The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W8-2023
Philippine Geomatics Symposium (PhilGEOS) 2023, 6–7 December 2023, Diliman, Quezon City, Philippines

DEVELOPMENT OF LAND-USE REGRESSION MODELS FOR PARTICULATE MATTER ESTIMATION IN NATIONAL CAPITAL REGION, PHILIPPINES

R. A. B. Torres [1 *], R. V. Ramos [1, 2], B. A. B. Recto [1], A.M. Tamondong [1, 2]

[1] Training Center for Applied Geodesy and Photogrammetry, University of the Philippines, Diliman, Quezon City, Philippines – rbtorres@alum.up.edu.ph, bbrecto@up.edu.ph
[2] Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City, Philippines – (rvramos, amtamondong)@up.edu.ph

KEY WORDS: Air quality, Land-Use Regression, Particulate Matter, BuildLUR, ApplyLUR.

ABSTRACT:

Regression models are commonly used to estimate unknown variables, such as environmental parameters. Multiple Linear Regression (MLR) is one of the techniques used to model air quality and measure air pollutant concentrations. Specifically, a technique called Land-Use Regression (LUR) enables the user to generate air pollutant models using geographical layers as input parameters. The study aims to generate models for fine and coarse particulate matter ($PM_{2.5}$ and PM_{10}, respectively) using LUR for the National Capital Region in 2019. Independent variables considered in this study are road network, traffic count, Normalized Difference Vegetation Index (NDVI), population density, and elevation. The final model results showed significant estimates based on the model parameters. For $PM_{2.5}$, the model resulted in high values of R^2 and adjusted R^2 and an RMSE of 0.77 µg/m^3. For PM_{10}, model parameters showed that the generated final model for PM_{10} was significant with a 55% R^2 value. Maps were then generated using the final LUR models of $PM_{2.5}$ and PM_{10}. The models can be improved by adding more types of input variables and longer observation periods.

1. INTRODUCTION

Statistical methods are used to model unknown variables and determine their values by identifying their relationship to another variable. The most common technique is regression modeling, where the relationship between a dependent and independent variable is explained by fitting the values to a line or curve, which can then be used to estimate the values of the dependent variable. Regression models are also useful for predicting outcomes and understanding the effect of one or more independent variables by controlling each of the input variables based on the researcher's needs. The most basic type of regression analysis is linear regression analysis.

Linear regression analysis examines the relationship between a dependent variable and one independent variable by fitting the values along a straight line. It is affected by the sample size, missing data, and nature of the sample used. Specifically, a small sample size might not be enough to determine the relationship between variables with a weak relationship (Ali and Younas, 2021). Multiple linear regression (MLR) is just a variation of linear regression analysis using multiple independent variables instead of only one. Assumptions of multicollinearity, homoscedasticity, and normality of residuals must be taken into account to accurately assess the models. Multicollinearity is the close correlation of the independent variables in a regression model, while homoscedasticity is the case where the variance of the error or residual is constant (Aarthi et al., 2020).

Land-Use Regression (LUR) is a type of MLR used to analyze pollution patterns and estimate pollutant concentrations in a heavily urbanized area. Moreover, LUR is an algorithm that utilizes geographic and urban setting predictors in estimating and analyzing ambient air pollution for applications such as human pollution exposure assessment and public health studies (Hoek et al., 2008; Ryan and LeMasters, 2007). LUR relies on Geographic Information Systems (GIS) for the input geographical and environmental variables. LUR models involve various input variables such as road networks, traffic volumes, population data, land-use, physical geography, and meteorology to generate air quality parameter estimates accurately based on the characteristics of the target area (Masiol et al., 2018).

Various studies have shown the results of regression techniques for measuring air pollutant concentrations. In 2019, a study used Moderate Resolution Imaging Spectroradiometer (MODIS)-derived aerosol optical depth (AOD) data at 3 km and 10 km, together with meteorological parameters such as planetary boundary layer, surface temperature, and surface wind speed, to estimate $PM_{2.5}$ concentrations over Turkey (Zeydan and Wang, 2019). The best model resulted in an R^2 of 0.61 with $p < 0.001$ and an RMSE of 0.337 µg/m^3. In 2020, a study tested simple linear regression, multiple linear regression, log-linear regression, and conditional-based MLR to estimate $PM_{2.5}$ concentrations from satellite-derived AOD in Agra and Rourkela, India, from 2015 to 2019 (Gogikar et al., 2021). Results showed that the models generated were all significant using Model II or MLR. Specifically, calculated R resulted in being statistically significant for both sites during training and validation.

For LUR, a study used road length, vehicle density, land use, and population density as input features of regression modeling to determine the NO, NO_2, $PM_{2.5}$, and light absorption estimates in Vancouver, Canada (Henderson et al., 2007). Specifically, the researchers used two road types (highways and major roads), two vehicle types (automobiles and trucks), and five land use types (residential, commercial, governmental, industrial, and open). The resulting adjusted R^2 of the models showed values of 0.39 to 0.62, with NO maps showing a more heterogeneous distribution than NO_2. Another study gathered predictor variables that were divided into data categories: (1) weather parameters; (2) atmospheric sounding indices; (3) land use; (4) road traffic density; (5) emission sources of marine and power stations; (6) natural geography; (7) and urban surface form (Shi Y. et al., 2018). For the study, MODIS AOD products from 2003 to 2015

* Corresponding author

were used, together with an extensive LUR variable database containing 294 variables. The resulting improved annual and seasonal AOD-LUR models showed significant improvement in the model-adjusted coefficient of determination by approximately 20-30%. Calibrated AOD-$PM_{2.5}$ models showed an adjusted R^2 value of 0.72 using the geographically and temporally weighted regression (GTWR).

LUR includes several pre-processing and modeling processes that can be tedious and time-consuming. One of the tools that is reliable and efficient for performing LUR modeling is the XLUR Tool, implemented through Python scripts. A research in 2021 demonstrated the use of XLUR tool in generating PM_{10} models in Greater Manchester (Motler and Lindley, 2021). The study showed how the model was built through standard publishing protocols and how accurate the models were. Moreover, the researchers used input variables such as Corine land-use, Openstreetmap land-use, population density, road length (major and minor), vehicle traffic, and altitude. The final model included the number of buses on the nearest major road, the number of heavy vehicles on the nearest major road * inverse distance to the nearest major road, the Corine natural land use area in a 1000 m buffer, the Y coordinate, and the Corine High density residential land use in a 100 m buffer.

The study aims to perform LUR analysis to generate $PM_{2.5}$ and PM_{10} models and maps for the National Capital Region in 2019 using road network, traffic count, NDVI, population density, and elevation as the independent variables.

2. METHODOLOGY

2.1 Study Area

Figure 1 shows the National Capital Region located in the Philippines using © Microsoft Bing VirtualEarth (2023) and the air quality monitoring stations within the region. The National Capital Region (NCR) includes the capital city of the Philippines and is home to the residential, commercial, and industrial centers in the country. Specifically, major parts of the region involve urban development activities and industrial activities, while other areas are suitable for crop cultivation, such as the Marikina Valley areas. The region near Manila Bay called the Coastal Margin includes activities for offshore fisheries, fishponds, and reclamation projects, while the Marikina River provides water for industrial uses and discharge.

According to the Coronas classification, NCR is identified as Climate Type 1, which means the area has two distinct seasons: dry and wet. The dry season lasts from November to April, while the wet season lasts from May to October, with the months of June to September as the maximum rain period (PAGASA, 2014; Tolentino et al., 2016).

A study done by Greenpeace, a global organization that aims to address and solve environmental problems through peaceful protest and creative confrontation, showed that the average pollution level in the region reached 17.6 μg/m³ in 2019, which even peaked at 117 μg/m³ on the New Year's Eve. Moreover, the study showed that the region's air pollution contributed to at most 27,000 deaths in 2018 (Greenpeace, 2020).

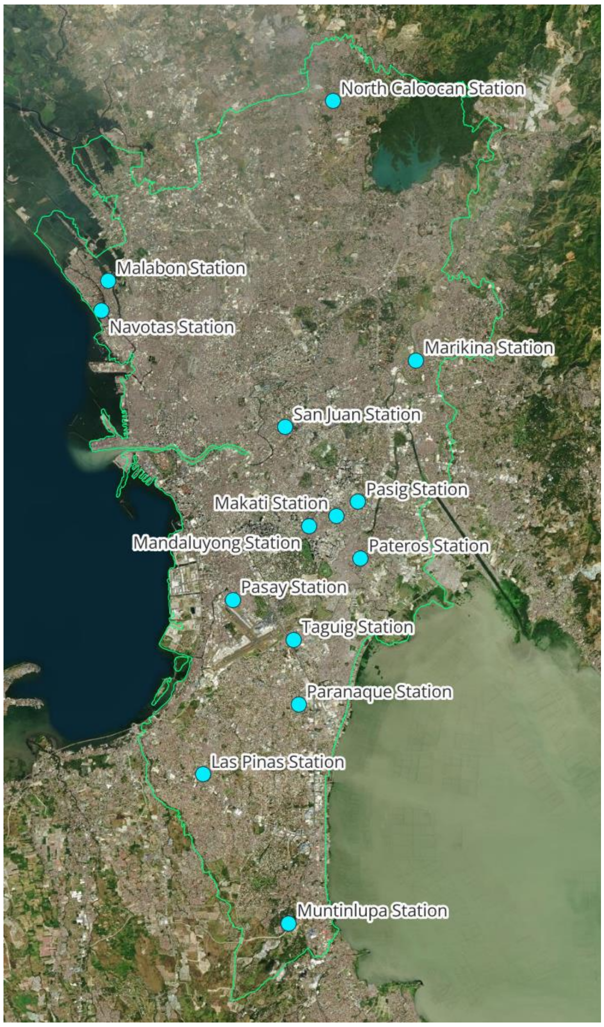

Figure 1. The National Capital Region and PM monitoring stations of the region (Base map source: Microsoft Bing VirtualEarth, 2023).

2.2 Model Variables

XLUR allows the use of both raster and vector data for the input variables. Five datasets were used as the independent variables for the LUR modeling. The road network data line shapefile and population density raster were gathered from an online database. Roads were categorized based on their classification as primary, secondary, or tertiary roads. The traffic count was derived from the annual average daily traffic (AADT) file obtained from the Metropolitan Manila Development Authority (MMDA). Traffic count was categorized based on the vehicle type, namely: cars, public utility jeepneys (PUJ), utility vehicles (UV), taxis, public utility buses (PUB), trucks, trailers, motorcycles (MC), and tricycles. The normalized difference vegetation index (NDVI) was derived from Sentinel-3 images, which indicate how healthy and dense vegetation is in urban areas. Elevation data was gathered from the elevation models of the National Mapping and Resource Information Authority (NAMRIA).

2.3 Buffered-based Analysis

Figure 2 shows the general methodology for the estimation of $PM_{2.5}$ and PM_{10} using LUR for NCR, 2019. For raster-type input variables, the value of the raster cells is directly extracted and used for the generation of the model; however, vector data

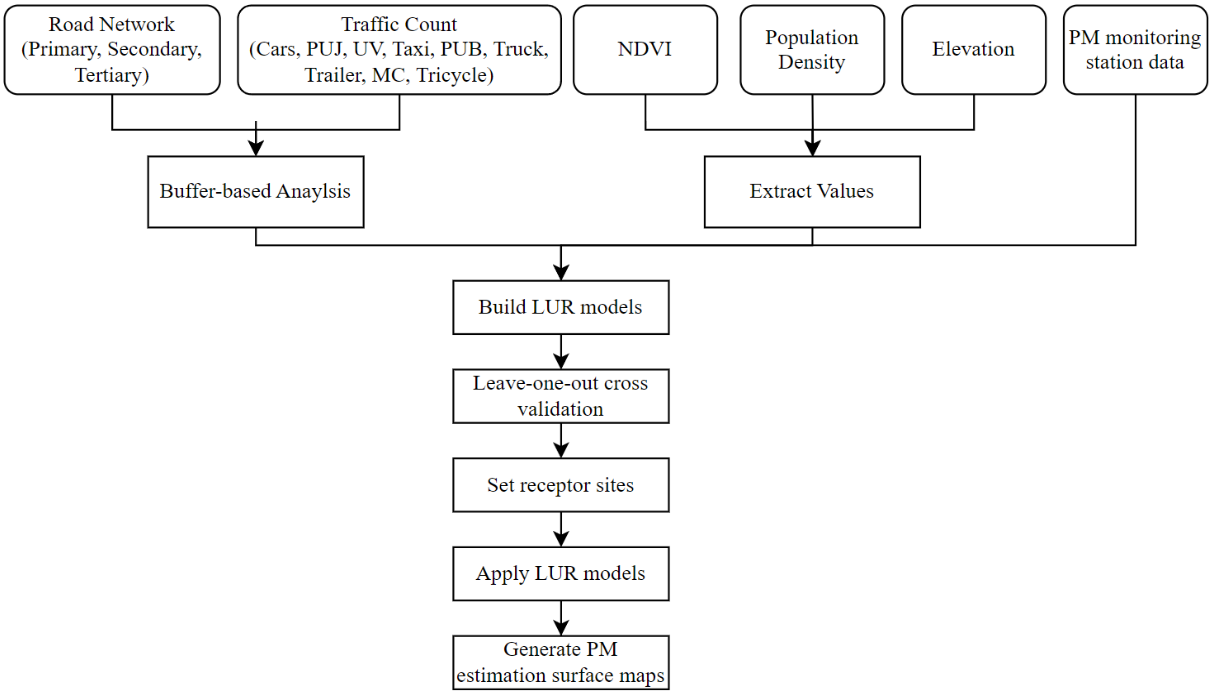

Figure 2. General methodology for the generation of LUR models for $PM_{2.5}$ and PM_{10}.

predictor variable values are extracted through buffers or nearest-distance set by the user. The vector file should contain a category field based on the feature and may contain numeric fields. The aggregation methods can be total length, length-weighted value, the product of length and value, sum of values, or mean or median of values. If the chosen aggregation method is anything besides total length, a numeric field is required. In this study, buffer-based predictors were derived from the road network and traffic count line shapefiles. Buffers of 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 meters were generated. For the road network, the total length for each buffer and each category was calculated and used as input parameters. For the traffic count, the length-weighted value was calculated by multiplying the count value of the line by the ratio of the length of the line inside the buffer over the total length of the line.

2.4 Building LUR models

The LUR modeling workflow was accomplished using XLUR Wizard. XLUR Wizard is a tool within the ArcGIS software that provides step-by-step functions for building and applying LUR models. It is a Python-based LUR modeling toolkit that consists of two main scripts, the BuildLUR and the ApplyLUR scripts.

In general, the BuildLUR tool specifies the input data that was provided by the user, containing monitored data points, study area polygons, and potential predictors. The script also creates a new file geodatabase to avoid overwriting the original data. Afterwards, BuildLUR will perform spatial analyses set by the user, such as buffer, intersect, spatial join, and nearest distance, which are exported to a SQLite database. Then, outputs are summarized and structured into formats suitable for statistical analysis. MLR with supervised variable selection based on the ESCAPE methodology is then performed using the statmodels package (Eeftens et al., 2012). Finally, final models are stored in the SQLite database together with a processing log, descriptive statistics, and model diagnostics saved in the user-defined output folder.

Specifically, the first dialog box of the BuildLUR Wizard is concerned with the directories of input and output geodatabases, coordinate system assignment, setting of the study area, and project name. The user must assign a coordinate system by inputting the respective WKID number. The study area must be a polygon shapefile. The second box of the wizard is dedicated to the outcomes. In this box, the dependent variable shapefile is selected, and the concentration field is specified. Each observation point must be a unique location, or no spatial duplicates should be present within the extent of the study area. Predictor variables for the X coordinate and Y coordinate of each site are also automatically added as variables. Missing, zero, or negative values will result in a warning message that can be accepted by the user. At least eight values or observation points are required to proceed to the next step. The dependent variable or monitoring station data were obtained from the Department of Natural Resources and Environment – Environmental Management Bureau (DENR-EMB). Eight and thirteen observation points for the ground fine $PM_{2.5}$ and coarse PM_{10} data, respectively, were gathered from the Continuous Air Monitoring Stations (CAMS) of the department. Yearly aggregates for each station were computed from the hourly ground monitoring station data for 2019.

The next step specifies the input variables to be used in the LUR modeling. In this step, the buffer-based, distance-based, or raster-based predictors are selected. Moreover, for each predictor variable, the direction of effect is set to either a positive or negative direction of effect based on the user's priori assumption for each predictor variable. The direction of effect is important in the model selection criteria in the statistical analysis. Afterwards, the model can be generated based on the type of model the user selected.

A summary of the input variables used in building the LUR models can be seen in Table 1. The road network is assumed to have a positive direction of effect, as more roads around a point imply more mobile transport, causing an increase in air pollutant concentrations. Similarly, traffic count is assumed to

Variable	Category	Unit	Buffers (m)	Assumed direction of effect
Road Network	Primary Secondary Tertiary	Meters	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000	Positive
Traffic Count	Cars PUJ UV Taxi PUB Truck Trailer MC Tricycle	No. of vehicles	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000	Positive
NDVI	-	-	-	Negative
Population Density	-	People per sq. km	-	Positive
Elevation	-	Meters	-	Negative
$PM_{2.5}$ and PM_{10}	Dependent Variable	μg/m^3	-	-

Table 1. Input variables for the LUR modeling of $PM_{2.5}$ and PM_{10}.

have a positive direction of effect. Population density is given a positive direction of effect as highly populated areas commonly result in air pollution hotspots. On the other hand, NDVI is assumed to have a negative direction of effect, as a more vegetative area indicates lower air pollutant concentration. Lastly, it is assumed that the air pollutants are more concentrated near the roads where vehicle exhausts are present.

2.5 Applying LUR models

The ApplyLUR tool starts with the user specifying locations where predicted values will be extracted, called receptor sites. Receptor sites can come from specified points defined by the user, generated randomly or at regular intervals. The final models stored in the database are then inspected to carry out the necessary spatial analyses based on the predictor variables of the final models. Outputs are exported to a new SQLite database, where they are summarized and structured once again into the necessary predictor variables. The final models are then applied to predict a value for each receptor site. Finally, predicted values can be mapped and accessed in the file geodatabase.

Specifically, the first dialog box of the ApplyLUR Wizard is for the specification of the generated model by loading the file geodatabase as well as the SQLite database and selecting the LUR model inside it. After setting the output name, the receptor points must be identified. The points can be loaded using a feature class created by specifying the horizontal and vertical distances between the points or randomly scattered within the study area. In this study, a 1000-meter distance between the receptor points was selected. Afterwards, the model can be applied, and estimates for the receptor points will be generated.

3. RESULTS AND DISCUSSION

LUR models were generated for $PM_{2.5}$ and PM_{10} for 2019 using road network, traffic count, NDVI, population density, and elevation as the independent variables. A total of 121 independent variables were used to determine the best model to estimate PM concentrations. Table 1 shows the model parameters for the starting model and final model for $PM_{2.5}$ estimation using LUR. The coefficient of determination (R^2) is a statistical parameter that shows the goodness of fit of the data to the regression model through the proportion of variance in the dependent variable that is explained by the independent variable. On the other hand, adjusted R^2 only considers the effect of significant independent variables on the regression model and shows if the model is still accurate even when adding multiple independent variables. The closer the value of R^2 is to 1, the better the fit. F-statistic is a measure used to check if the regression coefficients in the models are significant. The p-value indicates if the regression model fits the data better than a model with no independent variables; hence, a p-value less than the determined significance level means that the independent variables in the model improved the fit. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are criteria to determine which model fits the data the best by testing which model stayed accurate even with the addition of independent variables to the model. The lower the AIC or BIC, the better the model.

For $PM_{2.5}$, the starting model resulted in an R^2 value of 0.768 and an adjusted R^2 of 0.729 with the best starting predictor of tertiary-type roads with a 300-meter buffer. Sixteen (16) intermediate models were then tested and evaluated until there were no significant changes and all p-values were less than 1. In general, the resulting parameters for the final model showed better fit than the starting model. The final model results showed significant estimates based on the model parameters, as shown in Table 2. Specifically, the model resulted in high values of R^2 and adjusted R^2 and an RMSE of 0.77 μg/m^3. Overfitting might be present as the R^2 values were very high; however, when residuals were inspected for each monitoring station, low residual values with magnitudes of 0.001 to 0.29 μg/m^3 were observed. P-values less than 0.01 mean that the independent variables are significant with a significance level of 1%. The AIC and BIC of the final models are much lower than the starting model, indicating that the final model is the best model for estimating $PM_{2.5}$. The final predictor variables for $PM_{2.5}$ are tertiary-type road networks with 300m buffer, population density, secondary-type road networks with 100m buffer, and elevation. All predictor variables showed a positive correlation with $PM_{2.5}$ and acceptable VIF of less than 10, which indicates that multicollinearity is not present in the model and each individual parameter is not closely related to each other.

Parameter	Starting Model	Final Model
R^2	0.768	1
Adj. R^2	0.729	0.999
F-statistic	19.85	1585
p-value (F-statistic)	0.00431	0.000631
AIC	49.79	3.193
BIC	49.95	3.67

Table 2. Starting and final model parameters for $PM_{2.5}$ using LUR.

Table 3 specifically shows the final predictor variables along with their coefficient, standard error, t-value, and p-value. The T-value and p-value are statistical measures to determine if the remaining independent variables have a significant effect on the model. The table shows that all independent variables in the final model are significant at 5% significance level.

	Coefficient	Std. Error	t	P>|t|
Intercept	626.1976	34.447	18.179	0.003
Road Network - Tertiary: 300 m	0.0119	0	38.758	0.001
Population Density	0.0001	2.06E-05	6.035	0.026
Road Network - Secondary : 1000 m	0.0019	0	17.766	0.003
Y coordinate	-0.0004	2.11E-05	- 17.908	0.003
Elevation	-0.1685	0.018	-9.118	0.012

Table 3. Final model coefficient parameters for $PM_{2.5}$ using LUR.

Figure 3 shows the predicted vs. observed plot of $PM_{2.5}$ using Leave One Out Cross-Validation (LOOCV). LOOCV is a cross-validation technique where each individual observation is used as a validation set while the rest of the observations are used as a training set. The LOOCV shows high R^2 and very low RMSE results, indicating an accurate model.

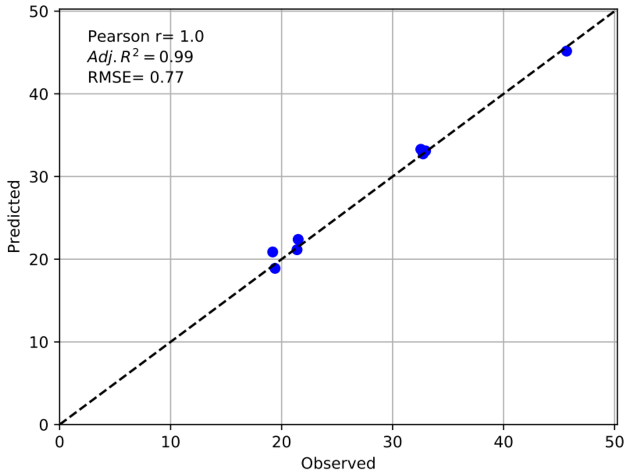

Figure 3. Predicted vs. Observed $PM_{2.5}$ concentrations using LUR and LOOCV.

Table 4 shows the final model parameters for PM_{10} using LUR. The starting PM_{10} model resulted in an R^2 of 0.552 and an adjusted R^2 of 0.51 with the starting predictor variable of population density. One hundred eighty-one (181) intermediate models were tested, with the resulting intermediate model including population density, secondary-type roads with a 300-meter buffer, and elevation as predictor variables. However, only population density was significant to the model with a p-value less than 0.1; therefore, the starting model is the final model. Model parameters showed that the generated final model for PM_{10} was significant with a 55% R^2 value. Residual values were checked for each monitoring station. Residual values resulted in magnitudes relatively low for PM_{10} with a range of 0 to 100 $\mu g/m^3$. As low as 0.08 $\mu g/m^3$ residual was observed in the Pateros station; however, high residual values can be observed in the Pasay and Pasig stations.

Parameter	Value
R^2	0.552
Adj. R^2	0.551
F-statistic	13.56
p-value (F-statistic)	0.00361
AIC	94.42
BIC	95.55

Table 4. Final model parameters for PM_{10} using LUR.

Table 5 shows the coefficients of the generated best model for PM_{10}. The table showed a significant independent variable at significance level of 1%. Figure 4 shows the LOOCV predicted vs. observed PM_{10}. It shows an adjusted R^2 of 37% and an RMSE of 9.01.

	Coefficient	Std. Error	t	P>|t|
Intercept	26.2568	6.367	4.124	0.002
Population Density	0.0007	0	3.682	0.004

Table 5. Final model coefficient parameters for PM_{10} using LUR.

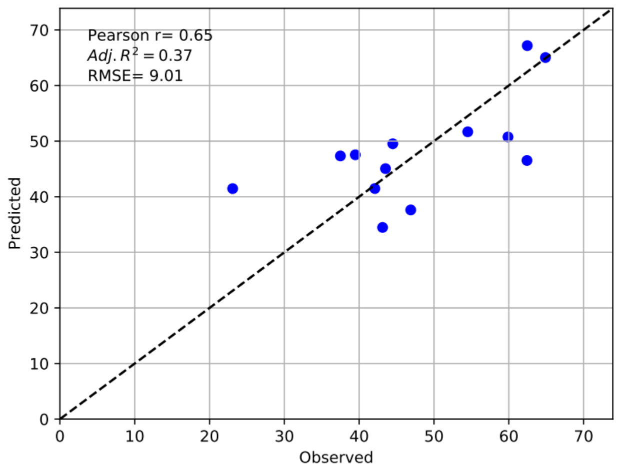

Figure 4. Predicted vs. Observed PM_{10} concentrations using LUR and LOOCV.

Maps were then generated using the final LUR models of $PM_{2.5}$ and PM_{10} as shown in Figures 5 and 6. High concentrations of PM are shown in bright red to white, while low concentration values are shown in dark red-violet to black. Even though the generated PM_{10} map was only estimated using population density, the map still showed a similar pattern to the $PM_{2.5}$, displaying the same areas with the highest concentration. Maps showed high concentrations in the Manila area (in Figure 7), Quezon City Circle area (in Figure 8), Navotas fish port (in Figure 9), and the roads near National Road 1 and Skyway (in Figure 10). These are areas where there are highly dense population, commercial and industrial activity, and mobile transport. Specifically, in 2019, the satellite navigation company, TomTom NV, determined Manila as the second world's most traffic-congested city (Esguerra, 2020).

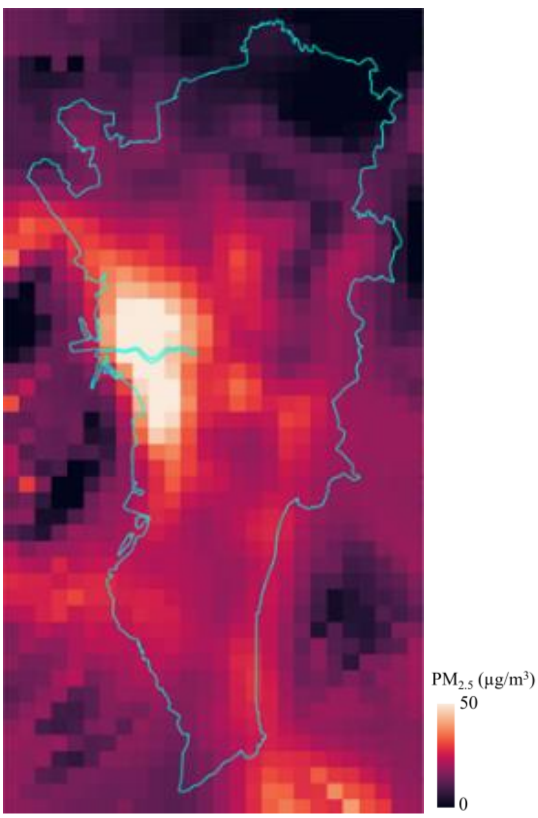

Figure 5. Generated $PM_{2.5}$ using LUR modeling.

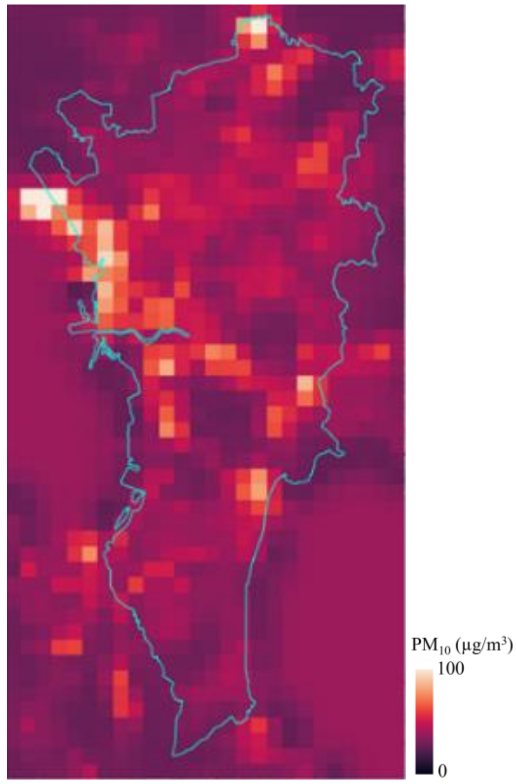

Figure 6. Generated PM_{10} map using LUR modeling.

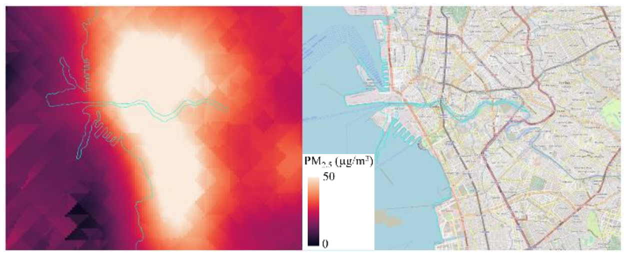

Figure 7. Generated $PM_{2.5}$ map (left) and OpenStreetMap View (right) of Manila City area.

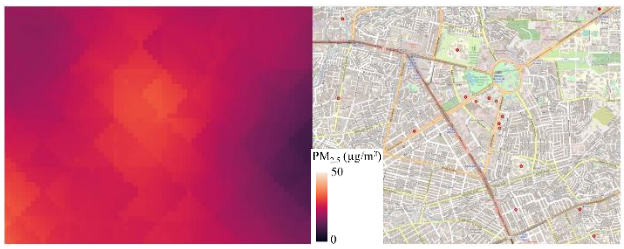

Figure 8. Generated $PM_{2.5}$ map (left) and OpenStreetMap View (right) of Quezon City area.

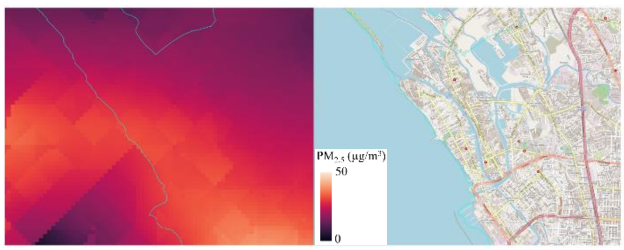

Figure 9. Generated $PM_{2.5}$ map (left) and OpenStreetMap View (right) of Navotas Fishport area.

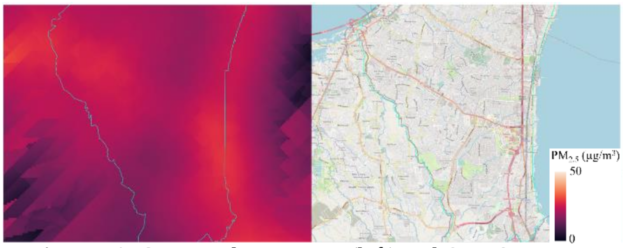

Figure 10. Generated $PM_{2.5}$ map (left) and OpenStreetMap View (right) of Alabang area.

4. CONCLUSIONS

$PM_{2.5}$ and PM_{10} models and maps were generated using LUR modeling with input population density, elevation, NDVI, annual average daily traffic, and road network as predictor variables. $PM_{2.5}$ final model resulted in a high R^2 value, a low RMSE, and very low residuals when compared to the observed $PM_{2.5}$ values of the monitoring stations. Final model independent variables include road network – tertiary: 300m, road network – secondary: 1000m, population density, and elevation. PM_{10} resulted in a lower R^2 value but still showed relatively small RMSE values. Moreover, the resulting final model independent variable only includes population density. This might imply that the PM_{10} model has difficulty taking into account the larger range of PM_{10} which is greater than 100 µg/m^3. The maps generated can be used to initially determine areas with high concentrations and hotspots. Generated PM maps showed that the area of Manila City had the highest concentration in the region, which might be due to its nature of being highly urbanized and densely populated. Additional input parameters, such as meteorological data and land-use, as well as additional observation periods, can be used for future studies to improve the models.

ACKNOWLEDGEMENTS

This research was done as part of the Ambient Air Remote Sensing, Modeling, and Visualization Environment (Project AiRMoVE). The Project was implemented by the University of the Philippines Training Center for Applied Geodesy and Photogrammetry (TCAGP), through the support and funding of the Department of Science and Technology (DOST) of the Republic of the Philippines and the Philippine Council for Industry, Energy, and Emerging Research and Development (PCCIERD).

REFERENCES

Aarthi, A., Gayathri, P., Gomathi, N. R., Kalaiselvi, S., Gomathi, Dr. V., 2020. Air Quality Prediction Through Regression Model. *International Journal of Scientific and Technology Research*, 9(3), 923–928.

Ali, P., Younas, A., 2021. Understanding and interpreting regression analysis. *Evidence Based Nursing*, 24(4), 116–118. https://doi.org/10.1136/ebnurs-2021-103425.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., … Hoek, G., 2012. Development of land use regression models for PM2.5, pm2.5 absorbance, pm10 and pmcoarse in 20 European study areas; results of the escape project. *Environmental Science & Technology*, 46(20), 11195–11205. https://doi.org/10.1021/es301948k.

Esguerra, D. J., 2020. Metro Manila Traffic is 2nd worst in the world-report. INQUIRER.net. https://newsinfo.inquirer.net/1221425/metro-manila-traffic-is-2nd-worst-in-the-world-report.

Gogikar, P., Tripathy, M. R., Rajagopal, M., Paul, K. K., Tyagi, B., 2020. PM2.5 estimation using multiple linear regression approach over industrial and non-industrial Stations of India. *Journal of Ambient Intelligence and Humanized Computing*, 12(2), 2975–2991. https://doi.org/10.1007/s12652-020-02457-2.

Henderson, S. B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic-related Nitrogen Oxides and fine particulate matter. *Environmental Science & Technology*, 41(7), 2422–2428. https://doi.org/10.1021/es0606780.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of Outdoor Air Pollution. *Atmospheric Environment*, 42(33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

Korek, M., Johansson, C., Svensson, N., Lind, T., Beelen, R., Hoek, G., Pershagen, G., Bellander, T., 2016. Can dispersion modeling of air pollution be improved by land-use regression? an example from Stockholm, Sweden. *Journal of Exposure Science & Environmental Epidemiology*, 27(6), 575–581. https://doi.org/10.1038/jes.2016.40.

Masiol, M., Zíková, N., Chalupa, D. C., Rich, D. Q., Ferro, A. R., Hopke, P. K., 2018. Hourly land-use regression models based on low-cost PM monitor data. *Environmental Research*, *167*, 7–14. https://doi.org/10.1016/j.envres.2018.06.052.

Mölter, A., Lindley, S. 2021. Developing land use regression models for environmental science research using the XLUR tool – more than a one-trick pony. *Environmental Modelling and Software*, 143, 105108. https://doi.org/10.1016/j.envsoft.2021.105108.

Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA), 2014. Climate of the Philippines. pagasa.dost.gov.ph (1 October 2023).

Ryan, P. H., LeMasters, G. K., 2007. A review of land-use regression models for characterizing Intraurban Air Pollution Exposure. *Inhalation Toxicology*, 19(sup1), 127–133. https://doi.org/10.1080/08958370701495998.

Shi, Y., Ho, H. C., Xu, Y., Ng, E., 2018. Improving satellite aerosol optical depth-PM2.5 correlations using land use regression with microscale geographic predictors in a high-density urban context. *Atmospheric Environment*, 190, 23–34. https://doi.org/10.1016/j.atmosenv.2018.07.021.

Tolentino, P. L., Poortinga, A.s, Kanamaru, H., Keesstra, S., Maroulis, J., David, C. P., Ritsema, C. J. 2016. Projected impact of climate change on hydrological regimes in the Philippines. PLOS ONE, 11(10). https://doi.org/10.1371/journal.pone.0163941.

Zeydan, Ö., Wang, Y., 2019. Using modis derived aerosol optical depth to estimate ground-level PM2.5 concentrations over

Turkey. *Atmospheric Pollution Research*, 10(5), 1565–1576. https://doi.org/10.1016/j.apr.2019.05.005.