



*Supplement of*

**Development and intercity transferability of land-use regression models for predicting ambient PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> and O<sub>3</sub> concentrations in northern Taiwan**

**Zhiyuan Li et al.**

*Correspondence to:* Steve Hung Lam Yim ([yimsteve@gmail.com](mailto:yimsteve@gmail.com))

The copyright of individual parts of the supplement might differ from the article licence.

**Table S1** Description of potential predictor variables.

GIS dataset	Predictor variable	Unit	Buffer size (m)	Direction of effect	
				All <sup>a</sup>	O <sub>3</sub>
Land use data	Agriculture	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-	+
	Forest	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-	+
	Transportation	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+	-
	Water bodies	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-	+
	Built-up	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+	-
	Public utilities	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+	-
	Recreation	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-	+
	Mining or salt production	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+	-
	Others	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	-	+
Normalized difference vegetation index	NDVI value	NU <sup>b</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5001	-	+
Traffic	Length of the local road	m	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5000	+	-
	Length of the major road	m	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5001	+	-
	Length of the expressway	m	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5002	+	-
	Area of the local road	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5003	+	-
	Area of the major road	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5004	+	-
	Area of the expressway	m <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5005	+	-
	Distance to the nearest major road	m	NA	-	+
	Distance to the nearest expressway	m	NA	-	+
Population density	Population density	Person/km <sup>2</sup>	50, 100, 300, 500, 700, 1000, 2000, 3000, 4000, 5005	+	+/-
Digital elevation data	Elevation	m	NA	-	+
Meteorology	Temperature	°C	NA	+/-	+
	Wind speed	m/s	NA	-	-
	Relative humidity	%	NA	+/-	+/-

Note: a. all air pollutants except ozone. b. No unit for NDVI value.

**Table S2** Comparison of this study with previous LUR studies in Taiwan.

Study area	PM <sub>10</sub>	PM <sub>2.5</sub>	NO <sub>2</sub>	O <sub>3</sub>	Reference
Taipei–Keelung metropolitan area	$R^2$ : 0.80	$R^2$ : 0.72	$R^2$ : 0.91	$R^2$ : 0.80	This work
	LOOCV $R^2$ : 0.72	LOOCV $R^2$ : 0.53	LOOCV $R^2$ : 0.88	LOOCV $R^2$ : 0.72	
Taipei and New Taipei City	-	$R^2$ : 0.75	-	-	Ho et al., 2015
	-	EV $R^2$ : 0.62	-	-	
Taipei and New Taipei City	-	$R^2$ : 0.90	-	-	Wu et al., 2017
	-	EV $R^2$ : 0.83	-	-	
Taipei and New Taipei City	-	-	$R^2$ : 0.74	-	Lee et al., 2014
	-	-	LOOCV $R^2$ : 0.63	-	
Taipei and New Taipei City	$R^2$ : 0.87	$R^2$ : 0.95	-	-	Lee et al., 2015
	LOOCV $R^2$ : 0.74	LOOCV $R^2$ : 0.91	-	-	
Entire Taiwan region	-	-	-	$R^2$ : 0.74	Hsu et al., 2019
	-	-	-	10-fold CV $R^2$ : 0.70	
Entire Taiwan region	-	$R^2$ : 0.66	-	-	Wu et al., 2018
Entire Taiwan region	-	-	$R^2$ : 0.90	-	Chen et al., 2020
	-	-	10-fold CV $R^2$ : 0.87	-	

**Note:**

LOOCV: leave-one-out cross validation

EV: external validation

10-fold CV: 10-fold cross validation

25 **Table S3** Description of the 2019 annual average LUR models for ambient PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub> using Taipei data.

Air pollutant	Variables	Coefficient	Standard error	p	VIF	Predictive accuracy
PM <sub>10</sub>	Constant	39.1	2.66	< 0.001	NA	$R^2 = 0.91$ ; RMSE = 1.52; LOOCV $R^2 = 0.82$ ; LOOCV RMSE = 2.21.
	LU2_100	-1.00E-03	1.28E-04	< 0.001	2.8	
	LU3_50	1.84E-03	2.91E-04	< 0.001	1.1	
	LU7_2000	-8.34E-06	1.76E-06	< 0.001	2.4	
	POP_500	-2.15E-04	7.20E-05	0.011	1.5	
PM <sub>2.5</sub>	(Constant)	8.43	1.5	< 0.001	NA	$R^2 = 0.64$ ; RMSE = 1.73; LOOCV $R^2 = 0.48$ ; LOOCV RMSE = 2.07.
	LU3_300	7.13E-05	1.87E-05	0.002	1.0	
	R2_100	3.32E-04	1.45E-04	0.038	1.0	
NO <sub>2</sub>	(Constant)	19.3	1.9	< 0.001	NA	$R^2 = 0.89$ ; RMSE = 2.14; LOOCV $R^2 = 0.83$ ; LOOCV RMSE = 2.64.
	LU2_3000	-8.39E-07	1.24E-07	< 0.001	1.2	
	LU3_50	2.27E-03	3.93E-04	< 0.001	1.1	
	LU7_1000	-1.52E-05	4.76E-06	0.007	1.2	
O <sub>3</sub>	(Constant)	41.9	2.66	< 0.001	NA	$R^2 = 0.76$ ; RMSE = 2.81; LOOCV $R^2 = 0.64$ ; LOOCV RMSE = 3.45.
	LU3_700	-2.75E-05	6.85E-06	0.001	1.1	
	LU3_50	-1.95E-03	5.10E-04	0.002	1.1	

**Note:**

LU2\_100, LU2\_3000: the area of forest in buffer sizes of 100 m and 3000 m

LU3\_50, LU3\_300, and LU3\_700: the area of transportation land in buffer sizes of 50 m, 300 m, and 700 m

LU7\_1000 and LU7\_2000: the area of recreational land in buffer sizes of 1000 m and 2000 m

30 POP\_500: the number of population in a buffer size of 500 m

R2\_100: the area of major roads in a buffer size of 100 m

VIF: the variance inflation factor

LOOCV: leave-one-out cross validation

RMSE: root mean square error

35 NA: not available

**Table S4** Description of the 2019 annual average LUR models for ambient PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub> using New Taipei & Keelung data.

Air pollutant	Variables	Coefficient	Standard error	p	VIF	Predictive accuracy
PM <sub>10</sub>	Constant	-964	287	0.007	NA	$R^2 = 0.63$ ; RMSE = 3.08; LOOCV $R^2 = 0.35$ ; LOOCV RMSE = 4.29.
	Lat	39.7	11.5	0.006	1.0	
	LU4_1000	-8.83E-06	3.12E-06	0.018	1.0	
PM <sub>2.5</sub>	(Constant)	13.3	0.28	< 0.001	NA	$R^2 = 0.65$ ; RMSE = 0.61; LOOCV $R^2 = 0.54$ ; LOOCV RMSE = 0.70.
	LU2_100	-1.33E-03	3.15E-04	0.002	2.0	
	LU8_5000	1.11E-05	4.95E-06	0.049	2.0	
NO <sub>2</sub>	(Constant)	0.34	1.26	0.800	NA	$R^2 = 0.95$ ; RMSE = 1.58; LOOCV $R^2 = 0.87$ ; LOOCV RMSE = 2.60.
	LU3_2000	7.80E-06	8.56E-07	< 0.001	2.6	
	LU3_50	2.31E-03	3.52E-04	< 0.001	1.3	
	LU7_2000	-1.00E-05	1.97E-06	0.001	2.9	
O <sub>3</sub>	(Constant)	41.7	0.95	< 0.001	NA	$R^2 = 0.93$ ; RMSE = 1.27; LOOCV $R^2 = 0.88$ ; LOOCV RMSE = 1.62.
	LU3_3000	-2.13E-06	2.08E-07	< 0.001	1.0	
	LU8_2000	-4.39E-05	1.41E-05	0.012	1.0	

**Note:**

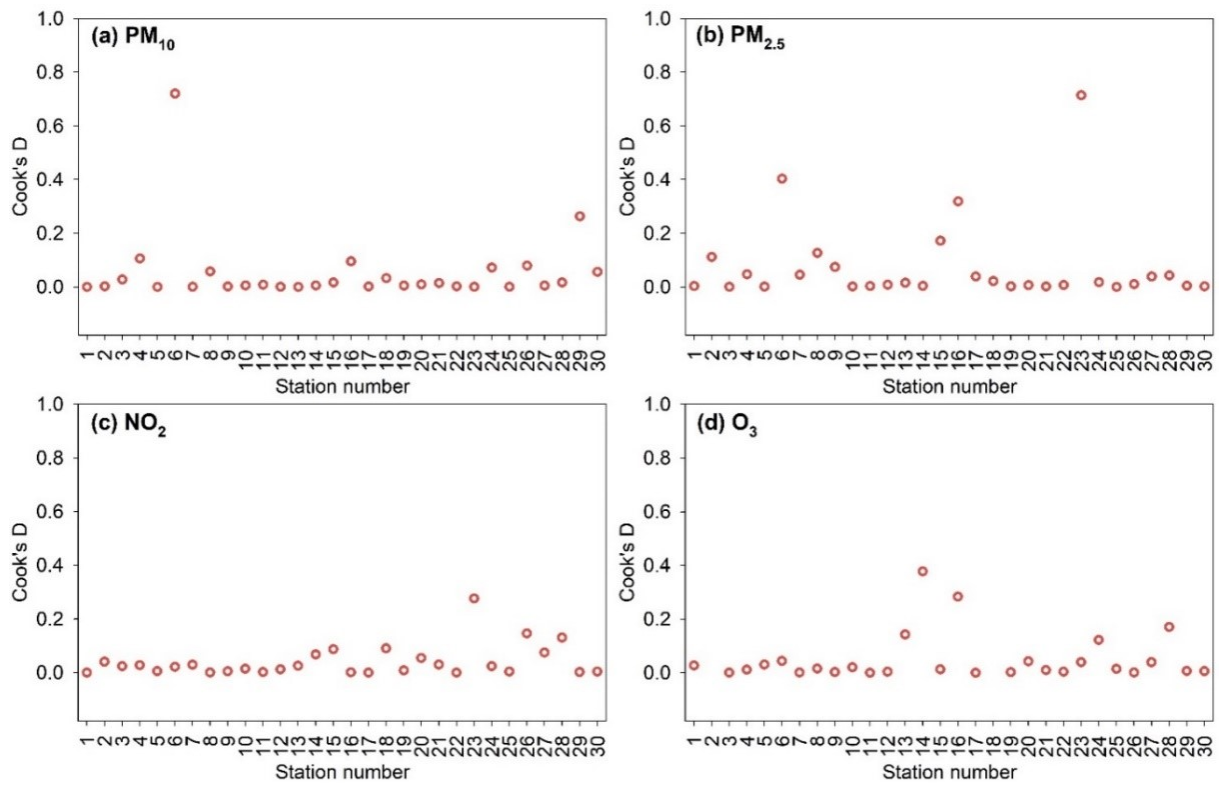
Lat: Latitude

- 40 LU4\_1000: the area of waterbody land in a buffer size of 1000 m  
 LU2\_100: the area of forest in a buffer size of 100 m  
 LU8\_2000, LU8\_5000: the area of mining or salt production land in buffer sizes of 2000 m and 5000 m  
 LU3\_50, LU3\_2000, and LU3\_3000: the area of transportation land in buffer sizes of 50 m, 2000 m, and 3000 m  
 LU7\_2000: the area of recreational land in a buffer size of 2000 m
- 45 VIF: the variance inflation factor  
 LOOCV: leave-one-out cross validation  
 RMSE: root mean square error  
 NA: not available

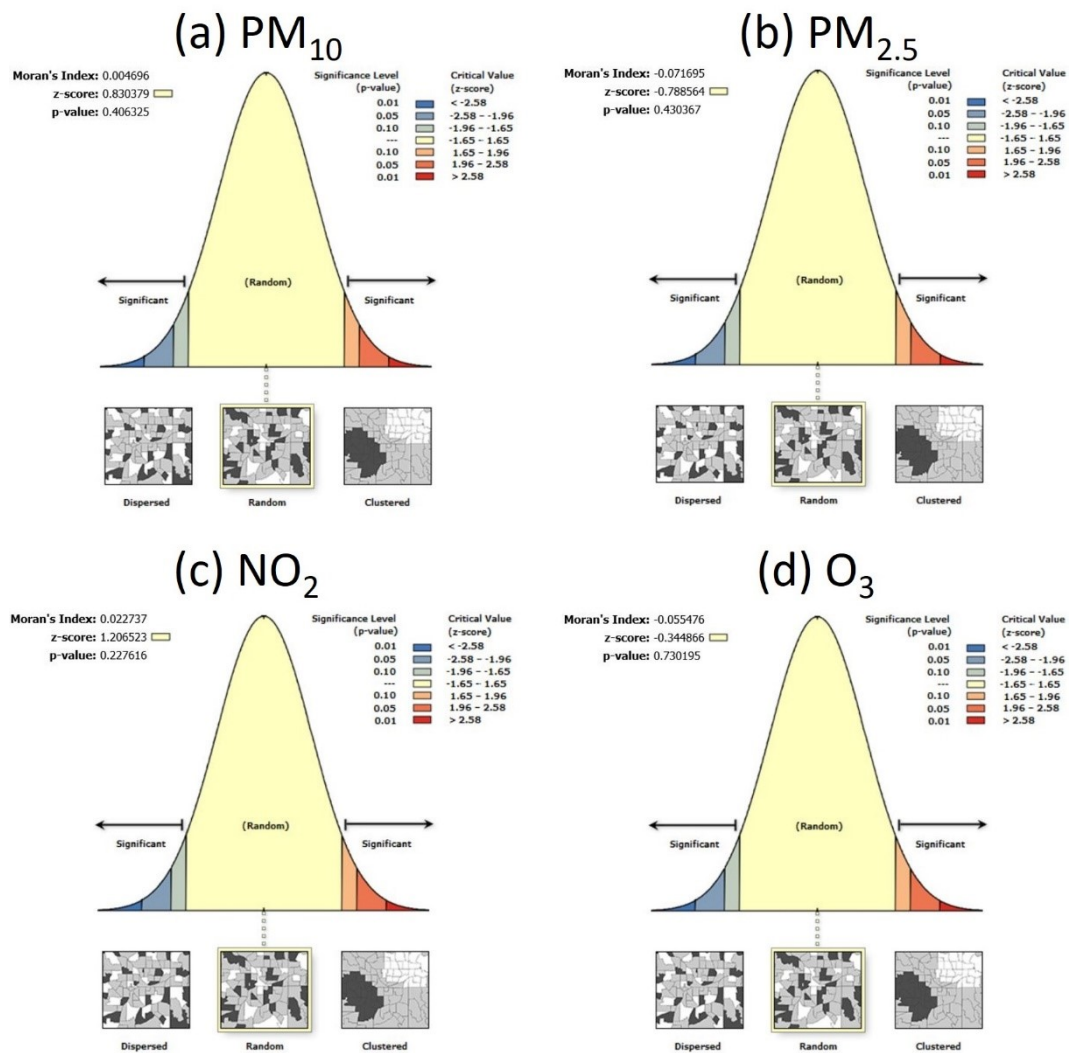
50 **Table S5** Statistical description of PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub> exposure estimates using nearby-station measurements and LUR-model-based estimates.

<b>Method</b>	<b>Nearby</b>	<b>PM<sub>10</sub> (µg m<sup>-3</sup>)</b>	<b>PM<sub>2.5</sub> (µg m<sup>-3</sup>)</b>	<b>NO<sub>2</sub> (ppb)</b>	<b>O<sub>3</sub> (ppb)</b>
<b>Nearby station measurements</b>	Mean	27.7	13.8	16.3	28.6
	SD	2.19	1.01	2.47	1.65
	Min	22.3	10.6	2.90	15.2
	Max	40.3	21.3	32.2	42.2
<b>LUR model-based estimates</b>	Mean	36.0	14.2	18.0	29.2
	SD	3.84	1.54	3.94	2.98
	Min	13.0	6.96	0.70	17.5
	Max	45.2	19.9	32.2	44.0
	Difference	8.23	0.41	1.73	0.60

Note: SD means the standard deviation; Min and Max refer to the minimum and maximum values of the air pollutant concentrations, respectively.



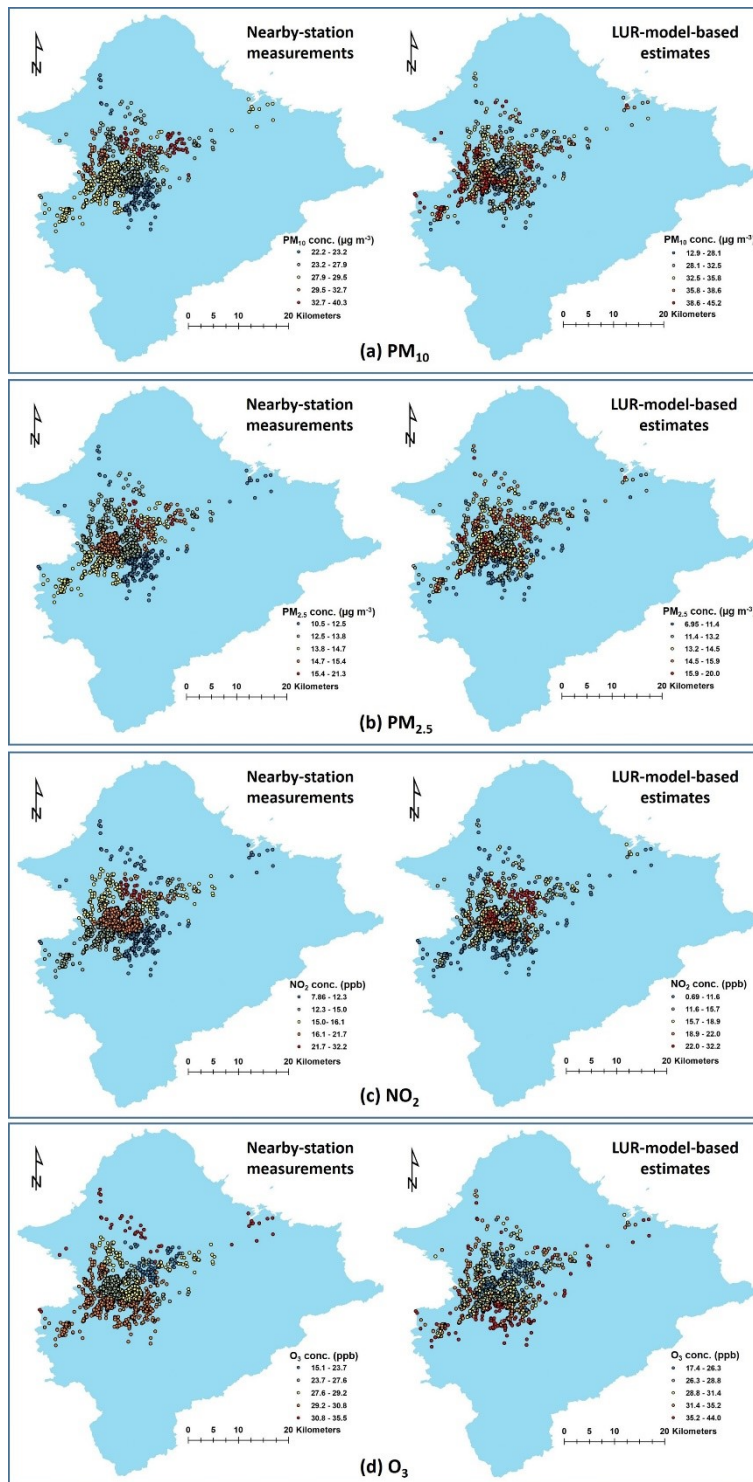
**Figure S1.** Cook's distance values of all the stations in establishing LUR models. (a)  $PM_{10}$ , (b)  $PM_{2.5}$ , (c)  $NO_2$ , and (d)  $O_3$ .



55

Figure S2. Moran's *I* index values of concentration residuals of LUR models. (a) PM<sub>10</sub>, (b) PM<sub>2.5</sub>, (c) NO<sub>2</sub>, and (d) O<sub>3</sub>.





**Fig. S3.** Spatial distributions of (a) PM<sub>10</sub>, (b) PM<sub>2.5</sub>, (c) NO<sub>2</sub>, and (d) O<sub>3</sub> exposure estimates for a cohort study using LUR-model-based estimates and nearby-station measurements.

## 60 References

- Chen, T.H., Hsu, Y.C., Zeng, Y.T., Lung, S.C.C., Su, H.J., Chao, H.J., and Wu, C.D.: A hybrid kriging/land-use regression model with Asian culture-specific sources to assess NO<sub>2</sub> spatial-temporal variations. *Environ. Pollut.* 259, 113875, 2020.
- 65 Ho, C.C., Chan, C.C., Cho, C.W., Lin, H.I., Lee, J.H., and Wu, C.F.: Land use regression modeling with vertical distribution measurements for fine particulate matter and elements in an urban area. *Atmos. Environ.* 104, 256-263, 2015.
- Hsu, C.Y., Wu, J.Y., Chen, Y.C., Chen, N.T., Chen, M.J., Pan, W.C., Lung, S.C.C., Guo, Y.L., and Wu, C.D.: Asian culturally specific predictors in a large-scale land use regression model to predict spatial-temporal variability of ozone concentration. *Int. J. Environ. Res. Public Health* 16(7), 1300, 2019.
- 70 Lee, J.H., Wu, C.F., Hoek, G., de Hoogh, K., Beelen, R., Brunekreef, B., and Chan, C.C.: Land use regression models for estimating individual NO<sub>x</sub> and NO<sub>2</sub> exposures in a metropolis with a high density of traffic roads and population. *Sci. Total Environ.* 472, 1163-1171, 2014.
- Lee, J.H., Wu, C.F., Hoek, G., de Hoogh, K., Beelen, R., Brunekreef, B., and Chan, C.C.: LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction. *Sci. Total Environ.* 514, 178-184, 2015.
- 75 Wu, C.D., Chen, Y.C., Pan, W.C., Zeng, Y.T., Chen, M.J., Guo, Y.L., and Lung, S.C.C.: Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM<sub>2.5</sub> spatial-temporal variability. *Environ. Pollut.* 224, 148-157, 2017.
- Wu, C.D., Zeng, Y.T., and Lung, S.C.C.: A hybrid kriging/land-use regression model to assess PM<sub>2.5</sub> spatial-temporal variability. *Sci. Total Environ.* 645, 1456-1464, 2018.