Estimation of Roadside Particulate Matter Using Traffic and Meteorological Conditions

Mary Grace Malana[1*], Edgar Vallar[2], Maria Cecilia Galvez[2], and Joel Ilao[1,*]
[1] *Computer Technology Department, De La Salle University*
[2] *Physics Department, De La Salle University*
mary_grace_malana@dlsu.edu.ph, joel.ilao@dlsu.edu.ph

Abstract: In the National Capital Region of the Philippines, air pollution levels are 70% higher than what is considered safe by WHO air quality guidelines. One main cause of this is the pervasive use and the increasing number of vehicles in the region. In other countries, monitoring stations and emissions models have aided air quality planning and policy making. However, acquiring and maintaining air quality monitoring stations are costly. Usage of an existing vehicle emissions model for quantifying high-quality emissions measurements requires in-depth adjustments and calibration of the model. In this paper, the development of a roadside air quality estimation system that considers the context of Philippine land transport is explored. In order to do this, statistical and machine learning techniques were used to model $PM_{2.5}$. The vehicle counter and classifier of the Vision-based Vehicle Monitoring System (VEMON) project, and meteorological and air quality measurements along the De La Salle University-Manila, Taft Avenue, Manila were used for the modeling.

Key Words: Air Quality Estimation; Particulate Matter; Roadside Pollution; Traffic Emissions

1. Introduction

An estimated three million people die yearly due to air pollution, according to the World Health Organization (2016). The air quality in many countries has greatly deteriorated posing numerous health risks to people. Because of this, people are at risk of developing several diseases which are linked with exposure to ambient air pollution. Among the countries in Asia, the Philippines has the highest road transport PM emissions per capita (Clean Air Asia, 2012). The problem in air pollution is apparent in urban areas in the country; in Metro Manila, the annual average Particulate Matter (PM) concentration is 70% more than WHO's recommended PM concentration level (Department of Environment and Natural Resources, 2017). A recent study showed that vehicle fleet and volume greatly influence black carbon, a toxic air pollutant, mass concentrations in Metro Manila (Alas et al., 2018). Because of the air conditions, various policies that affect road automobiles were ratified in the country (Philippine Congress, 1999; Department of Environment and Natural Resources, 2015; Department of Transportation, 2017;). In formulating these kinds of policies, resources on air quality are valuable. Rich and extensive air quality data can aid in making data-backed decisions and interventions.

There are different approaches and techniques for air pollution monitoring. One way is through air quality monitoring stations. Monitoring stations, in general, produce relatively accurate air pollution data but acquiring and maintaining them are costly; hence, they are sparsely distributed and may not be representative of the air quality condition that most of the citizens are exposed to (Doering, 2011). Another air pollution monitoring approach is by mobile sensors that collect air quality data around the city. This approach can be logistically challenging, and the sensors can still be costly. Air quality monitoring can be enhanced by using emission models. Traffic emissions models are used for calculating the amount and rate of air contaminants released on the road by vehicles. The challenge of using an existing traffic emissions model in the context of the Philippines is that it would require in-depth adjustments and calibration (H. B. Liu, Wang, Chen, & Han, 2013).

Recently, there is a trend of using data-driven models in many applications. Statistical and machine learning techniques can be utilized for air pollution monitoring. These techniques derive estimates of the functional relationships among variables from available data, unlike existing traffic emissions models which have strict data requirements that may not be readily available and easily accessible. With a problem that is apparent worldwide and a growing stack of statistical and machine learning techniques, there have been numerous studies on air quality prediction and estimation. Despite the availability of these studies and the urgency of the problem in the Philippines, there are only a few studies that used statistical and machine learning methods in the context of the urban air quality estimation in the country (Kecorius et al., 2017).

Traffic data can be obtained through different means. One way is with an Intelligent Transport System (ITS), which takes advantage of available technologies to solve transportation problems, among which is to provide transportation-related data for users and other stakeholders. In De La Salle University, a vision-based ITS called Vision-Based Vehicle Counter for Traffic Monitoring (VEMON) is currently being developed. One of the modules in the system detects, counts, and classifies vehicles along Philippine roadways (Batingal, Camering, Naguit, Dy, & Ilao, 2016). Traffic data collected from existing ITS presents an opportunity to estimate air pollution levels. Chang et al. (2013) explored the use of data derived from an ITS to estimate real-time carbon dioxide emissions from traffic.

In this paper, in order to estimate air quality data at a low cost and leverage on the existing network of roadside surveillance technologies, the development of a roadside air quality estimation system which is appropriate for the Philippine land transport context using features of an existing vehicle monitoring system was explored.

2. METHODOLOGY

2.1 Data Collection

Street view videos from cameras positioned in front of the DLSU Manila campus were collected from the University's live streaming facility, Archer's Eye. The videos were processed using the vehicle classifier of VEMON. Air quality and meteorological data were acquired using DLSU EARTH Laboratory sensors. More details on the data is shown in Table 1. Missing data were noted on weather and air quality data collected on June 5-6, and June 21-22.

Table 1. Dataset Description

Data	Description	Dates covered
Weather and Air Quality	Includes temperature, PM_1, $PM_{2.5}$, and PM_{10}.	June 5-30, 2018. 24 hours with missing data
VEMON Vehicle Classification Results	Includes Vehicle counts per class: Bus, Jeep, Motor, SUV, Sedan, Truck, Van, Non-vehicle	June 4-30, 2018, 6 AM-6 PM

Since the $PM_{2.5}$ levels estimator should work using readily available data such those that were automatically generated by existing systems, we purposely trained the model from VEMON generated vehicle classification data. VEMON can only process daytime videos; hence, the vehicle count data is only limited from a 6 AM-to-6 PM time window.

2.2 Data Preparation

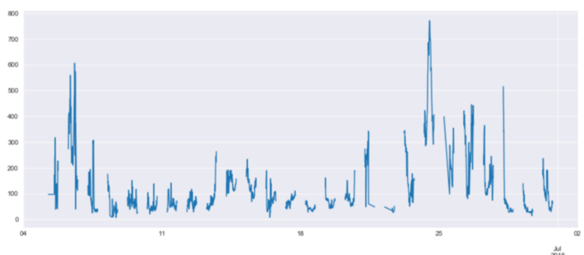

Fig. 1. Actual $PM_{2.5}$ values from January 5 to June 30

Prior to training for the air quality estimation model, the datasets must undergo different data preprocessing steps. The data were cleaned, normalized, and aggregated. The final dataset contained data from January 5 to June 30, 2018, 6 AM-6 PM intervals. Fig. 1 shows the actual $PM_{2.5}$ levels for a 3-week interval after data cleaning.

2.3 Implementation

We explored the use of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) in building the air quality models. MLR is a statistical approach used to describe simultaneous associations of several predictors to the response variable under the assumption of linear relationship (Rencher & Schaalje, 2008). Unlike MLR, SVR can handle nonlinear relationships. It also simultaneously optimizes models in terms of minimizing prediction error and complexity (Sánchez, Nieto, Fernndez, del Coz Daz, & Iglesias-Rodrguez, 2011). As such, a study conducted in Spain was developed to predict air quality index using meteorological and air quality (Sánchez et al., 2011).

In this study, Python libraries and modules were used to build MLR and SVR models.

2.4 Validation

The performance of each model was validated using Pearson's R and Mean Square Error.

3. RESULTS AND DISCUSSION

In this section, model results are discussed.

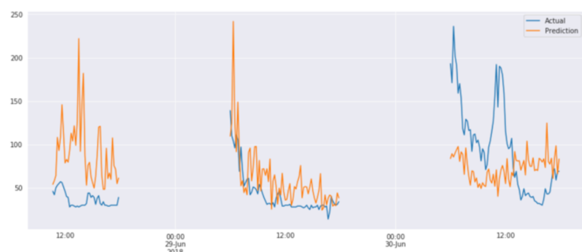

Fig. 2. Actual and Predicted $PM_{2.5}$ values using MLR

Figure 2 shows the predicted values using a trained MLR model. In Fig. 3, predicted values using a trained SVR model are shown.

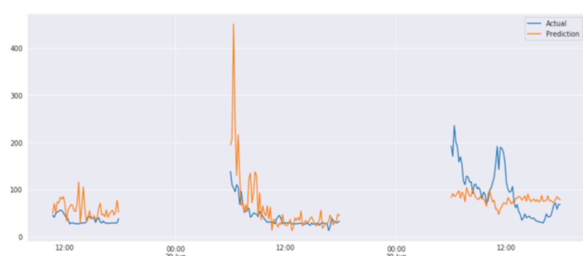

Fig. 3. Actual and Predicted $PM_{2.5}$ values using SVR

The predicted results using MLR has a 0.1 correlation with respect to actual sensor data and has 2,676 MSE. The predicted results using SVR yielded 0.38 correlation with respect to actual sensor data and a 2,490 MSE.

In terms of both metrics, SVR performed better than MLR. The SVR model predicted values correlate more with the actual values and has less error than the one generated by the MLR model.

4. CONCLUSION

This study shows the possibility of using statistical and machine learning techniques for estimating ambient Particulate Matter in Metro Manila. The SVR model performed better than the MLR model.

Various factors could be attributed the low performance of the models. A primary factor is the limited dataset. To reduce the noise of the data, a higher aggregation level can be used. However, it would make the dataset smaller. Another possible reason is the inexact classification of the vehicles. There were a lot of rainy days in June 2018, which resulted in unclear footages and inaccurate detection. Different vehicles produce varied emissions, especially those in different vehicle classifications; thus, accurate identification of vehicle type is important.

There are possible improvements that can be made to enhance the models. One is the collection for a longer period. A larger dataset would help in gaining more insight on the air pollution along DLSU. Another one is having nighttime classification of vehicles to help in training the model. It may also help to use a manually annotated vehicle classification data. With the annotated data, the error that the VEMON vehicle classifier introduces can be reduced and isolated. Lastly, other algorithms that make use of the temporal order of the data can be explored to improve estimation performance.

5. ACKNOWLEDGMENTS

This work was completed through a research grant provided by the Department of Science and Technology through the project TITAN, and ERDT Scholarship given to the first author. Special thanks to the Environment And RemoTe sensing researcH (EARTH) Laboratory and Universidad de Manila CHED DARETO project for providing technical advice and air quality and weather data gathering support. Gratitude is also given to the De La Salle University - Information Technology Services for giving access to street view videos in front of DLSU-M. And lastly, sincerest gratitude to other VEMON project members.

6. REFERENCES

Alas, H. D., Müller, T., Birmili, W., Kecorius, S., Cambaliza, M. O., Simpas, J. B. B., ... & Wiedensohler, A. (2018). Spatial characterization of black carbon mass concentration in the atmosphere of a Southeast Asian megacity: An air quality case study for Metro Manila, Philippines. Aerosol and Air Quality Research, 18(9), 2301-2317+.

Batingal, T. N., Camering, A. R., Naguit, R., Dy, T. R., & Ilao, J. (2016). Vision-Based Vehicle Detection, Counting, and Classification in Philippine Roadways. In DLSU Research Congress, Manila, Philippines

Chang, X., Chen, B. Y., Li, Q., Cui, X., Tang, L., & Liu, C. (2013, March). Estimating real-time traffic carbon dioxide emissions based on intelligent transportation system technologies. IEEE Transactions on Intelligent Transportation Systems, 14(1), 469-479. DOI: 10.1109/TITS.2012.2219529

Clean Air Asia. (2012). Accessing Asia: Air pollution and greenhouse gas emissions indicators for road transport and electricity [PDF file]. Retrieved from http://cleanairasia.org/wp-content/uploads/portal/files/documents/Accessing_Asia_2012_Edition_MAIN_REPORT_0.pdf

Department of Environment and Natural Resources. (2015). DENR administrative order no 2015-04. Department of Environment and Natural Resources.

Department of Environment and Natural Resources. (2017, Sep). Air Pollution: A Public Health Concern in the Philippines. Department of Environment and Natural Resources. Retrieved from https://www.denr.gov.ph/news-and-features/latest-news/3295-air -pollution-a-public-health-concern-in-the-philippines.html

Department of Transportation. (2017). DOT department order no. 2017-011. Department of Transportation.

Kecorius, S., Madueno, L., Vallar, E., Alas, H., Betito, G., Birmili, W., … & Wiedensohler, A. (2017). Aerosol particle mixing state, refractory particle number size distributions and emission factors in a polluted urban environment: Case study of Metro Manila, Philippines. Atmospheric Environment, 170, 169 - 183. Retrieved from http://www.sciencedirect.com/science/article/pii/S1352231017306362 DOI: https://doi.org/10.1016/ j.atmosenv.2017.09.037

Liu, H., Chen, X., Wang, Y., & Han, S. (2013). Vehicle emission and near-road air quality modeling for shanghai, china: Based on global positioning system data from taxis and revised moves emission inventory. Transportation Research Record, 2340(1), 38-48.

Rencher, A. C., & Schaalje, G. B. (2008). Linear models in statistics. John Wiley & Sons.

Sánchez, A. S., Nieto, P. G., Fernández, P. R., del Coz Díaz, J. J., & Iglesias-Rodríguez, F. J. (2011). Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). Mathematical and Computer Modelling, 54(5-6), 1453-1466.

Philippine Congress. (1999). R.A 8749: The Philippine Clean Air Act.

World Health Organization. (2016). Ambient air pollution: a global assessment of exposure and burden of disease. In Ambient air pollution: a global assessment of exposure and burden of disease [PDF file]. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/250141/9789241511353-eng.pdf