DEVELOPMENT AND EVALUATION OF LAND-USE REGRESSION MODELS USING MODELED AIR QUALITY CONCENTRATIONS

V. Isakov[a], M. Johnson[b], J. Touma[a], H. Özkaynak[c]

[a] Atmospheric Modeling and analysis Division, National Exposure Research Laboratory, Office of Research and Development, U.S. EPA, Research Triangle Park, North Carolina, 27711

[b]Air Health Science Division, Water Air and Climate Change Bureau, Health Canada, Ottawa, Ontario, Canada K1A 0K9

[c]Office of Research and Development, U.S. EPA, Research Triangle Park, North Carolina, 27711

1. INTRODUCTION

There is a growing body of literature linking proximity to roadways and traffic intensity with adverse respiratory health effects (Health Effects Institute, 2010). Pollutants of interest include nitrogen dioxide (NO_2), fine particular matter ($PM_{2.5}$), and elemental carbon (EC) since each has been linked to vehicular emissions and respiratory health. Land-use regression (LUR) models have emerged as a preferred methodology for estimating individual exposure to ambient air pollution in epidemiologic studies in absence of subject-specific measurements. Despite their advantages of being easy to apply, simple, practical, and widely used by researchers, LUR models have several limitations. Among these are that they require accurate monitoring data and at a large number of sites, especially in highly-industrialized urban areas with many types of emission sources. In these areas, monitoring data collection is expensive and time consuming. Other limitations are: 1) these models are typically not transferable from one urban area to the another; 2) they lack the ability to connect specific sources of emissions to concentrations for developing pollution mitigation strategies; and 3) they are not typically designed to address multi-pollutant aspects of air pollution (e.g., they usually deal with one pollutant at a time). LUR methodology, however, is relatively new and many of these issues are currently being examined or addressed. In contrast, air quality models have been used for many years in air quality management but only more recently applied in exposure assessments to provide improved spatial exposure estimates (Jerrett et al., 2005). Recent developments include the hybrid air quality modeling that combines dispersion models to provide local-scale concentrations with photochemical grid models to provide regional background concentrations (Isakov et al. (2009).

2. METHODS

First, we predicted air quality concentrations of $PM_{2.5}$, NO_x, and benzene using hybrid modeling techniques based on CMAQ and AERMOD model results. $PM_{2.5}$ was selected as a criteria pollutant that is mostly regional and is largely formed in the atmosphere due to secondary reactions. NO_x was selected because it is strongly influenced by local combustion sources. Benzene was selected as representative of air toxics pollutants that are mostly emitted from mobile sources. We focused on a 20 by 20 km area that includes the impacts from a majority of emission sources. Next, we used these modeled concentrations to develop and evaluate LUR models. The LUR models were then evaluated to examine the implications of varying the number of

training sites (25, 50, 75, 100, 125, 150, 200, and 285 sites) used to build the models on LUR fit and performance.

Spatially-resolved hourly concentrations for multiple pollutants were estimated based on a hybrid air quality model approach using the AERMOD dispersion model to predict local concentration gradients (e.g., within a few meters to few kilometers from the source) and the CMAQ regional model to provide concentrations due to photochemical reactions and transport from sources outside the domain, e.g., background (Isakov et. al., 2009, Touma et. al., 2006). Spatial maps of modeled concentrations, shown in Figure 1 for all three pollutants, reveal spatial variability due to impact of mobile and stationary sources. Predicted concentration levels were highest in the city center and near port areas and highways.

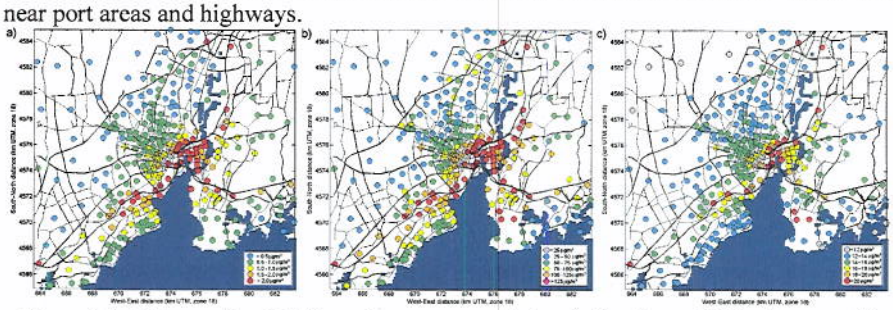

Figure 1. Spatial maps of modeled 2-month average concentrations in New Haven, CT for a) benzene, b) NO_x, and c) $PM_{2.5}$

To build a LUR model, the measured values of air quality concentrations at sampling locations are regressed against GIS variables. The dependent variables for LUR development were modeled pollutant concentrations generated by air quality models. The independent variables were traffic data, emission source information, and land-use variables. Two-month (July-August of 2001) summer time average of hourly modeled concentrations of $PM_{2.5}$, NO_x, and benzene were estimated at 318 census block group centroids in New Haven, CT. These receptors can be viewed as pseudo-monitoring sites (instead of field measurements) for the purpose of their use in the LUR model.

3. RESULTS

The LUR models for benzene, NO_x, and $PM_{2.5}$ were similar in terms of general model characteristics and performance. Benzene models typically had the greatest number of independent predictors per model, followed by NO_x and $PM_{2.5}$. We then evaluated the fitted LUR models using various approaches, including LOOCV and Hold-Out evaluation. Finally, we examined whether the LUR models performed similarly for different pollutants. For this comparison, we used the following evaluation metrics: predicted versus observed correlation (adjusted model R^2) and "Mean Residuals", calculated as mean predicted minus observed concentrations. The results of model performance or explanatory power of LUR models (measured by adjusted model R^2 values) in training datasets are shown in Figure 2. Mean and variability in adjusted R^2 values (shown here as mean and the inter-quartile range or 25^{th} – 75^{th} percentiles of the

mean adjusted R^2 values derived from 100 different iterations) for LUR models were inversely associated with the number of sites in the training set. Mean adjusted R^2 values for benzene ranged from 0.89 for training datasets with 25 sites to 0.67 for training datasets with 285 sites.

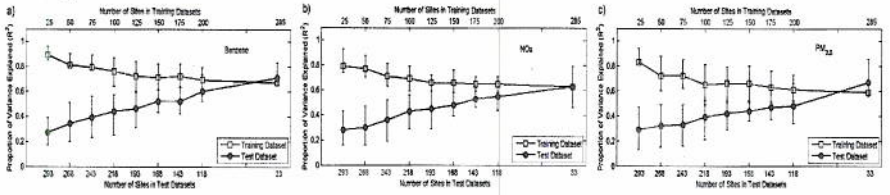

Figure 2. Mean and inter-quartile range of adjusted R^2 values for LUR models as a function of number of sites in both training (upper horizontal axis) and test datasets (lower horizontal axis) for a) Benzene, b) NOx and c) $PM_{2.5}$

For NO_x, mean adjusted R^2 values ranged from 0.79 for training datasets with 25 sites to 0.63 for those with 285 sites. Mean adjusted R^2 values for $PM_{2.5}$ ranged from 0.83 for training datasets with 25 sites to 0.59 for training datasets with 285 sites. Mean standardized prediction residuals approached zero and root mean square of standardized prediction residuals approached one as the number of training sites used to develop the models increased. These results were consistent for all three pollutants.

Figure 2 also shows the results from evaluation of LUR models based on the correlations between observed and predicted pollutant concentrations at test sites that were not used for model development. Training sites were selected randomly for each iteration of site selection, model development, and evaluation (e.g., N= 100 iterations for summer benzene models based on 25 training sites); and test sites included all sites that were not used in model development (e.g., 293 test sites for a training data set of 25 sites and 33 test sites for a training set of 285 sites). The individual test sites used to evaluate the LUR models also varied with each LUR model set. Although adjusted R^2 values for LUR models in the training datasets decreased with number of training sites used to develop the models, model performance and robustness in the test datasets used for hold-out evaluation improved with increased number of training sites . Specifically, the correlation between observed pollutant concentrations and LUR estimates in sites that were not used to develop the models (adjusted R^2 predicted versus observed in the test dataset) was positively associated with number of training sites used to develop the models. The adjusted R^2 for LUR models in the training datasets and adjusted R^2 for predicted versus observed in test datasets began to converge at approximately 125 sites. These results were observed in LUR models for benzene, NO_x, and $PM_{2.5}$.

4. SUMMARY

In this study we linked land-use regression modeling approaches with combined regional-local scale air quality models in order to evaluate and improve LUR techniques. The air quality modeling results were first evaluated against available monitoring data to assure their reliability and later used to develop and evaluate LUR models in a hierarchical fashion using an iterative site selection approach. We evaluated the fitted LUR models in several different ways and examined the implications of

varying the number of training sites used to develop the LUR models on LUR model performance for multiple pollutants.

Our results confirm the challenges facing the LUR community when attempting to fit empirical response surfaces to spatially and temporally varying urban pollution levels. Even at the 2-month averaging periods considered in this initial research, it is clear that complex emissions and atmospheric processes due to meteorological, transport, diffusion and chemical mechanisms can substantially limit the predictive power of most straightforward LUR based models. Clearly, the greater the number of sites selected for building these models (e.g., above N=100) the fit of these models can improve up to a certain level. Unfortunately, the locations where these sites are selected within the airshed could introduce additional uncertainties and potential for exposure misclassification in community based air pollution epidemiology studies. We have shown that the ambient concentration prediction errors greatly increase (nearly double) over the wide range and type of LUR models we evaluated. The variations that we observed in LUR model performance (e.g., across different models, pollutants and sample sizes) are most likely reflected in the diverse range of LUR model fits reported in the published literature, especially for NO_x and $PM_{2.5}$.

Based on these results, future work should examine best ways to augment basic LUR models with site-specific source-receptor information generated from air quality models. Despite their known limitations (e.g., need for detailed emissions and meteorological information and uncertainties due to model inputs, algorithms and outputs), air quality models have several features that can be useful in improving LUR model applications. For example, air quality models can reliably provide temporal (hourly) and spatial (at hundreds of locations) estimates, and have a long history of use by regulatory agencies in multi pollutant mitigation strategies. Air quality models are also based on physical and chemical principles, are widely used and tested in permit applications, and are readily and freely available with extensive user support tools, such as user's guides and manuals. Air quality models often undergo extensive peer review so model improvements are continuously made to enhance scientific credibility.

REFERENCES

V. Isakov, J. Touma, J. Burke, D. Lobdell, T. Palma, A., Rosenbaum, H., Özkaynak, 2009, Combining Regional and Local Scale Air Quality Models with Exposure Models for Use in Environmental Health Studies. *J. Air & Waste Manage. Assoc.* 59, 461-472.

M. Jerrett, Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Environ. Epidemiol.* 15, 185-204.

M. Johnson, V. Isakov, J. Touma, S. Mukerjee, H., Özkaynak, 2010, Evaluation Of Land Use Regression Models Used To Predict Air Quality Concentrations In An Urban Area. *Atmospheric Environment,* 44, 3660–3668.