Article

Development of Land Use Regression Model for Seasonal Variation of Nitrogen Dioxide (NO$_2$) in Lahore, Pakistan

Saimar Pervez [ID], Ryuta Maruyama, Ayesha Riaz and Satoshi Nakai *[ID]

Department of Artificial Environment, Graduate School of Environment and Information Sciences, Yokohama National University, Yokohama 240-8501, Japan; saimar-pervez-pz@ynu.jp or saim3035@gmail.com (S.P.); maruyama-ryuta-pw@ynu.jp (R.M.); ayesha.riaz.env@gmail.com (A.R.)
* Correspondence: nakai-satoshi-dc@ynu.ac.jp; Tel.: +81-45-339-4364

Abstract: Ambient air pollution and its exposure has been a worldwide issue and can increase the possibility of health risks especially in urban areas of developing countries having the mixture of different air pollution sources. With the increase in population, industrial development and economic prosperity, air pollution is one of the biggest concerns in Pakistan after the occurrence of recent smog episodes. The purpose of this study was to develop a land use regression (LUR) model to provide a better understanding of air exposure and to depict the spatial patterns of air pollutants within the city. Land use regression model was developed for Lahore city, Pakistan using the average seasonal concentration of NO$_2$ and considering 22 potential predictor variables including road network, land use classification and local specific variable. Adjusted explained variance of the LUR models was highest for post-monsoon (77%), followed by monsoon (71%) and was lowest for pre-monsoon (70%). This is the first study conducted in Pakistan to explore the applicability of LUR model and hence will offer the application in other cities. The results of this study would also provide help in promoting epidemiological research in future.

Keywords: ambient air pollution; exposure assessment; Pakistan; land use regression; vehicle workshops

1. Introduction

The increase in number of road traffic and rapid growth of urbanization pose a great health hazard for the surrounding environment and public health. Ambient air pollution is an important environmental issue [1]. Based on global burden of death (GBD), 4.9 million deaths are attributed in the world because of air pollution exposure [2]. Air pollution is one of the biggest concerns in the modern era because of improvement in the lifestyle, which requires more energy and exploration of resources, putting pressure on the generation of toxic air pollutants in the atmosphere. The emission of these air pollutants affect both climate and human health [3,4]. Studies have mentioned the effects of air pollution on human health, such as cardiovascular, respiratory and chronic diseases [5,6]. One of the Swedish cohort studies reports that exposure to long term air pollution may cause diabetes [7]. Poor air quality is a serious issue in developing countries because of overpopulation, urbanization and industrialization [8].

Pakistan is a South Asian country having a population crossing the figure of 200 million [9]. The rapid increase in the population and unplanned urbanization with the recent development in the industrial units, has worsened the condition of ambient air in the country [10,11]. Transportation is another source of air pollution emitting 25 times more carbon monoxide (CO) and carbon dioxide (CO$_2$), and 3.5 times higher sulfur dioxide (SO$_2$) as compared to the automobiles in the United States [12]. Pakistan is the second, among the top 10 polluted countries in the world accounting 22,000 premature deaths and 163,432 disability-adjusted life years (DALYS) lost [13]. A study conducted in the cities of Pakistan (Islamabad, Lahore, Rawalpindi) reported the levels of nitrogen oxides (NO$_x$) and particulate matter (PM$_{10}$) higher than WHO guidelines [14]. Pakistan Environmental

Protection Agency (Pak-EPA) has monitored the level of NO_2 in different cities of Pakistan, and estimated that maximum and minimum concentrations were 37.02 ppb and 14.61 ppb in Karachi and Islamabad, whereas, another study found that the maximum and minimum concentrations were 37.46 ppb and 2.48 ppb, respectively [14,15]. Urban air pollution costs the country a loss of about Rs. 65 billion, from total annual loss to the environmental damages which is Rs. 365 billion [13]. The financial loss occurs from morbidity and mortality linked with cardiovascular and respiratory diseases, lower respiratory illness (LRI) in children, however, if air pollution related problems are taken into consideration on education, malnutrition and earnings, the financial loss could be higher [16].

Road traffic is classified as one of the reasons behind the deterioration of air quality in urban areas [17]. Increase in the number of vehicles not only causes traffic congestion and greenhouse gas emissions, but imposes significant health impacts, and is a source of tropospheric ozone formation [18]. Different air pollutants are present in the atmosphere of urban environment like particulate matter (PM), nitrogen dioxide (NO_2), carbon monoxide and dioxide (CO and CO_2) and ozone (O_3), but the pollutant which correlates well with traffic densities, and an important photochemical oxidant is nitrogen dioxide [19,20]. Nitrogen dioxide (NO_2) has a major role for the generation of secondary air pollutants (SAP) and its concentration is correlated with photochemical smog, acid deposition and ozone variations [21,22]. After particulate matter, NO_2 is the second most abundant and dangerous air pollutant in Pakistan [14], therefore it is necessary to have the knowledge of different types of pollution sources and the factors affecting the concentration of pollutant within the city.

Different models and techniques are available, and have been used to access and model the air pollution concentration, such as dispersion models (DM), chemistry transport model (CTM) and other techniques like inverse distance weighting (IDW), ordinary kriging (OK), machine learning (ML), but the problem lies with the application of these models and techniques, is the high demand of data requirement, costs and the complexity [23–25]. The issue lies with the applicability of interpolation techniques is the assumption that variation in the pollution is dependent on the distance between the sites, which may lead to error in estimating pollution [26]. In contrast, land use regression (LUR) model has gained the attention as an easy and effective approach, to provide spatial distributions of air pollutants at intra-urban scale built on a specific number of monitoring sites and predictor variables values, gathered by geographic information system (GIS), [27,28], providing a reliable solution for air pollution exposure assessment for a developing country like Pakistan [29].

LUR is a statistical method of air pollution modeling. It is commonly used to estimate variations in air pollutant concentrations for population exposure assessment. The technique links spatially heterogeneous air quality measurements with geospatial predictors. LUR models provide a comparatively robust method for spatial prediction, while having a lower sampling effort compared to geo-statistical models and a lower data requirement than dispersion models [30].

LUR model incorporates air pollutants data, geographic predictors such as land use, population density and road traffic network data around the monitoring points and after multiple linear regression, it provides spatial annual or seasonal pollution level, at unmonitored locations [31]. The model provides a simple and cost-effective method for air pollution exposure assessment at regional and intra-urban level by substituting expensive dispersion models [30,32].

Recently, LUR model is mostly applied in developed countries in North America and Europe, and the exposure assessment of different air pollutants have been successfully conducted [33–36]. Still, developed countries remain the focus of conducting air pollution studies related to exposure assessment and its health effects [37]. Therefore, it is necessary to have studies related to exposure from air pollution, which can be helpful for epidemiological evidences in developing country conditions [38], having air pollution-related health problems.

Challenges and constraints in the development of LUR models in these situations, includes inadequate availability of GIS data, emissions from air pollution sources not well interrelated and deficiency of routine monitoring concentration data. Pollutant's concentration data gained from the national monitoring locations are used to characterize the air pollution of the entire city which would lead to evaluation error in public air pollution exposure [39]. It would be meaningful to perform LUR studies to explore the performance of model, thus an efficient and economical air pollutant concentration model can be obtained.

To address the need of air pollution exposure assessment, this study will explore the applicability of LUR model to predict spatial variation of NO_2 for Lahore, Pakistan in which emission sources include local industries, household fuel use and automobiles. This would not only provide the evidence of developing LUR models for different air pollutants in Pakistan, but also offer important application of exposure assessment with high representativeness. This study will provide a basic systematic LUR method to promote the wider use in other cities of Pakistan.

The aim of the present study was to develop an LUR model based on seasonal variation (pre-monsoon (April, May and June), monsoon (July, August), post-monsoon (September, October and November)) of NO_2 concentration for Lahore city due to the lack of exposure data. To have the knowledge of different pollution variables effect on the concentration of pollutant for each season, and to depict its spatial distribution, LUR was applied in Pakistan, which will also provide the long-term epidemiological studies in air pollution in future. It is necessary to compare the three period models for better understanding of local emission sources and the effect of different potential predictor variables for each season.

2. Materials and Methods

2.1. Study Area

Lahore is the second largest city in Pakistan after Karachi and known as the capital of Punjab province. It has an area of 1772 km^2 and population is 11.13 million based on 2017 census [9]. Lahore is at 74°19′45.75″ longitude, 31°34′55.36″ latitude, at an elevation of 217 m above sea level with a semi-arid climate and yearly precipitation of 628.8 mm. The average humidity is 39.8% and the main wind direction is north. The lowest and highest temperature in the city is between 19.8 and 40.4 °C with a yearly average of 30.16 °C [15]. The country's oldest and longest road known as Grand Trunk (GT) road passes through the city, considered a source of pollution and causing higher levels of air pollutants throughout the year. The study area and sampling locations are shown (Figure 1).

2.2. Air Pollution Data

The seasonal average concentration data of nitrogen dioxide (NO_2) were obtained from a previous study [40]. Fifteen field surveys were conducted for NO_2 on the periodic basis and the samples were collected, analyzed and the results were managed. The concentration of NO_2 was estimated using the diffusion tubes, and the sampling time was bi-weekly, recording the start and end time of the exposure. The sampling campaign was conducted during the year of 2006 from April to November, dividing the sampling work into three phases (pre-monsoon, monsoon and post-monsoon). From the available data, 18 different monitoring stations and their respective concentration data were gathered (Table S1).
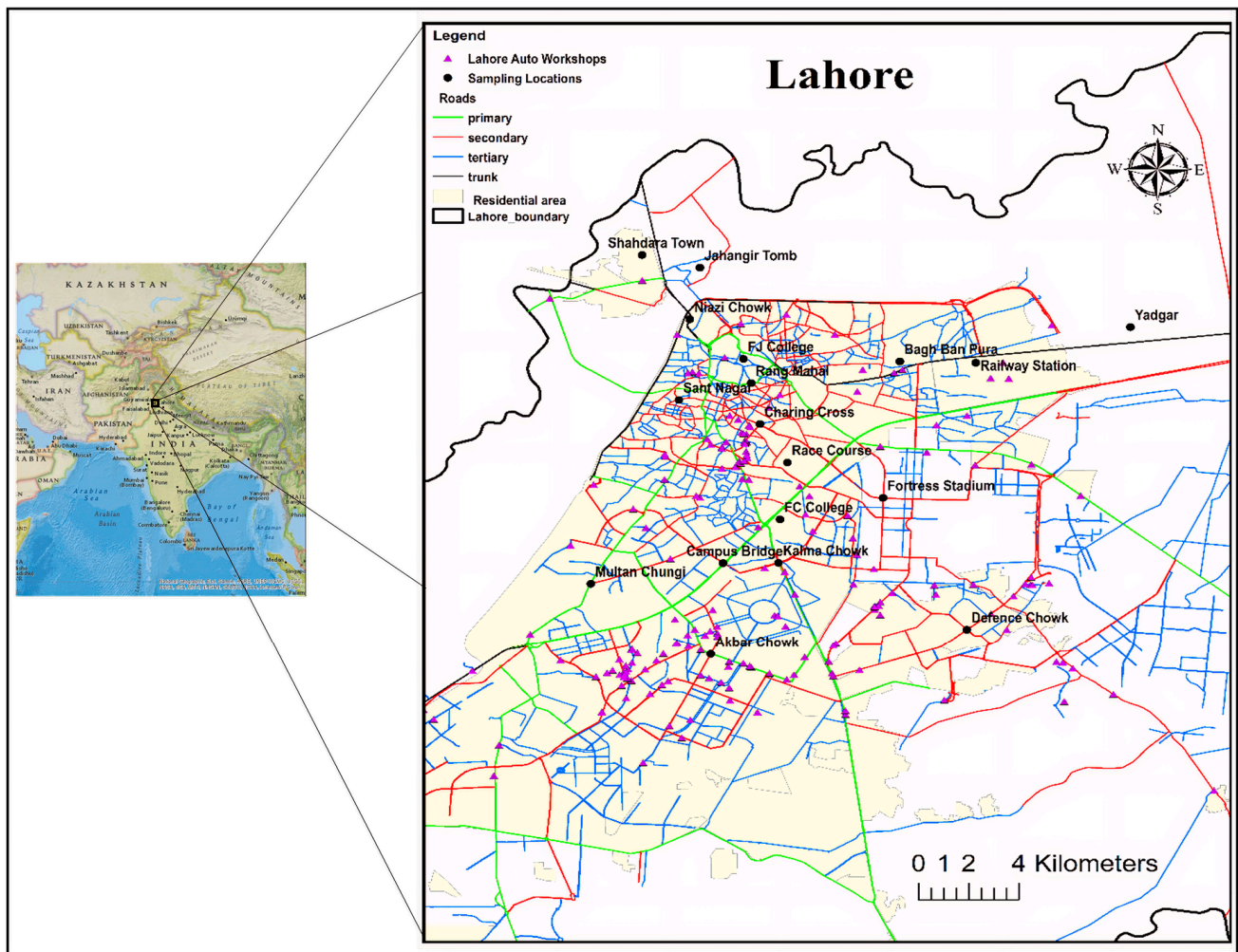
Figure 1. Location of the study area.

2.3. Potential Predictor Variables

For the development of LUR models, 22 potentially predictor variables were selected. The selected predictor variables were grouped into 4 main categories: road length, land use, distance variable, geographic location of monitoring stations and divided into 18 sub-categories (Table 1). Road length describes the length of different types of roads (primary, secondary, tertiary and trunk) and the values were calculated in the buffers of 25, 50, 100, 300, 500 and 1000 m, whereas land use represents the 9 types of land use classification (residential, commercial, educational area, etc.), and the area of each specific classification was calculated in the buffers of 100, 300, 500 and 1000 m around the monitoring stations. Further, the distance variable describes the distance of each station to the nearby local specific variable (vehicle maintenance workshop). Finally, the geographic information includes the elevation of monitoring stations in meters above the sea level and longitude and latitude, respectively.

Table 1. Brief description of potential predictor variables and 'a priori' definitions.

Variable Name (Unit)	Variable Type	Symbol Used	Buffer Size (m)	Input Rationale	A Priori [a]
road length (m)	Length of:		25, 50, 100, 300, 500, 1000	Road length is directly associated with traffic density, causing pollutant emission from traffic	
	Primary road	PR			+
	Secondary road	SR			+
	Tertiary road	TR			+
	Residential road	RR			+
	Trunk road	TRR			+
land use (m²)	Area of:		100, 300, 500, 1000	Different classification of land use generates different variety of pollutants or transportation activities	
	Residential area	LURA			+
	Commercial area	LUCA			+
	Educational area	LUEA			+
	Hospital area	LUHA			+
	Park area	LUPA			-
	Recreational area	LURCA			-
	Religious area	LUREA			+
	Health facility area	LUHFA			+
	Industrial area	LUIA			+
distance variable (m^{-1}, m^{-2})	Workshop distance to monitoring locations	DV	25, 50, 100, 300, 500, 1000	Vehicle maintenance workshops emits pollutant emissions	+
geographic location (m, decimal degrees)	Elevation	Elev.	NA		NA
	Latitude	Lat.	NA		NA
	Longitude	Long.	NA		NA

[a] a priori definition based on the definition used in the ESCAPE study [41].

Geographic information system (GIS) shapefiles were obtained from the local department (Urban Unit), and GIS analysis for the extraction of potential predictor variables were performed using ArcGIS ver. 10.2, following the ESCAPE [41] procedure. The buffer radii were selected based on the previous studies ranging from 25, 50, 100, 300, 500 and 1000 m for road network, and ranges from 100, 300, 500 and 1000 m for land use data. All the predictor variables values were extracted using these buffer radii (Table 1).

Specific Local Data Survey (SLDS)

For the collection of local specific data, such as vehicle maintenance workshops (VMW), a survey was conducted around the monitoring stations, and the locations of the VMW were manually geocoded, noted the longitude and latitude with a GPS device, and then imported to google earth. Survey was helpful in determining the locations of the VMW. There were 198 VMW that were selected in this study because they were present at the time of sampling year. Inverse distance and inverse squared distance were calculated for distance variable [41].

2.4. Land Use Regression Model Development

The average concentrations of seasonal variations, the predictor variables including GIS parameters and local specific data for 18 monitoring stations, were used for the development of LUR model. The average concentration was used as dependent variable, whereas all other variables were as independent. Linear regression models were established using forward step regression approach [34]. Following the European Study of Cohorts

for Air Pollution Effects (ESCAPE) methodology, a priori definition was assigned to each variable type, based on the assumption of air pollution dynamics, e.g., increase in the length of road would increase the traffic emissions (positive direction), while park area would decrease the concentration because of the absorption effect of plants (negative direction). Positive direction of effect was given to the local specific source, e.g., VMW, considered as a point pollution source [42].

For the development of the model, univariate regression analysis was performed with all potential predictor variables. The factor provided the maximum adjusted explained variance (R^2) was selected, keeping in mind that the direction of effect was as pre-defined. The left over variables were included in the model based on the four criteria: (1) the direction of effect of the variable was as pre-defined; (2) predictor variables already present in the model, keep their original direction of effect; (3) the *p*-values of all the variables were below 0.1; (4) the variation inflation factor (VIF) of the predictor variables were considered acceptable if less than 5 [41]. To evaluate the influence of each season, models were developed for each season using measured concentrations for each sampling campaign (i.e., pre-monsoon, monsoon and post-monsoon).

2.5. Model Performance Evaluation and Mapping

To evaluate the performance of the developed model, leave-one-out cross validation (LOOCV) method was used, in which each site was consecutively omitted from the model and the model was developed using n – 1 site, n being the total number of sites [43]. The process was repeated for each site, and the mean difference obtained from linear regression between measured and predicted values for the left-out sites, based on adjusted R^2 and root mean square error (RMSE). This was determined using JMP software version 13.2.1 (Japan), using the ESCAPE procedure [41].

For the evaluation of spatial autocorrelation, Moran's I was calculated on the residuals of the final model and the results were explained by Z-score values. Grid dimensions of 500 m * 500 m were used, and 8172 grids were created by ArcGIS ver. 10.2 and the spatial distribution of predicted NO_2 concentration were done by using developed models on the relevant grids.

3. Results

3.1. Variation of Measured NO_2 Concentration

The seasonal variation (pre-monsoon, monsoon, post-monsoon) of measured nitrogen dioxide concentration showed in (Table 2). The descriptive statistics of concentration showed that the lowest concentration (47.74 ppb) was observed in summer (pre-monsoon), and the highest concentration (91.75 ppb) was seen in winter days (post-monsoon).

Table 2. Measured and predicted NO_2 levels.

Time Period	Measured NO_2 (ppb)		Predicted NO_2 (ppb)	
	Mean (S.D) [a]	Range [b]	Mean (S.D) [a]	Range [b]
pre-monsoon	68.34 (7.29)	[47.74; 77.48]	65.71 (7.00)	[57.71; 78.56]
monsoon	73.91 (7.18)	[53.66; 83.70]	71.89 (7.54)	[60.07; 84.30]
post-monsoon	79.01 (7.46)	[60.19; 91.75]	76.67 (7.85)	[64.02; 89.70]

Abbreviations: [a] Mean (standard deviation), [b] [min; max].

3.2. Variation of Predicted NO_2 Concentration

The seasonal variation (pre-monsoon, monsoon, post-monsoon) of predicted nitrogen dioxide concentration is shown in (Table 2). A similar trend was shown in the predicted concentration as compared to measured concentration, which showed that the lowest concentration (57.71 ppb) was observed in summer (pre-monsoon), and the highest concentration (89.70 ppb) was seen in winter days (post-monsoon). The scatter plots of measured vs. predicted values of NO_2 showed in the (Figure 2).
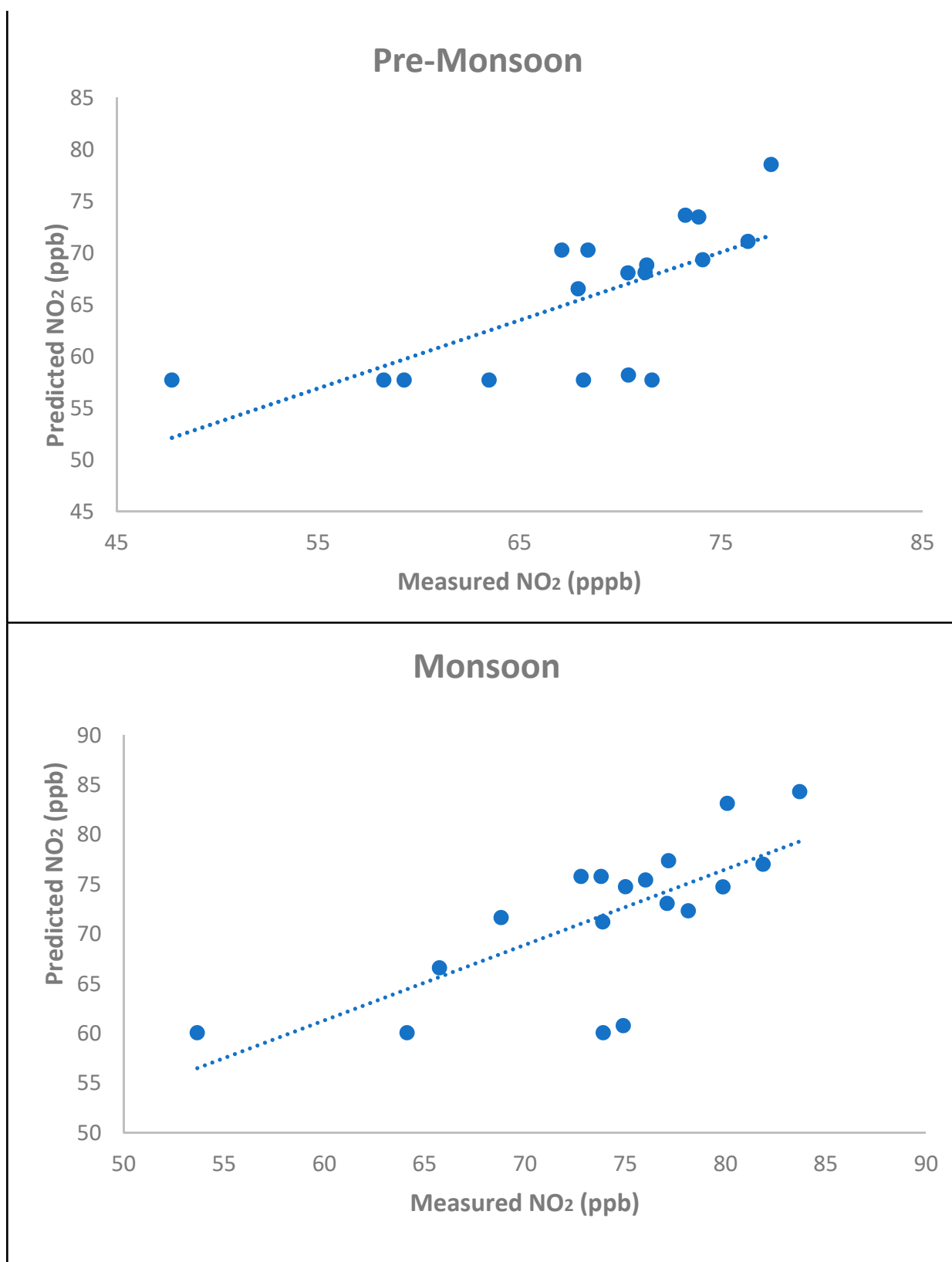
Figure 2. *Cont.*
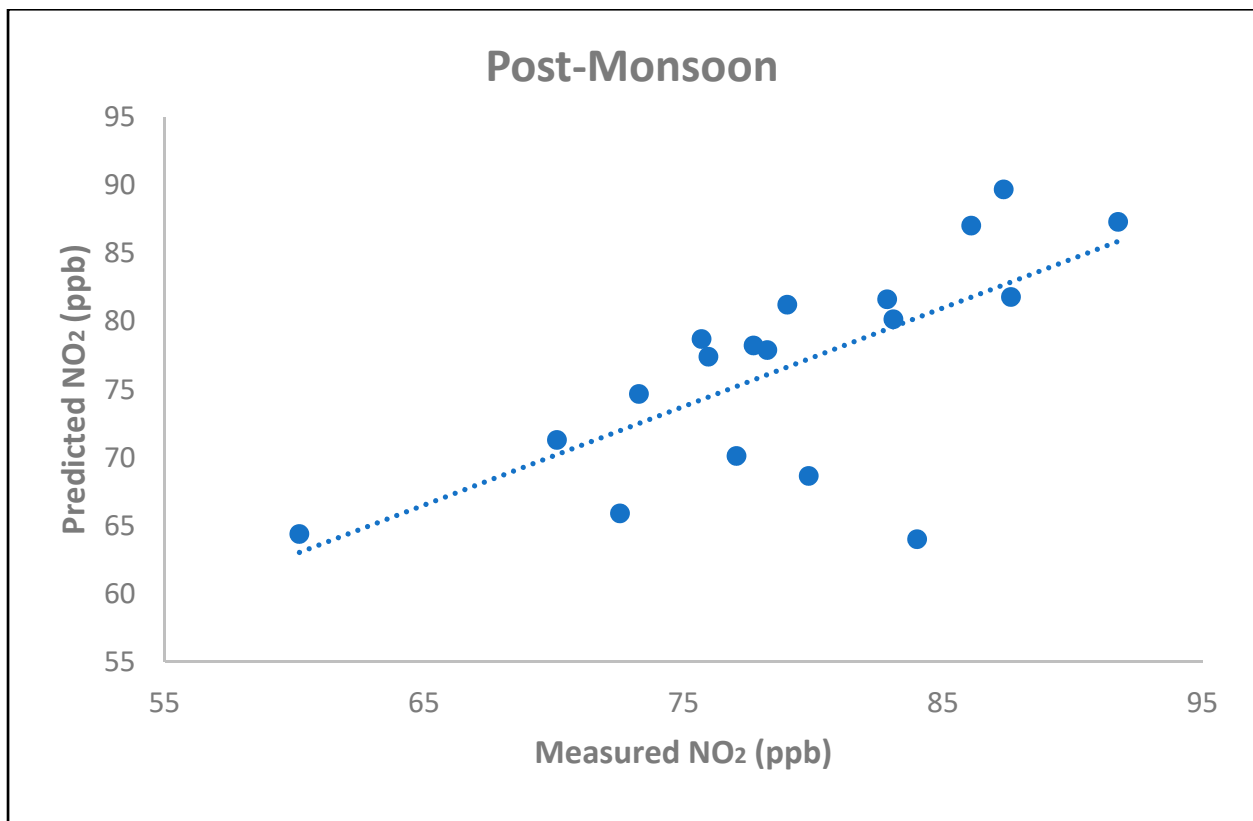
Figure 2. Scatter plots showing the measured vs. predicted values of NO$_2$ (pre-monsoon, monsoon and post-monsoon).

3.3. LUR Models

Final LUR models for seasonal variation of NO$_2$ were presented in (Tables 3–5). The adjusted R^2 and overall fit of LOOCV of LUR models were described in (Table 6).

Table 3. Summary of final land use regression model predicting NO$_2$.

Time Period	Variable	Coefficient	t	VIF
pre-monsoon	Intercept	57.7146	28.95	-
	Tertiary road length within 50 m	2.1×10^{-1}	2.25	1.12
	Residential area within 100 m	4×10^{-4}	2.50	1.16
	Distance variable within 300 m	1.3×10^{-3}	2.26	1.05

Table 4. Summary of final land-use regression model predicting NO$_2$.

Time Period	Variable	Coefficient	t	VIF
monsoon	Intercept	60.0727	25.18	-
	Tertiary road length within 100 m	8.08×10^{-2}	1.71	1.31
	Residential area within 100 m	5×10^{-4}	2.94	1.45
	Hospital area within 1000 m	5×10^{-5}	0.90	1.15
	Distance variable within 300 m	1.9×10^{-3}	3.26	1.27

Table 5. Summary of final land-use regression model predicting NO_2.

Time Period	Variable	Coefficient	t	VIF
	Intercept	63.0076	20.81	-
	Secondary road length within 1000 m	7×10^{-4}	0.96	1.17
post-monsoon	Tertiary road length within 300 m	8.8×10^{-3}	0.96	1.15
	Residential area within 100 m	5×10^{-4}	2.38	1.28
	Distance variable within 300 m	2.2×10^{-3}	3.26	1.22

Table 6. Adjusted R^2 of LUR model and overall fit of LOOCV.

Time Period	Model Performance Evaluation		LOOCV	
	Adj-R^2	RMSE (ppb)	R^2	RMSE (ppb)
pre-monsoon	0.7	4.35	0.6	6.11
monsoon	0.71	4.09	0.5	6.19
post-monsoon	0.77	3.87	0.57	6.34

Abbreviations: RMSE, root mean square error; LOOCV, leave-one-out cross validation.

3.3.1. Pre-Monsoon Model

In the final model, three significant factors were identified, including length of tertiary road within 50-m buffer, area of residential within 100-m buffer and distance variable within 300-m buffer. All the three influencing predictor variables were found to be positively associated, showing increase in the concentration of NO_2. VIF was less than 5.

The adjusted R^2 and RMSE were 0.70 and 4.35 ppb. The LOOCV R^2 and LOOCV RMSE were 0.60 and 6.11 ppb, respectively. Results of residual spatial autocorrelation analysis were presented in (Table 7). The Z-score was 0.70, the pattern appears to be random which shows the consistency with the hypothesis of spatial error independence.

Table 7. Spatial autocorrelation results of LUR model residuals.

Time Period	Moran's Index	z-Score	*p*-Value
pre-monsoon	0.02	0.70	0.48
monsoon	−0.04	0.11	0.91
post-monsoon	−0.09	−0.29	0.77

3.3.2. Monsoon Model

Four influencing factors were entered in the final model, including tertiary road length within 100-m buffer, residential area within 100-m buffer, hospital area within 1000-m buffer and distance variable within 300-m buffer. All the four influencing predictor variables were found to be positively associated, showing the increase in the concentration of NO_2. VIF were less than 5.

The adjusted R^2 and RMSE were 0.71 and 4.09 ppb. The LOOCV R^2 and LOOCV RMSE were 0.50 and 6.19 ppb, respectively. Results of residual spatial autocorrelation analysis were presented in (Table 7). The Z-score was 0.11, the pattern appears to be random, showing the consistency with the hypothesis of spatial error independence.

3.3.3. Post-Monsoon Model

In the final model, four influencing factors were identified, including secondary road length within 1000-m buffer, length of tertiary road within 300-m buffer, area of residential within 100-m buffer and the distance variable within 300-m buffer. All the four influencing factors were found to be positively associated with the NO_2 concentration. VIF was less than 5.

The adjusted R^2 and RMSE were 0.77and 3.87 ppb. The LOOCV R^2 and LOOCV RMSE were 0.57 and 6.34 ppb, respectively. Results of residual spatial autocorrelation analysis were presented in (Table 7). The Z-score was −0.29, the pattern appears to be random, showing the consistency with the hypothesis of spatial error independence.

LUR models showed moderate to good variance for all the seasons. Adjusted explained variance of the LUR models was highest for post-monsoon (77%), followed by monsoon (71%) and was lowest for pre-monsoon (70%). As for the overall fit of LOOCV, the R^2 was highest for pre-monsoon (61%), followed by post-monsoon (57%) and was the lowest for monsoon (50%).

3.4. Spatial Characteristics and Regression Maps of Predicted NO_2

The maps of seasonal variation of predicted NO_2 were done using 8172 uniformly distributed grids for the Lahore city (Figure 3A–C). Higher concentration was observed around the roads almost during all the seasons, indicating road traffic as a source of pollution while, lower concentration was seen around the sub-urban areas where the road network was not so strong, having open spaces and agricultural land (Figure S1).
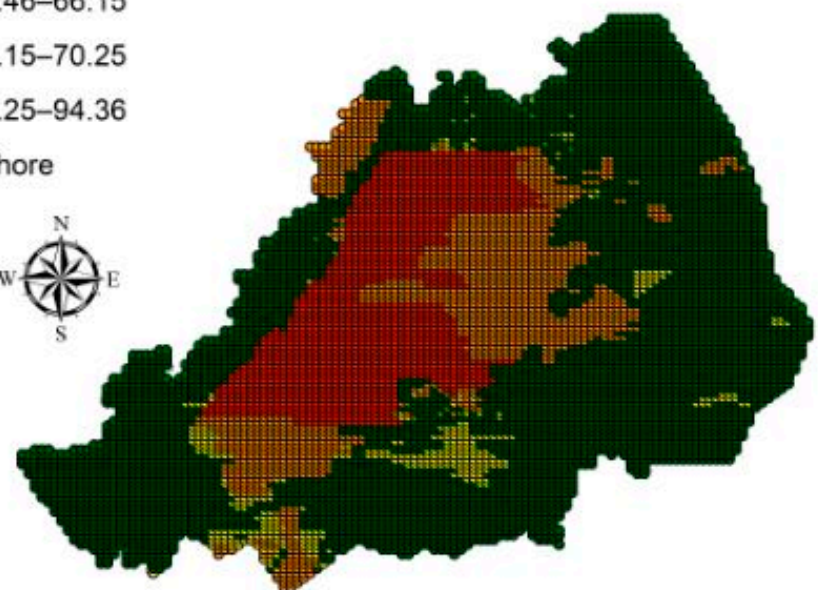

(**A**)

Figure 3. *Cont.*

NO2 (ppb)

Monsoon

Predicted concentration

- 57.71–62.53
- 62.53–66.19
- 66.19–70.85
- 70.85–75.74
- 75.74–94.36
- Lahore

(**B**)

NO2 (ppb)

Post- monsoon

Predicted concentration

- 57.71–64.40
- 64.40–67.66
- 67.66–73.78
- 73.78–86.10
- 86.10–94.36
- Lahore
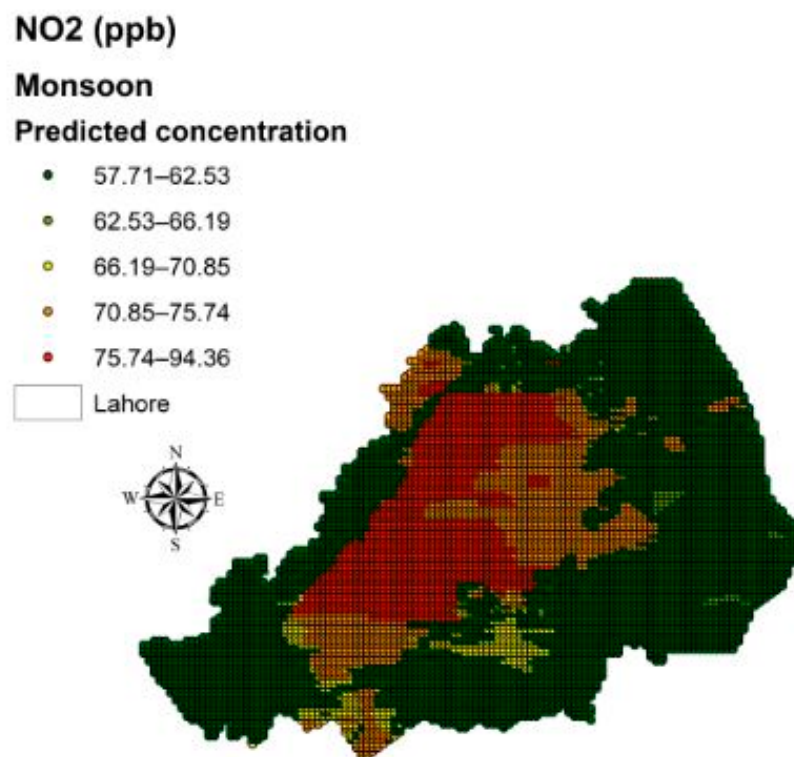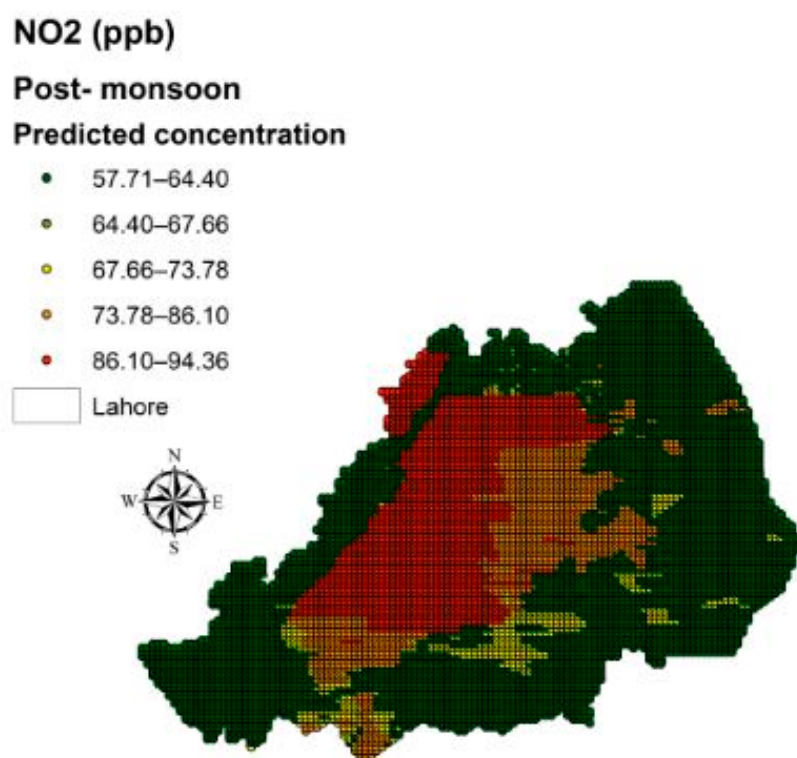
(**C**)

Figure 3. Regression mapping of predicted concentration ((**A**) pre-monsoon, (**B**) monsoon, (**C**) post-monsoon).

4. Discussion

Land use regression model has been applied in the developed countries, but still there is lack of application of LUR model in developing countries [37,44]. To the best of our knowledge, this is the first attempt to apply LUR model in Pakistan's urban area setting for a city, although LUR has been used in Pakistan at a national level for ambient $PM_{2.5}$ exposure [29]. LUR models were developed for seasonal variation (pre-monsoon, monsoon and post-monsoon) mean concentration of NO_2 pollutant, based on the collected data of 18 monitoring locations in the Lahore city, Pakistan. The final developed LUR models performed well, showing the reliability with high accuracy and spatial heterogeneity.

Previous LUR models in the literature have showed the values of R^2 ranging from lowest (0.41) in the startup model to highest (0.73) in the final model, achieving an R^2 of 0.68 for the winter model and 0.59 for the summer model [45]. The study conducted in Xian, China reported that the value of R^2 was greater than 0.8, indicating that the heating season had the best simulation effect [46]. The study conducted in Nanjing, China reported the R^2 value of 0.7 for NO_2 model [39]. The LUR models developed in this study, shows that the values of R^2 (0.5–0.61), like other studies conducted in the literature, thus indicating that it is feasible to develop these models for developing countries and can use for exposure assessment studies. The values of the model R^2 was close to the LOOCV R^2, showing the robustness of the LUR models for all the seasons [44].

Data collected from manually surveying the study area, proved helpful to increase the performance of LUR models. Such type of survey can provide us the valuable specific feature of potential predictor variables in the study area, that would improve the overall fit of LUR model. In this study, vehicle maintenance workshops data were collected by doing the manual survey of the study area, which can be a source of NO_2, thus highlighting the importance of culture or site-specific land use classes [47,48]. Distance to vehicle maintenance workshop entered in the final models of all the seasons was found to be one of the influencing factors for the source contribution to NO_2.

Different potential predictor variables, entered in the final developed models, showed that the different factors have influence on the pollutant concentrations, although same predictor variables were used to develop the LUR models. The road network (road length) was the influencing factor, with NO_2 being known as the traffic-based air pollutant, showed the positive association with the road length factor in the final LUR models for all the seasons. The residential area seems to be another effective factor, showing a positive association with the concentration of NO_2. The reason was that people in the household use the fuel-based generators for the electricity and combustion processes, and also because people travel frequently around residential area, causing increase in the concentration of NO_2 due to vehicle exhaust emissions [49]. Distance variable also showed the contribution in all the models, highlighting the importance of local automobile workshops in the emission of NO_2. Hospital area showed the contribution of NO_2 from a generators facility used in the vicinity of the hospital area, and local automobile workshop.

Electricity generation and its demand is a serious issue in developing countries, especially in Pakistan, which leads to the usage of fuel-based generators to produce electricity, in the households and hospitals, causing an increase in the pollutant concentration. The residential area also contributed as a source of NO_2 pollution, due to the burning of fossil fuels in the households, especially during the heating season (post-monsoon) resulting in the highest concentration during all the seasons. Specific local data survey (SLDS) was helpful in determining another factor, which was vehicle maintenance workshops, surrounded around the roads and residence area, cause of increase in the concentration of pollutant due to maintenance of vehicles.

The significant predictor variables identified in this study were like other studies conducted previously, showing the length of roads and residential land area as common influencing factor [50,51]. The study conducted in Taipei city, Taiwan also reflected that the high concentrations attributed to road length, as one of the influencing factors and to the dense road network [52]. Another study, also showed one of the influencing factors

was major roads and traffic influence, included in the models for heating and non-heating seasons [53]. The study conducted in Nanjing, China indicated that the residential area within 100m and 5000m buffer, entered in the final model, proved to be a significant predictor variable [39]. The developed models based on the influencing predictor variables in this study comparable with the other studies conducted previously, supporting the LUR models in the urban settings of Lahore, Pakistan.

Among the three seasons, predicted concentration regression maps showed the similar spatial characteristics with high concentration in the city center, where the residential land area is high due to the population. Next to residential area land feature, the maximum concentration can be observed around the road network, which is mainly distributed in the center and to the west part of the city. The vehicle maintenance workshops also surrounded around the roads, and nearby the residential area and hospital area, showing the public living nearby those areas could have a negative effect on their health, due to exposure to the pollutants. Since NO_2 pollutant concentration relates to traffic intensity and residential area, this spatial distribution is reasonable.

There were some limitations in this study. The selected measurements sites may not be able to capture the pollutant concentration distribution across the whole study area, because these sites were mostly in the city area, and therefor lack the coverage resolution of rural areas, which can be considered in the future studies. Although, NO_2 concentration related to the industrial emissions, but due to limited availability of industrial data, did not enter in the final model, which can be considered by using the specific local data survey to capture the industrial sites.

5. Conclusions

Land use regression models for seasonal variation of NO_2 were developed based on different predictor variables in Lahore, Pakistan. The R^2 values showed the precision of our generated models. This study has shown that the higher concentration can be seen around the areas surrounded by roads, residential areas and vehicle maintenance workshops, whereas lower concentration was observed around the parts of city, having not strong road network. People living nearby such areas are more affected and exposed to air pollutants which can increase the possibility of bad health outcomes. The developed models would contribute more to the broader use of LUR models in Pakistan, and if connected with the people's health indicators, this research can be helpful in providing support to provide epidemiological-based studies in the future.

References

1. World Health Organization. *Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease*; WHO: Geneva, Switzerland, 2016.
2. Tan, X.; Han, L.; Zhang, X.; Zhou, W.; Li, W.; Qian, Y. A review of current air quality indexes and improvements under the multi-contaminant air pollution exposure. *J. Environ. Manag.* **2021**, *279*, 111681. [CrossRef] [PubMed]
3. Ashfaq, A.; Sharma, P. Environmental effects of air pollution and application of engineered methods to combat the problem. *iControl Pollut.* **2012**, *29*, 25–28.
4. Pecl, G.T.; Araújo, M.B.; Bell, J.D.; Blanchard, J.; Bonebrake, T.C.; Chen, I.-C.; Clark, T.D.; Colwell, R.K.; Danielsen, F.; Evengård, B. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **2017**, *355*, eaai9214. [CrossRef]
5. Karl, T.R.; Melillo, J.M.; Peterson, T.C.; Hassol, S.J. *Global Climate Change Impacts in the United States*; Cambridge University Press: Washington, DC, USA, 2009.
6. Kelishadi, R.; Poursafa, P. Air pollution and non-respiratory health hazards for children. *Arch. Med Sci. AMS* **2010**, *6*, 483. [CrossRef] [PubMed]
7. Eze, I.C.; Schaffner, E.; Fischer, E.; Schikowski, T.; Adam, M.; Imboden, M.; Tsai, M.; Carballo, D.; von Eckardstein, A.; Künzli, N. Long-term air pollution exposure and diabetes in a population-based Swiss cohort. *Environ. Int.* **2014**, *70*, 95–105. [CrossRef]
8. Mannucci, P.M.; Franchini, M. Health effects of ambient air pollution in developing countries. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1048. [CrossRef]
9. Rana, I.A.; Bhatti, S.S. Lahore, Pakistan–Urbanization challenges and opportunities. *Cities* **2018**, *72*, 348–355. [CrossRef]
10. UNEP. *UNEP Year Book 2011: Emerging Issues in Our Global Environment*; UNEP/Earthprint: Nairobi, Kenya, 2011.
11. Rasheed, A.; Aneja, V.P.; Aiyyer, A.; Rafique, U. Measurement and analysis of fine particulate matter (PM2. 5) in urban areas of Pakistan. *Aerosol Air Qual. Res.* **2015**, *15*, 426–439. [CrossRef]
12. Barber, N. *World in Focus-Focus on Pakistan*; World Almanac Library: Stamford, CT, USA, 2008.
13. Ameratunga, S.; Hijar, M.; Norton, R. Road-traffic injuries: Confronting disparities to address a global-health problem. *Lancet* **2006**, *367*, 1533–1540. [CrossRef]
14. Colbeck, I.; Nasir, Z.A.; Ali, Z. The state of ambient air quality in Pakistan—A review. *Environ. Sci. Pollut. Res.* **2010**, *17*, 49–63. [CrossRef]
15. Shabbir, Y.; Khokhar, M.F.; Shaiganfar, R.; Wagner, T. Spatial variance and assessment of nitrogen dioxide pollution in major cities of Pakistan along N5-Highway. *J. Environ. Sci.* **2016**, *43*, 4–14. [CrossRef] [PubMed]
16. Sánchez-Triana, E.; Enriquez, S.; Afzal, J.; Nakawaga, A.; Khan, A.S. *Air Pollution in Pakistan*; World Bank Group: Islamabad, Pakistan, 2014.
17. Mage, D.; Ozolins, G.; Peterson, P.; Webster, A.; Orthofer, R.; Vandeweerd, V.; Gwynne, M. Urban air pollution in megacities of the world. *Atmos. Environ.* **1996**, *30*, 681–686. [CrossRef]
18. Colvile, R.; Hutchinson, E.J.; Mindell, J.; Warren, R. The transport sector as a source of air pollution. *Atmos. Environ.* **2001**, *35*, 1537–1565. [CrossRef]
19. World Health Organization. *Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide*; World Health Organization: Geneva, Switzerland, 2006.
20. Hewitt, C.N. Spatial variations in nitrogen dioxide concentrations in an urban area. *Atmos. Environ. Part B Urban Atmos.* **1991**, *25*, 429–434. [CrossRef]
21. Yu, B.; Chen, F.; Shirazi, Z. Impact of biomass fires on tropospheric nitrogen dioxide in South and Southeast Asia for the years from 2005 to 2014. *Appl. Geogr.* **2017**, *86*, 92–101. [CrossRef]
22. Wang, W.; Wang, T. On the origin and the trend of acid precipitation in China. *Water Air Soil Pollut.* **1995**, *85*, 2295–2300. [CrossRef]
23. Briggs, D. *The Use of GIS to Evaluate Traffic-Related Pollution*; BMJ Publishing Group Ltd.: London, UK, 2007.
24. Ren, X.; Mi, Z.; Georgopoulos, P.G. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. *Environ. Int.* **2020**, *142*, 105827. [CrossRef]
25. Šimić, I.; Lovrić, M.; Godec, R.; Kröll, M.; Bešlić, I. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. *Environ. Pollut.* **2020**, *263*, 114587. [CrossRef]
26. Mulholland, J.A.; Butler, A.J.; Wilkinson, J.G.; Russell, A.G.; Tolbert, P.E. Temporal and spatial distributions of ozone in Atlanta: Regulatory and epidemiologic implications. *J. Air Waste Manag. Assoc.* **1998**, *48*, 418–426. [CrossRef]
27. de Hoogh, K.; Korek, M.; Vienneau, D.; Keuken, M.; Kukkonen, J.; Nieuwenhuijsen, M.J.; Badaloni, C.; Beelen, R.; Bolignano, A.; Cesaroni, G. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environ. Int.* **2014**, *73*, 382–392. [CrossRef]
28. 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]
29. Shi, Y.; Bilal, M.; Ho, H.C.; Omar, A. Urbanization and regional air pollution across South Asian developing countries–A nationwide land use regression for ambient PM2. 5 assessment in Pakistan. *Environ. Pollut.* **2020**, *266*, 115145. [CrossRef]
30. Hoek, G.; Beelen, R.; De Hoogh, K.; 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]

31. Jerrett, M.; Arain, M.; Kanaroglou, P.; Beckerman, B.; Crouse, D.; Gilbert, N.; Brook, J.; Finkelstein, N.; Finkelstein, M. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J. Toxicol. Environ. Health Part A* **2007**, *70*, 200–212. [CrossRef]

32. Ghassoun, Y.; Ruths, M.; Löwner, M.-O.; Weber, S. Intra-urban variation of ultrafine particles as evaluated by process related land use and pollutant driven regression modelling. *Sci. Total Environ.* **2015**, *536*, 150–160. [CrossRef]

33. Novotny, E.V.; Bechle, M.J.; Millet, D.B.; Marshall, J.D. National satellite-based land-use regression: NO_2 in the United States. *Environ. Sci. Technol.* **2011**, *45*, 4407–4414. [CrossRef] [PubMed]

34. Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe–the ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10–23. [CrossRef]

35. Cattani, G.; Gaeta, A.; di Bucchianico, A.D.M.; De Santis, A.; Gaddi, R.; Cusano, M.; Ancona, C.; Badaloni, C.; Forastiere, F.; Gariazzo, C. Development of land-use regression models for exposure assessment to ultrafine particles in Rome, Italy. *Atmos. Environ.* **2017**, *156*, 52–60. [CrossRef]

36. 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]

37. Tonne, C.; Salmon, M.; Sanchez, M.; Sreekanth, V.; Bhogadi, S.; Sambandam, S.; Balakrishnan, K.; Kinra, S.; Marshall, J.D. Integrated assessment of exposure to PM2. 5 in South India and its relation with cardiovascular risk: Design of the CHAI observational cohort study. *Int. J. Hyg. Environ. Health* **2017**, *220*, 1081–1088. [CrossRef]

38. Ma, J.; Simonich, S.; Tao, S. *New Discoveries to Old Problems: A Virtual Issue on Air Pollution in Rapidly Industrializing Countries*; ACS Publications: Washington, DC, USA, 2017.

39. Huang, L.; Zhang, C.; Bi, J. Development of land use regression models for PM2. 5, SO_2, NO_2 and O_3 in Nanjing, China. *Environ. Res.* **2017**, *158*, 542–552. [CrossRef] [PubMed]

40. Mirza, A.I.; Mayo, S.; Aziz, A.; Sharif, M. Spatial-Temporal Variations of Nitrogen Dioxide in and Around Lahore Metropolitan Areas Using "GIS" Techniques. *Pak. J. Sci.* **2013**, *65*, 371.

41. Beelen, R.; Hoek, G. *ESCAPE Exposure Assessment Manual*; IRAS: TD Utrecht, The Netherlands, 2010.

42. Saucy, A.; Röösli, M.; Künzli, N.; Tsai, M.-Y.; Sieber, C.; Olaniyan, T.; Baatjies, R.; Jeebhay, M.; Davey, M.; Flückiger, B. Land use regression modelling of outdoor NO_2 and PM2. 5 concentrations in three low income areas in the western cape province, South Africa. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1452. [CrossRef]

43. Saraswat, A.; Apte, J.S.; Kandlikar, M.; Brauer, M.; Henderson, S.B.; Marshall, J.D. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. *Environ. Sci. Technol.* **2013**, *47*, 12903–12911. [CrossRef] [PubMed]

44. Meng, X.; Chen, L.; Cai, J.; Zou, B.; Wu, C.-F.; 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]

45. Muttoo, S.; Ramsay, L.; Brunekreef, B.; Beelen, R.; Meliefste, K.; Naidoo, R.N. Land use regression modelling estimating nitrogen oxides exposure in industrial south Durban, South Africa. *Sci. Total Environ.* **2018**, *610*, 1439–1447. [CrossRef]

46. Liu, Z.; Guan, Q.; Luo, H.; Wang, N.; Pan, N.; Yang, L.; Xiao, S.; Lin, J. Development of land use regression model and health risk assessment for NO2 in different functional areas: A case study of Xi'an, China. *Atmos. Environ.* **2019**, *213*, 515–525. [CrossRef]

47. 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 PM2. 5 spatial-temporal variability. *Environ. Pollut.* **2017**, *224*, 148–157. [CrossRef] [PubMed]

48. Smargiassi, A.; Brand, A.; Fournier, M.; Tessier, F.; Goudreau, S.; Rousseau, J.; Benjamin, M. A spatiotemporal land-use regression model of winter fine particulate levels in residential neighbourhoods. *J. Expo. Sci. Environ. Epidemiol.* **2012**, *22*, 331–338. [CrossRef]

49. Martin, A.N.; Boulter, P.G.; Roddis, D.; McDonough, L.; Patterson, M.; del Barco, M.R.; Mattes, A.; Knibbs, L.D. In-vehicle nitrogen dioxide concentrations in road tunnels. *Atmos. Environ.* **2016**, *144*, 234–248. [CrossRef]

50. Madsen, C.; Gehring, U.; Håberg, S.E.; Nafstad, P.; Meliefste, K.; Nystad, W.; Carlsen, K.C.L.; Brunekreef, B. Comparison of land-use regression models for predicting spatial NOx contrasts over a three year period in Oslo, Norway. *Atmos. Environ.* **2011**, *45*, 3576–3583. [CrossRef]

51. Aguilera, I.; Sunyer, J.; Fernández-Patier, R.; Hoek, G.; Aguirre-Alfaro, A.; Meliefste, K.; Bomboi-Mingarro, M.T.; Nieuwenhuijsen, M.J.; Herce-Garraleta, D.; Brunekreef, B. Estimation of outdoor NO x, NO_2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environ. Sci. Technol.* **2008**, *42*, 815–821. [CrossRef]

52. Lee, J.-H.; Wu, C.-F.; Hoek, G.; de Hoogh, K.; Beelen, R.; Brunekreef, B.; Chan, C.-C. Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. *Sci. Total Environ.* **2014**, *472*, 1163–1171. [CrossRef]

53. Chen, L.; Bai, Z.; Kong, S.; Han, B.; You, Y.; Ding, X.; Du, S.; Liu, A. A land use regression for predicting NO_2 and PM10 concentrations in different seasons in Tianjin region, China. *J. Environ. Sci.* **2010**, *22*, 1364–1373. [CrossRef]