Spatial Modelling of Air Pollutants in the City of Calgary and Surrounding Areas

Isabelle Couloigner[1], Stefania Bertazzon[1], Fox Underwood[1], Markey Johnson[2], Keith Van Ryswyk[2]

[1]University of Calgary, Department of Geography, 2500 University Dr. NW, Calgary, AB, Canada, T2N 1N4
Email: {icouloig, bertazzs, feunderw}@ucalgary.ca

[2]Health Canada, Air Health Science Division, 269 Laurier Ave West, Ottawa, ON, Canada, K1A 0K9
Email: {markey.johhnson, keith.vanryswyk}@hc-sc-gc.ca

Abstract

This study presents a spatial analysis of air quality over the City of Calgary and surrounding areas to support local air zone management strategies and air pollution health studies. Land use regression (LUR) models were estimated for nitrogen dioxide (NO_2), particulate matter (PM_{10}, $PM_{2.5}$), and a group of volatile organic compounds (BTEX) in summer 2015 and winter 2016. Both ordinary least squares (OLS) regression and geographically weighted regression (GWR) methods were applied. The results showed that GWR consistently performed statistically better and provides a more appropriate tool to model the spatial variability of air pollutant at the intra-urban level and across the urban-rural continuum. LUR models are consistent over a 5-year interval.

Background and Relevance

Air pollution is a health concern, due to its association with respiratory and cardiovascular, among other health conditions (Rückerl et al., 2011). In Canada, air pollution and exposure levels vary within urban areas, across urban areas, and over the urban-rural continuum (Ryan and LeMasters, 2007; Weijers et al., 2004; Westerdahl et al., 2005). This study addresses the spatial variability of selected air pollutants over an area that includes a large metropolitan area and a surrounding more rural zone. Land use regression (LUR) is used to identify land use variables associated with measured pollution levels and estimate air pollution at fine spatial resolution. The mechanics of LUR models are like any other regression model; however, the dependent variable is an air pollutant recorded over an urban area, whereas the independent variables represent land use information, as well as elevation, wind speed and direction, population densities, traffic volumes, etc. (Henderson et al., 2007).

This study builds on a previous study conducted by our group, by following up after a 5-year interval, *i.e.* sampling campaigns were conducted in summer 2010 and winter 2011, with a follow-up campaign during corresponding periods in summer 2015 and winter 2016 (Bertazzon et al., 2015; Bertazzon *et al.*, 2016). In comparison with the 2010-2011 study, the 2015-2016 study featured a larger set of sampling points and extended over a larger geographic area, including rural zones and satellite communities, in addition to the metropolitan area.

Air pollution estimates generated by this study will be used in a variety of health studies to examine associations between air pollution and various adverse health outcomes, for e.g. it will be used in a health impact assessment by CRAZ (Calgary Regional Airshed Zone) as part of its local air zone management strategies. General principles drawn from this study will be used to develop recommendations that can be broadly applied in other communities.

Methods and Data

Two-week integrated measurements were collected, at 125 sites, of nitrogen dioxide (NO_2), a group of volatile organic compounds (BTEX), particulate matter (PM_{10}, $PM_{2.5}$), black carbon (BC), and PM2.5 components during summer 2015 (August 5 - 19) and winter 2016 (January 20 - February 3). BTEX, comprising benzene, toluene, ethylbenzene and (m+p) xylenes, provide a well-rounded picture of VOCs that are present in most urban areas. The study area was centered on the metropolitan area of Calgary, and extended to Airdrie to the north, Chestermere to the east, Okotoks to the south, and Rocky View County to the west, including the areas in-between, with the exception of First Nations land managed under Treaty 7.

Land use data were collected through official sources: the Calgary Region Open Data catalogue, the City of Calgary data, and Rocky View County data. Topography information for the study site was acquired from AltaLIS lidar. Road and Rail network data for Alberta was acquired from the National Road Network (NRN) distributed by NRCan Geogratis and from DMTI Spatial. NRN uses the same road classification than the Province and Municipalities (*i.e.* 1. Highway; 2. Expressway; 3. Arterial; 4. Collector; and 5. Local) and provide a more accurate representation of Traffic Volume. Industrial point source emissions were acquired from the interactive maps of the Canadian Environmental Sustainability Indicators (CESI, 2015). Predictor variables (Table 1) were then defined on circular buffers of variable sizes from each sampling point.

Table 1. Land use, Industrial and Environmental Variables

Variables	***Unit or description***	***Circular buffers (meters)***
Elevation	Elevation at the sampling site in meters.	
Local roads	Total length of road segments within buffer, in meters.	100,200,... , 500,750, 1000
Collector roads		
Arterial roads		
Expressways and Highways		
DMTI Secondary highways		
Rail network	Total length of rail segment within buffer in meters.	
Land use: residential	Zoning category	100,200,...,500,750, 1000
Land use: industrial		
Land use: commercial		
Land use: institutional		
Land use: parks		
Industrial PM25 emissions	Reported emitting points.	1000 to 6000 every 1000
Industrial NOx emissions		
Industrial VOCs emissions		
Distance to airports (Calgary, Springbank, small) and heliports.	Distance in meters from the sites to the airport of interest.	

Air pollution data were combined with land use information to develop LUR models. Cross-correlation analyses, stepwise selection and subsets regression methods (Olejnik et al., 2000;

Frost, 2014; Loh, 2011) were used to help in identifying best predictors. Getis-Ord Gi and global Moran's I spatial statistical tests were conducted to assess spatial clustering and autocorrelation in the pollutants. Ordinary least squares (OLS) regression (Burt et al., 2009) was applied, as individual pollutants displayed different degrees of spatial variability. Geographically weighted regression (GWR) (Fotheringham et al., 2002; Bivand, 2015; Gollini et al., 2015) was applied to analyze pollutants displaying substantial spatial variability and self-similarity over short distances.

Results

For each of the pollutants, the values measured in summer and winter were summarized by descriptive statistics (Figure 1), prior to calculating summer and winter LUR models.

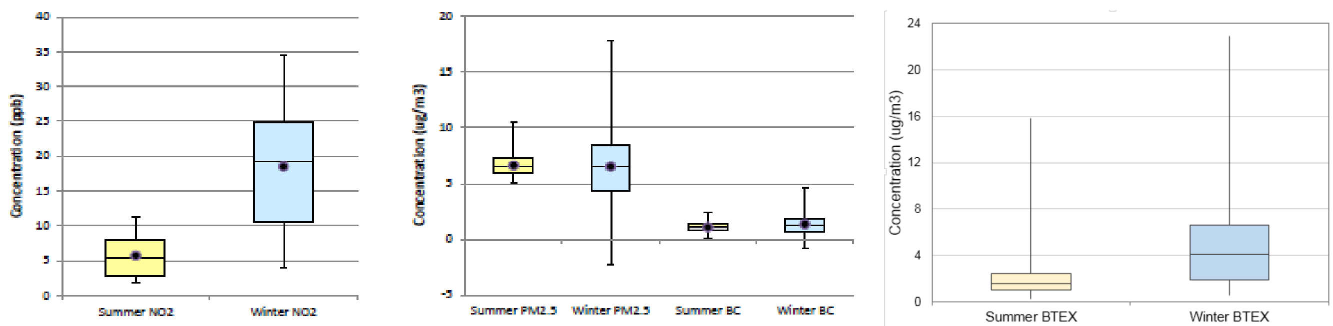

Figure 1. Seasonal observed values for NO2, PM2.5, BC and BTEX.

The plots suggest that most pollutants exhibit greater variability in the winter, since some emission sources vary by season (e.g., due to seasonal variation in heating or idling vehicles in cold temperature).

Figure 2 presents recorded values (a), along with results for OLS (b) and GWR (c) predictions of NO2 for summer 2015 as an example.

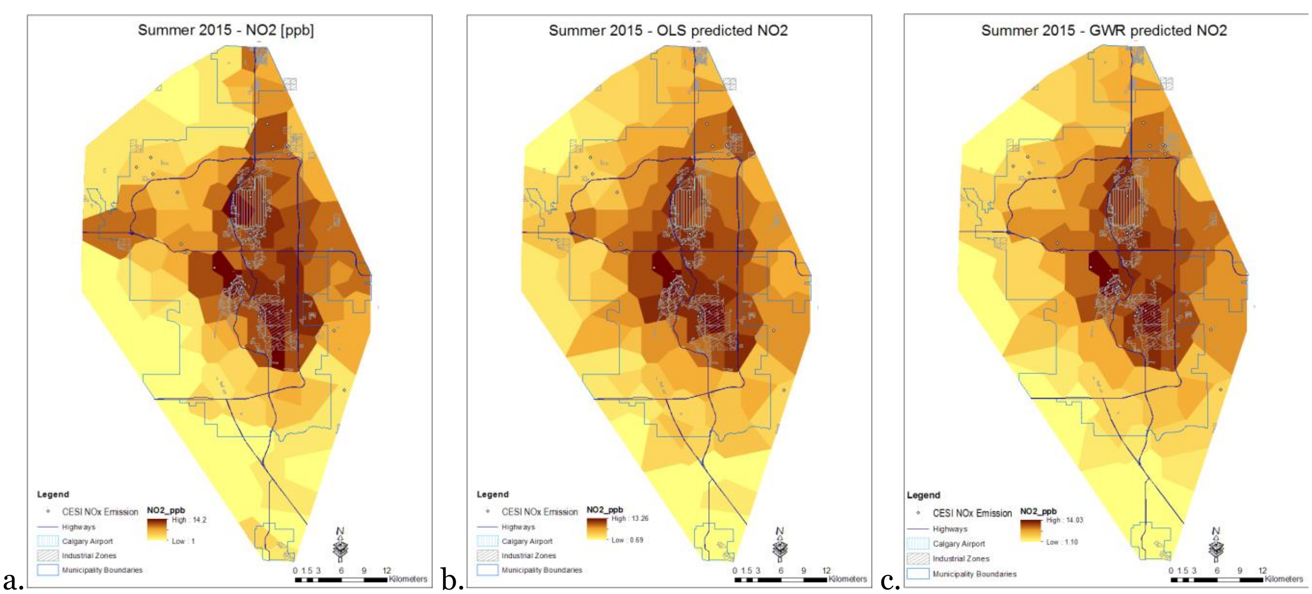

Figure 2. Predicted NO2 for summer campaign with OLS (b) and GWR (c) methods compared to Observed NO2 (a).

Most pollutants exhibited spatial patterns characterized by higher values in the eastern part of the study area, where more industrial facilities are located (Figure 2); this pattern was likely affected by the topography and prevailing winds as industrial areas are located in the flat part of Calgary and the wind mostly blows from W, NNW or SW. Most pollutants exhibited significant seasonal variations (Figure 1).

Table 2 summarizes regression results for selected pollutants.

Table 2. Standardized OLS regression predictors and GWR corresponding model results for the summer and winter campaigns

Predictors	**Summer NO2**	**Winter NO2**	**Summer PM2.5**	**Winter PM2.5**	**Summer BTEX**
Industrial land use 1000 m	0.23	0.18	0.23	0.19	
NOx/VOC emitters 5000 m	0.31	0.16			0.48
PM emitters 3000 m (S)/6000 m(W)			0.17	0.25	
Elevation	-0.25	-0.48	-0.37	-0.60	-0.28
Rail 750 m (S) / Rail 1000 m (W)	0.19	0.15			
Distance to Calgary airport	-0.38	-0.44	-0.43	-0.39	-0.14
Local roads 1000 m / 200 m (PM)		0.22		-0.07	0.14
Collector 100 (NO2) / 200 m (PM)	0.10		0.11		
DMTI Secondary highways 1000 m		-0.07		-0.07	
Arterial 750 m			0.15		
Expressways+Highways 750 m	0.16				
Commercial land use 1000 m				0.08	
OLS R^2 (adj.)	80.7 (79.4)	87.3 (86.5)	62.5 (60.1)	82.8 (81.7)	54.8 (53.2)
GWR R^2 (adj.)	89.7 (85.7)	91.8 (88.7)	N/A	91.2 (87.0)	64.7 (59.2)
OLS Breusch-Pagan test (p-value)	18.8 (0.009)	9.5 (0.22)	3.8 (0.7)	1.74 (0.88)	12.3 (0.015)
GWR F2-test (p-value)	2.14 (0.002)	1.64 (0.03)	N/A	1.84 (0.007)	2.30 (0.005)
OLS AICc	398.42	587.92	168.22	344.78	266.75
GWR AICc	373.26	584.96	N/A	328.67	255.66
OLS Res. Sum of Sq.	196.92	1217.2	26.97	113.72	61.28
GWR Res Sum of Sq.	105.58	788.63	N/A	58.4	47.84

As shown in Table 2, OLS models seemed to yield a good fit for all pollutants and seasons but summer BTEX ($R^2 = 0.55$). For all cases, the residual values were normally distributed, and there was no evidence of multicollinearity among predictors. However the Breusch-Pagan (1979) test infers evidence of heteroscedasticity in the models for Winter NO2 and PM2.5 in both seasons. Heteroscedasticity is likely associated with spatial autocorrelation and/or non-stationary of the residuals. Hence, GWR models were tested, as Getis-Ord Gi and Moran's I tests showed spatial clustering for most pollutants. In all cases, GWR models (Table 2) provided a statistically significant improvement over OLS as their R^2 increased while their AICc (Akaike, 1974) and Residuals Sum of Squares decreased. Topography (elevation), pollutant emitting sources, and distance to airport were significant predictors in all cases. Roads were significant predictors but predictive road classes varied by pollutant and season. Heteroscedasticity, or spatial heterogeneity, may also represent differences in the urban/rural continuum, which are modelled more reliably by the spatially varying GWR coefficients. Despite some differences, the overall model results were consistent with the 2010-2011 models (OLS and GWR for NO2), both in terms of predictors and fit (Bertazzon et al., 2015; 2015; Bertazzon et al., 2016). The 2015-2016 models benefitted from a larger sample size, and a more diverse sampling sites.

Conclusions

This research yields a localized version of land use regression modeling, which increases the reliability and accuracy of air quality estimates in the rural/urban continuum, and provides a detailed analysis of the significance of each predictor at the local level. While most of the times a simple LUR model such as OLS provides a satisfactory fit, GWR yields spatially varying coefficients, which can account for intra-urban and urban/rural spatial variability of the pollutants, thereby providing more reliable predictions, as well as a more accurate model of air pollution at the intra-urban and rural/urban levels. Finally, this research suggest that LUR models in Calgary were consistent over a 5-year interval. Future analyses will include temporal modelling along with the specification of spatially autoregressive (SAR) analysis.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/tac.1974.1100705

Bertazzon, S., Johnson, M., Eccles, K., & Kaplan, G. G. (2015). Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and Spatio-temporal Epidemiology, 14-15*, 9-21. doi:10.1016/j.sste.2015.06.002

Bertazzon, S., Underwood, F., Johnson, M., Zhang, J. Land Use Regression of Particulate Matter in Calgary, Canada. Miller et al., Eds., *GIScience 2016. The Ninth International Conference on Geographic Information Science. Short Paper Proceedings*. 21-24.

Bivand, R. (2015, September). Geographically weighted regression. Retrieved November, 2015, from http://cran.rproject.org/web/packages/spgwr/vignettes/GWR.pdf

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), 1287. doi:10.2307/1911963

Burt, J. E., Barber, G. M., & Rigny, D. L. (2009). *Elementary statistics for geographers* (3rd ed.). New York: Guilford Press.

CESI - Canadian Environmental Sustainability Indicators. Retrieved November, 2015, from https://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=En

Fotheringham, A. S., Brundson, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, West Sussex, England: Wiley.

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Los Angeles: Sage.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Los Angeles: Sage.

Frost, J. (2014, December). Which Is Better, Stepwise Regression or Best Subsets Regression? Retrieved January, 2016, from http://blog.minitab.com/blog/adventures-in-statistics/which-is-better,-stepwise-regression-or-best-subsets-regression

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software, 63*(17). doi:10.18637/jss.v063.i17

Henderson, S., Beckerman, B., Jerrett, M., & Braurer, M. (2007). Application of land use regression to estimate long-term concentration of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science and Technology*, *41*(7), 2422–2428.

Leung, Y., Mei, C., & Zhang, W. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A, 32*(1), 9-32. doi:10.1068/a3162

Loh, W. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 14-23. doi:10.1002/widm.8

Olejnik, S., Mills, J., & Keselman, H. (2000). Using Wherry's Adjusted R2 and Mallow's Cp for Model Selection From All Possible Regressions. *The Journal of Experimental Education, 68*(4), 365-380. doi:10.1080/00220970009600643

Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., & Peters, A. (2011). Health effects of particulate air pollution: A review of epidemiological evidence. *Inhalation Toxicology, 23*(10), 555-592. doi:10.3109/08958378.2011.593587

Weijers, E., Khlystov, A. Y., Kos, G. P. A., & Erisman, J. W. (2004). Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit. *Atmospheric Environment, 38*(19), 2993-3002. doi:10.1016/j.atmosenv.2004.02.045

Westerdahl, D., Fruin, S., Sax, T., Fine, P. M., & Sioutas, C. (2005). Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment, 39*(20), 3597-3610. doi:10.1016/j.atmosenv.2005.02.034