PROJECT PROFILE FUNDED BY WNI WxBunka Foundation

VOTE NUMBER GENERATED AT UNIVERSITI TEKNOLOGI MALAYSIA: R.J130000.7352.4B406

Estimating and Forecasting PM2.5 Concentrations in Malaysia

Research Area: Basic Science
Name of Project Leader: Prof Dr Kasturi Devi Kanniah
Research Alliance : Sustainability
Universiti Teknologi Malaysia
E-mail: kasturi@utm.my

ABSTRACT (120 words)

Compared to PM10, PM2.5 can penetrate deeper into the lungs, and can cause aggravation of cardiovascular and systemic inflammation. Recently, only 65 stations are available over whole Malaysia territory to measure $PM_{2.5}$. These Limited stations provide insufficient data to describe the spatial variations in PM2.5 and to forecast its concentrations. Therefore, remotely sensed satellite data were used in this study to estimate $PM_{2.5}$ on a daily basis over a large spatial domain i.e. entire Malaysia territory. Mapping highly polluted locations with PM2.5 will help the public health organizations and environmental protection agencies to estimate mortality rate attributed to ambient air pollution and the future $PM_{2.5}$ concentration prediction can help local authorities to perform preventative measures.

1. INTRODUCTION

In Malaysia, Particulate Matter <10 micron (PM10) were measured as a main parameter for Air Pollutant index (API) from 1995-2017. Previously, Malaysian API became an issue when people notice the stark difference in air pollution reading between Malaysia (specifically in Sothern Peninsular) and Singapore even though the haze was evidently bad (New Straits Times., 2015). Thus, starting mid-August 2018, DOE improved the calculation of API by using Particulate Matter <2.5 micron (PM2.5). However, only 65 operational stations are still insufficient to cover the whole Malaysian territory with an area of 330,290 km^2. Establishing more air quality monitoring stations is very costly and a certain station is only capable to satisfactorily represent the pollutant concentrations within a radius of about 15 km (Ibrahim, Ismail, & Yong, 2012). Alternatively, remote sensing data encourages more studies to be conducted on atmospheric particulates and air quality, since satellite-based technology provides aerosol optical depth (AOD) as a key predictor of PM over a large area (Van Donkelaar *et al*., 2010). During the last 20 years, several studies on PM estimation using AOD products from various satellite sensors have been conducted in polluted regions such as India, China, SEA and elsewhere (Dey et al., 2012; Gupta et al., 2006; Sinha et al., 2015; Wang & Christopher, 2003). In order to extend the spatial coverage to the whole country (both Peninsular and Island Malaysia) and for improving the accuracy of $PM_{2.5}$ estimates, the present study integrates hourly AOD products from Himawari-8 satellite sensor, along with meteorological parameters and other gaseous pollutants using machine learning techniques i.e., random forest (RF) and Support Vector Regression (SVR).

2. RESEARCH METHODOLOGY

Data

The prime data that was used to obtain $PM_{2.5}$ over a large spatial domain is the Aerosols Optical Depth (AOD) data. AOD corresponds to the total columnar aerosol, whereas PM refers to particle concentration at the surface. These two parameters are strongly correlated and the correlation depends on the vertical distribution of aerosols and several other factors (described below). AOD data used in this study is provided by the Himawari-8 satellites. Daily data with 5km spatial resolution covering years 2018 to 2019 were used. PM2.5, meteorological and gases data obtained from the Department of Environment (DOE) Malaysia. Hourly PM2.5 concentrations collected from 65 stations around Malaysia were used for an empirical model development. In order to develop a robust prediction model, it is necessary to consider potential factors that affect the PM2.5 concentrations as much as possible. Atmospheric gases data such as Carbon monoxide, Nitrogen dioxide, Ozone and Sulphur dioxide also will be used as the combination of these precursor gases are able to form $PM_{2.5}$, while meteorological parameters such as wind speed, wind direction, ambient temperature, and relative humidity can affect the concentration and distribution of $PM_{2.5}$.

Methods

PM2.5 will be estimated using an empirical model relating PM2.5 measured in the field with AOD (from Himawari-8 satellite data) coupled with meteorological and chemical parameters. A machine learning techniques called Random Forest (RF) and Support Vector Regression (SVR) were used as it can provide insight into the spatial-temporal distribution of PM2.5.

Once PM2.5 is estimated for entire Malaysia, forecasting of PM2.5 for the next day was performed. In the forecasting model, the average daily PM2.5 is considered as dependent variable and wind speed, wind direction, ambient temperature, relative humidity, rainfall CO, NO_2, O_3 and SO_2 as independent variables. All the parameter for the year 2018-2019 were used in order to gain a better understanding of $PM_{2.5}$ temporal variability. The results of the study ($PM_{2.5}$ estimation) was validated using cross validation (CV) technique in “e1071” and “Random Forest” in the R program. The developed model were validated by using 10 fold cross-validation techniques. During the validation of the second model (forecast model), the input parameters of the days before were used and the predicted $PM_{2.5}$ was compared with actual $PM_{2.5}$ value. The accuracy of the models will be evaluated through performance indicators i.e. Root Mean Square Error, Mean Bias Error (MBE) and Nash-Sutcliffe efficiency (NSE).

3. LITERATURE REVIEW

Significant progress has been made in developing and establishing various techniques for estimating PM concentrations at local, regional and global scales. In early 2000s, most researchers predicted PM using only AOD, by means of simple linear regression techniques (J. A. Engel-Cox, Holloman, Coutant, & Hoff, 2004; Wang & Christopher, 2003). Later, more advanced techniques were developed to incorporate AOD and other important parameters that may influence PM distribution spatially and temporarily, starting from multiple linear regressions (Benas, Beloconi, & Chrysoulakis, 2013; Chitranshi, Sharma, & Dey, 2015; Gupta & Christopher, 2009a; Schaap, Apituley, Timmermans, Koelemeijer, & Leeuw, 2009; Zaman, Kanniah, & Kaskaoutis, 2017), chemical transport models (CTM) (Crouse et al., 2016; J. Engel-Cox, Kim Oanh, van Donkelaar, Martin, & Zell, 2013; Liu et al., 2004; Van Donkelaar, Martin, & Park, 2006), mixed effect models (MEM) (Beloconi, Kamarianakis, & Chrysoulakis, 2016; Kloog, Koutrakis, Coull, Lee, & Schwartz, 2011; Lee, Liu, Coull, Schwartz, & Koutrakis, 2011; Xie et al., 2015), artificial neural networks (ANN) (Di et al., 2016; Grivas & Chaloulakou, 2006; Gupta & Christopher, 2009b; Wu, Guo, Zhang, & Li, 2011; Zaman et al., 2017), geographic weighted regression (GWR) (Bai et al., 2016; Z. Hu, 2009; Ma, Hu, Huang, Bi, & Liu, 2014; You et al., 2016) and generalized additive models (GAM) (Liu, Paciorek, & Koutrakis, 2009; Paciorek, Liu, Moreno-Macias, & Kondragunta, 2008; Song et al., 2015; Zou, Chen, Zhai, Fang, & Zheng, 2016). These techniques are limited to capture the non-linear relationships that exist between the variables. Consequently, complex techniques have been developed by combining two or more statistical techniques; for instance, merging MEM and GWR (X. Hu et al., 2014) or incorporating MEM into GAM (Ma et al., 2016) etc. Machine learning techniques such as deep neural network (DNN), support vector regression (SVR) and random forest (RF) are also able to capture the complex relationships between parameters and have greater performance in estimating $PM_{2.5}$ (Danesh Yazdi et al., 2020; Li et al., 2018). Nowadays, machine learning techniques are increasingly used in air quality studies (Gholami, Mohamadifar, Sorooshian, & Jansen, 2020; Gholami, Mohammadifar, Bui, & Collins, 2020; Kleine Deters, Zalakeviciute, Gonzalez, & Rybarczyk, 2017; Shin et al., 2020).

4. FINDINGS

For the overall model, SVR calibration performed slightly better than RF with R^2 =0.69 and RMSE= 10.62 µg m^{-3} against measured $PM_{2.5}$ concentrations. Whilst, for the spatial models the RF validation performed slightly better than SVR for both urban/industrial and suburban/rural areas with statistical indicators of R^2 =0.76, RMSE= 11.47 µg m^{-3} and R^2 =0.64, RMSE=10.76 µg m^{-3}, respectively. Therefore, both RF and SVR displayed slightly higher performance for $PM_{2.5}$ estimations at urban/industrial sites with higher levels of AOD and air pollutants. Furthermore, the estimation accuracy of SVR and RF models was lower in the wet (November-March) and inter-monsoon (April-May) seasons compared to the dry (June-September) season. Based on the model accuracy and variable importance analysis, CO was always the most influential predictor variable for $PM_{2.5}$ estimations in Malaysia, followed by AOD, O_3, NO_2, SO_2 and meteorological parameters, but with different order depending on the dataset and model. An important finding was the very weak correlation and contribution of the meteorological variables to $PM_{2.5}$ estimations, as well as the low correlation between $PM_{2.5}$ and columnar AOD, indicating that surface pollution follows a different temporal pattern than AOD, and the presence of a significant aerosol layer aloft due to transported smoke plumes from wildfires in southeast Asia.

5. CONCLUSION

The current results show that the use of machine-learning techniques for $PM_{2.5}$ estimations over Malaysia is promising as these models can satisfactorily represent the values and temporal evolution of $PM_{2.5}$ concentrations over both urban/industrial and suburban/rural sites. In a next step, we will include gases pollutants from satellite remote sensing observations in order to estimate $PM_{2.5}$ concentrations over large areas aiming to cover the whole Malaysian territory with high spatial resolution.

6. PROBLEM/ CONSTRAIN

Due to COVID-19, we cannot disseminate knowledge to the public via workshops. Even though, we have prepare all the paperwork.

7. RESEARCH OUTPUT

(i) Kanniah, K. D., Zaman, N. A. F. K., Kaskaoutis, D. G., & Latif, M. T. (2020). COVID-19's impact on the atmospheric environment in the Southeast Asia region. *Science of the Total Environment*, 139658. Impact factor: 6.551 (Q1)

(ii) Zaman, N. A. F. K., & Kanniah, K. D. 2020, Spatio-temporal assessment of Aerosol Optical Depth from Himawari-8 satellite data over Malaysia. In *IOP Conference Series: Earth and Environmental Science* (Vol. 540, No. 1, p. 012053). IOP Publishing.

(iii) Kasturi Devi Kanniah and Nurul Amalin Fatihah Kamaruzaman, 2021, Remotely Sensed Particulate Matter Estimation in Malaysia during the Biomass Burning Season in Southeast Asia Published by CRC Taylor & Francis (Chapter in Press)

(iv) Nurul Amalin Fatihah Kamarul Zaman, Kasturi Devi Kanniah, Dimitris G. Kaskaoutis and Mohd Talib Latif, 2021, $PM_{2.5}$ estimations in Malaysia using machine learning techniques, submitted to Atmospheric Research (Q1 journal)

8. HUMAN CAPITAL DEVELOPMENT

(i) Suman A/L Selvaraju (Masters student)

(ii) Nurul Amalin Fatihah binti Kamarul Zaman (PhD student)

REFERENCES

Bai, Y., Wu, L., Qin, K., Zhang, Y., Shen, Y., & Zhou, Y. (2016). A geographically and temporally weighted regression model for ground-level PM2. 5 estimation from satellite-derived 500 m resolution AOD. *Remote Sensing, 8*(3), 262.

Beloconi, A., Kamarianakis, Y., & Chrysoulakis, N. (2016). Estimating urban PM10 and PM2. 5 concentrations, based on synergistic MERIS/AATSR aerosol observations, land cover and morphology data. *Remote Sensing of Environment, 172*, 148-164.

Benas, N., Beloconi, A., & Chrysoulakis, N. (2013). Estimation of urban PM10 concentration, based on MODIS and MERIS/AATSR synergistic observations. *Atmospheric Environment, 79*, 448-454.

Chitranshi, S., Sharma, S. P., & Dey, S. (2015). Satellite-based estimates of outdoor particulate pollution (PM10) for Agra City in northern India. *Air Quality, Atmosphere & Health, 8*(1), 55-65.

Crouse, D. L., Philip, S., Van Donkelaar, A., Martin, R. V., Jessiman, B., Peters, P. A., . . . Burnett, R. T. (2016). A new method to jointly estimate the mortality risk of long-term exposure to fine particulate matter and its components. *Scientific reports, 6*, 18916.

Danesh Yazdi, M., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., . . . Schwartz, J. (2020). Predicting Fine Particulate Matter (PM2. 5) in the Greater London Area: An Ensemble Approach using Machine Learning Methods. *Remote Sensing, 12*(6), 914.

Dey, S., Di Girolamo, L., van Donkelaar, A., Tripathi, S., Gupta, T., & Mohan, M. (2012). Variability of outdoor fine particulate (PM2. 5) concentration in the Indian Subcontinent: A remote sensing approach. *Remote Sensing of Environment, 127*, 153-161.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM2. 5 exposures with high spatiotemporal resolution across the continental United States. *Environmental science & technology, 50*(9), 4712-4721.

Engel-Cox, J., Kim Oanh, N. T., van Donkelaar, A., Martin, R. V., & Zell, E. (2013). Toward the next generation of air quality monitoring: Particulate Matter. *Atmospheric Environment, 80*, 584-590. doi: https://doi.org/10.1016/j.atmosenv.2013.08.016

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment, 38*(16), 2495-2509.

Gholami, H., Mohamadifar, A., Sorooshian, A., & Jansen, J. D. (2020). Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran. *Atmospheric Pollution Research*.

Gholami, H., Mohammadifar, A., Bui, D. T., & Collins, A. L. (2020). Mapping wind erosion hazard with regression-based machine learning algorithms. *Scientific reports, 10*(1), 1-16.

Grivas, G., & Chaloulakou, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment, 40*(7), 1216-1229.

Gupta, P., & Christopher, S. A. (2009a). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research: Atmospheres, 114*(D14).

Gupta, P., & Christopher, S. A. (2009b). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *Journal of Geophysical Research: Atmospheres, 114*(D20).

Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment, 40*(30), 5880-5892.

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., . . . Puttaswamy, S. J. (2014). Estimating ground-level PM2. 5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment, 140*, 220-232.

Hu, Z. (2009). Spatial analysis of MODIS aerosol optical depth, PM 2.5, and chronic coronary heart disease. *International journal of health geographics, 8*(1), 27.

Ibrahim, M. Z., Ismail, M., & Yong, K. H. (2012). Mapping the Spatial Distribution of Criteria Air Pollutants in Peninsular Malaysia Using Geographical Information System (GIS).

Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., & Rybarczyk, Y. (2017). Modeling PM2. 5 urban pollution using machine learning and selected meteorological parameters. *Journal of Electrical and Computer Engineering, 2017*.

Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved PM2. 5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment, 45*(35), 6267-6275.

Lee, H., Liu, Y., Coull, B., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM 2.5 concentrations. *Atmospheric Chemistry & Physics Discussions, 11*(3).

Li, L., Chen, B., Zhang, Y., Zhao, Y., Xian, Y., Xu, G., . . . Guo, L. (2018). Retrieval of Daily PM2. 5 Concentrations Using Nonlinear Methods: A Case Study of the Beijing–Tianjin–Hebei Region, China. *Remote Sensing, 10*(12), 2006.

Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM2. 5 concentrations using satellite data, meteorology, and land use information. *Environmental health perspectives, 117*(6), 886.

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level PM2. 5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research: Atmospheres, 109*(D22).

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2. 5 in China using satellite remote sensing. *Environmental science & technology, 48*(13), 7436-7444.

Ma, Z., Hu, X., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y., . . . Liu, Y. (2016). Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013. *Environmental health perspectives, 124*(2), 184.

Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM2. 5. *Environmental science & technology, 42*(15), 5800-5806.

Schaap, M., Apituley, A., Timmermans, R., Koelemeijer, R., & Leeuw, G. d. (2009). Exploring the relation between aerosol optical depth and PM 2.5 at Cabauw, the Netherlands. *Atmospheric Chemistry and Physics, 9*(3), 909-925.

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., & Quackenbush, L. J. (2020). Estimating ground-level particulate matter concentrations using satellite-based data: a review. *GIScience & Remote Sensing, 57*(2), 174-189.

Sinha, P., Gupta, P., Kaskaoutis, D., Sahu, L., Nagendra, N., Manchanda, R., . . . Sreenivasan, S. (2015). Estimation of particulate matter from satellite-and ground-based observations over Hyderabad, India. *International Journal of Remote Sensing, 36*(24), 6192-6213.

Song, Y.-Z., Yang, H.-L., Peng, J.-H., Song, Y.-R., Sun, Q., & Li, Y. (2015). Estimating PM2. 5 concentrations in Xi'an City using a generalized additive model with multi-source monitoring data. *PloS one, 10*(11), e0142149.

Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2. 5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres, 111*(D21).

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2. 5 mass: Implications for air quality studies. *Geophysical Research Letters, 30*(21).

Wu, Y., Guo, J., Zhang, X., & Li, X. (2011). *Correlation between PM concentrations and aerosol optical depth in eastern China based on BP neural networks.* Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International.

Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., & Bai, Y. (2015). Daily estimation of ground-level PM2. 5 concentrations over Beijing using 3 km resolution MODIS AOD. *Environmental science & technology, 49*(20), 12280-12288.

You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., & Wang, W. (2016). National-scale estimates of ground-level PM2. 5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Remote Sensing, 8*(3), 184.

Zaman, N. A. F. K., Kanniah, K. D., & Kaskaoutis, D. G. (2017). Estimating Particulate Matter using satellite based aerosol optical depth and meteorological variables in Malaysia. *Atmospheric Research, 193*, 142-162.

Zou, B., Chen, J., Zhai, L., Fang, X., & Zheng, Z. (2016). Satellite based mapping of ground PM2. 5 concentration using generalized additive modeling. *Remote Sensing, 9*(1), 1.