atmosphere

MDPI

Article

Land Use Regression Modeling of $PM_{2.5}$ Concentrations at Optimized Spatial Scales

Liang Zhai [1], Bin Zou [2,3,*], Xin Fang [2], Yanqing Luo [2], Neng Wan [4] and Shuang Li [1]

1 National Geographic Conditions Monitoring Research Center, Chinese Academy of Surveying and Mapping, Beijing 100830, China; zhailiang@casm.ac.cn (L.Z.); ls02020029@163.com (S.L.)
2 School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; xinfang@csu.edu.cn (X.F.); 472978832@csu.edu.cn (Y.L.)
3 Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, Changsha 410083, China
4 Department of Geography, University of Utah, Salt Lake, UT 84112, USA; wanneng99@gmail.com
* Correspondence: 210010@csu.edu.cn; Tel.: +86-731-8883-6502

Academic Editor: Robert Talbot
Received: 28 November 2016; Accepted: 20 December 2016; Published: 23 December 2016

Abstract: Though land use regression (LUR) models have been widely utilized to simulate air pollution distribution, unclear spatial scale effects of contributing characteristic variables usually make results study-specific. In this study, LUR models for $PM_{2.5}$ in Houston Metropolitan Area, US were developed under scales of 100 m, 300 m, 500 m, 800 m, and 1000–5000 m with intervals of 500 m by employing the idea of statistically optimized analysis. Results show that the annual average $PM_{2.5}$ concentration in Houston was significantly influenced by area ratios of open space urban and medium intensity urban at a 100 m scale, as well as of high intensity urban at a 500 m scale, whose correlation coefficients valued −0.64, 0.72, and 0.56, respectively. The fitting degree of LUR model at the optimized spatial scale (adj. R^2 = 0.78) is obviously better than those at any other unified spatial scales (adj. R^2 ranging from 0.19 to 0.65). Differences of $PM_{2.5}$ concentrations produced by LUR models with best-, moderate-, weakest fitting degree, as well as ordinary kriging were evident, while the LUR model achieved the best cross-validation accuracy at the optimized spatial scale. Results suggested that statistical based optimized spatial scales of characteristic variables might possibly ensure the performance of LUR models in mapping $PM_{2.5}$ distribution.

Keywords: $PM_{2.5}$; LUR; air pollution; spatial scale; GIS

1. Introduction

Fine particulate matter ($PM_{2.5}$) in air pollution has become a significant threat to global human health. Due to its minuscule diameter (≤2.5 microns) $PM_{2.5}$ is inhaled and penetrates into the circulatory, respiratory, and immune systems, triggering cancer, mutagenesis, and other skin diseases [1–3]. $PM_{2.5}$ refers to the solid or liquid fine particulate that is characterized by irregular shapes, strong enrichment effects, and the absorption of abundant hazardous substances [4]. Various measures have been attempted to reduce $PM_{2.5}$ pollution, such as improvements of vehicle technology and energy use efficiency, yet global $PM_{2.5}$ pollution levels still remain at a harmful level due to increasing fuel consumption and urbanization.

A team from US (United States) NASA (National Aeronautics and Space Administration) utilizing the MODIS (MODerate Resolution Imaging Spectroradiometer)/MISR (Multiangle Imaging SpectroRadiometer) based aerosol optical depth (AOD) data and the GEOS-Chem (Geostationary Ocean Color Imager) chemical transmission model identified that most of the world experienced annual average $PM_{2.5}$ concentrations that exceeded the WHO defined safety limit (i.e., 10 μg·m^{-3}) [5].

Mean $PM_{2.5}$ concentrations were greater than 50 µg·m^{-3} and particularly high in North Africa and East Asia [6]. The Global Burden of Disease Study 2010 reported that $PM_{2.5}$ pollution caused 3.2 million premature deaths and a loss of 76 million healthy life years annually around the world [7]. There were several previous studies focused on the components and health effects (e.g. mortality, emergency hospital admissions, emergency department visits) of $PM_{2.5}$ related in Houston [8–10]. This situation suggests that public health has suffered serious risks associated with $PM_{2.5}$ pollution. Therefore, clearly and correctly understanding the spatial-temporal characteristics of $PM_{2.5}$ distribution is essential to effectively evaluate and decrease human exposure risks.

To accurately simulate the spatial and temporal distribution of $PM_{2.5}$ concentrations, several methods, including spatial interpolation, air pollution dispersion modeling, MODIS remote-sensing retrieval, land use regression (LUR), geographically weighted regression (GWR), timely structure adaptive model (TSAM), and artificial neural network [11–17], have been proposed to estimate $PM_{2.5}$ concentrations. LUR utilizes observed concentrations as well as characteristic variables at air quality monitoring sites within a certain area, and can be used to predict the air pollution concentration of spatial locations in the area [18]. This method has been considered an ideal proxy for $PM_{2.5}$ estimation because of the comprehensive element consideration, acceptable simulation accuracy, spatial resolution, and wide applicability in simulating $PM_{2.5}$ distribution ins situation where currently there is no clear physical-chemical dispersion mechanism of $PM_{2.5}$ [13,16].

Since its introduction in 1997, the LUR method has been widely applied in globally distributed air pollution simulation studies of NO_2 (nitrogen dioxide), NO (nitric oxide), PM_{10} (inhalable particles), and $PM_{2.5}$, including in Britain, United States, Netherlands, Canada, and China [19–24]. In these studies, the adjusted fitting degree (R^2) of reported LUR models ranged from 0.17 to 0.73. One of the most important factors that has contributed to the accuracy differences of the LUR models is the different buffering radius ranging from 20 m–30 km used to measure value of characteristic variable [25–28]. However, to the best of our knowledge, an effective method for determining the reasonable spatial scale of a characteristic variable is still lacking due to the complex physical-chemical mechanism of $PM_{2.5}$ pollution.

However, fortunately, statistical experience analysis has been proven as the reasonable way to preliminarily detect the relationship between two factors with possible association, while the true interactive mechanism of these factors in the real world is not clear [29–31]. Therefore, this study aims to explore the spatial scale dependence of associated characteristic variables on the observed $PM_{2.5}$ concentrations at monitoring sites, and further evaluate whether the performance of the LUR model with characteristic variables at optimized spatial scale can be enhanced without the integration of a clear physical-chemical mechanism of $PM_{2.5}$ pollution. The research results could provide a theoretical basis for assessing the contribution of characteristic variables to $PM_{2.5}$ concentrations at surrounding spatial locations. More importantly, this study is going to discuss about the spatial scale dependence of LUR modeling, and will greatly promote the reliability and stability of the LUR method in urban/regional $PM_{2.5}$ mapping in terms of spatial scale optimization.

2. Data and Method

2.1. Study Area and Data Collection

Houston, Texas, USA is a typical urban pollution area with stable geographic and meteorological environment, high air pollution level, and comparatively intensive urban $PM_{2.5}$ monitoring sites. As the fourth largest metropolitan area in the US, Houston displays significant characteristics that are relevant to urban air pollution. Flat and built on former swampland, the city has a subtropical climate with 1224 mm of precipitation annually and an average temperature of 20.7 °C. It is well known for its petroleum industry, high economic development, and 26% population growth from 2000–2010. In 2011, 17 $PM_{2.5}$ monitoring sites, which including federal reference monitors (FRM) and federal equivalent method (FEM) monitors which provide measurements on days when FRMs are not recording and at locations without FRMs, were installed across greater Houston. Due to the local industrial production and traffic emission, the annual mean of observed particulate matter concentration ranged from 9.87 $\mu g \cdot m^{-3}$ (minimum) to 14.24 $\mu g \cdot m^{-3}$ (maximum), and the average value was 11.66 $\mu g \cdot m^{-3}$, while, there was only one station within the WHO $PM_{2.5}$ concentration safety limit (10 $\mu g \cdot m^{-3}$) in this region.

Land use (e.g., fraction of built, forest, water, and grass), road traffic, road (e.g., road length, distance to the nearest road), coast (e.g., distance to the nearest coast), population distribution, geographical location, and climate characteristics were considered to be the general factors associated with $PM_{2.5}$ emission and dispersion in previous LUR research findings [16,23,27,28,32–35]. Data collected for LUR modeling in this study therefore contains annual average $PM_{2.5}$ concentration [36], land use/cover in 2011 [37], road network in 2011 [38], and census data in 2010 [39]. The basic geographical data and $PM_{2.5}$ monitoring sites distribution within the Houston area are shown in Figure 1.

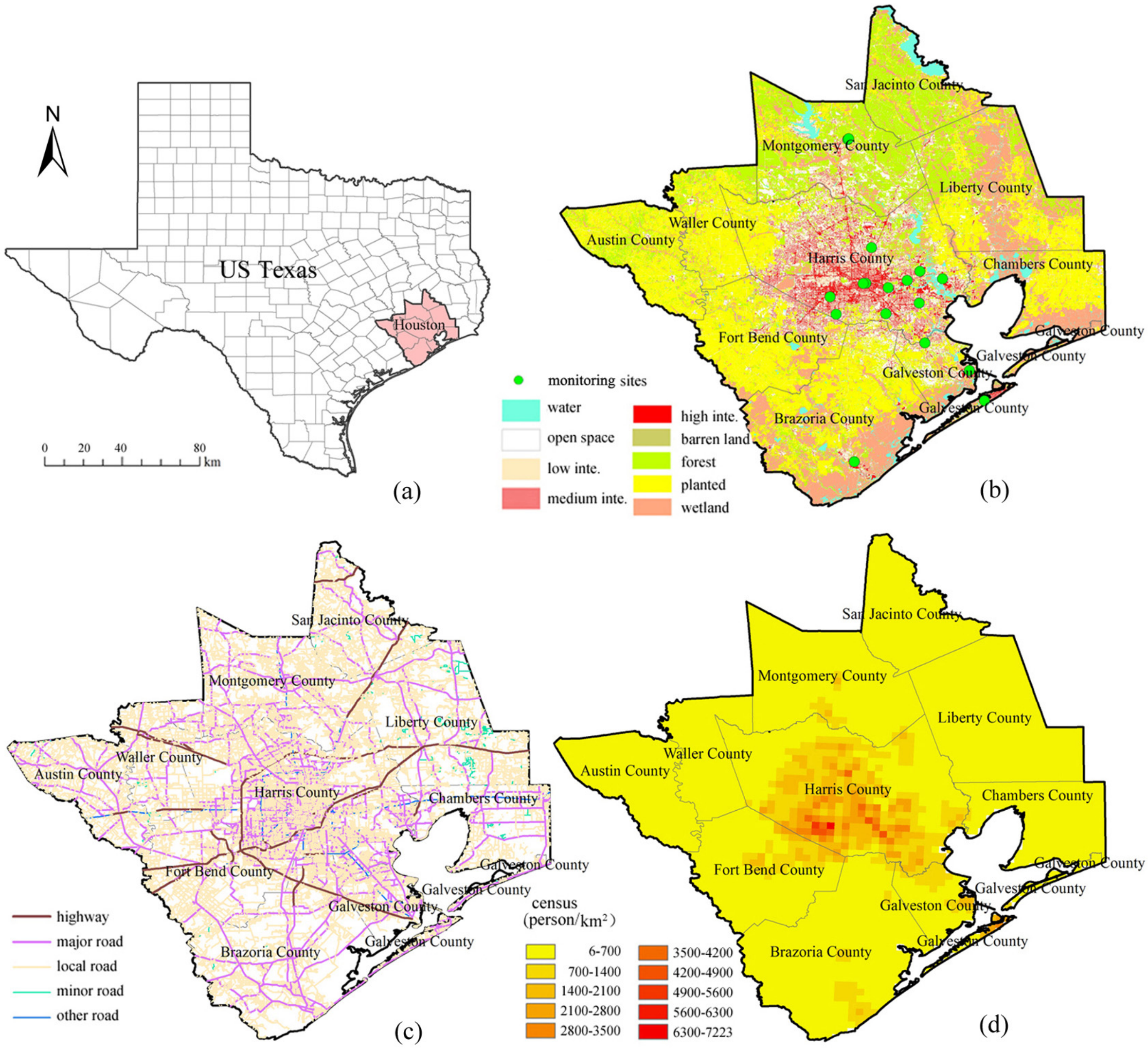

Figure 1. Study area and $PM_{2.5}$ monitoring site: (**a**) Study area; (**b**) $PM_{2.5}$ monitoring site and land use/cover; (**c**) road network; (**d**) census data.

2.2. Study Design

As shown in Figure 2, this study was divided into three parts including extraction of characteristic variables, correlation analysis, and impact analysis of spatial scale on LUR modeling and mapping.

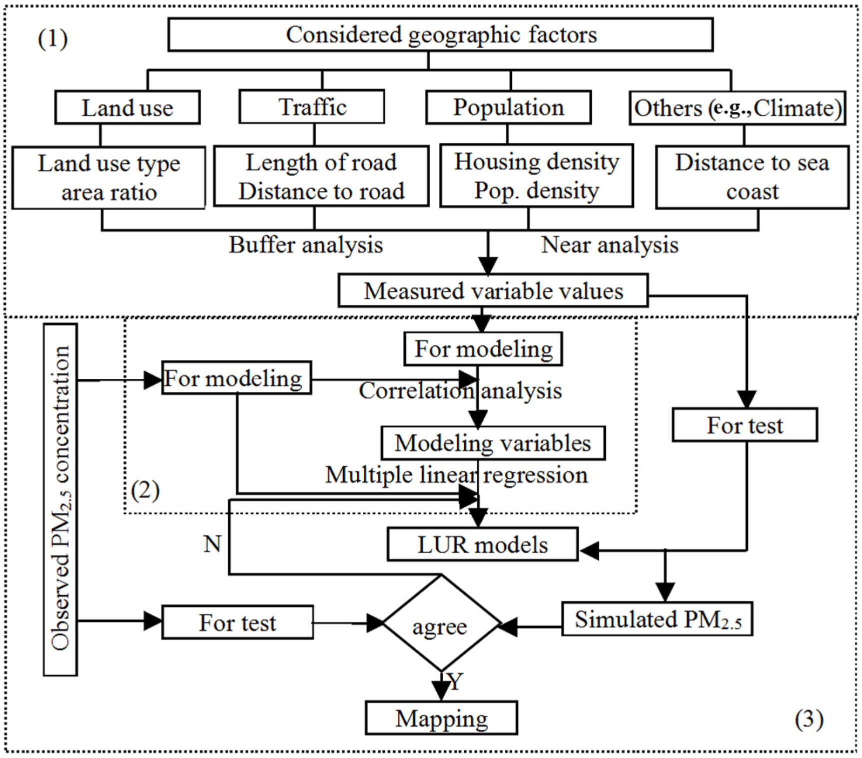

Figure 2. Framework of study procedure. LUR, land use regression. (1) Variables extraction; (2) Variables screen; (3) LUR model fitting and cross-validation.

2.2.1. Extraction of Characteristic Variables

As the rules mentioned above, characteristic variables utilized for LUR modeling in this study included area ratio of land use, total road length, distance to nearest road, population density, housing density, and distance to sea coast. All these factors, except distance to nearest road, had obvious spatial scale effects. That is to say, the measured values would vary with the changes of the buffering radius of $PM_{2.5}$ monitoring sites. The buffering radiuses were set as 100 m, 300 m, 500 m, 800 m, and 1000–5000 m with intervals of 500 m, according to previous research findings [27,28]. For the area ratio (%) of a specific land use type of $PM_{2.5}$ monitoring site, it was implemented by measuring the area of this land use type and then dividing it by the total area of all land use types within the certain buffering radius of this site. In this process, the original land use types were reclassified into "forest" ($Forest_{11}$), "open space urban" ($O\text{-}urban_{12}$), "medium intensity urban" ($M\text{-}urban_{13}$), "high intensity urban" ($H\text{-}urban_{14}$), and "barren land" ($Barren_{15}$) based on the similarity of reducing or increasing $PM_{2.5}$ concentration diffusion. For characteristic variables of "total road length" ($T\text{-}length_{21}$) and "distance to nearest road" ($D\text{-}road_{22}$), the measured values (unit: km) were computed for the length based on all level roads including highway, major road, local road, minor road, and other road, within certain buffering radius. Similarly, "population density" ($P\text{-}density_{31}$, unit: person/km^2) and "housing density" ($H\text{-}density_{32}$, unit: house/km^2) were calculated by counting the number of populations and houses, respectively, and then dividing them by the area of each buffering radius. Additionally, spatial scale free variable of "distance to sea coast" ($D\text{-}coast_{41}$, unit: km) was also extracted to indirectly represent the possible influences of other geographical and climate characteristic factors (e.g., wind speed, temperature, and humidity).

2.2.2. Correlation Analysis

Based on the clear characteristic variables of the model, the aforementioned step for LUR modeling was used to extract the 'measured values' of these variables at different preset buffering radiuses. However, these measured values usually varied with the spatial scales, as shown in Table 1, and reported LUR models were plagued on account of a lack of reasonable methods to determine the ideal spatial scales of these measured values [31,40,41]. Therefore, this study attempted to develop a way to initially discern the measured values of characteristic variables at an ideal buffering radius (i.e., optimized spatial scale) to improve the performance of LUR. This procedure was conducted by conducting correlation analyses between all the measured values of characteristic variables at preset various buffering radiuses. The annual mean $PM_{2.5}$ concentrations were calculated based on the observed measurements from the regulatory monitoring stations. As a result, the measured value of each of the characteristic variables at a relatively ideal buffering radius with regards to the maximum Pearson coefficient could be kept. In contrast, those measured values of variables at irrelevant buffering radiuses would be screened out due to the statistically weaker values of "Pearson coefficient" [42].

2.2.3. Impact Analysis of Spatial Scale on LUR Modeling and Mapping

To validate the feasibility of statistically determining the ideal spatial scale of a characteristic variable in LUR modeling using correlation analysis and its impact on the accuracy of LUR mapping, this study developed all the LUR models both at the ideal buffering radius and relatively irrelevant buffering radiuses. These LUR models were built using SAS analysis (SAS Institute, Cary, NC, USA) and backward multi linear regression (MLR) with non-spatial variables (i.e., distance to sea coast, distance to nearest road). The significant level of t tests less than 0.05 and VIF (Variance Inflation Factors) values less than 5, which were used to control the collinearity between modeling variables, were used as the additional conditions for characteristic variables to determine whether they were introduced into the LUR model or not. Differences of simulation results among LUR models were firstly validated by comparing predicted annual average $PM_{2.5}$ concentrations with observed concentrations at monitoring sites using the N-1 cross validation strategy. Consequently, in order to demonstrate the outperformance of the LUR model at the optimized spatial scale, annual average $PM_{2.5}$ concentration surfaces of Houston were produced by LUR models with best, moderate-, and weakest fitting degrees, respectively, as well as ordinary kriging, which is a preferred geostatistical method in air pollution modeling [14]. In this process, a *Levene*'s test [43] and an F test [44] were also employed to verify the difference between the concentrations extracted from above surfaces with a regular grid size of 3 km $\times$ 3 km.

3. Results

3.1. Preliminary Identification of $PM_{2.5}$ Related Characteristic Variables

Figure 3 demonstrates that the Pearson correlation coefficients between the characteristic variables and annual average $PM_{2.5}$ concentrations varied with the changes of buffering radiuses (i.e., spatial scales). These correlation coefficients ranged from -0.64 to 0.72 for land use class (Figure 3a), from 0.10 to 0.46 for total road length (Figure 3b), and from -0.26 to 0.14 for population- and housing density (Figure 3c). More importantly, the cross-scale comparison of correlation coefficients identified unique spatial scale effects of different variables. For instance, the area ratios of forest ($Forest_{11}$) and open space urban ($O\text{-}urban_{12}$) were negatively correlated with annual average $PM_{2.5}$ concentrations, peaking at 100 m and 5000 m, respectively. The correlations of medium intensity urban ($M\text{-}urban_{13}$), high intensity urban ($H\text{-}urban_{14}$), and barren land ($Barren_{15}$) with annual average $PM_{2.5}$ concentrations were most influenced by the scales of 100 m, 500 m, and 3 km. The total road length ($T\text{-}length_{21}$) was positively correlated with annual average $PM_{2.5}$ concentration, particularly at the 100 m scale, while the correlation coefficients decreased rapidly as the buffering radius increased to 500 m. However, this decreasing trend fluctuated after the buffering radius of 500 m and remained relatively stable at about

0.20. The correlation coefficients of population- (*P-density*$_{31}$) and housing density (*H-density*$_{32}$) with annual average $PM_{2.5}$ concentrations varied greatly within 2000 m and decreased thereafter, while the optimized scales for them were 100 m and 2 km, respectively.

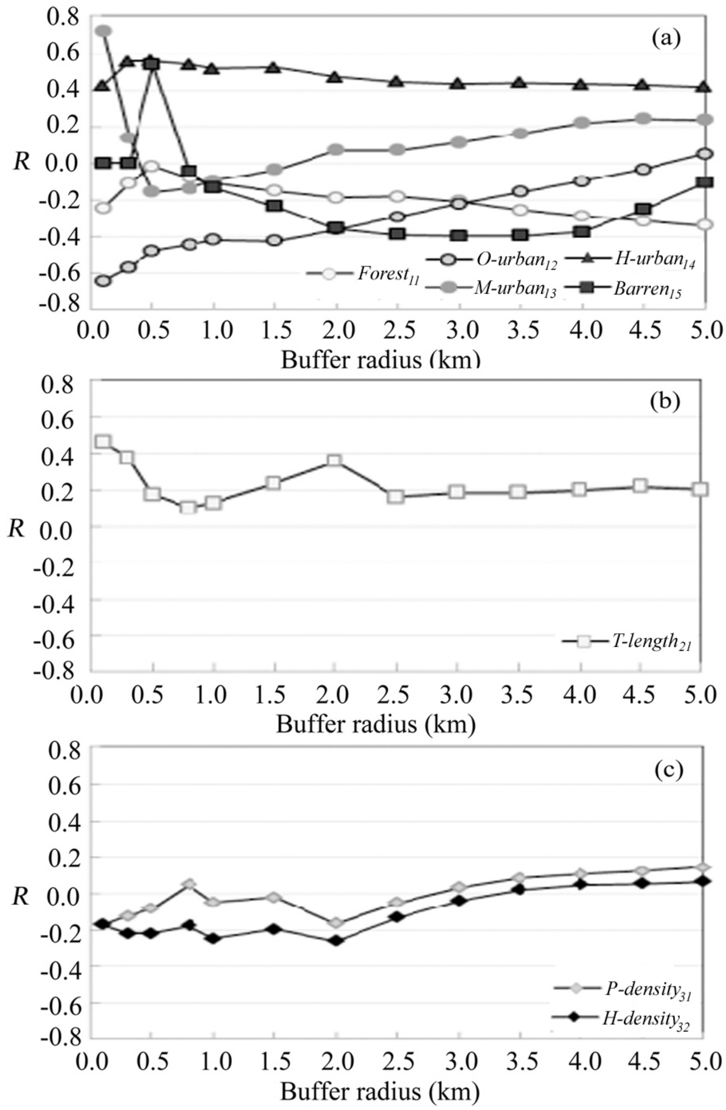

Figure 3. Correlation coefficients between characteristic variables and annual average $PM_{2.5}$ concentrations: (**a**) land use; (**b**) road traffic; (**c**) population and housing density.

Table 1. Statistics of "measured values" (mean (min, max), Units: as listed in Section 2.1).

Variables	Measured Values	Variables	Measured Values	Variables	Measured Values	Variables	Measured Values
$Forest_{11}$-5000	31.95 (0.16, 73.18)	$M\text{-}urban_{13}$-5000	21.07 (5.37, 45.72)	$Barren_{15}$-5000	0.64 (0.00, 3.48)	$P\text{-}density_{31}$-5000	574.35 (155.31, 1719.81)
$Forest_{11}$-4500	30.25 (0.09, 70.08)	$M\text{-}urban_{13}$-4500	22.09 (5.36, 46.08)	$Barren_{15}$-4500	0.49 (0.00, 2.19)	$P\text{-}density_{31}$-4500	568.81 (159.48, 1691.27)
$Forest_{11}$-4000	28.41 (0.10, 66.52)	$M\text{-}urban_{13}$-4000	22.48 (5.59, 44.89)	$Barren_{15}$-4000	0.40 (0.00, 2.25)	$P\text{-}density_{31}$-4000	543.95 (163.63, 1609.78)
$Forest_{11}$-3500	26.30 (0.05, 63.76)	$M\text{-}urban_{13}$-3500	22.47 (5.58, 42.55)	$Barren_{15}$-3500	0.37 (0.00, 2.54)	$P\text{-}density_{31}$-3500	554.16 (143.52, 1700.21)
$Forest_{11}$-3000	23.86 (0.05, 62.41)	$M\text{-}urban_{13}$-3000	22.71 (5.69, 39.13)	$Barren_{15}$-3000	0.33 (0.00, 2.34)	$P\text{-}density_{31}$-3000	706.87 (128.20, 2536.17)
$Forest_{11}$-2500	21.46 (0.05, 56.23)	$M\text{-}urban_{13}$-2500	22.93 (7.63, 41.70)	$Barren_{15}$-2500	0.36 (0.00, 2.22)	$P\text{-}density_{31}$-2500	552.94 (108.41, 1859.69)
$Forest_{11}$-2000	17.98 (0.00, 44.52)	$M\text{-}urban_{13}$-2000	23.65 (10.23, 43.70)	$Barren_{15}$-2000	0.36 (0.00, 2.07)	$P\text{-}density_{31}$-2000	686.97 (85.69, 2558.86)
$Forest_{11}$-1500	12.96 (0.00, 35.46)	$M\text{-}urban_{13}$-1500	24.25 (10.1, 42.78)	$Barren_{15}$-1500	0.41 (0.00, 2.70)	$P\text{-}density_{31}$-1500	609.19 (83.05, 1744.41)
$Forest_{11}$-1000	9.95 (0.00, 30.46)	$M\text{-}urban_{13}$-1000	24.57 (7.87, 44.43)	$Barren_{15}$-1000	0.31 (0.00, 3.26)	$P\text{-}density_{31}$-1000	687.06 (94.90, 1652.37)
$Forest_{11}$-800	8.36 (0.00, 26.11)	$M\text{-}urban_{13}$-800	25.56 (7.23, 47.13)	$Barren_{15}$-800	0.24 (0.00, 2.39)	$P\text{-}density_{31}$-800	635.68 (77.56, 1958.88)
$Forest_{11}$-500	5.80 (0.00, 18.82)	$M\text{-}urban_{13}$-500	25.17 (10.78, 50.59)	$Barren_{15}$-500	0.02 (0.00, 0.31)	$P\text{-}density_{31}$-500	589.57 (24.73, 1359.31)
$Forest_{11}$-300	3.48 (0.00, 19.32)	$M\text{-}urban_{13}$-300	23.55 (9.13, 41.76)	$Barren_{15}$-300	0.00 (0.00, 0.00)	$P\text{-}density_{31}$-300	555.93 (24.87, 1400.62)
$Forest_{11}$-100	0.75 (0.00, 10.45)	$M\text{-}urban_{13}$-100	21.79 (0.00, 48.78)	$Barren_{15}$-100	0.00 (0.00, 0.00)	$P\text{-}density_{31}$-100	509.19 (24.67, 1389.24)
$O\text{-}urban_{12}$-5000	32.22 (11.06, 55.30)	$H\text{-}urban_{14}$-5000	13.49 (3.05, 36.42)	$T\text{-}length_{21}$-5000	458.93 (202.46, 1104.14)	$H\text{-}density_{32}$-5000	201.90 (57.11, 669.74)
$O\text{-}urban_{12}$-4500	33.04 (12.87, 55.22)	$H\text{-}urban_{14}$-4500	14.12 (2.92, 38.63)	$T\text{-}length_{21}$-4500	390.49 (174.75, 977.80)	$H\text{-}density_{32}$-4500	197.99 (56.30, 640.30)
$O\text{-}urban_{12}$-4000	33.79 (12.48, 54.16)	$H\text{-}urban_{14}$-4000	14.91 (2.51, 41.99)	$T\text{-}length_{21}$-4000	306.33 (123.91, 765.44)	$H\text{-}density_{32}$-4000	187.03 (48.61, 592.81)
$O\text{-}urban_{12}$-3500	34.78 (10.95, 52.14)	$H\text{-}urban_{14}$-3500	16.07 (2.36, 45.92)	$T\text{-}length_{21}$-3500	238.98 (80.52, 605.11)	$H\text{-}density_{32}$-3500	188.16 (40.68, 618.35)
$O\text{-}urban_{12}$-3000	36.45 (10.60, 56.73)	$H\text{-}urban_{14}$-3000	16.65 (2.34, 49.67)	$T\text{-}length_{21}$-3000	188.69 (53.10, 489.08)	$H\text{-}density_{32}$-3000	240.04 (37.25, 913.73)
$O\text{-}urban_{12}$-2500	38.16 (10.45, 64.74)	$H\text{-}urban_{14}$-2500	17.09 (3.01, 55.39)	$T\text{-}length_{21}$-2500	131.09 (30.94, 328.16)	$H\text{-}density_{32}$-2500	181.78 (33.11, 645.51)
$O\text{-}urban_{12}$-2000	40.16 (10.92, 71.96)	$H\text{-}urban_{14}$-2000	17.85 (3.25, 59.95)	$T\text{-}length_{21}$-2000	88.05 (23.51, 214.96)	$H\text{-}density_{32}$-2000	225.16 (28.27, 872.09)
$O\text{-}urban_{12}$-1500	43.62 (11.07, 75.23)	$H\text{-}urban_{14}$-1500	18.77 (3.25, 68.21)	$T\text{-}length_{21}$-1500	52.19 (18.15, 128.37)	$H\text{-}density_{32}$-1500	183.94 (27.67, 565.37)
$O\text{-}urban_{12}$-1000	46.22 (6.75, 79.01)	$H\text{-}urban_{14}$-1000	18.95 (2.92, 77.30)	$T\text{-}length_{21}$-1000	23.27 (9.53, 60.66)	$H\text{-}density_{32}$-1000	211.81 (31.75, 560.28)
$O\text{-}urban_{12}$-800	47.20 (6.50, 77.84)	$H\text{-}urban_{14}$-800	18.64 (1.83, 79.98)	$T\text{-}length_{21}$-800	14.90 (5.76, 39.44)	$H\text{-}density_{32}$-800	183.35 (24.03, 448.72)
$O\text{-}urban_{12}$-500	51.72 (6.49, 77.28)	$H\text{-}urban_{14}$-500	17.29 (1.09, 76.29)	$T\text{-}length_{21}$-500	5.65 (1.46, 15.22)	$H\text{-}density_{32}$-500	186.44 (7.97, 450.35)
$O\text{-}urban_{12}$-300	56.43 (8.44, 88.25)	$H\text{-}urban_{14}$-300	16.55 (2.62, 72.68)	$T\text{-}length_{21}$-300	1.98 (0.22, 5.67)	$H\text{-}density_{32}$-300	176.91 (8.02, 460.55)
$O\text{-}urban_{12}$-100	62.56 (0.00, 100.00)	$H\text{-}urban_{14}$-100	14.90 (0.00, 69.78)	$T\text{-}length_{21}$-100	0.21 (0.00, 0.68)	$H\text{-}density_{32}$-100	169.69 (7.95, 456.81)
$D\text{-}road_{22}$	79.67 (0.18, 279.51)	$D\text{-}coast_{41}$	55.39 (1.38, 125.15)				

3.2. Performance Validation of LUR Models under Different Spatial Scales

Table 2 illustrates the $PM_{2.5}$ LUR models built both at ideal buffering radius (optimized spatial scale) and less correlated (non-optimized spatial scale) buffering radiuses, assisted with variables without spatial scale effects but had strong correlations. It can be observed that the LUR model based on variables' optimized spatial scale measured values obtained the best fitting result (adj. $R^2 = 0.78$). This was followed by models based on variables' measured values at other less correlated scales of 4 km (adj. $R^2 = 0.65$), 4.5 km (adj. $R^2 = 0.62$), 5 km and 3.5 km (adj. $R^2 = 0.61$), 500 m (adj. $R^2 = 0.51$), and 100 m (adj. $R^2 = 0.48$). Other LUR models had a comparatively lower fitting degree for adjusted R^2 ranging from 0.19 to 0.39 at scales from 1 km to 3 km. Moreover, the LUR models in Table 2 also obviously indicated the fluctuations of the predictive variables. Under smaller spatial scales the predictive variables favored medium- (*M-urban*$_{13}$) and high intensity urban ratios (*H-urban*$_{14}$), total road length (*T-length*$_{21}$), as well as distance to nearest road (*D-road*$_{22}$). The contributions of housing density (*P-density*$_{31}$), high intensity urban ratios (*H-density*$_{32}$), and distance to sea coast (*D-coast*$_{41}$) increased gradually with the increase in the spatial scales.

Table 2. Predictors and adjusted R^2 of LUR models for $PM_{2.5}$ concentration simulation.

Model ID	Spatial Scale	Model Predictors	Model R^2
1	Best scale	*M-urban*$_{13}$-100, *P-density*$_{31}$-100, *Forest*$_{11}$-5000	0.78
2	100 m	*M-urban*$_{13}$-100	0.48
3	300 m	*T-length*$_{21}$-300, *H-urban*$_{14}$-300, *D-road*$_{22}$	0.45
4	500 m	*T-lenegth*$_{21}$-500, *H-urban*$_{14}$-500, *D-road*$_{22}$	0.51
5	800 m	*T-length*$_{21}$-800, *H-urban*$_{14}$-800	0.39
6	1000 m	*H-urban*$_{14}$-1000	0.21
7	1500 m	*D-coast*$_{41}$, *O-urban*$_{12}$-1500, *P-density*$_{31}$-1500	0.19
8	2000 m	*H-density*$_{32}$-2000, *O-urban*$_{12}$-2000, *Forest*$_{11}$-2000	0.30
9	2500 m	*H-density*$_{32}$-2500, *H-urban*$_{14}$-2500	0.38
10	3000 m	*H-density*$_{32}$-3000, *H-urban*$_{14}$-3000	0.34
11	3500 m	*H-density*$_{32}$-3500, *D-coast*$_{41}$, *H-urban*$_{14}$-3500	0.61
12	4000 m	*H-density*$_{32}$-4000, *D-coast*$_{41}$, *H-urban*$_{14}$-4000	0.65
13	4500 m	*H-density*$_{32}$-4500, *D-coast*$_{41}$, *H-urban*$_{14}$-4500	0.62
14	5000 m	*H-density*$_{32}$-5000, *D-coast*$_{41}$, *H-urban*$_{14}$-5000	0.61

To avoid the col-linearity in the MLR process, residual analyses of the six LUR models with relative higher fitting degrees were also conducted in this study. The results in Figure 4 show that all standardized residuals were stochastically distributed, roughly falling at the horizontal zonal area ($|r| \leq 2$) without any potential trend. Additionally, the comparison of mean error rate (MER) and root mean squared error (RMSE) for LUR models of $PM_{2.5}$ simulation concentration in Table 3 further confirmed the reliability of LUR models, with MER under 20%. The model that was established by variables at the optimized spatial scale (Model 1) had the smallest MER of 11.84% and the RMSE value of 1.43.

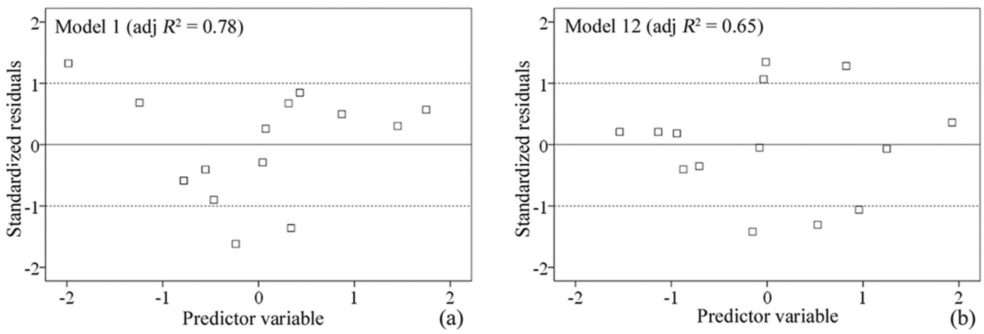

Figure 4. *Cont.*

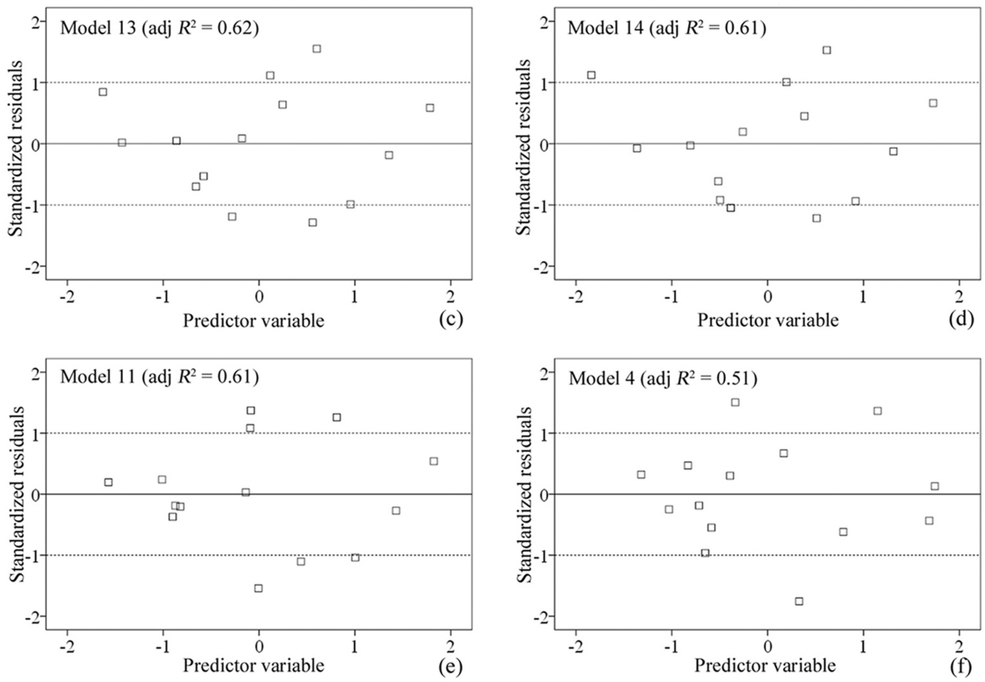

Figure 4. Standardized residual error map of LUR models: (**a**) Model 1; (**b**) Model 12; (**c**) Model 13; (**d**) Model 14; (**e**) Model 11; (**f**) Model 4.

Table 3. Comparison of mean error rate (MER) and root mean squared error (RMSE) for LUR models of $PM_{2.5}$ concentration simulation.

Model ID	MER (%) [1]	RMSE ($\mu g \cdot m^{-3}$)
1	11.84	1.43
2	17.22	2.65
3	16.73	2.45
4	16.78	1.97
5	19.93	3.13
6	28.26	4.16
7	28.37	4.35
8	27.32	3.87
9	19.30	3.26
10	26.32	3.69
11	14.37	1.72
12	15.03	1.80
13	15.58	1.87
14	13.21	1.58

[1] MER = |Observed concentration − Simulated concentration|/Observed concentration × 100%.

3.3. $PM_{2.5}$ Concentration Surfaces Mapped by LUR Models and Ordinary Kriging

As an implementation of LUR modeling, mapping performance is particularly important for correctly understanding the $PM_{2.5}$ pollution pattern of an area, as illustrated by Figure 5 in this study. Figure 5 shows that there was clearly different annual average $PM_{2.5}$ concentration surfaces for Houston produced by the LUR models with best- (i.e., Model 1), moderate- (i.e., Model 2), weakest (i.e., Model 7) adjusted R^2, as well as ordinary kriging. These differences mean the relative large biases of mapping results of Model 2, Model 7, and ordinary kriging based on the performance validation results of LUR models in Section 3.2. Specifically, for models 1 and 2 the higher annual mean $PM_{2.5}$ concentrations (i.e., >10 $\mu g \cdot m^{-3}$) were distributed in urban Harris County and the surrounding area, except that these high level $PM_{2.5}$ polluted areas in Model 2 were greater than those in Model 1.

However, on the other hand, the results of Model 7 disclosed that almost the entire annual average $PM_{2.5}$ concentrations in the Houston area were less than 9 μg·m^{-3}, which were inconsistent with the observed $PM_{2.5}$ concentration values from the regulatory monitoring sites. In addition, results from *Levene*'s test and *F* test with *p* values less than 0.05 in this study echoed these significant differences demonstrated in Figure 5, which indirectly confirms the reliability of the LUR model built at the optimized spatial scale.

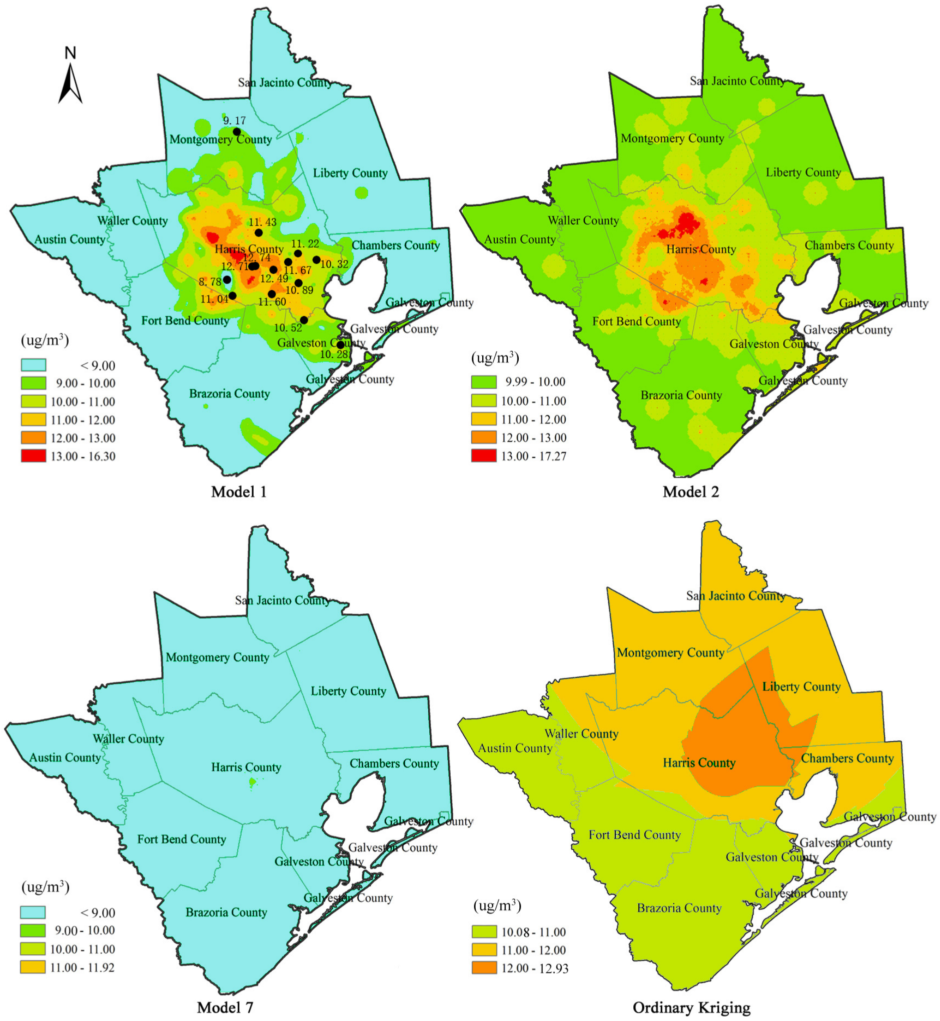

Figure 5. $PM_{2.5}$ concentration surfaces of Houston based on LUR models and ordinary kriging.

4. Discussion

Using LUR-based $PM_{2.5}$ concentration simulation in Houston, US as a case study, this study explored for the first time the influences of spatial scales of characteristic variables on LUR modeling by employing the idea of statistically optimized analysis. We found that the accuracy of LUR models changed significantly with different spatial scales. The model based on the optimized spatial scale achieved a much higher "fitting degree" (adj. R^2 = 0.78) than at any other scales (adj. R^2 range from 0.19 to 0.65), which performed better than previous similar study [45]. However, further improvements are needed to broaden the applicability of these research results.

4.1. Results Analysis

Our results demonstrated that land use and road traffic were more related with annual average $PM_{2.5}$ concentration than population distribution and distance to sea coast. Medium-, high intensity-, and open space urban had the strongest land use correlations; for road traffic, total road length within the 100 m buffer and distance to nearest road were the main correlations. The reason may lie in the fact that $PM_{2.5}$ in Houston predominately comes from industrial production and transportation emissions [46,47]. Places with greater urbanization and intensive road traffic generally experienced more serious $PM_{2.5}$ pollution, resulting in the stronger correlations between medium-, high intensity urban and $PM_{2.5}$ concentrations. However, the impact of population and distance to sea coast was much weaker. $PM_{2.5}$ from transport emissions diffused slowly and accumulated near roads because of low-lying terrain and building obstructions which increased the correlations between total road lengths, distance to nearest road, and annual average $PM_{2.5}$ concentrations. This exemplifies why road traffic is the major factor affecting urban $PM_{2.5}$ pollution worldwide.

Variations of correlation coefficients between the characteristic variables and annual average $PM_{2.5}$ concentration under different spatial scales identified how the spatial scale setting can significantly influence correlations. At different spatial scales the correlation coefficients changed in both direction and value. Optimized spatial scales differed from different variables, including 5 km for forest, 100 m for open space urban and medium intensity urban, 500 m for high intensity urban, 3 km for barren land, 100 m for total road length, and 2 km for housing density. The variability reflects how diverse geographical factors have different influencing radiuses for $PM_{2.5}$ pollution. For example, road traffic is an important emission source of $PM_{2.5}$ and sites closer to road will be exposed to more serious $PM_{2.5}$ pollution levels that cause a stronger correlation at smaller spatial scales. Similarly, open space urban, medium-, and high intensity urban land use space would have less impact on $PM_{2.5}$, leading to a smaller influencing radius. However, since only a large amount of forests can significantly reduce $PM_{2.5}$ pollution dispersion, they may influence $PM_{2.5}$ pollution at a larger spatial scale (i.e., the optimized spatial scale of forest area ratio in this study was 5 km).

Comparatively, the LUR model in our study based on variables at the optimized spatial scales achieved an impressive R^2 (0.78), mean error rate (11.84%), and RMSE (1.43). These results were not only better than those based on variables at other spatial scales in this study (i.e., at the non-optimized spatial scale, the fitting degree ranged between 0.19 and 0.65; maximum mean error rate and RMSE reached to 28.37% and 4.35, respectively), but they also significantly outperform some previous reported adjusted R^2 of $PM_{2.5}$ LUR models for New York, El Paso, and California. The values of adjusted R^2 for those studies were 0.64, 0.49, and 0.65, respectively [26,48,49]. In addition, the annual average $PM_{2.5}$ concentration maps, which were separately produced by the LUR models with the best-, moderate-, weakest adjusted R^2, showed that the LUR model with weakest fitting degree could not simulate the distribution of $PM_{2.5}$ concentrations well, while models with moderate and best fitting degree both showed better simulation results than ordinary kriging with wider concentration scope. This result actually again confirms the significance of the identification of the optimized spatial scale in LUR modeling, which means the $PM_{2.5}$ distribution disclosed by LUR Model 1 with the best adjusted R^2 was more similar with the true scenario.

4.2. Limitations

In Houston, $PM_{2.5}$ primarily originates from industrial and vehicle emissions, occasional biomass burning, and floating dust. Though this study emphasized several factors (land use, road traffic, population distribution, distance to sea coast, and other geographical features) during LUR modeling and achieved a surrounding annual average $PM_{2.5}$ concentration simulation, further improvements on the coverage of geographical factors are still required. For example, variables such as real industrial emissions and urban morphologies of microenvironments (e.g., urban ecological landscape index, street canyon, vegetation index, etc.) can be incorporated into $PM_{2.5}$ LUR modeling [26,50,51]. These variables may provide additional representative descriptions of $PM_{2.5}$ pollution. Additionally, a recent study shows $PM_{2.5}$ emissions from unscheduled maintenance, startup, or shutdown activities continue to increase in recent years [52]. LUR models have limited ability to capture such emission events from industries in Houston.

Based on previous research findings, this study took multiple spatial scales of variables into account. While analyses proved the feasibility of deriving an optimized spatial scale in a statistical manner under the currently unclear physical-chemical dispersion mechanism of $PM_{2.5}$, isometric discrete spatial scales might fail to continuously identify the spatial scale dependence of characteristic variables because it is a relatively crude scheme. Additionally, the spatial scale range from 100 m to 5 km, though it covered the influence radiuses of most variables, may not be able to fully reflect the relationship between some variables with a scope for very small or very large influence (e.g., road traffic, $PM_{2.5}$ pollution source, distance to chimney, forest, etc.) on $PM_{2.5}$ pollution [53]. Therefore, future spatial scale dependence analyses could expand or narrow the spatial scale range of the current study and consider differences in variables' physical and chemical dispersion mechanisms of $PM_{2.5}$ pollution with more abundant data.

This study applied MLR to establish the $PM_{2.5}$ LUR model. Although MLR is the most popular LUR regression model with reliable simulation effect [54], it assumes that variables make the same spatial contributions to $PM_{2.5}$ pollution in any location in the modeling area. However, under real situations, geographical factors have different levels of spatial heterogeneity (except for spatial correlation). Therefore, future LUR model applications can consider adding spatial weights (e.g., establish a geographical weighted model) into existing models to enhance the simulation accuracy of LUR. Meanwhile, the semi-parametric regression model, which takes linear and non-parametric variables into consideration at the same time, might also be a promising way to improve the accuracy of LUR [55].

In addition, the training sample size might be another important factor influencing the accuracy of LUR models. Although the number of monitoring sites used as training samples in the previously reported $PM_{2.5}$ LUR models ranged from 13 to more than 100 and the surrounding simulation results also had been achieved under few monitoring data [20,53], the results in this study have to be cautiously explained due to the limited monitoring sites employed. Further validation work in regards to considerations surrounding area through the use of many more sampling sites will greatly promote the exploration of the relationship between monitoring sites and the LUR model's accuracy, which is a critical problem having not been fully considered in LUR modeling field so far.

5. Conclusions

This study represents the first time that a systematic exploration of the influence of spatial scales of characteristic variables on LUR-based $PM_{2.5}$ concentration estimation modeling was carried out in a statistical manner. It used LUR-based $PM_{2.5}$ concentration estimation modeling in Houston, US as an example to illustrate how the challenge of the spatial scale clarity can be investigated. Results indicate that statistical based identification of optimized spatial scales of characteristic variables is necessary to ensure the performance of LUR models in mapping $PM_{2.5}$ distribution without current clearly understood physical-chemical dispersion mechanisms. LUR models at optimized spatial scales were observed to perform better than unified spatial scales. More importantly, this study provides a scientific basis for the spatial scale selection of characteristic variables in future LUR based air pollution mapping.

Acknowledgments: This research work was supported by the National Key Research and Development Program (2016YFC0206205), the National Natural Science Foundation of China (No. 41201384), the Key Laboratory of Geo-informatics of State Bureau of Surveying and Mapping (No. JC201503), the Open Fund of University Innovation Platform, Hunan under Grant 15K132, and the National Geographic Conditions Monitoring of Hunan (HNGQJC201503). Program for the 2016 Young Academic and Technological Leaders of NASG, funded by Key Laboratory of Geo-informatics of NASG. Thanks Guoqing Sun for his contributions in the paper revision.

Author Contributions: Liang Zhai and Bin Zou conceived and designed the experiments; Xin Fang and Yanqing Luo performed the experiments; Liang Zhai, Neng Wan, and Bin Zou analyzed the data; Shuang Li contributed reagents/materials/analysis tools; Bin Zou and Yanqing Luo wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dominici, F.; McDermott, A.; Zeger, S.L.; Samet, J.M. National maps of the effects of particulate matter on mortality: Exploring geographical variation. *Environ. Health Perspect.* **2003**, *111*, 39–44. [CrossRef] [PubMed]
2. Dominici, F.; Peng, R.D.; Bell, M.L.; Pham, L.; McDermott, A.; Zeger, S.L.; Samet, J.M. Fine particulate air pollution and hospital admission for cardiovascular and respiratory and respiratory diseases. *J. Am. Med. Assoc.* **2006**, *295*, 1127–1134. [CrossRef] [PubMed]
3. Kam, W.; Delfino, R.J.; Schauer, J.J.; Sioutas, C. A comparative assessment of $PM_{2.5}$ exposures in light-rail, subway, freeway, and surface street environments in Los Angeles and estimated lung cancer risk. *Environ. Sci.: Processes Impacts* **2013**, *15*, 234–243.
4. Dergham, M.; Billet, S.; Verdin, A.; Courcot, D.; Cazier, F.; Shirali, P.; Garçon, G. Toxicological impact of air pollution particulate matter ($PM_{2.5}$) collected under urban, industrial or rural influence: Occurrence of Oxidative stress and inflammatory reaction in BEAS-2B human bronchial epithelial cells (corrected version). *Adv. Mater. Res.* **2011**, *324*, 489–492. [CrossRef]
5. NASA (2010) New Map Offers a Global View of Health-Sapping Air Pollution. Available online: http://www.nasa.gov/topics/earth/features/health-sapping.html (accessed on 11 January 2013).
6. van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from Satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 848–855. [CrossRef] [PubMed]
7. The Lancet (2012) Global Burden of Disease Study 2010. Available online: http://www.thelancet.com/themed/global-burden-of-disease (accessed on 26 April 2013).
8. Liu, S.; Zhang, K. Fine particulate matter components and mortality in Greater Houston: Did the risk reduce from 2000 to 2011? *Sci. Total Environ.* **2015**, *538*, 162–168. [CrossRef] [PubMed]
9. Liu, S.; Ganduglia, C.M.; Li, X.; Delclos, G.L.; Franzini, L.; Zhang, K. Short-term Associations of fine particulate matter components and emergency hospital admissions among a privately insured population in greater Houston. *Atmos. Environ.* **2016**, *147*, 369–375. [CrossRef]
10. Liu, S.; Ganduglia, C.M.; Li, X.; Delclos, G.L.; Franzini, L.; Zhang, K. Fine particulate matter components and emergency department visits among a privately insured population in greater Houston. *Sci. Total Environ.* **2016**, *566–567*, 521–527. [CrossRef] [PubMed]
11. Zou, B.; Pu, Q.; Bilal, M.; Weng, Q.; Zhai, L.; Nichol, J.E. High-resolution satellite mapping of fine particulates based on geographically weighted regression. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 495–499. [CrossRef]
12. Fang, X.; Zou, B.; Liu, X.; Troy, S.; Zhai, L. Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. *Remote Sens. Environ.* **2016**, *186*, 152–163. [CrossRef]
13. Hoek, G.; Beelen, R.; Hoogh, K.D.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578. [CrossRef]
14. Zou, B.; Wilson, J.G.; Zhan, F.B.; Zeng, Y. Air pollution exposure assessment methods utilized in epidemiological studies. *J. Environ. Monit.* **2009**, *11*, 475–490. [CrossRef] [PubMed]
15. Lee, H.J.; Liu, Y.; Coull, B.A.; Schwartz, J.; Koutrakis, P. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmos. Chem. Phys.* **2011**, *11*, 7991–8002. [CrossRef]
16. Zou, B.; Luo, Y.; Wan, N.; Zheng, Z.; Sternberg, T.; Liao, Y. Performance comparison of LUR and OK in $PM_{2.5}$ concentration mapping: A multidimensional perspective. *Sci. Rep.* **2015**, *5*. [CrossRef] [PubMed]
17. Zou, B.; Wang, M.; Wan, N.; Wilson, J.G.; Fang, X.; Tang, Y. Spatial modeling of $PM_{2.5}$ concentrations with a multifactorial radial basis function neural network. *Environ. Sci. Pollut. Res.* **2015**, *22*, 10395–10404. [CrossRef] [PubMed]
18. Gilliland, F.; Avol, E.; Kinney, P.; RJerrett, M.; Dvonch, T.; Lurmann, F.; Buckley, T.; Breysse, P.; Keeler, G.; de Villiers, T.; et al. Air pollution exposure assessment for epidemiologic studies of pregnant women and children: Lessons learned from the centers for children's environmental health and disease prevention research. *Environ. Health Perspect.* **2005**, *113*, 1447–1454. [CrossRef] [PubMed]
19. Briggs, D.J.; Collins, S.; Elliot, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pryl, K.; Reeuwijk, H.V.; Smallbone, K.; der Veen, A.V.; et al. Mapping urban air pollution using GIS: A regression-based approach. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 699–718. [CrossRef]

20. Henderson, S.B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **2007**, *41*, 2422–2428. [CrossRef] [PubMed]
21. Mavko, M.E.; Tang, B.; George, L.A. A sub-neighborhood scale land use regression model for predicting NO_2. *Sci. Total Environ.* **2008**, *2*, 68–74. [CrossRef] [PubMed]
22. Chen, L.; Bai, Z.; Su, D.; You, Y.; You, Y.; Li, H.; Quan, L. Application of land use regression to simulate ambient air PM_{10} and NO_2 concentration in Tianjin City. *China Environ. Sci.* **2009**, *29*, 685–691.
23. de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dėdelė, A.; et al. Development of land Use regression models for particle composition in twenty study areas in Europe. *Environ. Sci. Technol.* **2013**, *47*, 5778–5786. [CrossRef] [PubMed]
24. Meng, X.; Chen, L.; Cai, J.; Zou, B.; Wu, C.; Fu, Q.; Zhang, Y.; Liu, Y.; Kan, H. A land use regression model for estimating the NO_2 concentration in Shanghai, China. *Environ. Res.* **2015**, *137*, 308–315. [CrossRef] [PubMed]
25. Hochadel, M.; Heinrich, J.; Gehring, U.; Morgenstern, V.; Kuhlbusch, T.; Link, E.; Wichmann, H.E.; Kramer, U. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmos. Environ.* **2006**, *40*, 542–553. [CrossRef]
26. Ross, Z.; Jerrett, M.; Ito, K.; Tempalski, B.; Thurston, G.D. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmos. Environ.* **2007**, *41*, 2255–2269. [CrossRef]
27. Ho, C.; Chan, C.; Cho, C.; Lin, H.; Lee, J.; Wu, C. Land use regression modeling with vertical distribution measurements for fine particulate matter and elements in an urban area. *Atmos. Environ.* **2015**, *104*, 256–263. [CrossRef]
28. Lee, J.; Wu, C.; Hoek, G.; de Hoogh, K.; Beelen, R.; Brunekreef, B.; Chan, C. LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction. *Sci. Total Environ.* **2015**, *514*, 178–184. [CrossRef] [PubMed]
29. Wan, N.; Zou, B.; Strenberg, T. A 3-step floating catchment area method for analyzing spatial access to health services. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1073–1089. [CrossRef]
30. Wan, N.; Zhan, F.B.; Zou, B.; Chow, E. A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Appl. Geogr.* **2012**, *32*, 291–299. [CrossRef]
31. Zou, B.; Peng, F.; Wan, N.; Wilson, J.G.; Xiong, Y. Sulfur dioxide exposure and environmental justice: A multi-scale and source-specific perspective. *Atmos. Pollut. Res.* **2014**, *5*, 491–499. [CrossRef]
32. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved $PM_{2.5}$ exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [CrossRef]
33. Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van Donkelaar, A.; Lamsal, L.; Martin, R.; Jerrett, M.; et al. Creating national air pollution models for population exposure assessment in Canada. *Environ. Health Perspect.* **2011**, *119*, 1123–1129. [CrossRef] [PubMed]
34. Tunno, B.J.; Michanowicz, D.R.; Shmool, J.L.C.; Kinnee, E.; Cambal, L.; Tripathy, S.; Gillooly, S.; Roper, C.; Chubb, L.; Clougherty, J.E. Spatial variation in inversion-focused vs 24-h integrated samples of $PM_{2.5}$ and black carbon across Pittsburgh, PA. *Expo. Sci. Environ. Epidemiol.* **2016**, *26*, 365–376. [CrossRef] [PubMed]
35. Xu, G.; Jiao, L.M.; Zhao, S.L.; Yuan, M.; Li, X.M.; Han, Y.Y.; Zhang, B.; Dong, T. Examining the impacts of land use on air quality from a spatio-temporal perspective in Wuhan, China. *Atmosphere* **2016**, *7*, 62. [CrossRef]
36. U.S. EPA (2011). Available online: http://www.epa.gov/airquality/airdata/ad_data_daily.html (accessed on 12 January 2013).
37. USGS (2011). Available online: http://seamless.usgs.gov (accessed on 13 January 2013).
38. ESRI. Available online: http://www.openstreetmap.org/#map=5/51.500/-0.100 (accessed on 22 December 2016).
39. U.S. census bureau (2010). Available online: http://www.census.gov/main/www/access.html (accessed on 10 February 2013).
40. Mao, L.; Qiu, Y.; Kusano, C.; Xu, X. Predicting regional space-time variation of $PM_{2.5}$ with land-use regression model and MODIS data. *Environ. Sci. Pollut. Res.* **2012**, *19*, 128–138. [CrossRef] [PubMed]
41. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dėdelė, A.; Dons, E.; de Nazelle, A.; et al. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environ. Sci. Technol.* **2012**, *46*, 11195–11205. [CrossRef] [PubMed]

42. Pearson, K. On lines and planes of closest fit to systems of points is space. *Philos. Mag. J. Sci.* **1901**, *62*, 559–572. [CrossRef]
43. Levene, H. Robust tests for equality of variances. In *Contributions to Probability and Statistics*; Olkin, I., Ed.; Stanford University Press: Palo Alto, CA, USA, 1960; pp. 278–291.
44. Duncan, D.B. Multiple range and multiple F-test. *Biometrics* **1995**, *11*, 1–42. [CrossRef]
45. Zhang, K.; Larson, T.; Gassett, A.; Szpiro, A.; Daviglus, M.; Burke, G.; Kaufman, J.; Adar, S. Characterizing spatial patterns of airborne coarse particulate ($PM_{10-2.5}$) mass and chemical components in three cities: The Multi-Ethnic Study of Atherosclerosis. *Environ. Health Perspect.* **2014**, *122*, 823–830.
46. Fraser, M.P.; Yue, Z.; Buzcu, B. Source apportionment of fine particulate matter in Houston, TX, using organic molecular markers. *Atmos. Environ.* **2003**, *37*, 2117–2123. [CrossRef]
47. Zou, B.; Zheng, Z.; Wan, N.; Qiu, Y.H.; Wilson, J.G. An optimized spatial proximity model for fine particulate matter air pollution exposure assessment in areas of sparse monitoring. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1–21. [CrossRef]
48. Olvera, H.A.; Garcia, M.; Li, W.; Yang, H.; Amaya, M.A.; Myers, O.; Burchiel, S.W.; Berwick, M.; Pingitore, N.E., Jr. Principal component analysis optimization of a $PM_{2.5}$ land use regression model with small monitoring network. *Sci. Total Environ.* **2012**, *425*, 27–34. [CrossRef] [PubMed]
49. Beckerman, B.S.; Jerrett, M.; Martin, R.V.; van Donkelaar, A.; Ross, Z.; Burnett, R.T. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. *Atmos. Environ.* **2013**, *77*, 172–177. [CrossRef]
50. Yu, H.; Wang, C.; Liu, M.; Kuo, Y. Estimation of fine particulate matter in Taipei using land use regression and bayesian maximum entropy methods. *Int. J. Environ. Res. Public Health* **2011**, *8*, 2153–2169. [CrossRef] [PubMed]
51. Xu, S.; Zou, B.; Pu, Q.; Guo, Y. Impact Analysis of Land Use/Cover on Air Pollution. *J. Geogr. Sci.* **2015**, *3*, 287–297. (In Chinese)
52. Luong, C.; Zhang, K. An Assessment of Emissions Events Trends within the Greater Houston Area during 2003-2013. *Air Qual., Atmos. Health* **2016**. [CrossRef]
53. Chen, C.; Wu, C.; Yu, H.; Chan, C.; Cheng, T. Spatiotemporal modeling with temporal-invariant variogram subgroups to estimate fine particulate matter $PM_{2.5}$ concentrations. *Atmos. Environ.* **2012**, *54*, 1–8. [CrossRef]
54. Ryan, P.H.; LeMasters, G.K. A review of land-use regression models for characterizing intra-urban air pollution exposure. *Inhalation Toxicol.* **2007**, *19*, 127–133. [CrossRef] [PubMed]
55. Szyszkowicz, M.; Mahmud, M.; Tremblay, N. A semi-parametric regression model to estimate variability of NO_2. *Environ. Pollut.* **2013**, *2*, 46–50. [CrossRef]