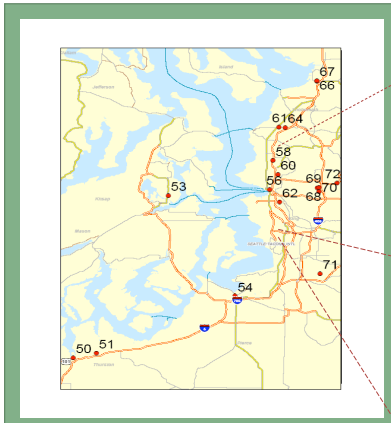
Development of a Spatio-Temporal Model for Predicting CO levels

Anne Roubal; DePaul University; University of Washington Summer Intern in Environmental Health and Occupational Health Sciences
Sun Young Kim PhD.; University of Washington, Department of Epidemiology
Lianne Sheppard PhD.; University of Washington, Departments of Environmental Health Sciences and Biostatistics

Abstract

The association between increasing air pollution and adverse health effects has been demonstrated by numerous studies in the fields of epidemiology and environmental health. Carbon monoxide (CO) is an air pollutant commonly released through vehicle emissions. It can serve as an indicator of air pollution levels in specific locations. This project is interested in predicting CO levels at homes of women in the Seattle area in order to determine whether they are predictive of carboxyhemoglobin (COHb) levels in pregnant females. One approach for a CO prediction model is Land Use Regression (LUR). A LUR predictor model will be fit to the data. A smooth curve will also be generated in order to see trends in the data. However, this project has focused mostly on developing an improved model which includes both space and time coefficients. This model is known as a spatio-temporal model. The data for model fitting was obtained from the Puget Sound Clean Air Database. Average monthly CO levels were calculated across monitor locations in the greater Seattle area for the period 1996-2006. The geographical predictors of distance to major roadway, population density, and street density within a 500m buffer were calculated using 2000 census data. The smooth curves added to the data visually appear to fit better than the LUR predictions. A spatio-temporal model will be fit and used to better capture variation in time and space at each site as well as make future predictions. This model is expected to predict the CO levels at women's homes more accurately than simply using LUR. Future work will relate CO concentration to human absorbance using the biomarker COHb. COHb levels are predictive of adverse pregnancy outcomes in women. Furthermore, the process of generating the model for CO levels can be applied to other air pollutants.




Site 61


Site 60


Site 62

Figure 2. Map of CO monitors in the Greater Seattle Area and satellite images of Site 61(top), Site 60, and Site 62 (bottom).

Site Variation

Each site has unique GIS covariates. This is exemplified by the differences in Sites 60, 61, 62.

•Site 61 is in Lynnwood. An A2 road (highway) is present in the bottom right corner of the image. Its population and street density are about average.

•Site 60 resides on 45th Street in the University District. It has the highest population density of any of the sites.

•Site 62 is a monitor on Beacon Hill. The monitor rests near a golf course and park. Its location contributes to its unusually low variation because of the lack of vehicles which pass directly by the monitor.

These evident differences in site location make it necessary to take each site into account and not generalize across them unless their covariates are similar.

CO Concentrations and Predictions across Sites

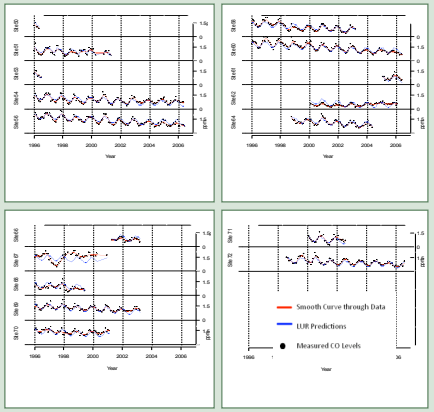


Figure 4. Graphs of CO Data, Smoothing line, and LUR predictions

A smooth curve and the LUR predictions are included with the data above. The difference between these two curves at different sites and the difference of the LUR curve from the data indicate the need for a more sophisticated model. The spatio-temporal model will be constructed in order to obtain predictions for time periods and sites where there are no data.

CO in the Atmosphere and Its Impact on Health

•The main method of CO release is through vehicle emissions. 67% of the 103 tons released last year was caused by vehicle emissions.

•CO binds to hemoglobin forming COHb, a biomarker of adverse effects (see figure).

•Long-term exposure to air pollution is associated with higher COHb levels, increased incidence of heart failure, and damage to the nervous system.

•Exposure to increased CO levels in utero is connected with preterm delivery, intrauterine growth restriction, and reduced birth weight in diverse geographic settings.[1]

•Short-term increases in CO exposure can lead to fatigue of skeletal muscles and respiratory infections.

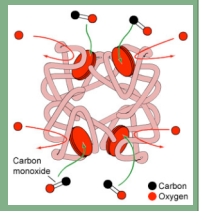


Figure 1. Carboxyhemoglobin[2]: When a high concentration of CO is present in the environment CO molecules will bind to the iron in a hemoglobin molecule causing the release of oxygen molecules and the formation of COHb.

Methods

•Average monthly CO levels were calculated using data at 17 western Washington EPA sites.

•The data were plotted and a smooth curve was fit through the points to show the overall trend at each site.

•A Land Use Regression model was fit in order to make long-term predictions from the model.

•For comparison, a spatio-temporal model will be developed to provide predictions. This model allows predictions to vary over space and time.

•CO predictions will be made at homes of women and we will assess their relationship with maternal prenatal COHb levels, an important biomarker of prenatal exposure to CO.

Land Use Regression Steps

•Calculate GIS covariates: distance to major roadways (classified as A1 – freeways, A2 – highways, and A3 – secondary roads), population density, and street density within a 500 meter buffer at the site.

•Categorize distance to roadways and street density.

•Develop the LUR regression model to obtain predicted CO values from the following equation:

$C_{predicted} = \alpha + \beta_1$ (population density) + β_2 (street density) + β_3 (distance to A1 or A2 road) + β_4 (distance to A3 road) + β_5 (month indicator) + β_6 (year indicator)

Table 1. Summary of LUR covariates

CO model predictor term	No. site-months (no. sites)	β (Standard Error) in ppm
Street density within a 500 m buffer (km/km^2)		
0-3.87	417 (9)	0
3.88-6.79	667 (8)	0.11 (0.03)
Distance to the nearest A1 or A2 road (m)		
1-410	458 (6)	0
411-815	435 (6)	0.03 (0.02)
>816	191 (5)	-0.23 (0.02)
Distance to the nearest A3 road (m)		
0.01-50	490 (6)	0
50.01-100	376 (5)	-0.29 (0.02)
>100.01	218 (6)	-0.16 (0.03)
Population density Per 1000 people/km^2	1084 (17)	0.000017 (1.02e-6)

Spatio-temporal Model

This is a model that allows the long-term average and temporal trends to vary across site. Conceptually it is a multiple regression model with a spatially varying intercept (β_{0s}) and a spatially varying coefficient for the trend (β_{1s}). Predicted CO can be written $C_{predicted}(s,t) = \beta_{0s} + \beta_{1s} f(t)$ where *f(t)* is the trend function, estimated from a smoothed singular value decomposition (SVD) of the data.

Approach

•Choose sites with over 50 months of data and format appropriately.

•Run the SVD and smooth the first singular vector to obtain trend component *f(t)*.

•Fit a multiple linear regression at each site to obtain preliminary estimates of the coefficients for each site.

•The trend coefficients will be predicted for sites with little or no data.

•The relationship between the LUR covariates and the SVD coefficients will be examined.

•When the spatio-temporal model is complete CO predictions can be made at women's homes.

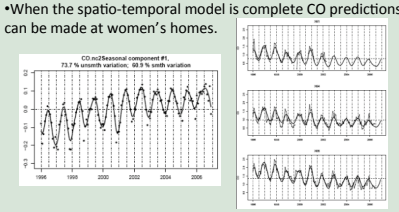


Figure 3. Trend base function for 1 SVD trend component and empirically estimated trend fits across three sites (Site numbers 51, 54, and 56)

Discussion and Conclusions, Future Steps

•The LUR predictions miss important features of the data. This is especially evident in site 62 and site 67. The LUR predictions lie far from the actual data points and would therefore be unreliable if used for predictions of CO levels.

•The spatio-temporal model does a better job of capturing how temporal variation differs between sites. This is because the curve is adjusted differently based on the unique characteristics of each site.

•More work is needed to understand how to incorporate differences across sites simply in the spatio-temporal model. It is proving difficult to fit a complex model when there are a limited number of sites with long time series.

References and Acknowledgements

1.Rudra, Carole, Williams, Michelle A, Sheppard, Lianne, et al. Whole blood carboxyhemoglobin in relation to ambient carbon monoxide exposure estimated using land use regression. Submitted

2.Carboxyhemoglobin. Orthosmoke.com 8-10-09. *http://www.orthosmoke.org/index.php/pt/Carbon%20Monoxide*

3.Kind assistance and code from Paul Sampson, Adam Szpiro

SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON
Department of Environmental and Occupational Health Sciences

Supported by National Institute of Environmental Health Sciences (NIEHS) grant 5R25ES016150-02 and National Institutes of Health (R01ES014716).