 4250 m; [d] Residential area within a radius of 4000 m.

3.3. PM_{10} and NO_2 Concentrations Variations

Figures 4 and 5 represent the annual average concentration of PM_{10} and NO_2, respectively, for the entire study period when applying the GWR model. Blue to red and green to red on the maps represent the respective levels from low to high of PM_{10} and NO_2 concentrations. Both figures illustrate that the southwestern portion of Surabaya City had the highest PM_{10} and NO_2 concentrations throughout the predicted period.
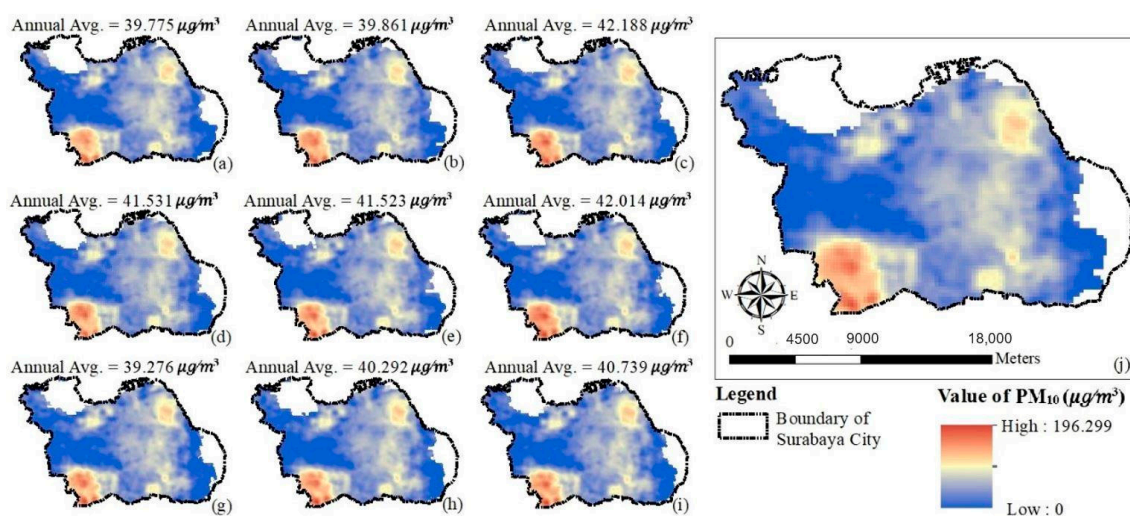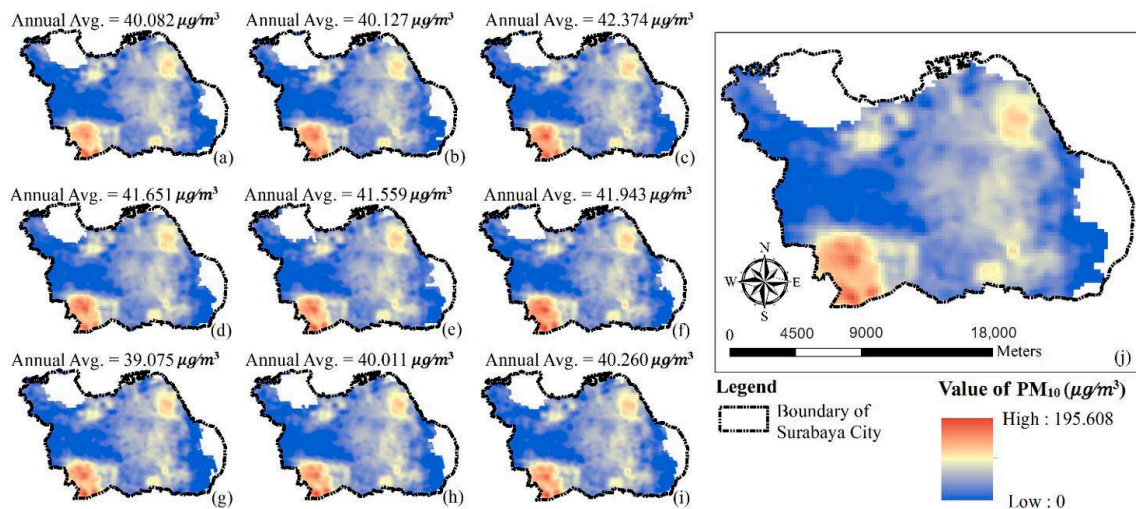

Figure 4. Prediction maps of PM_{10} concentration variations using GWR model: (**a**) 2010 (**b**) 2011 (**c**) 2012 (**d**) 2013 (**e**) 2014 (**f**) 2015 (**g**) 2016 (**h**) 2017 (**i**) 2018 (**j**) average from 2010 to 2018.

Annual Avg. = 9.397 $\mu g/m^3$ Annual Avg. = 9.897 $\mu g/m^3$ Annual Avg. = 15.010 $\mu g/m^3$

Annual Avg. = 10.458 $\mu g/m^3$ Annual Avg. = 15.136 $\mu g/m^3$ Annual Avg. = 13.159 $\mu g/m^3$

Annual Avg. = 9.136 $\mu g/m^3$ Annual Avg. = 9.421 $\mu g/m^3$ Annual Avg. = 19.466 $\mu g/m^3$

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Legend
Boundary of Surabaya City

Value of NO_2 ($\mu g/m^3$)
High : 86.578
Low : 0
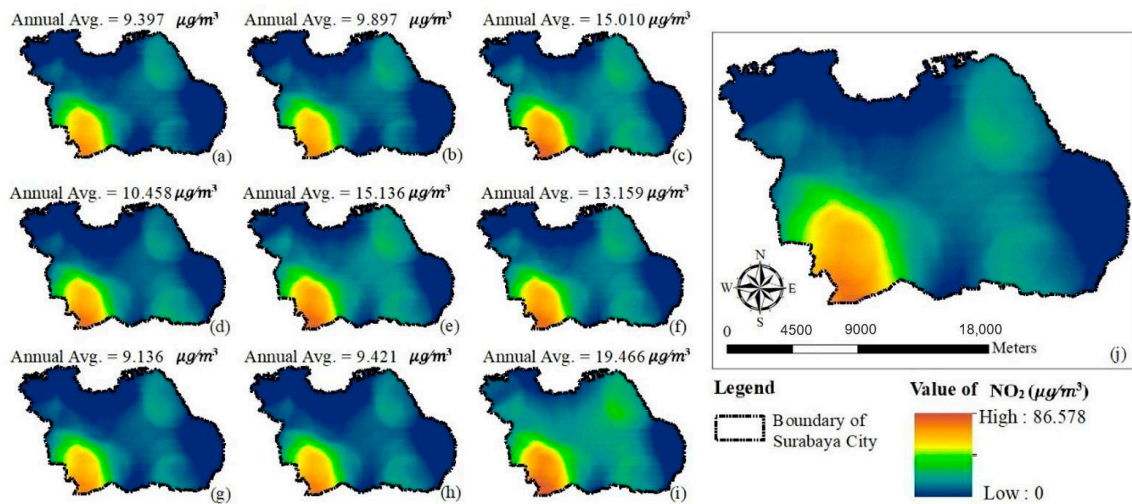
Figure 5. Prediction maps of NO_2 concentration variations using GWR model: (**a**) 2010 (**b**) 2011 (**c**) 2012 (**d**) 2013 (**e**) 2014 (**f**) 2015 (**g**) 2016 (**h**) 2017 (**i**) 2018 (**j**) average from 2010 to 2018.

Figures 6 and 7 illustrate the annual average concentration of PM_{10} and NO_2, respectively, for the entire study period, as simulated by the GTWR model. Blue to red and green to red on the maps represent the levels of respective PM_{10} and NO_2 concentrations from low to high. Both figures illustrate the southwest portion of Surabaya City had the highest PM_{10} and NO_2 concentrations throughout the predicted period.

Annual Avg. = 40.082 $\mu g/m^3$ Annual Avg. = 40.127 $\mu g/m^3$ Annual Avg. = 42.374 $\mu g/m^3$

Annual Avg. = 41.651 $\mu g/m^3$ Annual Avg. = 41.559 $\mu g/m^3$ Annual Avg. = 41.943 $\mu g/m^3$

Annual Avg. = 39.075 $\mu g/m^3$ Annual Avg. = 40.011 $\mu g/m^3$ Annual Avg. = 40.260 $\mu g/m^3$

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Legend
Boundary of Surabaya City

Value of PM_{10} ($\mu g/m^3$)
High : 195.608
Low : 0

Figure 6. Prediction maps of PM_{10} concentration variations using GTWR model: (**a**) 2010 (**b**) 2011 (**c**) 2012 (**d**) 2013 (**e**) 2014 (**f**) 2015 (**g**) 2016 (**h**) 2017 (**i**) 2018 (**j**) average from 2010 to 2018.
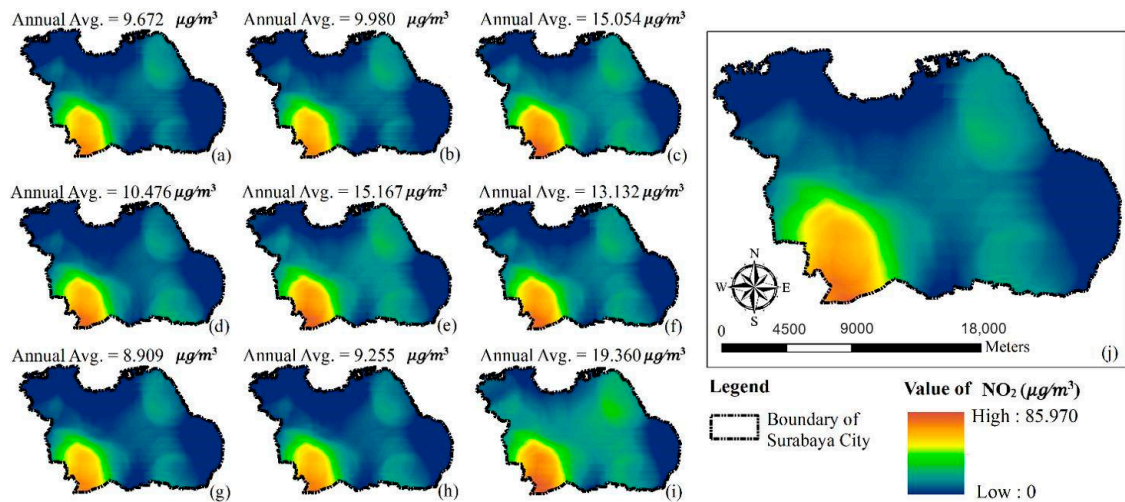
Figure 7. Prediction maps of NO$_2$ concentration variations using GTWR model: (**a**) 2010 (**b**) 2011 (**c**) 2012 (**d**) 2013 (**e**) 2014 (**f**) 2015 (**g**) 2016 (**h**) 2017 (**i**) 2018 (**j**) average from 2010 to 2018.

Figure 8 displays the comparison using the GWR method between model predictions and on-site observations. The R^2 was 0.49 for PM$_{10}$ and 0.459 for NO$_2$, with *p* values less than 0.05, confirming the robustness of established models.
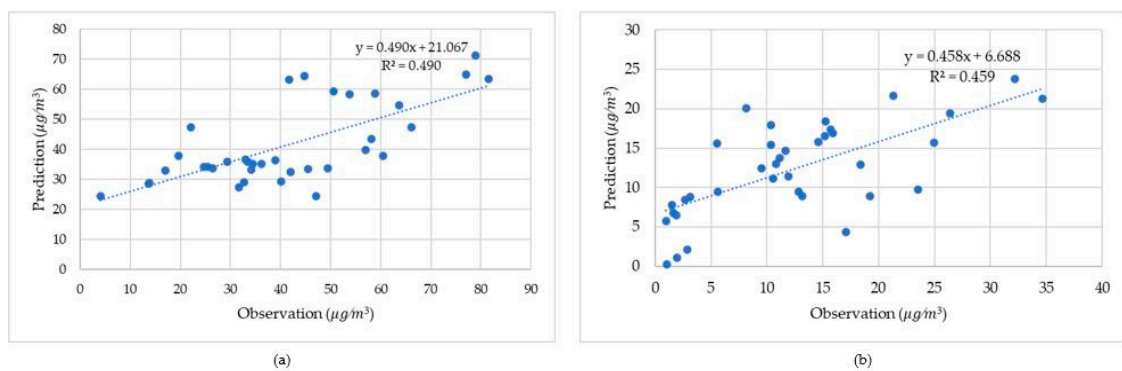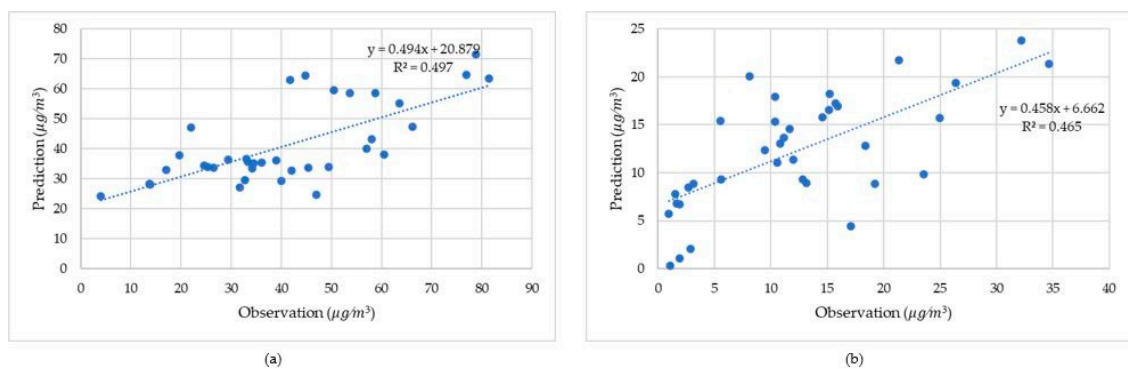

Figure 8. External verification using GWR based on observations regressed against predictions. (**a**) PM$_{10}$ (**b**) NO$_2$.

Figure 9 illustrates the comparison using the GTWR method between model predictions and on-site observations. The R^2 was 0.497 for PM$_{10}$ and 0.465 for NO$_2$, with *p* values less than 0.05, confirming the robustness of established models.

Figure 9. Verification using GTWR based on observations regressed against predictions. (**a**) PM$_{10}$ (**b**) NO$_2$.

Int. J. Environ. Res. Public Health **2020**, *17*, 8883

11 of 15

4. Discussion

To the best of our knowledge, this is the first study to estimate PM_{10} and NO_2 concentrations across Surabaya City, Indonesia, by employing three different methods (i.e., LUR, GWR and GTWR) with the use of land-use information and weather data in order to examine the determinants influencing air quality. According to the study results, land use type, (e.g., paddy fields, residential areas, industry and warehousing), climate conditions (e.g., rainfall and temperature), and greenness were determinants influencing air quality. The study results could be used for reference in developing countries similar to Indonesia for establishing policies or regulations focusing on air quality control.

Several studies have demonstrated meteorological factors such as humidity, temperature, wind speed and wind direction are highly correlated with air pollution [45,46]. Consistent with these studies, our results reflected a positive relationship between temperature and NO_2. In contrast, increased humidity and rainfall are protective factors against air pollution [47], which is consistent with the results of our study. In addition to meteorological factors that cannot be controlled, greenness plays a significant role in reduction of air pollutants [47,48]. Our study findings demonstrated greenness to have a protective effect against air pollutants. As such, it is suggested that local governments prioritize retaining a greater percentage of greenness when executing urban planning and design.

Consistent with previous studies [49–52], our results confirm that types of land use, such as public facility, industry and warehousing, residential area, and paddy field are positively associated with PM_{10} and NO_2 levels. Prior studies have indicated that higher levels of NO_2 may come from cooking fuels and biomass burning [49,50,53], helping to explain the relationship between air pollution and both residential areas and paddy fields. Therefore, it is suggested that every household should have an exhaust hood in order to cope with emissions of cooking oil fumes. Meanwhile, biomass burning should be banned in order to prevent producing an avoidable amount of harmful air pollutants. In addition, the insignificant impact from climate conditions (temperature and rainfall) on PM_{10} concentrations may be because other variables (e.g., paddy field and NDVI) influence the seasonal variation more than climate conditions.

According to the study results the southwest portion of Surabaya City has the highest PM_{10} and NO_2 concentrations. This may be because the southwest portion of the country has the highest residential density and road traffic. According to the findings of a previous study focusing on Surabaya City, the air quality of the city is related to its traffic volume, seeing as air quality has continually worsened as the number of vehicles on the road has increased [46]. Hence, developing a public transportation system and encouraging the use of electrical vehicles may be ways to alleviate the traffic burden and to improve air quality.

This study made a comparison between three methods to examine the influence of determinants in PM_{10} and NO_2 concentrations. Although solid analyses were employed, there are some limitations to this study. First, Surabaya only has seven monitoring stations and the monitoring stations are not evenly distributed throughout the city (e.g., there are no air pollution monitoring stations in the western portion of Surabaya City), reducing the generalizability of study results. However, the distribution of monitoring stations within a circular buffer covers a large percentage of the study site and was able to include major factors such as residential and industrial areas. Second, given that Surabaya City had more than two million vehicles in 2015 and the total number of vehicles steadily increased from 2011 to 2015 [54], traffic conditions should be included in the model. Third, using the IDW interpolation method to estimate meteorological conditions can result in specific errors. Nonetheless, there are no traffic intensity data available, limiting the degree of explanation of the final model. Lastly, the lack of MODIS aerosol optical depth is another limitation in this study since this was considered an important variable for estimating PM_{10} and NO_2 [55]. While a moderate model predictive power was obtained from this study, we suggest future studies consider the aforementioned predictor variables in order to improve model prediction performance.

5. Conclusions

This is the first study to implement spatio-temporal variations of pollution concentration in Surabaya City, Indonesia, by making a comparison between LUR, GWR, and GTWR models. The results show land-use type, climate conditions, and greenness were determinants influencing air quality. Moreover, by using data from a monitoring network which routinely monitors air quality with different characteristics in different areas, we confirmed that the models developed in this study can predict fine spatial variability in short-term and long-term outdoor PM_{10} and NO_2 concentrations. This study helps us to better understand the air quality in Surabaya City and, in effect, helps to provide a direction for future epidemiological studies.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-4601/17/23/8883/s1, Table S1: List of potential predictor variables collected from each database.

References

1. United States Environmental Protection Agency (USEPA). *Integrated Science Assessment for Particulate Matter*; United States Environmental Protection Agency (USEPA): Washington, DC, USA, 2009.
2. World Health Organization (WHO). *World Health Statistics*; World Health Organization: Luxembourg, 2015.
3. HEI. Outdoor Air Pollution and Health in the Developing Countries of Asia: A Comprehensive Review. *Spec. Rep.* **2010**, *18*, 284.
4. Oanh, N.T.K.; Bich, T.L.; Tipayarom, D.; Manadhar, B.R.; Prapat, P.; Simpson, C.; Liu, L.-J.S. Characterization of Particulate Matter Emission from Open Burning of Rice Straw. *Atmos. Environ.* **2010**, *45*, 493–502. [CrossRef] [PubMed]
5. González, L.O.; Zhang, Z.; Sánchez, B.N.; Zhang, K.; Brown, D.G.; Bracho, L.R.; Vargas, A.O.; Ortega, F.; Vadillo O'Neil, M.S. An Assessment of Air Pollutant Exposure Methods in Mexico City, Mexico. *J. Air Waste Manag. Assoc.* **2015**, *65*, 581–591. [CrossRef] [PubMed]
6. Achakulwisut, P.; Brauer, M.; Hystad, P.; Anenberg, S.C. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: Estimates from global datasets. *Lancet Planet Health* **2019**, *3*, e166–e178. [CrossRef]
7. Dockery, D.W.; Pope, C.A.; Xu, X.; Spengler, J.D.; Ware, J.H.; Fay, M.E.; Ferris, B.G.; Speizer, F.E. An Association between Air Pollution and mortality in Six U.S. Cities. *J. Med.* **1993**, *329*, 1753–1759. [CrossRef] [PubMed]
8. Pope, C.A.; Thun, M.J.; Namboodiri, M.M.; Dockery, D.W.; Evans, J.S.; Speizer, F.E.; Heath, C.W. Particulate air pollution as a predictor of mortality in a prospective study of U.S. *Adults. Am. J. Respir. Crit. Care Med.* **2015**, *151*, 669–674. [CrossRef] [PubMed]
9. Cao, J.; Yang, C.; Li, J.; Chen, R.; Chen, B.; Gu, D.; Kan, H. Association between long-term exposure to outdoor air pollution and mortality in China: A cohort study. *J. Hazard. Mater.* **2011**, *186*, 1594–1600. [CrossRef]
10. Gu, Y.; Lin, H.; Liu, T.; Xiao, J.; Zeng, W.; Li, Z.; Lv, X.; Ma, W. The Interaction between Ambient PM_{10} and NO_2 on Mortality in Guangzhou, China. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1381. [CrossRef]

11. Yudha, S.W. *Air Pollution in Indonesia-Challenges and Imperatives for Change*; The National Bureau of Asian Research: Washington, DC, USA, 2016; Available online: https://www.nbr.org/publication/air-pollution-in-indonesia/ (accessed on 5 July 2020).

12. Greenstone, M.B.; Fan, Q. *Indonesia's Worsening Air Quality and Its Impact on Life Expectancy*; Air Quality Life Index: Chicago, IL, USA, 2019.

13. Duki, M.I.Z.; Sudarmadi, S.; Suzuki, S.; Kawada, T.; Tri-Tugaswati, A. Effect of air pollution on respiratory health in Indonesia and its economic cost. *Arch. Environ. Health* **2003**, *58*, 135–143. [CrossRef]

14. Susanto, D.A. Air Pollution and Human Health. *Med. J. Indones.* **2020**, *29*, 8–10. [CrossRef]

15. Tri-Tugaswati, A. Review of Air Pollution and its Health Impact in Indonesia. In *Fourth International Symposium on Neurobehavioral Methods and Effects in Occupational and Environmental Health*; Academic Press: Tokyo, Japan, 1994; pp. 767–772. [CrossRef]

16. Mannucci, P.M.; Franchini, M. Health effects of ambient air pollution in developing countries. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1048. [CrossRef] [PubMed]

17. Peña, M.S.B.; Rollins, A. Environmental Exposures and Cardiovascular Disease: A Challenge for Health and Development in Low- and Middle- Income Countries. *Cardiol. Clin.* **2017**, *35*, 71–76. [CrossRef] [PubMed]

18. Liu, H.L.; Shen, Y.S. The impact of green space changes on air pollution and microclimates: A case study of the Taipei metropolitan area. *Sustainability* **2014**, *6*, 8827–8855. [CrossRef]

19. Weng, Q.; Yang, S. Urban air pollution patterns, land use, and thermal landscape: An examination of the linkage using GIS. *Environ. Monit. Assess.* **2006**, *117*, 463–489. [CrossRef] [PubMed]

20. Ostro, B.; Malig, B.; Broadwin, R.; Basu, R.; Gold, E.B.; Bromberger, J.T.; Carol, D.; Steven, F.; Gail, A.G.; Elizabeth, A.J.; et al. Chronic $PM_{2.5}$ exposure and inflammation: determining sensitive subgroups in mid-life women. *Environ Res.* **2014**, *132*, 168–175. [CrossRef]

21. Wu, C.-D.; Chen, Y.-C.; Pan, W.-C.; Zeng, Y.-T.; Chen, M.-J.; Guo, Y.L.; Lung, S.-C.C. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model $PM_{2.5}$ spatial-temporal variability. *Environ. Pollut.* **2017**, *224*, 148–157. [CrossRef]

22. Liu, W.; Li, X.; Chen, Z.; Zeng, G.; León, T.; Liang, J.; Huang, G.; Gao, Z.; Jiao, S.; He, X.; et al. Land use regression models coupled with meteorology to model spatial and temporal variability of NO_2 and PM_{10} in Changsha, China. *Atmos. Environ.* **2015**, *116*, 272–280. [CrossRef]

23. Bai, Y.; Wu, L.; Qin, K.; Zhang, Y.; Shen, Y.; Zhou, Y. A geographically and temporally weighted regression model for ground-level $PM_{2.5}$ estimation from satellite-derived 500 m resolution AOD. *Remote Sens.* **2016**, *8*, 262. [CrossRef]

24. Eeftens, M.; Meier, R.; Schindler, C.; Aguilera, I.; Phuleria, H.; Ineichen, A.; Künzli, N. Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. *Environ. Health* **2016**, *15*, 53. [CrossRef]

25. Guo, Y.; Tang, Q.; Gong, D.Y.; Zhang, Z. Estimating ground-level $PM_{2.5}$ concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sens. Environ.* **2017**, *198*, 140–149. [CrossRef]

26. Gaeta, A.; Cattani, G.; Bucchianico, A.D.M.D.; De Santis, A.; Cesaroni, G.; Badaloni, C.; Ancona, C.; Forastiere, F.; Sozzi, R.; Bolignano, A.; et al. Development of nitrogen dioxide and volatile organic compounds land use regression models to estimate air pollution exposure near an Italian airport. *Atmos. Environ.* **2016**, *131*, 254–262. [CrossRef]

27. Hsu, C.Y.; Zeng, Y.T.; Chen, Y.C.; Chen, M.J.; Lung, S.C.C.; Wu, C.D. Kriging-based land-use regression models that use machine learning algorithms to estimate the monthly BTEX concentration. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6956. [CrossRef] [PubMed]

28. Fotheringham, A.S.; Brunsdon, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley & Sons Ltd.: Chichester, UK, 2002; pp. 27–64.

29. Zhang, L.; Ma, Z.; Guo, L. An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *For. Sci.* **2009**, *55*, 533–548.

30. Wei, Q.; Zhang, L.; Duan, W.; Zhen, Z. Global and geographically and temporally weighted regression models for modelling $PM_{2.5}$ in Heilongjiang, China from 2015 to 2018. *Int. J. Environ. Res. Public Health* **2019**, *16*, 5107. [CrossRef]

Int. J. Environ. Res. Public Health **2020**, *17*, 8883

14 of 15

31. Ma, X.; Zhang, J.; Ding, C.; Wang, Y. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Comput. Environ. Urban Syst.* **2018**, *70*, 113–124. [CrossRef]

32. Peng, Y.; Li, W.; Luo, X.; Li, H. A geographically and temporally weighted regression model for spatial downscaling of MODIS land surface temperatures over urban heterogeneous regions. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5012–5027. [CrossRef]

33. BPS (Statistics Surabaya). *2014, Surabaya in Figures*, 1102001.35. ed.; Statistics Bureau of Surabaya: Surabaya, Indonesia, 2014.

34. BAPPEKO (City Development Planning Bureau of Surabaya). *Land Use Map of Suarabaya in 2014*; Surabaya City Government Press: Surabaya, Indonesia, 2014.

35. Rahayu, Y.E.; Ahyudanari, E.; Pratomoadmojo, N.A. Land Use Development and its Impact on Airport Access Road. *Procedia Soc. Behav. Sci.* **2016**, *227*, 31–37. [CrossRef]

36. Pemkot Surabaya. *Informasi Kinerja Pengelolaan Lingkungan Hidup Daerah Tahun 2016*; Pemerintah Kota Surabaya: Surabaya, Indonesia, 2017.

37. USEPA. Particulate Matter (PM) Pollution. Available online: https://www.epa.gov/pm-pollution/particulate-matter-pm-basics (accessed on 10 July 2020).

38. USGS. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006*. Available online: https://lpdaac.usgs.gov/products/mod13q1v006/ (accessed on 11 July 2020).

39. Zeng, Y.T.; Wu, C.D.; Chen, Y.C.; Hsu, C.Y.; Chen, M.J. Comparison of geospatial-temporal modeling approaches in air pollution estimations. In Proceedings of the 40th Asian Conference on Remote Sensing: Progress of Remote Sensing Technology for Smart Future, ACRS 2019, Daejeon, Korea, 14–18 October 2020.

40. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically Weighted Regression: A method for exploring spatial nonstationary. *Geogr. Anal.* **1996**, *28*, 52. [CrossRef]

41. Anselin, L. Local indicator of spatial association—LISA. *Geog. Anal.* **1995**, *27*, 93–115. [CrossRef]

42. Huang, B.; Wu, B.; Barry, M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 383–401. [CrossRef]

43. Hu, Z. Spatial analysis of MODIS aerosol optical depth, $PM_{2.5}$, and chronic coronary heart disease. *Int. J. Health Geogr.* **2009**, *8*, 27. [CrossRef] [PubMed]

44. Liu, J.; Zhao, Y.; Yang, Y.; Xu, S.; Zhang, F.; Zhang, X.; Shi, L.; Qiu, A. A mixed geographically and temporally weighted regression: Exploring spatial-temporal variations from global and local perspectives. *Entropy* **2017**, *19*, 53. [CrossRef]

45. Tai, A.P.; Mickley, L.J.; Jacob, D.J. Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change. *Atmos. Environ.* **2010**, *44*, 3976–3984. [CrossRef]

46. Syafei, A.D.; Fujiwara, A.; Zhang, J. Spatial and temporal factors of air quality in Surabaya city: An analysis based on a multilevel model. *Procedia-Soc. Behav. Sci.* **2014**, *138*, 612–622. [CrossRef]

47. Kwak, H.; Ko, J.; Lee, S.; Joh, C. Identifying the correlation between rainfall, traffic flow performance and air pollution concentration in Seoul using a path analysis. *Transp. Res. Procedia* **2017**, *25*, 3556–3567. [CrossRef]

48. Macnaughton, P.; Eitland, E.; Kloog, I.; Schwartz, J.; Allen, J. Impact of particular matter exposure and surrounding "Greenness" on chronic absenteeism in Massachusetts public schools. *Int. J. Environ. Res. Public Health* **2017**, *14*, 207. [CrossRef]

49. Quackenboss, J.J.; Spengler, J.D.; Kanarek, M.S.; Letz, R.; Duffy, C.P. Personal exposure to nitrogen dioxide: Relationship to indoor/outdoor air quality and activity patterns. *Environ. Sci. Technol.* **1986**, *20*, 775–783. [CrossRef]

50. Ryan, P.B.; Spengler, J.D.; Halfpenny, P.F. Sequential box models for indoor air quality: Application to airliner cabin air quality. *Atmos. Environ.* **1988**, *22*, 1031–1038. [CrossRef]

51. Han, L.; Zhou, W.; Li, W.; Meshesha, D.T.; Li, L.; Zheng, M. Meteorological and urban landscape factors on severe air pollution in Beijing. *J. Air Waste Manag. Assoc.* **2015**, *65*, 782–787. [CrossRef]

52. Park, S.H.; Ko, D.W. Investigating the effects of the built environment on $PM_{2.5}$ and PM_{10}: A case study of Seoul Metropolitan city, South Korea. *Sustainability* **2018**, *10*, 4552. [CrossRef]

53. Lasko, K.; Vadrevu, K. Improved rice residue burning emissions estimates: Accounting for practice-specific emission factors in air pollution assessments of Vietnam. *Environ. Pollut.* **2018**, 795–806. [CrossRef] [PubMed]

Int. J. Environ. Res. Public Health **2020**, *17*, 8883

15 of 15

54. Risk Factors and Early Symptoms Related to Respiratory Disease in Pedicab Drivers in Surabaya. *J. "Kesehatan Masyarakat* **2019**, *15*, 224–235.

55. Zhang, Z.; Wang, J.; Hart, J.E.; Laden, F.; Zhao, C.; Li, T.; Zheng, P.; Li, D.; Ye, Z.; Chen, K. National scale spatiotemporal land-use regression model for $PM_{2.5}$, PM_{10} and NO_2 concentration in China. *Atmos. Environ.* **2018**, *192*, 48–54. [CrossRef]