Comparison of Geospatial-Temporal Modeling Approaches in Air Pollution Estimations

Yu-Ting Zeng (1), Chih-Da Wu (1)(2), Yu-Cheng Chen (2), Chin-Yu Hsu (2), Mu-Jean Chen (2)

[1] National Cheng Kung Univ., No. 1, Daxue Rd., East Dist., Tainan City 701, Taiwan.
[2] National Health Research Inst., No. 35, Keyan Rd, Zhunan, Miaoli County 35053, Taiwan.
Email: k9a8r9a8n2i0@gmail.com; chidawu@mail.ncku.edu.tw; yucheng@nhri.edu.tw; gracecyhsu@nhri.org.tw; zeromagi@nhri.org.tw

KEY WORDS: air pollution; geographically and temporally weighted regression (GTWR); geographically weighted regression (GWR); kriging-based hybrid model; land-use regression (LUR).

Abstract: Recent advancements in the geographic information systems and remote sensing technology have supported the development of geospatial-temporal modeling approaches for air pollution. Particulate matter (PM_{10}) and ozone (O_3) are two pollutants of great concern in all pollutants. Previous studies estimated the spatial-temporal variability of PM_{10} and O_3 using a single model, but only a few studies considered exposure assessment using multiple models and compared model performance. In this study, PM_{10} and O_3 data during 2015 to 2018 were collected from specific industrial monitoring stations provided by the Taiwan Environmental Protection Agency. Three geospatial-temporal modeling approaches including land-use regression (LUR), geographically weighted regression (GWR), and geographically and temporally weighted regression (GTWR) were used to predict PM_{10} and O_3 exposure. Furthermore, the kriging-based hybrid model was integrated with these three geospatial-temporal models, and totally performs six models for each pollutant for our comparison. The results showed that integrating the GTWR and kriging-based hybrid models have the greatest performances compared to LUR, GWR, and the combination of both with kriging-based hybrid models. R^2 obtained from the GTWR coupled with kriging-based hybrid models for PM_{10} and O_3 was 0.96 and 0.92, respectively. Of all variables used, wind speed, pure residential area, manufacturing, park; rice field, orchard; and forest land were important predictors for PM_{10}. Whereas, wind direction, industrial area, dry farming, and orchard were variables selected to predict O_3.

1. BACKGROUND AND AIM

Particulate matter with an aerodynamic diameter between 10-2.5 μm (PM_{10}) and ozone (O_3) exposure has been identified as a significant risk factor for the development of lung cancer and adverse health outcomes from cardiovascular and respiratory causes (Brook et al. 2010; Pope and Dockery 2006; Aguilera et al. 2015; Pope et al. 2002; Sabaliauskas et al. 2015). As personal monitoring is not generally feasible for large cohorts, methods to assess accurately within-city variability in exposure to PM_{10} and O_3 are required (Jerrett et al. 2005; Wu et al. 2017). In the past period, there are many predictive methods for capturing ambient air pollution gradients. Spatial interpolation, such as Kriging interpolation (Bayraktar and Turalioglu 2005), predicted pollutant level in an area from a limited number of monitoring sites. Spatial autocorrelation, the statistical relationships of distance among the measured points were used to explain and predict the variation of air pollutants in the surface. However, intra-urban air pollution concentrations could vary due to proximity to industrial parks, road and traffic density, and other site characteristics, such as population and land use (Tunno et al. 2016). The lack of consideration about the local emission sources between monitoring sites could deteriorate the accuracy of predictions. Compared with spatial interpolation, land-use regression (LUR) has been proved to have more advantages on characterizing the spatial relationships between local emissions and intra-urban pollution variations (Clougherty et al. 2013; Hoek et al. 2008; Michanowicz et al. 2016). LUR normally combines distributed pollution measures at multiple sites with a set of potentially predictive geographic source covariates, to develop a multiple linear regression model that can be rendered in a Geographic Information System (GIS) to estimate air pollution levels at unmeasured areas (Wu et al. 2017). Geographic predictors include traffic patterns, surrounding land-use allocations, demographic characteristics, green space distribution, and micro climatic conditions (Aguilera et al. 2015; Su et al. 2010; Wang et al. 2013; Wu et al. 2017; Shi et al. 2016). Likewise, geographically weighted regression (GWR) and geographically and temporally weighted regression (GTWR) are also spatial predictive methods for modelling spatial-temporal variation of air pollution, and gain more and more attention in recent studies (Chu et al., 2018; Cui et al., 2019; Guo et al., 2017; Ma et al., 2018).

In this study, three geospatial-temporal modeling approaches including LUR, GWR), and GTWR were used to predict PM_{10} and O_3 exposure. Furthermore, the kriging-based hybrid model was integrated with these three geospatial-temporal models, and totally performs six models were performed for each pollutant for our comparison.

2. METHODS

The coastal areas of Kaohsiung City, in which several heavy metal industrial parks located on, were selected for the experimental study (Fig. 1). Monthly averaged concentrations from May 2015 to September 2018 of the two study pollutants, PM_{10} and O_3, were obtained from 9 specific industrial monitoring stations. There are four steps for data Analysis (Fig. 2). Environmental factors such as meteorological factors (relative humidity, and temperature etc.), co-pollutants (SO_2, and $PM_{2.5}$ etc.), topography (elevation), land-use distributions (residential areas, and surrounding greenness etc.), and community emission sources (temple and Chinese restaurant) were combined with the recorded concentrations and used as explanatory predictors to develop the conventional LUR models; Second, Kriging-based PM_{10} and O_3 estimations were further added into the explanatory variables pools for building the kriging-based hybrid models; In addition, predictors variables selected by the conventional LUR and hybrid models were integrated with GWR and GTWR models to predict PM_{10} and O_3 exposure as well, this earned a comparison of six models. In the final step, all models were used to illustrate the spatial variability of PM_{10} and O_3..

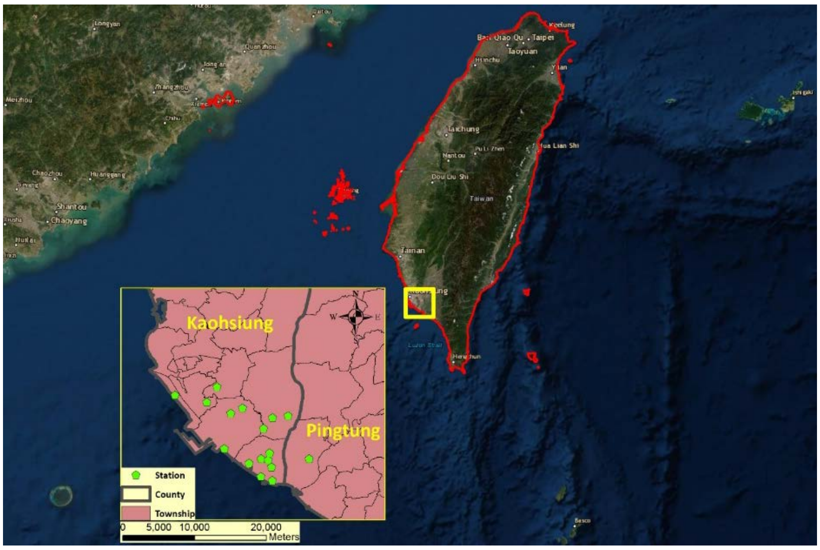

Fig. 1. Study area

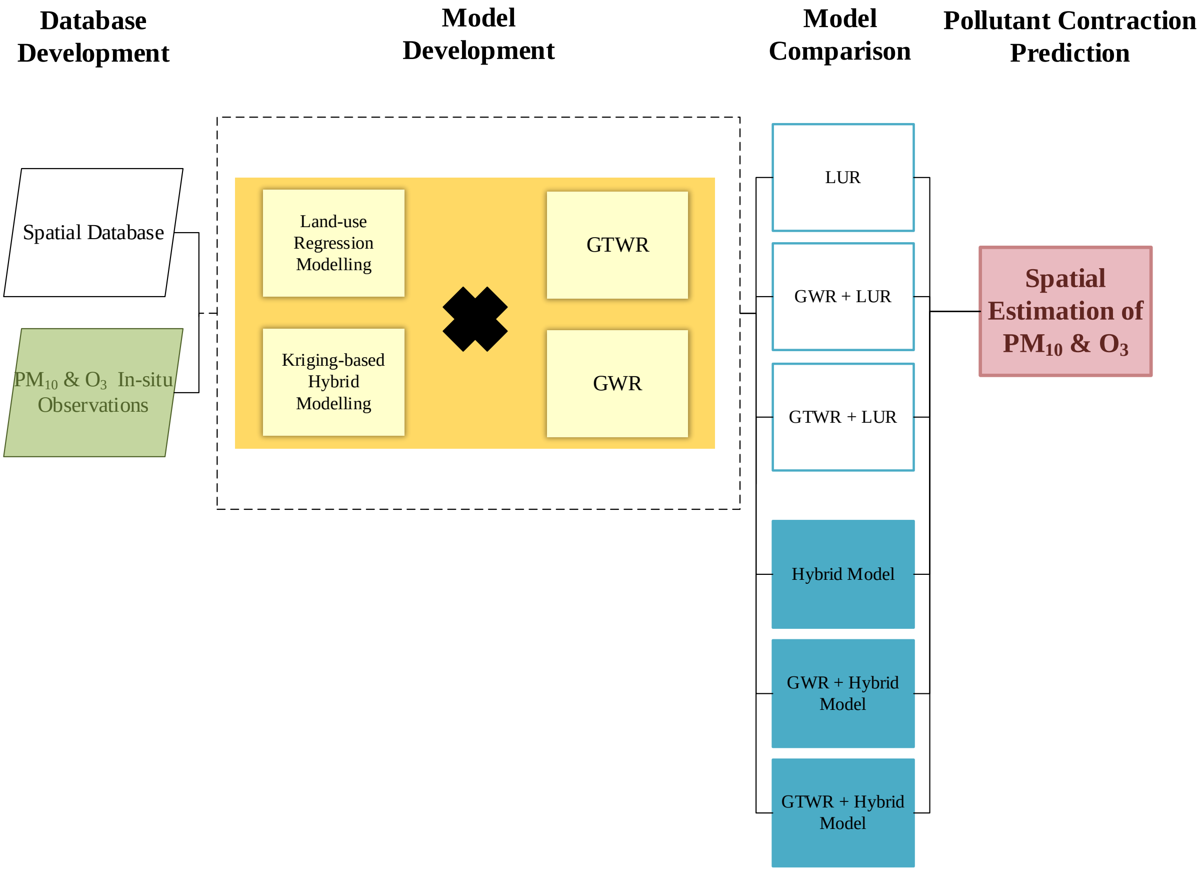

Fig. 2. Study flowchart

RESULTS

The results (Table 3) showed that, integrating the GTWR and predictor variables selected by kriging-based hybrid models have the greatest performances compared to LUR, GWR, and the combination of both with kriging-based hybrid model. R^2 obtained from the GTWR coupled with kriging-based hybrid models for PM_{10} and O_3 was 0.96 and 0.92, respectively. Of all variables used, wind speed, pure residential area, manufacturing, park; rain-ed crop, orchard; and forest land were important predictors for PM_{10}. Whereas, wind direction, industrial area, dry crop, and orchard were variables selected to predict O_3 (Table 2; Table 3). Then, estimating PM_{10} and O_3 concentration surfaces predicted by these 6 models. They all showed that located in the crowded and bustling place experienced higher PM_{10} levels, and O_3 concentration is higher near farmland (Fig. 3; Fig. 4).

Table 3. All model results for PM_{10}

Variable	LUR Model	LUR Model + GWR	LUR Model + GTWR
	Coefficient	Coefficient (Median)	Coefficient (Median)
(Intercept)	$+16.19^{***}$	-0.51	-2.72

NO_2	$+3.13^{***}$	3.34	+3.35
Rainfall	-0.39^{***}	-0.47	-0.47
Wind speed	-8.03^{***}	-2.14	-0.94
Fall	$+3.22^{***}$	+4.44	+4.28
Pure residential area $_{500m}$	$+9.08\times10^{-3**}$	+0.02	+0.01
Manufacturing $_{1250m}$	$+5.99\times10^{-3***}$	$+7.71\times10^{-3}$	$+7.17\times10^{-3}$
Funeral facility $_{nearest\ distance}$	-7.29×10^{-4}	-7.46×10^{-4}	-9.19×10^{-4}
Park$_{250m}$	-0.02^{**}	-0.03	-7.03×10^{-3}
Park$_{2500m}$	-0.20^{***}	-0.14	-0.16
Rain-ed crop$_{150m}$	$+7.53\times10^{-3*}$	$+7.59\times10^{-3}$	+0.01
Orchard$_{1250m}$	$+0.02^{**}$	+0.02	+0.02
All farmland$_{50m}$	$+6.84\times10^{-3**}$	$+3.88\times10^{-3}$	$+5.80\times10^{-3}$
Forest land$_{4000m}$	-0.04^{***}	-0.03	-0.03
Model Performance	**R^2 = 0.89** **ADJ R^2 = 0.89** **RMSE = 7.29**	**R^2 = 0.9** **ADJ R^2 = 0.9**	**R^2 = 0.91** **ADJ R^2 = 0.9**

Variable	**Hybrid Model**	**Hybrid Model + GWR**	**Hybrid Model + GTWR**
	Coefficient	Coefficient (Median)	Coefficient (Median)
(Intercept)	+1.55	+5.69	-6.23
PM_{10} (Kriging-based)	$+0.97^{***}$	+0.99	+0.99
Wind speed	-3.75^{***}	-0.93	-0.22
Pure residential area $_{500m}$	$+0.02^{***}$	+0.02	+0.01
Manufacturing $_{1250m}$	$+7.53\times10^{-3***}$	$+8.30\times10^{-3}$	$+7.42\times10^{-3}$
Park $_{2500m}$	-0.03^{*}	-0.02	-0.06
Rain-ed crop $_{150m}$	$+0.01^{***}$	+0.01	+0.01
Orchard $_{1250m}$	$+0.03^{***}$	+0.03	+0.04
Forest land $_{4000m}$	-0.02^{***}	-0.02	-0.03
Model Performance	**R^2 = 0.95** **ADJ R^2 = 0.95** **RMSE = 5.2**	**R^2 = 0.94** **ADJ R^2 = 0.94**	**R^2 = 0.96** **ADJ R^2 = 0.96**

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 3. All model results for O_3

Variable	LUR Model Coefficient	GWR Coefficient (Median)	GTWR Coefficient (Median)
(Intercept)	-73.71	-148.51	-2.34
Relative humidity	-0.45^{***}	-0.34	-0.49
Atmospheric pressure	$+0.13^{*}$	+0.20	+0.05
Wind direction	-0.02^{***}	-0.02	-0.02
Fall	$+8.23^{***}$	+8.76	+7.56
Industrial area $_{\text{nearest distance}}$	$+1.11\times10^{-3***}$	$+1.03\times10^{-3}$	$+7.86\times10^{-4}$
Dry $crop_{150m}$	$+0.01^{***}$	+0.01	+0.01
$Orchard_{50m}$	$+7.22\times10^{-3**}$	$+7.61\times10^{-3}$	.
$Orchard_{1000m}$	-0.03^{***}	-0.03	-0.02
$NDVImean_{5000m}$	$+8.74\times10^{-4**}$	$+1.10\times10^{-3}$	$+1.27\times10^{-3}$
Model Performance	**R^2 = 0.42 ADJ R^2 = 0.41 RMSE = 5.11**	**R^2 = 0.44 ADJ R^2 = 0.42**	**R^2 = 0.57 ADJ R^2 = 0.57**

Variable	Hybrid Model Coefficient	GWR Coefficient (Median)	GTWR Coefficient (Median)
(Intercept)	$+1.39^{**}$	+0.68	-0.07
O_3 (Kriging-based)	$+0.98^{***}$	+0.99	+0.99
Wind direction	-0.01^{***}	-8.65×10^{-3}	-5.39×10^{-3}
Industrial area $_{\text{nearest distance}}$	$+0.24\times10^{-3***}$	$+1.29\times10^{-3}$	$+1.26\times10^{-3}$
Dry $crop_{150m}$	$+8.50\times10^{-3***}$	$+9.69\times10^{-3}$	$+8.19\times10^{-3}$
$Orchard_{50m}$	$+8.64\times10^{-3***}$	$+9.18\times10^{-3}$	.
$Orchard_{1000m}$	-0.03^{***}	-0.03	-0.02
Model Performance	**R^2 = 0.87 ADJ R^2 = 0.87 RMSE = 2.4**	**R^2 = 0.87 ADJ R^2 = 0.87**	**R^2 = 0.92 ADJ R^2 = 0.92**

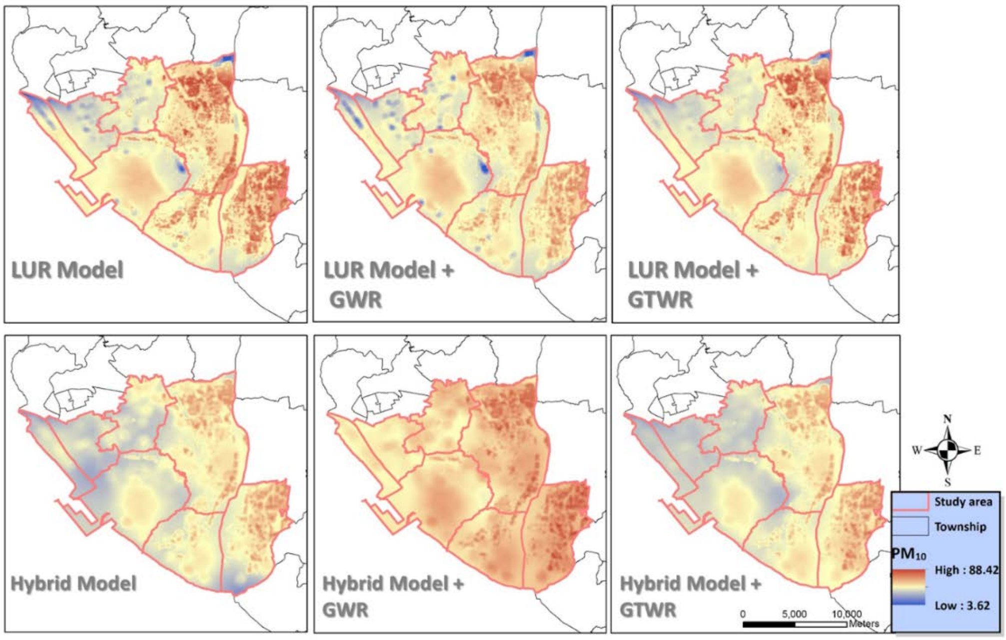

Fig. 3. Averaged PM_{10} concentration surfaces predicted by the all models

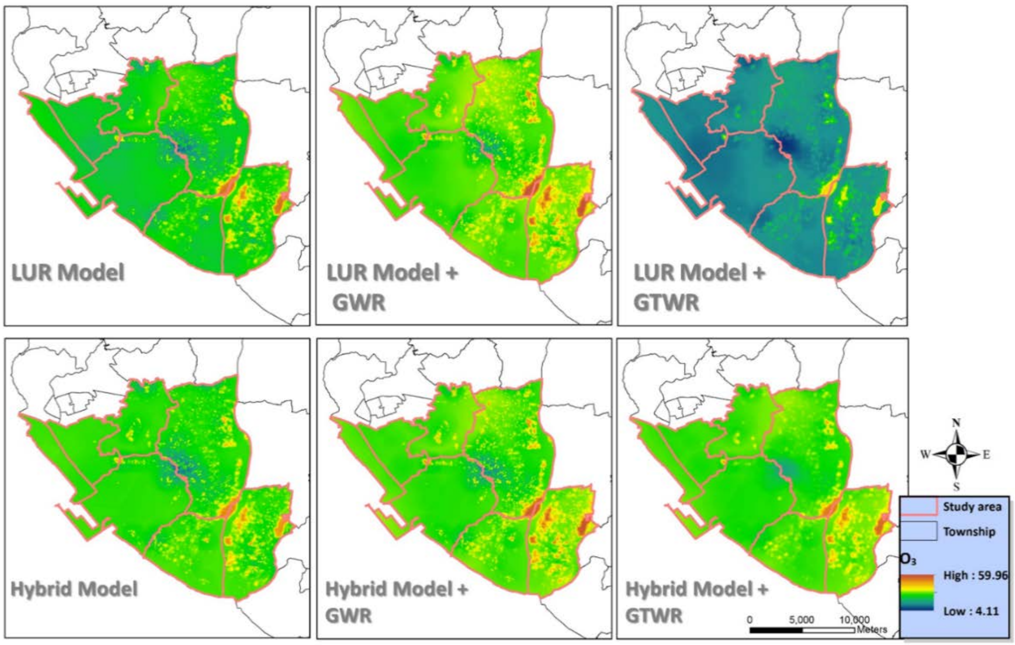

Fig. 4. Averaged O_3 concentration surfaces predicted by the all models

3. CONCLUSION

The estimated R^2 assured the robustness of the performance of integrated GTWR and kriging-based hybrid models on predicting temporal-spatial variability of PM_{10} and O_3 in this study.

ACKNOWLEDGEMENTS

This study was grant supported by the National Health Research Institutes (NHRI-108-EMGP02).

REFERENCE

Aguilera, I., Eeftens, M., Meier, R., Ducret–Stich, R. E., Schindler, C., Ineichen, A., et al. 2015. Land use regression models for crustal and traffic-related $PM_{2.5}$ constituents in four areas of the SAPALDIA study. Environmental Research, 140, pp. 377-384.

Bayraktar, H., Turalioglu, F. S., 2005. A Kriging-based approach for locating a sampling site-in the assessment of air quality. Stochastic Environmental Research and Risk Assessment, 9, pp. 301-305.

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., et al. 2010. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation, 121, pp. 2331-2378.

Chua, H.-J., Kong, S.-J., Chang, C.-H., 2018. Spatio-temporal water quality mapping from satellite images using geographically and temporally weighted regression. Int J Appl Earth Obs Geoinformation, 65, pp. 1-11.

Clougherty, J. E., Kheirbek, I., Eisl, H. M., Ross, Z., Pezeshki, G., Gorczynski, J. E., et al. 2013. Intra–urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS). Journal of Exposure Science and Environmental Epidemiology, 23(3), pp. 232-240.

Cui, L., Li, R., Zhang,Y., Meng,Y., Zhao,Y., Fu, H., 2019. A geographically and temporally weighted regression model for assessing intra-urban variability of volatile organic compounds (VOCs) in Yangpu district, Shanghai. Atmospheric Environment, 213, pp. 746-756.

Guo, Y., Tang, Q., Gong, D.-Y., Zhang Z., 2017. Estimating ground-level $PM_{2.5}$ concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. Remote Sensing of Environment, 198, pp. 140-149.

Hoek, G., Beelen, R., Hoogh, K. D., Vienneau, D., Gulliver, J., Fischer, P., et al. 2008. A review of land–use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment, 42, pp. 7561-7578.

Jerrett, M., Burnett, R. T., Ma, R., Pope, C. A., Krewski, D., Newbold, K. B., et al. 2005. Spatial Analysis of Air Pollution and Mortality in Los Angeles. Epidemiology 16(6), pp. 727-736.

Maa, X., Zhang, J., Ding, C., Wang, Y., 2018. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. Computers, Environment and Urban Systems, 70, pp. 113-124.

Michanowicz, D. R., Shmool, J. L., Cambal, L., Tunno, B. J., Gillooly, S., Hunt, M. J. O., et al. 2016. A hybrid land use regression/line–source dispersion model for predicting intra–urban NO_2. Transportation Research Part D: Transport and Environment 43, pp. 181-191.

Pope, C. A., Burnett, R. T., Thu, M. J., Calle, E. E., Krewski, D., Ito, K., et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA The Journal of the American Medical Association, 287, pp.1132-1141.
Pope, C. A., Dockery, D. W., 2006. Health effects of fine particulate air pollution: lines that connect. Journal of the Air & Waste Management Association, 56(6), pp. 709-742.
Sabaliauskas, K., Jeong, C. H., Yao, X., Reali, C., Sun, T., Evans, G. J., 2015. Development of a land-use regression model for ultrafine particles in Toronto, Canada. Atmospheric Environment, 110, pp. 84-92.
Shi, L., Zanobetti, A., Kloog, I., Coull, B. A., Koutrakis, P., Melly, S. J., et al. 2016. Low-concentration $PM_{2.5}$ and mortality: estimating acute and chronic effects in a population-based study. Environmental Health Perspectives, 124(1), pp. 46-52.
Su, J. G., Jerrett, M., Nazelle, A. D., Wolch, J., 2011. Does exposure to air pollution in urban parks have socioeconomic, racial or ethnic gradients? Environmental Research 111, pp. 319-328.
Tunno, B. J., Michanowicz, D. R., Shmool, J. L., Kinnee, E., Cambal, L., Tripathy, S., et al. 2016. Spatial variation in inversion-focused vs 24-h integrated samples of $PM_{2.5}$ and black carbon across Pittsburgh, PA. Journal of Exposure Science and Environmental Epidemiology, 26(4), pp. 365-376.
Wang, R., Henderson, S. B., Sbihi, H., Allen, R. W., Brauer, M., 2013. Temporal stability of land use regression models for traffic-related air pollution. Atmospheric Environment, 64, pp. 312-319.
Wu, C. D., Chen, Y. C., Pan, W. C., Zeng, Y. T., Chen, M. J., Guo, Y. L., et al. 2017. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model $PM_{2.5}$ spatial–temporal variability. Environmental Pollution, 224, pp. 148-157.