Land use regression models (LUR) for reliable estimation of air quality in Calgary

Stefania Bertazzon[1], Olesya Barrett[2], Markey Johnson[4], Kristin Eccles[3], Jue Yi Zhang[5]

1 Department of Geography, University of Calgary, bertazzs@ucalgary.ca
2 Department of Geography, University of Calgary, odelikan@ucalgary.ca
4 Air Health Science Division, Water Air and Climate Change Bureau, Health Canada, markey.johnson@hc-sc.gc.ca
3 Department of Geography, University of Calgary, keccles@ucalgary.ca
5 Air Health Science Division, Water Air and Climate Change Bureau, Health Canada, jue.yi.zhang@hc-sc.gc.ca

Abstract

Air quality, clean water, and land contamination are Canadians' top environmental concerns. This paper presents reliable land use regression models (LUR) to estimate air pollution and potentially adverse health outcomes at local geographical scale. Currently, most air pollution models involve interpolation of data measured at few locations over large areas. An emerging alternative is the estimation of pollutant concentrations using land use characteristics, such as traffic volumes, industrial sources, and population density. Still, land use regression models are prone to uncertainty induced by spatial phenomena: this paper yields accurate and reliable air quality estimates obtained through localized spatial analytical models. Within a project developed in collaboration with Health Canada, we conducted a summer and a winter monitoring campaign in the city of Calgary, analyzing a total of 13 pollutants, including nitrogen dioxides, sulfur dioxides, volatile organic compounds, polycyclic aromatic hydrocarbons, and particulate matter. In this paper we discuss the implementation of spatial analytical techniques to estimate reliable land use regression models.

Background and Relevance

Canadians are affected by a broad range of environmental issues, but their top concerns are air quality, clean water, and land contamination (Environment Canada; CNews, 2012). Air quality is an important factor of respiratory diseases. Over 3 million Canadians experience serious respiratory diseases (Public Health Agency of Canada, 2012). Indeed, Canadians are not simply concerned with environmental issues, but with the consequences these may have on their own lives.

Currently, most air pollution models involve interpolation of pollution data measured at a limited number of locations over large areas (Karppinen et al., 2000). Interpolation over large areas obscures local trends, resulting in inaccurate estimates especially in urban areas. A promising alternative is land use regression modeling, where pollutant concentrations are estimated at local scale using land use characteristics, such as traffic volumes, industrial sources, and population density [Gilbert et al., 2005; Ross et al., 2005; Nguyen et al., 2006; Henderson et al., 2007; Su et al., 2009). Still, land use regression models are prone to uncertainty induced by spatial phenomena: their accuracy and reliability can be enhanced by applying proper spatial analytical methods. The objective of this study is to provide accurate and reliable air quality estimates for various pollutants in the city of Calgary. Standard LUR models will be calculated for each of the recorded pollutants. Further, spatial analytical methods will be applied, to enhance the model reliability, over the whole city and locally.

Methods and Data

As part of a project run in collaboration with Health Canada, our research group conducted one summer and one winter campaign over a network of monitors to assess air quality in Calgary. To identify the best sampling period, a temporal analysis of historic data for each station was performed following the method described in (Henderson et al., 2007) and a search for the most stable period with the lowest variance was performed for each station. A network of 50 monitors was placed over the city of Calgary. About 40 sampling sites were identified using a modified version of the location-allocation method suggested by Kanaroglou and colleagues (2005) to optimize coverage and representativeness. A total of 13 pollutants were studied, including nitrogen dioxides, sulfur dioxides, volatile organic compounds, polycyclic aromatic hydrocarbons, and particulate matter. Data from active and passive sensors were extracted. Collection and refinement of explanatory variable is ongoing. Initial land use variables and traffic volumes were collected through official statistics and City of Calgary data; industrial point source emissions were acquired from the National Pollutant Release Inventory (Bertazzon et al., 2011). Speed and direction of prevailing was taken into consideration, by calculating windrose buffers for summer and winter.

LUR models typically utilize standard linear regression techniques which assume that observations are spatially independent. Nonetheless, intra-urban pollutant concentrations and land-use characteristics tend to display significant spatial variability and self-similarity over short distances, as confirmed by empirical observation and quantitative assessments, especially for traffic-related pollutants. Advanced methods exist to mitigate the effect of these properties, but neither method can cope with the simultaneous presence of the two properties. Spatial autoregressive analysis (SAR), is unreliable for non-stationary processes, as it fails to address the diversity of the study area (Anselin, 1988; Gatrell et al., 2004). Conversely, geographically weighted analysis (GWR) is unreliable for spatially dependent processes, as it fails to address local redundancies (Fotheringham et al., 1998). Hence, applying one method in the presence of the other problem, for example using GWR on spatially dependent data, may solve one problem, but aggravate the other one, yielding a model that is paradoxically worse than a non-spatial analysis.

Preliminary analysis has indicated that local methods (GWR) can enhance the estimate, by highlighting variables that are significant locally (Bertazzon et al., 2011). In this study we will extend the application of GWR methods to the whole set of pollutants, in order to inspect local variable significance, for each pollutant and for winter and summer, respectively. Following this local, exploratory analysis, we will implement spatially autoregressive methods, integrating the two analytical approaches. The models will be validated by statistical indices computed locally. The proposed method achieves greater accuracy and efficiency of LUR models. The identification of locally significant land use variables will lead to accurate air quality estimates, yielding a detailed and reliable map of air quality.

Results

For each of the pollutants, the values recorded in summer and winter are summarized by descriptive statistics and mapped, aided by interpolation techniques. Due to the high number of tables and maps, these summaries are not reported in this abstract. For each pollutant, a summer and a winter land use regression model is calculated. Summer and winter models are compared, as they generally differ both in terms of values and in terms of significant variables.

Most pollutants exhibit spatial patterns characterized by higher values in the eastern part of the city, where more industrial facilities are located; this pattern is also affected by prevailing winds. Most pollutants exhibit significant seasonal variations. LUR models achieve satisfactory results, as shown by model goodness of fit, and evaluated through various regression diagnostics.

Significant explanatory variables include local emissions, traffic volumes, elevation, as well as industrial, commercial, and residential land use zones. Seasonal variations are more pronounced for some pollutants.

Conclusions

Air quality and associated health effects are among the top environmental concerns of Canadians. The study presented in this paper provides concrete ways to improve modelling of air quality and potentially its association with health risks. Hence, this research can contribute to a more accurate public perception of environmental risks, more effective target policies, more efficient use of resources, and better tools for future planning. This research yields a localized version of land use regression modeling, which increases the reliability and accuracy of air quality estimates, and provides a detailed analysis of the significance of each predictor at the local level. The resulting models can be used to examine the association between air pollution and adverse health outcomes in future health studies. Improved air pollution estimates will improve exposure classification in these analyses, improving our ability to detect and quantify air pollution health effects and to support air pollution policy and guidelines. It will provide models and estimates calibrated on Calgary data. The methods will be directly applicable to other cities and provinces.

References

L Anselin, 1988. Spatial Econometrics: Methods and Models. New York: Kluwer.

S. Bertazzon, O. Elikan, M. Johnson, 2011. Land use regression to model air pollution and its health effects. Proceedings of the 17th European Colloquium on Quantitative and Theoretical Geography (ECQTG2011), Athens, 2-6 September 2011.

Cnews, May 22, 2012: Clean water tops Canadians' environment concerns. By Daniel Proussalidis, Parliamentary Bureau.

Environment Canada, Environmental Issues, Environmental Issues. www.ec.gc.ca.

Fotheringham, A., Brundson, C., Charlton, M., 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environ. & Planning A 30,1905–1927.

A Gatrell, J Popay, C Thomas, 2004. Mapping the determinants of health inequalities in social space: can Bourdieu help us? Health & Place. 10, 245—257

Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., & Jerrett, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. (TECHNICAL PAPER). Journal of the Air & Waste Management Association, 55(8), 1059(1055).

Henderson, S. B., 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. Environ.Sci.Technol.41, 2422-2428.

Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R. (2005). Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. 12th International Symposium, Transport and Air Pollution, 39(13), 2399-2409.

A. Karppinen, J. Kukkonen, T. Elolaehde, M. Konttinen, T. Koskentalo (2000) A modelling system for predicting urban air pollution: comparison of model predictions with the data of an urban measurement network in Helsinki. Atmospheric Environment 34:3735-3743.

Nguyen, H. T., & Kim, K.-H. (2006). Comparison of spatiotemporal distribution patterns of NO2 between four different types of air quality monitoring stations. Chemosphere, 65(2), 201-212.

Public Health Agency of Canada, 2012. Chronic Respiratory Diseases. www.publichealth.gc.ca

Ross, Z., English, P. B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., et al. (2005). Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses. 16(2), 106-114.

Su, J. G., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, J. K., & Ritz, B. (2009). Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. [Article]. Environmental Research, 109(6), 657-670.