Estimation of Nitrogen Dioxide Concentrations in the Inner Bangkok by Land Use Regression Model

Pannee Cheewinsiriwat, Pongsri Chanhow

Chulalongkorn University, Thailand

0143

The Asian Conference on Sustainability, Energy & the Environment 2013

Official Conference Proceedings 2013

Abstract

The Pollution Control Department (PCD) has long been responsible for an hourly measurement of Nitrogen Dioxide (NO_2) concentrations at its twelve stations located within the 430 square kilometer area of Inner Bangkok. In the past, to estimate NO_2 concentrations at any unmeasured location, the proximity model, interpolation model, or dispersion model were employed. These models used distance from a measured location as the sole determinant of any estimation. Toward the end of the 1990's, the more sophisticated Land Use Regression (LUR) model was introduced. This model with its built-in Geographic Information System (GIS) and multiple regression analysis enabled the inclusion of other important determining variables such as land use types, traffic volume and selected meteorological variables. This research aims to apply the LUR model for the estimation of NO_2 concentrations over the study area covering the Inner Bangkok. Monthly average NO_2 concentrations, traffic count, land use types, road area together with humidity, temperature, wind speed, and rainfall data, measured at or within the vicinities of the twelve PCD stations were input into the model. Only humidity, temperature, wind speed, residential land use, industrial land use, and rainfall are found to have influenced the NO_2 concentrations in the Inner Bangkok. The resulting coefficient of determination (R square) of 0.759 implies that seventy-six percent of the variations in NO_2 concentrations in the Inner Bangkok can be explained by model. The research will, however, continue to obtain more precise traffic volume data in terms of time scale to improve the model.

Keywords: Nitrogen Dioxide, Bangkok, GIS, Land Use Regression model, multiple regression analysis, air pollution

1. Introduction

Air pollution is one of the major environmental problems in urban areas especially big cities including Bangkok. Among the different air pollutants, Nitrogen dioxide (NO_2) has been rarely explored over coverage of Bangkok. It is harmful to human health particularly for the respiratory system. High concentration inhaled into human bodies can cause lung cancer. The implementation of air pollutant measurement stations is costly, requires high technical support and is usually limited to a small number of stations within a city. In the case of Bangkok, The Pollution Control Department (PCD) has long been responsible for an hourly measurement of NO_2 concentrations at its twelve stations located within the 430 square kilometer area of Inner Bangkok. Although the NO_2 concentration levels in Bangkok are still below the air quality standard level defined by the Ministry of Technology and Environment of Thailand, the levels need to be monitored to ensure the quality of life of people in Bangkok.

As every location cannot be measured, NO_2 concentrations at unmeasured locations can be obtained only by methods of estimation. In the past, to estimate NO_2 concentrations at any unmeasured location, the proximity model, interpolation model, or dispersion model were employed (Jerrett *et al.*, 2005). These models used distance from a measured location as a sole determinant of any estimation. Distance is, however, not the only factor that affects the concentrations of NO_2. Toward the end of the 1990's, the more sophisticated Land Use Regression (LUR) model was introduced. The model with its built-in Geographic Information System (GIS) and multiple regression analysis enabled the inclusion of other important determining variables such as land use types, traffic volume and selected meteorological variables (Briggs *et al.*, 1997). This is to fill the gaps of the sparse air pollution measurements stations.

The LUR model has been successfully implemented in several urban areas in different countries such as Sheffield in UK, Los Angeles in U.S.A. and Montreal in Canada (Briggs *et al.,* 1997; Gilbert *et al.,* 2005; Beelen *et al.*, 2009; Su *et al.*, 2009). The models implemented in different cities were based on the same concept but different factors or variables. This research aims to apply the LUR model to estimate NO_2 concentrations over the study area covering the Inner Bangkok. Monthly average NO_2 concentrations, traffic volume, land use area, road area together with relative humidity, temperature, wind speed, and rainfall, measured at or within the vicinities of the twelve PCD stations, are the input factors of the Bangkok LUR model.

In the next section, the applications of LUR and GIS on NO_2 concentration issues in previous researches are identified. Then, the study area and datasets used in the Bangkok model are described, followed by the analysis results and conclusions.

2. LUR and GIS Implementations on NO_2 Concentration Issues

As air pollution is one of the major problems in big cities, it has been studied in order to monitor, identify, and predict levels of risk to human health. Estimation methods such as proximity model, interpolation model, LUR model, and dispersion model, have been employed to generate a map of air pollution levels over the study area covering both locations with and without measurements (Jerrett *et al.,* 2005). The proximity model estimates air pollution concentrations based on distance from sources. Locations nearer to the sources are estimated to have higher concentrations. The proximity model considers distance as the sole factor, although the other factors may also affect the concentration levels. The interpolation model estimates concentrations by employing spatial statistics both deterministic and stochastic methods. The interpolation model uses a set of sample measured concentrations from a number of measurement stations. The accuracy of the results is based on numbers of measurements input to the model.

The dispersion model employs Gaussian plume equations to estimate concentrations. The dispersion model is good to generate the concentrations from sources over time. However, it is costly to obtain input data from sources for this kind of model. The LUR model estimates concentrations from pollution-related factors which are in or around a certain location. The pollution-related factors include traffic volume, land use, climate-related data, etc. At the same quality level of the output, the LUR model provides better results and lower cost than the other three models.

The LUR model has been applied in several researches on air pollution concentrations in big cities (Hoek *et al.,* 2008). Briggs et al. (1997) has introduced a regression-based method integrated with GIS, later named LUR model, to generate air pollution maps of three cities in Europe, which are Amsterdam (the Netherlands), Hudderfield (UK), and Prague (the Czech Republic). The input data of this research were road characteristics and traffic volume, land use, elevation, and measured NO_2 concentrations. It was found that NO_2 concentrations generated by road traffic affect areas within 100 meters from roads. The model was able to estimate NO_2 concentrations at unmeasured locations with high accuracy and precision (Briggs *et al.*, 1997). After that, LUR models were implemented in several cities including Los Angeles, Montreal, and Rome (Gilbert *et al.*, 2005; Rosenlund *et al.,* 2007; Hoek *et al.,* 2008; Su *et al.,* 2009).

In regression analysis, the concentrations measured at stations or sites are the samples that are vital in the estimation calculation for unmeasured sites (Jerrett *et al.*, 2005). The number of samples varies in each study area depending on size, from 20 in Texas to 459 in Europe (Hoek *et al.* 2008). The factors used for the study areas are rather similar, but different choices of the factors were selected for a particular study area based on the characteristics of the city. The factors include transportation network, road length, traffic volume, vehicle type, distance from roads, distance from the sea, number of population, population density, number of buildings, land use, physical characteristics, and meteorological data (Jerrett *et al.*, 2005; Arain *et al.,* 2007; Hoek *et al.*, 2008).

For Bangkok, relationships between NO_2 concentrations and related-environment factors were studied in six study areas distributed over Bangkok (Leong *et al.* 2003). The study reported that the levels of NO_2 concentrations relate to various factors included road type, road width, traffic volume and speed, vehicle type, as well as meteorological factors. It was also reported that the concentrations were influenced by seasonal climate; higher in winter and lower in summer. In 2004, a set of air pollution risk maps were also produced using a spatial interpolation technique based on the samples of measured NO_2 concentrations from a number of PCD stations (Pangsang, 2004). The risk maps revealed that the high risk districts (or so-called “Khet”) were Bang Rak, Bueng Kum, Pathum Wan, Lat Phrao, Sathon, and Watthana, which were areas of high density, business centers, and congested traffic.

LUR and GIS have been used to study NO_2 concentrations in several cities and the estimation results were accurate enough to be reliable for health and environment planning (Hoek *et al.*, 2008). Although most factors used in each model are quite similar, a model which is good for one city may not be appropriate for another city because of the differences in built-up areas and physical characteristics of the city (Jerrett *et al.*, 2005). As air pollution in Bangkok is considered one of the major problems of the city and the LUR model had not been previously applied to any area in Thailand, this research attempted to apply the LUR model to estimate NO_2 concentrations in the area of Inner Bangkok. This research included traffic volume, land use, and meteorological data as the input data to the model. A multiple regression equation based on the input data was expected as the result.

3. Study Area and Datasets

The study area is the Inner Bangkok covering 32 districts (Bangkok consists of 50 districts) or 430 square kilometers. Within the study area, 12 PCD stations are located to hourly measure air quality including NO_2 concentrations. In Figure 1, the districts within the study area as well as the PCD stations with their station IDs are shown. The data used in this study consists of the measured NO_2 concentrations, traffic volume, road area, land use, relative humidity, temperature, wind speed, and rainfall. The data were collected in the period of January to December 2010. The NO_2 concentrations, relative humidity, temperature, wind speed, and rainfall data were measured by the PCD, while traffic volume, road area, and land use were collected and digitalized by the Bangkok Metropolitan Authority (BMA).

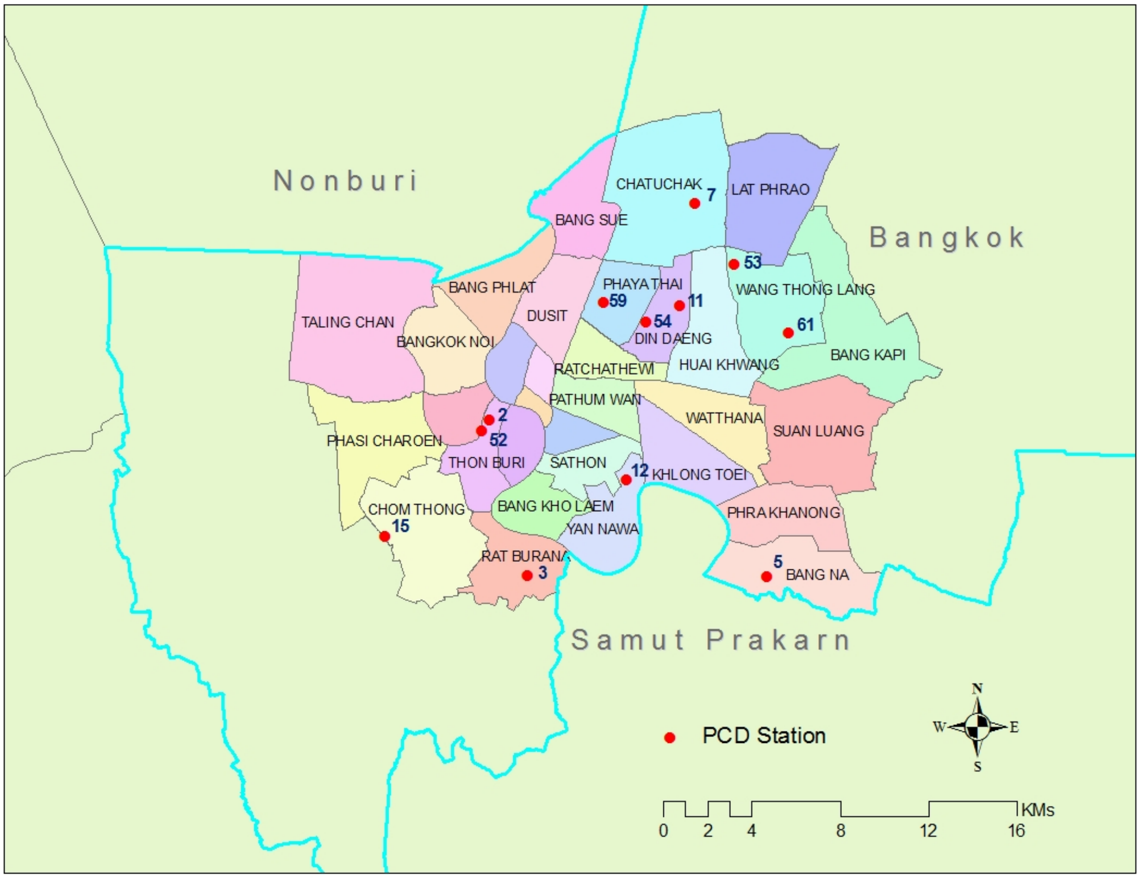

Figure 1 The districts within the study area and PCD stations with their station IDs.

Figure 2 shows that the averages hourly NO_2 concentrations of the 12 stations are of different levels, but mostly in the same direction. This can be interpreted that the NO_2 concentrations in Inner Bangkok are related to season, which corresponds to the previous research. In Figure 3, Traffic volume on each road was estimated based on the traffic count at 251 intersections. This research classified roads into four distinct categories: highway, main road, secondary road, and small road. The traffic counts were also classified according to the category of the road on which they were counted. Traffic counts on the same road category were sorted and the median values were selected to be the estimated traffic volume of each road category. Road areas were also calculated from the length of roads multiplied by their width.

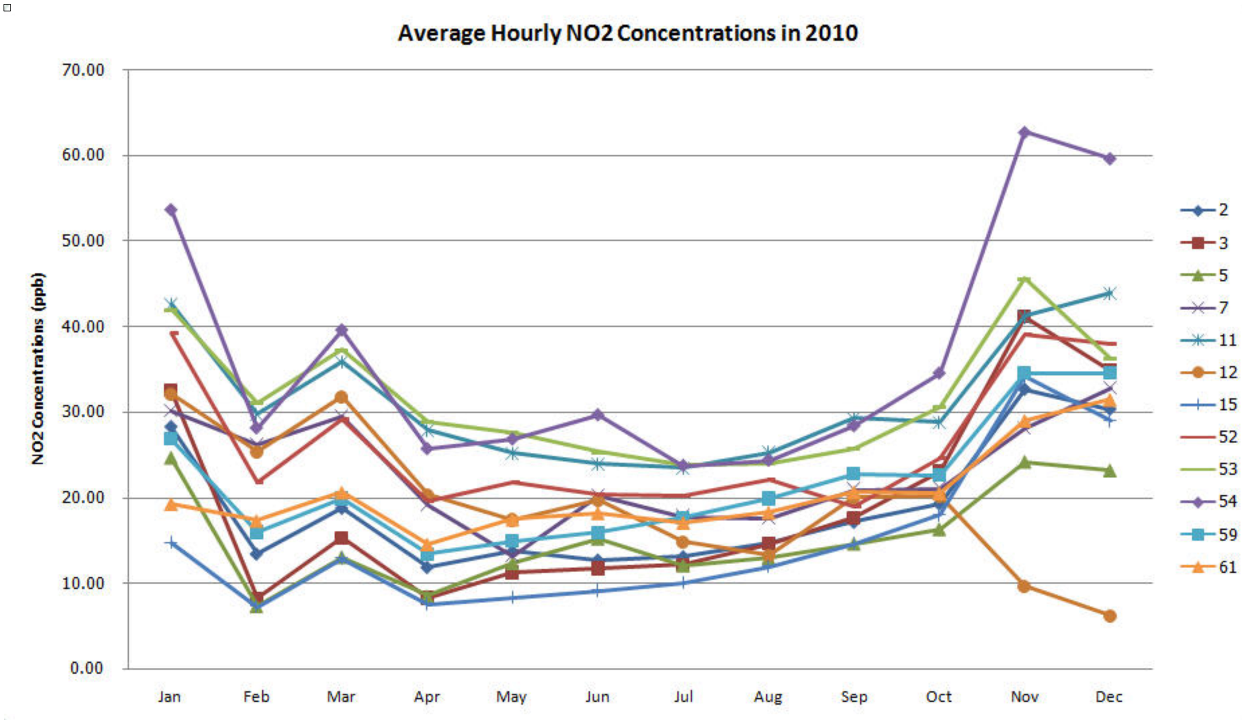

Figure 2 Average hourly NO_2 concentrations in 2010.

This research focused on only four land use categories: residential, commercial, industrial, and agriculture. In this research, building footprints were used as the land use factor. Land use areas were thus calculated from building footprint areas of each building use categories. In Figure 4, the map depicts the four distinct categories of buildings based on their usage over the study area. For climate data, percentages of relative humidity levels in Bangkok range from below 10 up to 100 for the year of 2010 (Figure 5), while temperatures were between 20 -35 degree Celsius (Figure 6). Bangkok is not in a windy area as the highest wind speed of most stations is about 4-5 meter/second (Figure 7). The maximum hourly rainfalls in Bangkok were about 40-50 mm (Figure 8).

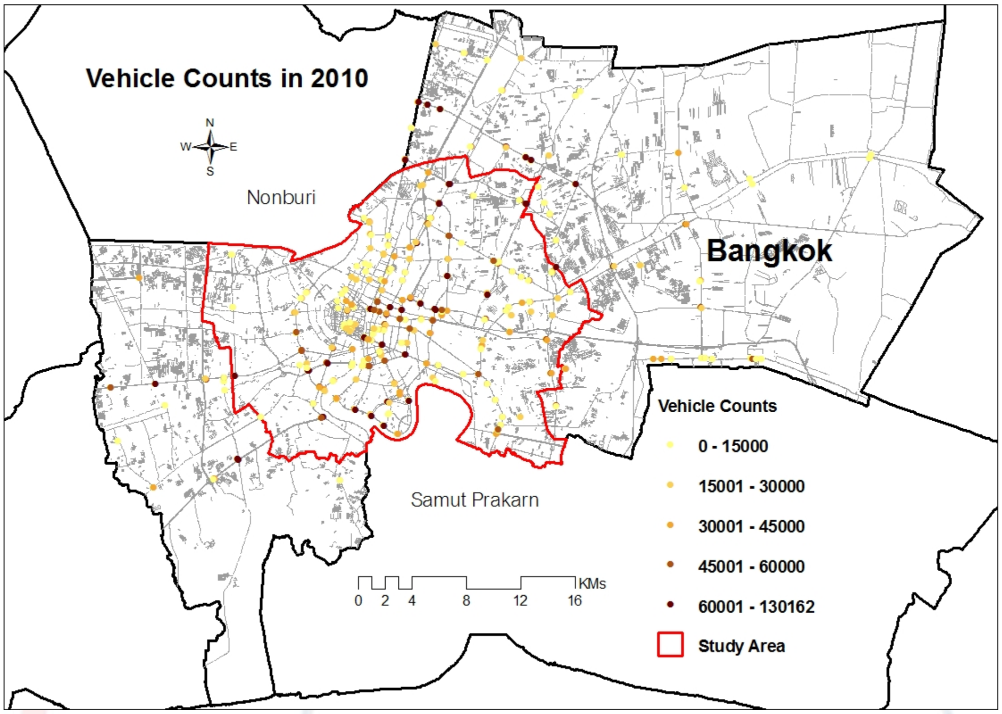

Figure 3 Traffic counts at 251 intersections in 2010.

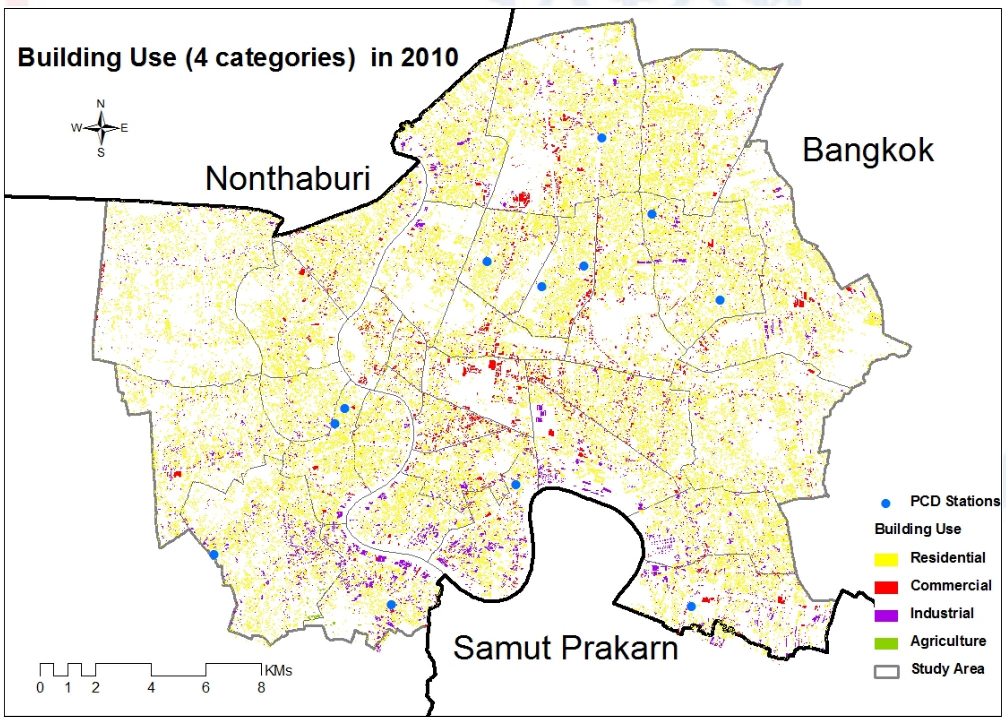

Figure 4 Building footprints categorized by building use within the study area.

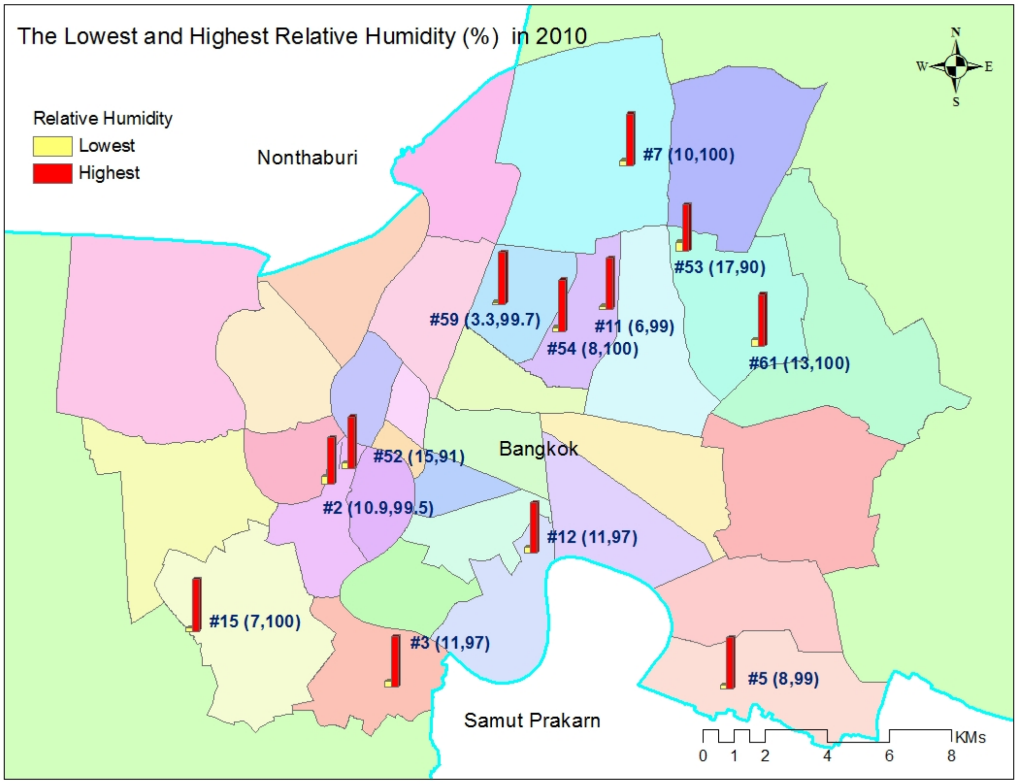

Figure 5 The lowest and highest relative humidity levels (%) in 2010.

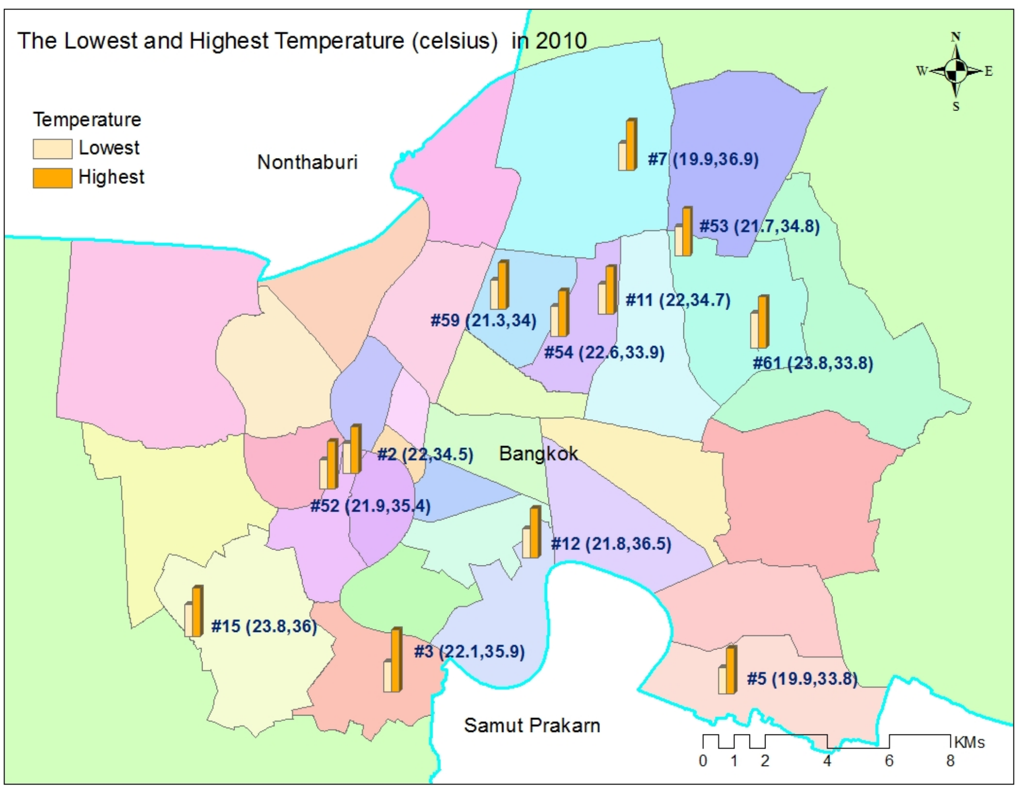

Figure 6 The lowest and highest temperature (celsius) in 2010.

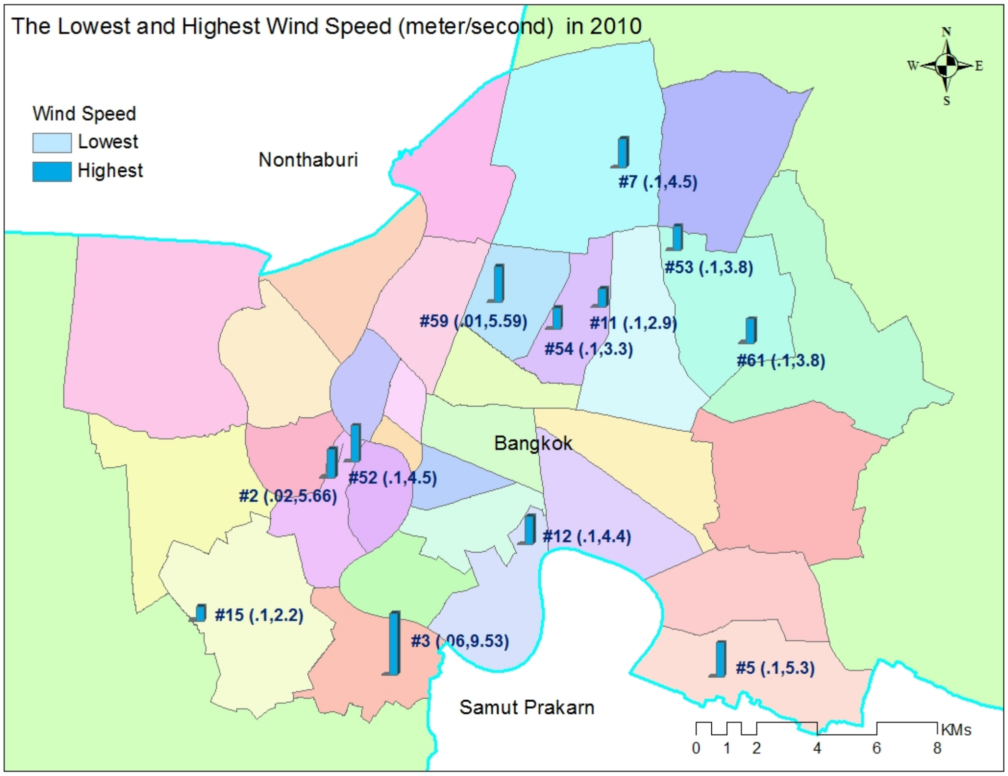

Figure 7 The lowest and highest wind speed (meter per second) in 2010.

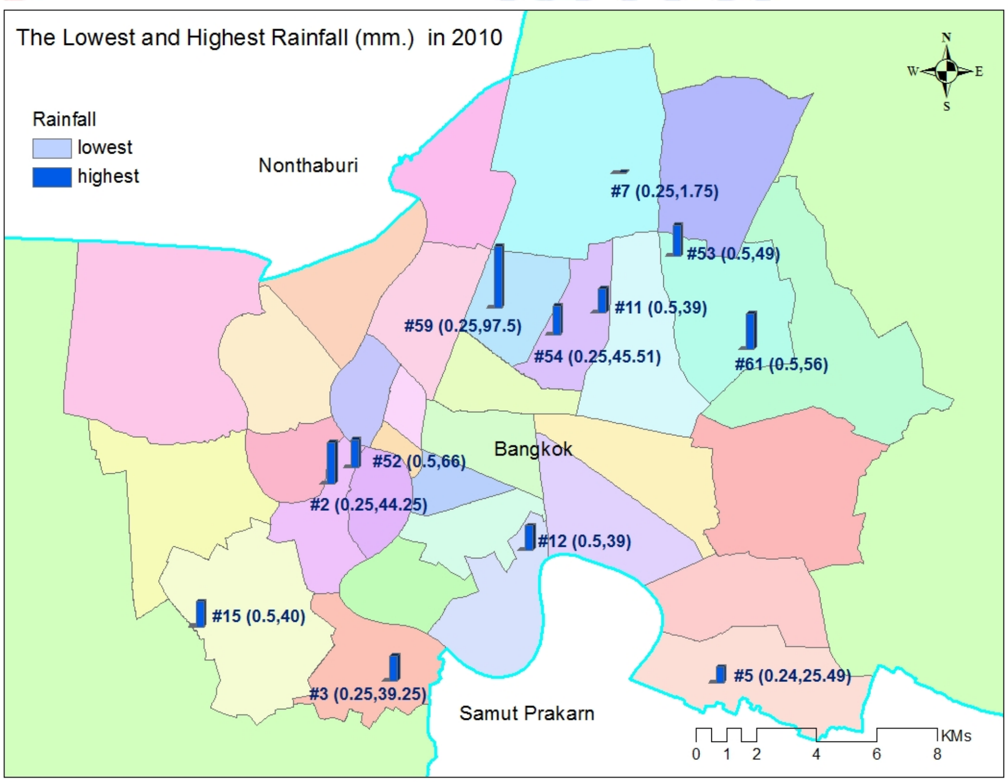

Figure 8 The lowest and highest rainfall (millimeters) in 2010.

4. Analysis and Results

In the analysis process, all data values used in the model are on monthly basis starting from January to December 2010. NO_2 concentrations are in average of hourly concentrations of each month. Values of traffic volume, road area, and land use are assumed unchanged over the year, so their values of each month are the same for each PCD station. The values of meteorological data are also in average of hourly measurements of each month. At the beginning of the analysis process, the input factors were tested from their correlation with hourly average NO_2 concentrations of each month at 12 PCD stations. If values of a factor and NO_2 concentrations had a correlation with a significant level greater than 0.05, the factor was then selected. And if the factor is not collinear with other factors, it is considered one of the input factors to the multiple regression model.

For traffic volume, road area, and land use, it was necessary to find out a distance from the 12 PCD stations that most affect the NO_2 concentrations at the stations. To achieve this, in GIS environment, the area of 1 kilometer around the stations was divided into 20 buffer zones (50 meters each). Values of each factor within 50, 100, 150, …, 1000 meters were calculated and then input to a statistical software package to find out the distance with the strongest correlation to the NO_2 concentrations. The values of each factor at the selected distance were then input to the multiple regression model.

4.1 Traffic Volume

Traffic volume on roads surrounding the 12 PCD stations were calculated based on estimated traffic counts on roads within a distance from the stations. The monthly volumes of a station were assumed the same throughout the year. It was found that traffic volume within 50 meters from the stations had the strongest correlation with NO_2 concentration at 0.391, with a signification level less than 0.01. Considering only correlation with a significant level greater than 0.05, the correlation indices were decreasing from 0.391 at 50 m to 0.204 at 150 m. At further distance, the indices were increasing from 0.158 at 400 m to 0.311 at 650 m, and then decreasing to 0.176 at 850 m. In conclusion, traffic volumes within 50 meters from the 12 PCD stations were included into the model.

4.2 Road Area

With limitation of the data received, road areas used in this research were also assumed static throughout the year. Thus, road area values of a station are the same for every month. The 50 meter buffer zones within 1 km around the 12 PCD stations were also applied with this factor to find out an appropriate distance that road areas considerably affect NO_2 concentrations. It was found that all of the distances have correlation with NO_2 concentrations with significant level greater than 0.05. The highest correlation index was found at the distance of 650 m with significant level less than 0.01.

4.3 Land Use

In this research, land use areas in 4 categories were selected: residential, commercial, industrial, and agriculture. The areas were calculated by summing up the building footprint areas in the same category. The twenty meter buffer zones were also applied to this factor. Thus, the areas of each category within each buffer zone were calculated, and then used to find out the appropriate distances for each land use category. For residential, the highest correlation index, 0.473, was found at 50 meters. For commercial, the highest correlation index was 0.489 at 150 meters. The industrial areas gave the negative correlation index at -0.312 at 650 meters. Compared to the other three land use categories, the agriculture area gave the least correlation. The highest correlation index for the agriculture was 0.267 at 600 meters. The significant levels of the correlation indices of all four categories were less than 0.01.

4.4 Relative Humidity

The averages of hourly relative humidity (%) in each month of the 12 PCD stations were calculated, and then input to the statistical calculation. It was revealed that the values of relative humidity negatively correlated with the NO_2 concentrations at the index of -0.388 with a significant level less than 0.01.

4.5 Temperature

The monthly average temperature values of the 12 PCD stations were calculated from the recorded hourly measurements. The averages, ranging from 27 to 33 degree Celsius, negatively correlated with the NO_2 concentrations at the index of -0.396 with a significant level less than 0.01.

4.6 Wind Speed

The averages of hourly wind speed (meter/second) for each month were calculated. It was found that wind speed also negatively correlated with the NO_2 concentrations at the index of -0.433 with a significant level less than 0.01.

4.7 Rainfall

The averages of hourly rainfall (mm) for each month were calculated, and then input to the statistical calculation. It was found that the rainfall also negatively correlated with the NO_2 concentration at the index of -0.211 with a significant level less than 0.01.

4.8 Multiple Regression Analysis

Since all factors correlated with the NO_2 concentrations significantly, all of them were input into the multiple regression analysis. The stepwise method was selected to perform the analysis, with the entry of 0.05 and removal of 0.10. The analysis gave the R-squared value of 0.768 and the adjusted R-squared value of 0.758. This means that 76.8% of the variation can be explained by the regression and the given input factors. The resultant regression coefficients are shown in

Table 1. The factors removed from the model are traffic volume, road area, commercial area, and agriculture area. The t-test shows that all included factors are statistically significant at the level of 0.05. Based on the resultant regression coefficients in Table 1, the regression equation was derived and shown in Equation 1.

Table 1 The resultant regression coefficients.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	160.943	10.550		15.255	0.00000
HUMIDITY	-0.648	0.079	-0.408	-8.242	0.00000
TEMPERATURE	-2.971	0.300	-0.414	-9.909	0.00000
WIND SPEED	-3.440	0.938	-0.170	-3.666	0.00035
RESIDENTIAL	0.008	0.001	0.527	12.008	0.00000
INDUSTRIAL	-0.00014	0.00002	-0.312	-7.001	0.00000
RAINFALL	-7.508	3.195	-0.119	-2.350	0.02021

$$NO_2 = 160.943 - (0.648 * Humitidy) - (2.971 * Temperature) - (3.44 * WindSpeed) + (0.008 * Residential\_D50) - (0.00014 * Industrial\_D650) - (7.508 * Rainfall) \qquad \textbf{Equation 1}$$

The regression coefficients show that relative humidity, temperature, wind speed, industrial area, and rainfall factors are negatively correlated to the NO_2 concentrations, while only the residential area factor is positively correlated. The standardized coefficients also show that the residential area factor has the most influence on the NO_2 concentrations, while wind speed and rain fall has the least influence.

5. Conclusions

The LUR model estimates the NO_2 concentrations based on a set of related NO_2-generated data at measured locations. The data may vary depending on the characteristic of the cities. Therefore, the concept of the LUR model can be applied to a city but not all the datasets may be useful. Bangkok is a big city and has not applied the LUR model for NO_2 concentration estimation yet. This research has investigated the LUR model concept and attempted to apply it on the area of Inner Bangkok. The LUR model of the inner Bangkok reveals that land use only residential and industrial area affect the NO_2 concentrations, while all meteorological data (relative humidity, temperature, wind speed, and rainfall). A few observations comparing to the previous researches are identified as follows. Firstly, while traffic volume was included in the LUR model of previous researches, it is removed from the Bangkok LUR model. This is possibly because the calculated traffic volume data used in this research was not accurate enough. More precise traffic volume data in terms of time scale should be collected to improve the model. Secondly, meteorological factors were not often used as input data from the previous researches, although a few researches suggested including them to improve the models. In this model, all four meteorological data were used and none of them were removed from the model. As the results, about seventy-six percent of the variations in NO_2 concentrations in the Inner Bangkok can be explained by the Bangkok LUR model.

Acknowledgement

The authors would like to thank Thailand Research Fund, Office of Higher Education, and Chulalongkorn University for their financial support. The authors also would like to express our appreciation to the Thai Meteorological Department, the Pollution Control Department, and the Ministry of Transport for the valuable digital data used in this project.

References

Arain, M. A., R. Blair, N. Finkelstein, J. R. Brook, T. Sahsuvaroglu, B. Beckerman, L. Zhang and M. Jerrett (2007). "The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies." Atmospheric Environment **41**: 3453-3464.

Beelen, R., G. Hoek, E. Pebesma, D. Vienneau, K. D. Hoogh and D. J. Briggs (2009). "Mapping of background air pollution at a fine spatial scale across the European Union." Science of the Total Environment **407**: 1852-1867.

Briggs, D. J., S. Collins, P. Elliott, P. Fisher, S. Kingham, E. Lebret, K. Pryl, H. V. Reeuwijk, K. Smallbone and A. V. D. Veen (1997). "Mapping urban air pollution using GIS: a regression-based approach." International Journal Geographical Information Science **11**(7): 699-718.

Gilbert, N. L., M. S. Goldberg, B. Beckerman, J. R. Brook and M. Jerrett (2005). Assessing Spatial Variability of Ambient Nitrogen Dioxide in Montreal, Canada, with a Land-Use Regression Model, Air & Waste Management Association. **55:** 1059-1063.

Hoek, G., R. Beelen, K. d. Hoogh, D. Vienneau, J. Gulliver, P. Fischer and D. Briggs (2008). "A review of land-use regression models to assess spatial variation of outdoor air pollution." Atmospheric Environment **42**: 7561-7578.

Jerrett, M., A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahsuvaroglu, J. Morrison and C. Giovis (2005). "A review and evaluation of intraurban air pollution exposure models." Journal of Exposure Analysis and Environmental Epidemiology **15**: 185-204.

Leong, S. T. and P. Laortanakul (2003). "Monitoring Framework for Nitrogen Dioxide Measurement in a Tropical City : A Bangkok Case Study." Environmental Engineering Science **20**: 557-567.

Pangsang, V. (2004). Geographic Information System for Determining Air Pollution Risk Area in the Bangkok Metropolis. Department of Geography, Faculty of Arts. Bangkok, Thailand, Chulalongkorn University. **Degree of Master of Arts in Geography:** 186.

Rosenlund, M., F. Forastiere, M. Stafoggia, D. Porta, M. Perucci, A. Ranzi, F. Nussio and C. A. Perucci (2007). "Comparison of regression models with land-use and emission data to predict the spatial distribution of traffic-related air pollution in Rome." Journal of Exposure Science and Environmental Epidemiology **17**.

Su, J. G., M. Jerrett, B. Beckerman, M. Wilhelm, J. K. Ghosh and B. Ritz (2009). "Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy." Environmental Research **109**: 657-670.

iafor