A Neural Network-based Land Use Regression Model to Estimate Spatial-temporal Variability of SO_2

Ya-Ping Hsiao (1), Chih-Da Wu (1, 2)*, Jen-Wei Huang (3), Tee-AnnTeo (4), Shih-Yuan Lin (5)

[1] Department of Geomatics, National Cheng Kung University, Tainan, Taiwan.
[2] National Institute of Environmental Health Sciences, National Health Research Institutes, Miaoli, Taiwan.
[3] Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan.
[4] Department of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan.
[5] Department of Land Economics, National Chengchi University, Taipei, Taiwan.
Email: hsiaoyapiau@mail.com.; chidawu@mail.ncku.edu.tw; jwhuang@mail.ncku.edu.tw; tateo@mail.nctu.edu.tw; syl@nccu.edu.tw

KEY WORDS: Air Pollution, Sulfur dioxide (SO_2); Land-use Regression (LUR); Deep Neural Network (DNN).

ABSTRACT: One of the air pollutants from industrial waste and transportation combustion is Sulfur Dioxide (SO_2). Previous studies have shown that SO_2 has a serious impact on human health, in particularly on respiratory problems. By taking Taiwan as the study area, this study aimed to assess the spatial-temporal variability of SO_2 using a neural network-based land use regression model. Daily SO_2 observations during 2000 to 2018 were obtained from 73 monitoring stations established by Taiwan Environmental Protection Agency (EPA). Totally, around 0.48 million observations were collected for our analysis. Several databases were used to collect the spatial predictor variables, including EPA environmental resources database, meteorological database, land-use inventory, landmark database, digital road network map, DTM, MODIS NDVI dataset, and thermal power plant distribution database. To establish the integrated approach, conventional land-use regression (LUR) was first used to identify the important predictors variables. After that, a deep neural network (DNN) algorithm was applied to fit the prediction model.

The results showed that, the adj-R^2 obtained from the conventional LUR approach was 0.37. Of the 15 variables selected by the stepwise variable selection procedure, PM10, nearest thermal power plants, and NO_2 are important variables that increased the SO_2 exposures with the explanatory ability up to 18%, 6%n and 4%, respectively. Compared to the conventional LUR approach, by combining DNN algorithm can improve the model explanatory ability up to 21% (adj-R^2=0.59). The results of 10-fold cross validation and external data verification confirmed that the value of the adj-R^2 after combining both approaches increased from 0.37 to 0.59, and RMSE decreased from 2.48 ppm to 2.01 Findings of this study confirm that the combination of LUR, and DNN algorithm can improve the prediction performance level and the explanatory abilities in assessing spatial-temporal variability of SO_2 exposure.

1. INTRODUCTION

The rapid development of economy and industrialization, accompanied by emissions from factories and transportation, has led to a deterioration in air quality. Among them, Sulfur Dioxide (SO_2) is one of the most common Sulfur Oxides in the air, and it is known for being irritant and toxic. The main sources of SO_2 are due to the combustion of industrial petrochemical fuel and coal burning, not to mention sulphuric acid and phosphate fertilizers that produce industrial waste gas, as well as exhaust emissions from motor vehicles. Many studies have shown long-term exposure to high concentrations of Sulfur Dioxide (Sulfur Dioxide, SO_2) can affect the function of the respiratory system, and promote allergic rhinitis, bronchitis and asthma morbidity (Chiang et al., 2016, Yuan et al., 2015a, Yuan et al., 2015b) , which has a serious impact on health.

Based on the restrictions on the number and distribution of air quality monitoring stations of the Environmental Protection Administration (EPA) of the Executive Yuan, the unique and complex regional pollution sources such as diverse catering patterns and temples on Taiwan Island are scattered in densely populated and busy communities. These areas all have possibility to increase the spatial variation of the concentration of air pollution. In order to solve this problem, several methods for estimating outdoor air pollution concentration have been developed internationally, such as Spatial Interpolation, Dispersion Model, and Land Use Regression (LUR) (Vardoulakis et al., 2003, Pfender et al., 2006; Wang et al., 2014; Wang et al., 2015; Wu et al., 2015; Weichenthala et al., 2016).The land use regression model uses GIS and Remote Sensing (RS) to obtain possible spatial concentration prediction variables such as population density, land use, traffic related variables and other related predictive variables.

The land use regression model combined with air quality monitoring data and statistical models, can find the best predictors of pollution concentration, and then establish a statistical relationship between land use and pollutant concentration to estimate the spatial variation of air pollutants. In recent years, with the development of Geographic Information Systems (GIS) technology, the problem of large-scale spatial data analysis and collection has been solved. Therefore, compared with other methods, the land use regression model has gradually. An important worldwide method for estimating wide range air pollution concentration. On the other hand, with advances in artificial intelligence (AI) and the rise of information technology, not only do they spark big changes on the original operating model within the industry, it also provides a new approach to big data analysis. While some studies have applied this technology to the estimation of air pollution, they mainly focused on the prediction of the concentration of air pollution monitoring stations within a short amount of hours, and the influencing factors and traits that they took into consideration were mostly meteorological related, topographic related and also have a lot to do with air pollution concentration spatio-temporal autocorrelation. These analyses currently do not include the distribution of surrounding pollution sources and their impact on urban air pollution changes. Also, land use regression and machine learning algorithms are not combined to estimate the spatial variability of island-wide resolution SO_2 under the influence of climate change.

Therefore, the purpose of this study is to collect the SO_2 concentration and the land use survey from the Central Meteorological Bureau as well as the road network devaluation map, the landmark database, the NDVI, the DTM, etc. of the Air Quality Monitoring Station of the Taiwan Environmental Protection Administration (EPA) as an example. A hybrid model can be established using land use regression and deep neural network algorithms to estimate the long-term calendar spatial and temporal distribution of sulfur dioxide in Taiwan Island.

2. MATERIALS AND METHODSTITLE

2.1 Study Area

Taiwan locates in the East Asia and neighbours with China to the west, Japan to the northeast, and the Philippines to the south. It stretches over a geographical area of 36,193 km^2.With a population of 23,476,640, the averaged population density of Taiwan is 649 people/km^2 ranked as the 17th most densely populated country in the world. Traffic emissions from more than 22 million registered motor vehicle contributes significantly to urban air pollution. Moreover, diverse culture-specific SO_2 emission sources such as traffic and industry with not only elevate the level of pollutants, but also increase the difficulty in predicting spatial-temporal variability of SO_2 and their constitutes in Taiwan.

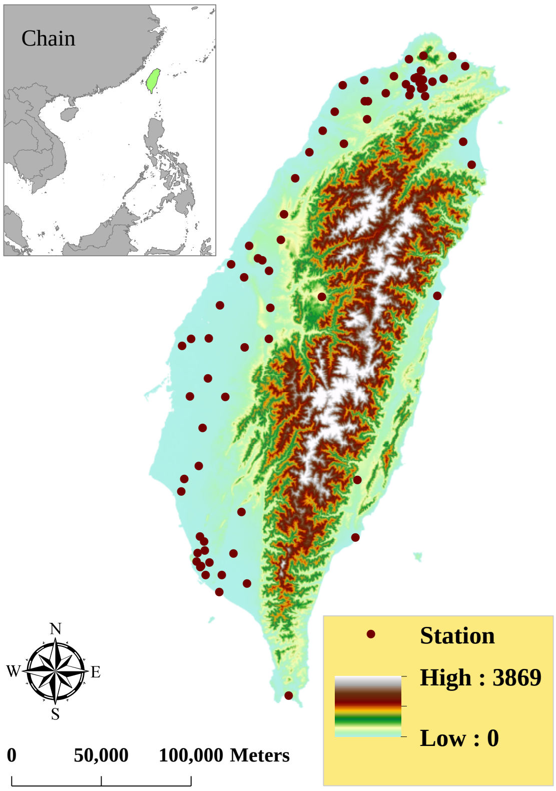

Figure 1. Spatial distributon of the seventy three monitiring staitons in Taiwan.

2.2 Databases

Daily SO_2 observations during 2000 to 2018 were obtained from 73 monitoring stations established by Taiwan EPA. Totally, around 0.48 million observations were collected for our analysis. Including several databases were used to collect the spatial predictor variables, including EPA environmental resources database, meteorological database, land-use inventory, landmark database, digital road network map, Digital Terrain Model with 20m resolution, landmark databases, and MODIS NDVI database, and thermal power plant distribution database.

2.3 Methodology

To establish the integrated approach, conventional land-use regression (LUR) was first used to identify the important predictors variables. Using Circular buffers surrounded to the monitoring sites were generated with the radius from 50m to 5000m. Land-use allocation within each buffer range were calculated, such as road density, distribution of residential areas, industrial parks, green spaces, temples, and Chinese restaurants. Pollutants levels of the monitoring stations were then regressed the land-use allocation information, and a supervised stepwise procedure referred to Wu et al. (2017) applied to develop the LUR models applied of a deep neural network (DNN) algorithm was applied to fit the prediction model was then to develop models for estimating the spatial-temporal variability of SO_2 constitutes across the main-island of Taiwan.

3. RESULTS AND DISCUSSION

Automatic Learning rate formula of the model was LearnRate_value × 1 /(1 + decay_value × epoch_value) and as soon as the LOSS is less than 0.9 mode, stop immediately.Epoch value of model development was set for 210, during each epoch descending gradient method was used for minimizing the error. To optimize the model performance, package ‘Adam’ was used for optimizing the model and the loss function package ‘mean_absolute_error’ was used for identifying the mean absolute error (MAE) between model predictions and observations within the model.

The results showed that, the adj-R^2 obtained from the conventional LUR approach was 0.37. Of the 15 variables selected by the stepwise variable selection procedure, PM10, nearest thermal power plants, and NO_2 are important variables that increased the SO_2 exposures with the explanatory ability up to 18%, 6%n and 4%, respectively. Compared to the conventional LUR approach, by combining DNN algorithm can improve the model explanatory ability up to 21% (adj-R^2=0.59).Findings of this study confirm that the combination of LUR, and DNN algorithm can improve the prediction performance level and the explanatory abilities in assessing spatial-temporal variability of SO_2 exposure.

Table 1. SO_2 model result

Variable	β	P-value	Partial R^2	VIF	LUR model performance	DNN model performance
intercept	-1.264	<0.05			R^2=0.37 Adj R^2=0.367 RMSE=2.482	R^2=0.59 Adj R^2=0.585 RMSE=2.015
PM_{10}	0.023	<0.05	0.188	1.698		
Thermal power plant distance	0.000	<0.05	0.066	1.547		
NO_2	0.125	<0.05	0.043	1.963		
port_distance	0.000	<0.05	0.033	1.160		
Landfill_distance	0.000	<0.05	0.021	1.114		
temperature	0.067	<0.05	0.020	1.649		
water_distance	0.001	<0.05	0.011	1.085		
humidity	-0.023	<0.05	0.004	1.193		
bus_distance	0.000	<0.05	0.003	1.060		
Incinerator_distance	0.000	<0.05	0.003	1.476		
summer	0.463	<0.05	0.002	1.603		
Press	0.004	<0.05	0.001	1.248		
Sandstone_distance	0.000	<0.05	0.001	1.114		
wind direction	0.088	<0.05	0.001	1.363		
localroad_width_distance	-0.001	<0.05	0.001	1.180		

4. CONCLUSIONS

Findings of this study confirm that the combination of LUR, and DNN algorithm can improve the prediction performance level and the explanatory abilities in assessing spatial-temporal variability of SO_2 exposure.

5. REFERENCES FOM Journals

Chiang, T. Y., Yuan, T. H., Shie, R. H., Chen, C. F., & Chan, C. C. 2016. Increased incidence of allergic rhinitis, bronchitis and asthma, in children living near a petrochemical complex with SO_2 pollution. Environment international, 96, pp.1-7.

Yuan, T. H., Chio, C. P., Shie, R. H., Pien, W. H., & Chan, C. C. 2016. The distance-to-source trend in vanadium and arsenic exposures for residents living near a petrochemical complex. Journal of Exposure Science and Environmental Epidemiology, 26 (3), pp.270.

Yuan, T. H., Shie, R. H., Chin, Y. Y., & Chan, C. C. 2015. Assessment of the levels of urinary 1-hydroxypyrene and air polycyclic aromatic hydrocarbon in $PM_{2.5}$ for adult exposure to the petrochemical complex emissions. Environmental research, 136, pp.219-226.

Vardoulakis, S., Fisher, B. E., Pericleous, K., & Gonzalez-Flesca, N. 2003. Modelling air quality in street canyons: a review. Atmospheric environment, 37 (2), pp.155-182.

Pfender, W., Graw, R., Bradley, W., Carney, M., & Maxwell, L. 2006. Use of a complex air pollution model to estimate dispersal and deposition of grass stem rust urediniospores at landscape scale. Agricultural and forest meteorology, 139 (1-2), pp.138-153.

Wang, M. R. Beelen, T. Bellander, M. Birk, G. Cesaroni, M. Cirach, J. Cyrys, K. D. Hoogh, C. Declercq, K. Dimakopoulou, M. Eeftens, K. T. Eriksen,F. Forastiere, C. Galassi, G. Grivas, J. Heinrich, B. Hoffmann, A. Ineichen, M. Korek, T. Lanki, S. Lindley, L. Modig, A. Mölter, P. Nafstad, M. J. Nieuwenhuijsen, W. Nystad, D. Olsson, O. Raaschou-Nielsen, M. Ragettli, A. Ranzi, M. Stempfelet, D. Sugiri, M.- Y. Tsai, O. Udvardy, M. J. Varró, D. Vienneau, G. Weinmayr, K. Wolf, T. Yli-Tuomi, G. Hoek and B. Brunekreef1. 2014 Performance of multi-city land use regression models for nitrogen dioxide and fine particles. Environmental Health Perspectives 122 (8), pp. 843-849.

Wang, M., U. Gehring, G. Hoek, M. Keuken, S. Jonkers, R. Beelen, M. Eeftens, D. S. Postma and B. Brunekreef (2015) Air pollution and lung function in dutch children: a Comparison of exposure estimates and associations based on land use regression and dispersion exposure modeling approaches. Environmental Health Perspectives 123 (8), pp. 847-851.

Weichenthala, S., K. V. Ryswyka, A. Goldsteinb, M. Shekarrizfardc and M. Hatzopoulouc (2016) Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. Environmental Pollution 208, pp. 241-248.