A Land Use Regression Road Map for the Burrard Inlet Area Local Air Quality Study

Prepared for the Greater Vancouver Regional District by:

Michael Brauer
Sarah B. Henderson
Julian Marshall

FINAL REPORT

December 22, 2006

Preface

This report has been prepared by its authors based on a scope of work prepared by the Greater Vancouver Regional District. While the report has been reviewed by representatives of the GVRD as sponsors of the study, the findings and conclusions expressed in this report are the opinion of the consultant authors, and do not necessarily represent the opinion of the GVRD.

Introduction to Land Use Regression

Background

Several recent studies have measured and reported considerable spatial variability in the concentrations of traffic-related pollutants within urban areas *(1-13)*. These "neighborhood scale" intra-urban differences tend not to be well-characterized by air quality monitoring networks, suggesting that exposure variation within the population is not well-characterized by regulatory monitoring networks. Land use regression (LUR) was first developed by public health researchers to address this misclassification of exposure, and the method has recently gained attention in the air quality management and urban planning communities.

There is no standard method for conducting LUR, but detailed descriptions of the general approach can be found elsewhere *(14-21)* and are summarized in this report. In brief, a pollutant is measured at multiple sites specifically selected to capture the complete intra-urban range of its concentrations. Geographic attributes that might be associated with those concentrations are measured around each site in a Geographic Information System (GIS). Typical geographic predictor variables describe site location, surrounding land use, population density, and traffic patterns. Linear regression is used to correlate measured concentrations with the most predictive variables, and the resulting equation can be used to estimate pollutant concentrations anywhere that all of the predictors can be measured. Concentration maps with high spatial resolution can be generated by rendering the regression model in GIS. Figure 1 summarizes the approach.

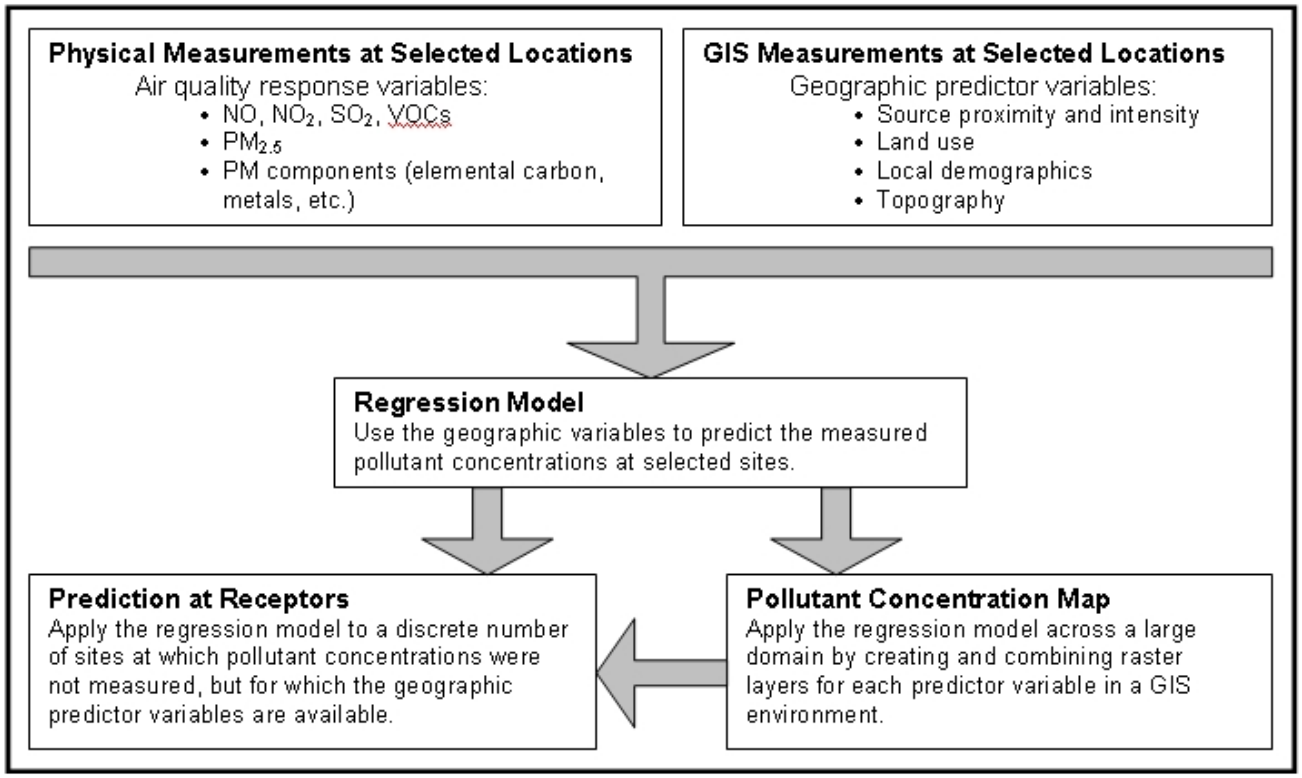


Figure 1. The LUR modeling procedure.

Literature Review

Previous Studies

Land use regression was initially developed in Europe to help estimate individual-level exposure to traffic-related air pollutants for epidemiological studies of large populations *(15, 18, 22-25)*. This need arose from (1) the infeasibility of collecting individual measurements for all subjects and (2) inaccuracies inherent to crude surrogates such as self-reported traffic exposure, distance to nearest road, or data from the nearest regulatory monitoring locations. With LUR, researchers were able to estimate individual exposures from statistical models that combined the predictive power of several surrogates based on their relationship with measured concentrations. Although interest in traffic-related health effects has favoured the development of LUR for traffic-related pollutants, the method is now being explored for other applications, such as mapping the spatial variability of residential wood smoke *(26)*.

The initial development and application of LUR was in 1993-1994 as part the SAVIAH (Small Area Variations In Air pollution and Health) studies, which focused on intra-urban variation in NO_2 within four European cities *(25)*. Models were built on a limited number of measurements with small sets of predictors. Beginning in 1999 the international TRAPCA (Traffic Related Air Pollution and Childhood Asthma) study extended this approach to airborne particulate matter. Substantial variability in annual average concentrations of NO_2, $PM_{2.5}$ and "soot" (a surrogate for elemental carbon) was measured at the 40 sites in three study locations. At least 62–85% of this variability was explained by the available predictor variables.

Since its inception in SAVIAH and TRAPCA, several researchers have used LUR to characterize NO_X and PM concentrations in Canadian, American and European cities. Results published in the peer-reviewed literature are summarized in Table 4 (page 16). While most of these studies were undertaken to provide exposure assessment for concurrent or future epidemiological research, there are two notable exceptions. Gonzales et al. *(17)* used LUR in El Paso, Texas to examine traffic-related pollution around the US-Mexico border and found that three variables – (1) elevation, (2) distance to a main highway, and (3) distance to a port of entry – explained 81% of the variability in NO_2 measurements. Sahsuvaroglu et al. *(27)* used LUR in the heavily industrialized city of Hamilton, Ontario to test its performance in the context of non-traffic-related pollution. They were able to explain 76% of the variability in measured NO_2 with variables describing traffic and industrial land use. Comparison of R^2 values across study areas and pollutant types in Table 4 suggests that LUR produces consistent results regardless of location, though models for the GVRD and Montreal are somewhat less predictive than those developed elsewhere. Like Montreal, the GVRD is surrounded by a complex series of waterways, the impact of which may not have been well-characterized by the geographic predictor variables used in regression analyses. Suggestions for improving the GVRD variable set with information about shipping and port traffic are made in Section Chapter 1222972.

LUR versus Dispersion Modeling

One alternative to LUR is dispersion modeling, where emissions parameters are input into models that use physical and chemical equations to predict pollutant concentrations at individual receptors. While this is a common approach in risk assessment and air quality management evaluation, it is rarely used for epidemiological studies because dispersion models require specific inputs. Data on traffic volume, motor vehicle fleet makeup, street configurations, industrial emissions, local meteorology, etc. may not be available for all areas. Even where complete input data exist, dispersion model operation requires considerable time, resources and expertise. Users who wish to produce high-resolution maps of

pollutant concentrations must usually (1) interpolate these results or (2) have access to the computing power necessary to run the models at a higher resolution.

In comparison, LUR allows flexibility in terms of inputs, resource requirements, and outputs. Land use regression models can be built on a location-by-location basis with whatever data are available. Sampling can be conducted at a flexible number of sites over a flexible period of time using a wide range of instrumentation. Once data collection is complete the analyses can easily be conducted by individuals with a background in statistics and GIS. Final models can be rendered into high-resolution pollution maps. Because LUR is a stochastic approach that uses actual measurements, model estimates tend to be realistic. Dispersion models use estimated emission factors that can result in considerable disparity between model output and actual concentrations. On the other hand, dispersion models can easily be used to evaluate different emissions scenarios – a limitation of LUR that is addressed in Section Chapter 1222972.

As part of the SAVIAH study Briggs et al. *(24)* compared LUR with other methods for estimating intra-urban spatial variability in air pollutant concentrations including the CAR and CALINE dispersions models. Their results are reproduced in Table 1.

Table 1. Comparison of the performance of NO_2 mapping methods*

Site	Statistic	CALINE-3	TIN-contouring	Kriging	Trend surface analysis	LUR
Amsterdam	R^2 S.E.E.	- -	0.39 (10) 7.51	- -	0.48 (10) 6.99	0.79 (10) 4.45
Huddersfield	R^2 S.E.E.	0.63 (8) 5.25	0.56 (7) 5.69	0.44 (8) 6.45	0.27 (8) 8.04	0.82 (8) 3.69
Prague	R^2 S.E.E.	- -	0.09 (9) 10.66	0.34 (9) 10.66	0.37 (9) 10.44	0.87 (10) 4.67

*Values in parentheses refer to the number of sites

Within the TRAPCA project, results for LUR and dispersion models of NO_2 concentrations were compared in Stockholm and Munich. In Stockholm the R^2 for estimates made with the AIRVIRO[1] model and measured concentrations of NO_2 was 0.69, with greater correlations observed for sites located in street canyons. The LUR model had an R^2 value of 0.76. The TRAPCA study concluded that AIRVIRO and LUR had similar predictive power, but the applicability of LUR in the absence of emission inventories was an attractive advantage. This finding was supported in a recent study by Cyrys et al. *(28)* that compared dispersion (IMMIS net[2]) and LUR estimates of NO_2 and $PM_{2.5}$ concentrations for their study population in Munich, Germany and concluded that both methods performed equally well in estimating exposures of their study population

Even more recently, Briggs et al. *(29)* compared LUR with a state-of-the-art dispersion model (ADMS-Urban) for NO_2 and PM_{10} at a limited number of measurement sites (N=18 for PM_{10}, N=8 for NO_2) in London, England. The LUR estimates had correlations (Pearson's coefficient, r) of 0.61 for NO_2 and 0.88 for PM_{10} compared to the annual mean. The ADMS estimates had correlations of 0.72 and 0.81 for NO_2 and PM_{10}, respectively. These results suggest that LUR pollutant concentration estimates are

[1] http://www.indic-airviro.smhi.se/
[2] http://www.ivu-umwelt.de/e/index.html

of equal or better accuracy than those from dispersion models, including advanced packages like ADMS. Beyond its aforementioned flexibility, another important advantage of LUR is its applicability to specific components of particulate matter, such as elemental carbon or source-specific tracers. In contrast, sophisticated dispersion models like ADMS and CALINE4 are only available for a limited set of pollutants such as NO_2 and PM_{10}.

History in the GVRD

Traffic-Related Nitrogen Oxides

One previous study has used LUR to estimate long-term ambient concentrations of nitrogen oxides across the GVRD *(21)*. In March and September of 2003 Henderson et al measured NO_X and NO_2 with passive Ogawa® samplers fixed at 116 sites for two weeks. One-hundred sites were identified by a location-allocation model *(30)* parameterized to optimize the variability in NO_2 concentrations. The others were manually selected to address specific interests of project stakeholders. Duplicate samples were collected at 15% of the sites, and 16 samplers were collocated with chemiluminescence monitors in the GVRD network.

All samples were extracted in water and analyzed by ion chromatography. Measurements for the spring and fall campaigns were averaged to estimate the annual mean concentrations of NO and NO_2 at each site. To model these results with linear regression 55 variables in five categories were generated to describe each site in terms of its surrounding street network, traffic intensity, land use, population density, and geography. Table 2 summarizes the variable set and Table 6 (in Section Chapter 1224220) provides a general description of how variables in each category can be generated.

Table 2. Description of LUR variables used for modeling traffic-related pollution in the GVRD.

Category (N variables)	**Description**	**Variable Sub-Categories**	**Buffer Radii in Meters**
Road Length (12)	Total length (in km) of two road types.	RD1 (Highways) RD2 (Major Roads)	100, 200, 300, 500, 750, 1000
Vehicle Density (12)	Density (in vehicles/ hectare) of two vehicle types during morning rush hour.	AD (Automobiles) TD (Trucks)	100, 200, 300, 500, 750, 1000
Land Use (20)	Total area (in hectares) of five land use types.	RES (Residential) COM (Commercial) GOV (Governmental) IND (Industrial) OPN (Open Area)	300, 400, 500, 750
Population Density (6)	Density (in persons/hectare) of the population.	POP (Persons)	750, 100, 1250, 1500, 2000, 2500
Location (5)	Variables describing specific attributes (in km) of site location.	ELEV (Elevation) X (Longitude) Y (Latitude) DIST (Distance to Highway) SHOR (Distance to Seashore)	N/A

Variables in the *Road Length* and *Vehicle Density* categories were treated as mutually exclusive traffic metrics and independently combined with the remaining 31 variables to build two models for both NO and NO_2. A detailed description of the model-building assumptions and algorithm can be found elsewhere[1]. The resulting R^2 values ranged from 0.56 to 0.62 with good agreement between models built using the two traffic metrics. Because variables with 100-meter buffers were more influential for the NO models than the NO_2 models it was concluded that LUR was sensitive to the distinction between primary and secondary traffic-related pollutants. A series of evaluation exercises produced R^2 values ranging from 0.31 to 0.79 for the relationship between predicted and measured concentrations

Traffic-Related Particulate Matter

Two previous studies in the GVRD have applied LUR to model fine particulate matter ($PM_{2.5}$) and its light absorbing coefficient (ABS), which is a good proxy for its elemental carbon content *(31-33)*.

In conjunction with the study described in Section Chapter 1224220, Harvard Impactors (Air Diagnostics and Engineering, Harrison, ME) and programmable pumps (SKC Inc., Model 224-PCXR8, Eighty Four, PA) were used to collect one-week samples of $PM_{2.5}$ at 25 sites subset from those identified by location-allocation. Five battery- and solar-powered units were rotated between the sites over eight weeks from March through May of 2003. A sixth unit was collated with the TEOM at GVRD station T18 in North Burnaby and data from the TEOM were used to adjust weekly measurements for temporal variability during the study period (refer to Section Chapter 1224220). The mass concentration of $PM_{2.5}$ was measured by microbalance and the ABS coefficient was measured with a Smokestain Reflectometer (Diffusion Systems Ltd. Model 43, Harwell, UK).

Variables generated for the NO and NO_2 models (Table 2) were also used for $PM_{2.5}$ and ABS. Both the *Road Length* and *Vehicle Density* models had R_2 values of 0.52 for $PM_{2.5}$, but their performance in evaluation exercises was poor. The values for ABS were 0.39 and 0.41, respectively, and evaluation performance was equally poor. Other studies have achieved better results from more sampling locations *(20, 23, 34)* and it was concluded that 25 sites is not adequate for LUR analyses on particulate matter in the GVRD.

In a 2005 follow-up, Larson et al used a mobile particle soot absorption photometer (PSAP, Radiance Research, Seattle WA)[2] to measure the real-time light absorbance of ambient particulate matter *(35)* at 39 of the 116 sites described in Section Chapter 1224220. Of these, 10 were also included in the 25 sites used for the ABS models described above (Pearson's correlation between measurements = 0.41). A central reference site was established at an intersection (41st and Cambie) for temporal adjustment of the measurements, and it was visited at least once during each sampling day. The same protocols and variables described above were used for the regression analyses, and R^2 values for the Road Length and Vehicle Density models were 0.56 and 0.65, respectively. Performance on evaluation exercises was consistent with that of the NO and NO_2 models. Maps of $PM_{2.5}$ and its elemental carbon content (as estimated from the absorbance coefficient) around the Burrard Inlet are found in Figure 2 on page 17.

Wood Smoke

Residential wood smoke can be an important local source of ambient particulate matter during winter months *(36)* but its distribution is often not well-characterized by regulatory monitoring networks due, in part to the sparsely-located sources in residential areas. During the winter of 2005 Larson et al. conducted a mobile monitoring campaign to map the impact of wood smoke across the GVRD using LUR – a novel application of the method at the time. Researchers first identified potential hotspots for

[1] http://www.cher.ubc.ca/PDFs/traffic_report_full.pdf
[2] http://www.cmdl.noaa.gov/aero/instrumentation/psap.html

wood-burning based on property assessment data, the results of a telephone woodburning survey and topography. A network of six battery and solar-powered Harvard Impactors was fixed at potential hotspot and control sites to collect two-week (using a duty cycle to collect the equivalent of a 48-hr sample during a two-week period) samples of $PM_{2.5}$ and levoglucosan, a biomass combustion tracer compound, between October 2004 and April 2005. These samples were analyzed for levoglucosan to confirm that local $PM_{2.5}$ concentrations were associated with wood smoke. On 19 cold, clear nights (9pm to 1am) between November 2004 and March 2005 researchers conducted mobile sampling in a vehicle equipped with a logging GPS and light-scattering nephelometer (Radiance Research M903, Seattle WA)[1]. The routes were pre-selected to (1) cover the north or south half of the domain, (2) traverse populated areas, and (3) circumnavigate the fixed-location monitoring sites.

These campaigns generated more than 12 000 pairs of geospatial coordinates and light-scattering coefficients (b_{sp}) that were temporally-adjusted and merged into a single, high-resolution file for LUR analysis. To generate data for linear regression the model domain was divided into ~50 air catchments, assuming that a given location is systematically downwind of uphill sources under stable meteorological conditions (e.g. cold, clear nights). The b_{sp} values and predictive variables were averaged at the catchment level, and all uphill catchments within an 8km radius were assumed to contribute to the mean b_{sp} of the downhill catchment. Variables describing the population, ethnic composition, economic status, buildings, and wood-burning appliance usage in each catchment produced an R^2 value of 0.64. A similar mobile monitoring campaign was also conducted in the Capital regional district and comparable model results were obtained.

Methodological Considerations

Sampling Site Selection

Site selection protocols for LUR continue to evolve. Initial studies used relatively unstructured approaches, while more recent work has developed and applied front-end models that systematically optimize sampler location. Measurement sites for the TRAPCA study were selected to maximize the variation among the traffic-related predictor variables in the locations of interest. Both "urban background" and "urban traffic" sites were identified in all study locations, and "rural" sites were included for The Netherlands and Sweden to further characterize their variability. The criterion for an "urban background" site was that no more than 3000 vehicles per day should pass by it within a radius of 50 meters. The "urban traffic" sites had no sources other than traffic nearby. Some 'open' and some 'street canyon'[2] streets were also included in each country.

All LUR studies in Canada have used location-allocation models *(30)* to identify sampling sites. First, a demand surface is built from available regulatory air quality data, land use coverage, and population density to estimate how concentrations of a pollutant are distributed in the study domain. Second, a constrained spatial optimization problem is solved to select a pre-specified number of sites so that they capture the complete range of concentrations while maximizing the between-site distances. Table 4 shows that LUR models built from sites chosen by location-allocation seem to have lower R^2 values than those built from sites selected by other methods. This may be associated with differences in approaches to site selection, but is more likely the product of fundamental differences between European and North American cities.

---

[1] http://www.cmdl.noaa.gov/aero/instrumentation/RR_neph.html

[2] Defined according to the Euroairnet criteria (Larssen S and Sluyter R *Criteria for Euroairnet, the EEA air quality monitoring and information network.*; European Environment Agency: Copenhagen, 1999) as a street for which the ratio of the distance from the buildings to the axis of the street and the height of the building was less than 1.5.

In the recent development of LUR for wood smoke, researchers targeted only those areas where they expected to find the pollutant. Based on property assessment (which includes information on wood-burning appliances) data and results of a telephone survey (to find out if people actually use their wood-burning appliances) GIS was used to predict wood smoke hot spots in the GVRD and Victoria's Capital Regional District (CRD). These sites were then used to identify the mobile monitoring routes that were most likely to capture a complete range of concentrations for wood smoke-related particulate matter. This approach could easily be adapted for fixed and mobile monitoring of other source-specific pollutants.

Number of Sampling Sites

The question of how many sites should be sampled for LUR has never been formally addressed. Table 4 shows a range from 18 to 114, but there is limited evidence to support choosing more or fewer sites. Clearly the decision should be influenced by local characteristics, expected variability in the measured concentrations, and the extent to which the tails of the distribution are to be characterized. Some simple LUR models have been built from just a few sampling locations (generally using data from regulatory monitoring networks) *(37)*, although high concentration estimates from such models are likely to be inaccurate. Without prior information we suggest that no fewer than 40 sampling locations should be used for models designed to describe the full range of concentrations.

Where measurements can be made with relatively inexpensive passive samplers (as is possible for most gases), simultaneous sampling at a large number of sampling sites is feasible. More expensive, high-maintenance equipment is required for particle sampling. In the TRAPCA study $PM_{2.5}$ samples were collected in four groups of ten sites and time between sampling periods allowed field technicians to collect the samplers from one site, check them into the laboratory, and then re-deploy them to the next site. The "urban traffic" and "urban background" site types were evenly distributed over the four sampling groups. A similar procedure was used in the GVRD LUR (Henderson et al, 2006), where $PM_{2.5}$ sampling was conducted at 25 locations with five sites being sampled simultaneously. For extended sampling periods (up to 14 days) programmable pumps may be used to (1) prevent filter overloading and to (2) negate the need for external power sources. For example, the GVRD LUR used programmable, battery-powered pumps (SKC Inc., Model 224-PCXR8, Eighty Four, PA) with solar chargers to collect an equivalent 24-hour sample over the seven day period. Mobile monitoring, as described above and in more detail in the following section offers another potential approach to collect particulate matter measurements at a larger number of sites, for use in LUR modeling.

Air Pollutant Measurements

Most LUR studies have modeled gaseous pollutants based on fixed-location measurements taken with simple and well-described passive sampling techniques. Specifically, diffusion samplers like the Palmes Tube *(38)* and the Ogawa® *(39)* badge have been used to measure NO_2, although the Ogawa® samplers can measure O_3, SO_2 and NH_3 as well. The TRAPCA and GVRD studies also collected particle samples with Harvard Impactors (Air Diagnostics and Engineering, Harrison, ME) and analyzed filters for mass concentrations and light absorbance (which is a surrogate for elemental carbon). Further non-destructive analyses could be used to measure metal concentrations, and destructive methods could be applied to ionic species and organic compounds.

At least two LUR studies, introduced above, have conducted sampling with mobile monitors. The flexibility of a mobile platform allows one to (1) acquire significant information about spatial variability, (2) investigate and quantify expected hot spots, and possibly (3) identify unexpected hot spots. Examples of pollutants and instruments that would be well-suited to this approach are given in

Table 3. Instruments in this table are available with a response time of one minute or less, and their use in mobile campaigns has been previously demonstrated *(40)*.

Beyond some pilot studies on data from regulatory networks there are no examples of using fixed, continuously-logging instruments to measure air pollutants for LUR. The GVRD is currently evaluating six real-time PM samplers to supplement its existing network, and possibly for use in a series of local air quality studies. Of the samplers under consideration the Met One E-Sampler, Turnkey TOPAS and Thermo ADR 1200S also have filters for capturing sampled particles. This feature is desirable in the context of LUR because secondary analyses on the filters can provide information about particle composition. By rotating a limited number of continuous monitors between LUR sampling sites (as described above) it would be possible to characterize the between-site spatial variability and the within-site temporal variability. Because most readily available geographic predictor variables have no temporal component it would be challenging to model this type of variability with LUR, but modeling its standard deviation at each site could prove valuable. For example, we expect that sites heavily impacted by road traffic would show a strong diurnal signal while those impacted by industrial point sources might not. Alternatively, continuous monitors could be used to focus LUR predictions on specific periods of the day or week when specific sources of interest may be dominant. For example, focusing on morning rush hour periods to assess the impact of general motor vehicle emissions or on weekday vs weekday periods to distinguish between light and heavy duty vehicles.

Table 3. Selected fast response instruments for mobile monitoring.

Pollutant	**Measurement technology**
PM mass concentration	Nephelometer
PM number concentration	Condensation particle counter (CPC), Scanning mobility particle sizer (SMPS)
Black carbon PM	Aethalometer, Particle soot absorption photometer (PSAP)
CO	Electro-chemical sensor
NO, NO_2, NOx	Chemiluminescense
PAHs	Photoelectric Aerosol Sensor (PAS)

Measurement Duration and Identification of Sampling Periods

In most fixed-site LUR studies, sampling durations of 1-2 weeks have been used to capture spatial variability while minimizing the influence of short-term meteorological events. To smooth the effect of seasonal variations, the SAVIAH study proposed that multiple (~4) 14-day samples spaced throughout one year will adequately characterizes differences in annual average NO_2 concentrations across sites *(25)*. The TRAPCA study collected four 14-day samples of NO_2 and PM throughout an 18 month study period. The 14-day means for $PM_{2.5}$ were compared with daily means at several "urban background" sites in The Netherlands. The ratio of the 14-day average to daily averaged ranged from 0.96 to 1.05 with a mean of 1.01, suggesting that this sampling approach introduced little error relative to consecutive 24-hour samples that are frequently reported for regulatory purposes.

Initial analysis in the GVRD indicated that the annual mean of NO_2 could be accurately estimated by two well-timed 14-day campaigns. Specifically, five years of NO_2 data from 15 regulatory (GVRD) monitoring stations were analyzed to identify optimal sampling periods. Starting on January 1st of each year the running two-week averages for the entire year were calculated, the means of diametric values

(i.e. those separated by 26 weeks) were taken, and the results were compared to the annual mean at each station. In 70 out of 75 cases the combined means for Feb-19 to Mar-4 and Aug-20 to Sep-2 were within 15% of the annual value. Sampling for NO, NO_2 and NO_X was conducted on these approximate dates in 2003, with subsequent analyses suggesting a strong 1:1 relationship between the campaign-specific and annual means (average $R^2 = 0.97$, average slope = 0.96). Measurements for the $PM_{2.5}$ campaign were not specifically timed, and were later adjusted to approximate the long-term mean as described in Section Chapter 1224220 below.

Of course different approaches to sampling duration are needed when using mobile monitors to collect data for LUR. This is a relatively untested application of the methodology, so there is a limited body of literature upon which to base new study designs. Given that high concentrations of wood smoke are expected on cold, calm winter evenings, mobile monitoring for the wood smoke study described above was only conducted under such conditions. During these targeted sampling periods other sources of air pollution (e.g. traffic) were expected to have little influence on PM concentrations, so mobile monitoring was a source-specific approach that could provide extensive spatial coverage. Mobile routes were traveled on multiple nights and in opposite directions to reduce the potential influence of temporal variability within or between sampling periods. Given that high concentrations of traffic-related particles are expected under heavy traffic conditions, mobile campaigns to measure particle absorbance were conducted on week days during the afternoon rush hour period. Each site was sampled by driving in a clover-leaf pattern around the surrounding blocks, with the mean absorbance value providing the measurement for that site. Measurements conducted during morning rush hour periods were much more variable due resulting from atmosphere conditions preventing adequate mixing *(41)*.

Adjustment for Temporal Variation

If fixed-site LUR measurements are not collected simultaneously, differences among the sites may occur due to temporal variation (as a result of meteorological conditions, for example). To ensure that measurements reflect only spatial variability they must be adjusted for the impact of temporal variability using data from a fixed site where continuous measurements were made *(23)*. Consider, for example, a study in which 10 samplers are rotated between 50 sites for 1-week samples over the course of eight weeks. For temporal adjustment of the measurements one would 1) calculate the 8-week mean at the continuous site; 2) calculate the five 1-week means that correspond with the actual sampling periods at the continuous site; 3) divide the overall mean by the five 1-week means to obtain five adjustment factor; and 4) multiply each of the 10 measurements taken during each sampling week against the adjustment factor for that week. Mobile monitoring should be conducted with co-temporal fixed-site measurements so that results can be similarly adjusted. The procedure for the wood smoke study was quite complex, with $PM_{2.5}$ TEOM measurements from 7 and 4 sites, respectively, being spatially interpolated across the GVRD and CRD for (1) the entire duration of sampling and (2) each of the sampling nights. All logged measurements were then multiplied against their night- and location-specific adjustment factors. For the mobile ABS study in Vancouver a central site at Cambie and 41st was visited at the beginning and end of each sampling session. Measurements for each session were then multiplied against adjustment factors reflecting the ratio of the study-specific over session-specific means at the central location.

Geographical Predictor Variables

Although the availability of geographic data depends upon local circumstances, most LUR studies have used variables that measure traffic intensity (sometimes vehicle-class-specific), road classification density, distances to roads, population/building density, areas of land use classifications, and topography. Each type of variable is measured within circular buffers of several different radii

surrounding each measurement site. To date, efforts to use non-circular buffers that may capture meteorological phenomena have not resulted in superior models, though this is an area of active research *(41, 42)*. Details about the availability and construction of geographic indicators in the GVRD are presented in Section Chapter 1224220.

Regression Modeling and Pollutant Mapping

The relationship between the geographic predictor variables and measured air pollution concentrations is modeled with multiple linear regression. No standard procedure has been defined for this step, but most approaches have used similar methods. The following algorithm was applied for NO_X and PM models in the GVRD:

(1) Rank all variables by the absolute strength of their correlation with the measured pollutant.
(2) Identify the highest-ranking variable in each group (variables of the same type, but with different measurement radii).
(3) Eliminate other variables in each group that are correlated (Pearson's $r \geq 0.6$) with the most highly-ranked variable.
(4) Enter all remaining variables into a stepwise linear regression (a process that will automatically identify the model with the highest R^2 value).
(5) Remove from the available pool any variables that have (a) insignificant t-statistics and/or (b) coefficients that are inconsistent with a priori assumptions about the effect direction (e.g. pollutant concentrations should *increase* with increasing traffic impact, but they should *decrease* with increasing distance from major roads).
(6) Repeat steps 4 and 5 to convergence and remove any variable that contributes less than 1% to the R^2 value for a parsimonious final model.

The end product is a multiple regression model of the form:

$$\text{Pollutant Concentration} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 - \beta_4 X_4$$

Where, for example: 1) α is the model intercept; 2) β_1 is the coefficient for X_1, which could be population density within 300 meters; 3) β_2 is the coefficient for X_2, which could be population density within 5000 meters; 4) β_3 is the coefficient for X_3, which could the truck intensity within 50 meters; and 5) β_4 is the coefficient for X_4, which could be elevation. Of course a model could have more than four variables, but most tend to have between three and six predictors. The model above can be rendered to a map of pollutant concentrations in GIS by multiplying all cells in the variable rasters (X_1,…, X_4) against their associated coefficients (β_1,…, β_4) and summing the resulting grids with the constant intercept α.

Model Evaluation

Various approaches can be used to evaluate models. In some cases *(19)* most sampling sites are used for building the model while a smaller subset is reserved for its evaluation. By reducing variability in the input measurements this approach is likely to affect model quality, so many investigators have resorted to other (perhaps less optimal) methods of evaluation. For the Vancouver LUR studies, model estimates were compared to annual averages measured at GVRD air quality monitoring sites. Since these locations are most representative of regional background air quality, additional comparisons were made with limited measurements collected during a 2002 pilot study *(43)*. In addition, the predictive error for all models was estimated with leave-one-out (LOO) cross-validation where each model is repeatedly parameterized on N - 1 data points and then used to predict the excluded measurement. The mean difference between predicted and measured values estimates the model error. A similar approach

was used in the TRAPCA study. While model evaluation was not a priority in the pioneering LUR studies, we recommend that it be an *a priori* consideration for all further work in the field.

Review of Strengths, Challenges and Unknowns

Strength: Measurement Flexibility

Land use regression can be applied to pollutant measurements made by fixed sampler arrays or mobile monitors. Where resources do not exist for LUR-specific sampling campaigns, data from regulatory monitors or previous studies can be used. Any type of measured pollutant can be modeled, though spatial variability in the measured concentrations is required. Measurements can be taken simultaneously or consecutively and they can be adjusted so that models can reflect short- or long-term trends. By timing measurements appropriately and identifying source tracers it is possible to use LUR to model different sources.

Strength: Variable Flexibility

The quality of geographic data varies considerably from region to region, but LUR can make use of whatever files are available. For example, variables reflecting traffic intensity can be generated where cars and trucks are systematically counted, but road classifications can be used as a surrogate without significantly reducing model quality. Furthermore, data processing techniques allow analysts to explore multiple relationships between pollutant measurements and a single geographic data file. Variables like 'distance to highway', 'density of highways within a 500 meter radius', and 'distance to nearest intersection of highways' are all derived from one source. Although variable generation should always be guided by theory, creative geoprocessing allows analysts to explore a variety of hypotheses.

Challenge: Accounting for Meteorology

Meteorology plays an important role in atmospheric chemistry and pollutant transport, but its influence has proved challenging to capture in LUR analyses. Some studies have calculated the distance to sources in upwind directions *(27, 44)* but these variables have not been strong predictors of measured concentrations. For variable generation within the GVRD researchers have attempted to use triangular wedges reflecting wind direction and magnitude *(41, 42)* instead of circular buffers, but have failed to produce more predictive models. Even if such approaches were to improve model quality, they would be difficult to implement for high-resolution pollutant mapping across large domains. Research is this area is ongoing and some further ideas for capturing wind influences are discussed in section Chapter 1224220.

Another meteorological challenge for LUR is posed by the formation of street canyons, which occur where buildings prevent the dispersion of pollution along roadways. These conditions are difficult to assess from publicly-available geographic data[1] and manual classification over large areas is infeasible. The SilverEye software package by GeoTango[2] (Toronto, ON) presented a promising solution by deriving building heights and footprints from high-resolution satellite imagery, but their technology was recently acquired by Microsoft. In future we recommend that LUR researchers record street canyon categories (or urban climate zones, as shown in Appendix 1) for all sites while sampling to better examine how important this variable is for predicting pollutant concentrations. In the TRAPCA study, addition of a street canyon variable, based on fieldworker characterization, did lead to small increases in explained variability in models describing measurements of $PM_{2.5}$ (increase of 7 and 13%

[1] rumors of privately-held building footprint and height files could not be confirmed
[2] http://www.geotango.com/

for Munich and Stockholm, respectively) and elemental carbon (increase of 4 and 6% for Munich and Stockholm, respectively).

Challenge: Capturing Spatiotemporal Variability

The equipment necessary to measure the temporal component of pollutant variability is expensive, and it is challenging to design LUR models that might reflect this additional information. Most readily available geographic predictor variables are updated infrequently (e.g. annually), meaning they do not co-vary with pollutant concentrations on the temporal scale. The one exception is meteorological data, which, as discussed above, seems to have little impact on the quality of LUR models. Although no published studies have used continuously logging real-time instruments, such data would make it possible to model pollutant concentrations over relevant time averages (e.g. rush hour, daytime) with temporally-static predictor variables.

Unknown: Utility for Source Contributions

There are few published examples of LUR being used to model source-specific air quality impacts. The NO_X measurements used for most studies in Table 4 are associated with vehicular traffic, but not specific to this source. Likewise, higher ABS coefficients can indicate more pollution from diesel engines, but they are not a unique marker. In a recent innovation, Ryan et al. *(45)* used a multivariate receptor model to estimate the percent contribution of traffic to elemental carbon (EC) measurements made at 24 sites. By multiplying EC concentrations against the traffic-related fraction and running LUR on the resulting values, they used non-source-specific measurements to model the specific impact of diesel vehicles. With a considerably different methodology Larson et al. *(26)* used LUR to model and map the distribution of wood smoke in the GVRD. This study focused on the impact of a specific source through (1) campaign timing; (2) *a priori* identification of areas with elevated concentrations; (3) theory-driven buffer identification; (4) source-specific variable selection; and (5) measurements of levoglucosan, a source-specific tracer. Neither study reported robust estimates of source contributions to the measured pollutants. However, their relative success suggests that, when combined with source-apportionment models and/or source-specific study design, LUR can be valuable for characterizing spatial variability in the contribution of specific sources.

Unknown: Utility for Prediction

Land use regression is a stochastic method that is, by nature, retrospective. While it is unlikely that LUR can ever be adapted for real-time prediction modeling, it may be valuable for evaluating the air quality impacts of changes to geographic predictor variables. For example, models developed for the GVRD used output from the EMME/2 transportation model run by TransLink. The effects of proposed infrastructure on traffic flow in the GVRD can be simulated with EMME/2 and, in turn, new output could be used to illustrate air quality under different scenarios. Researchers have proposed similar methods for reconstructing historic maps of pollutant concentrations for epidemiological studies, and retrospective analyses using dispersion models and historical emissions estimates have been conducted *(46, 47)*.

Unknown: Transportability

There is little evidence to support or refute the advisability of transporting an LUR model beyond the area for which it was developed. Because geographic data and, therefore, LUR predictor variables are different between regions it is often impossible to examine this problem with replicated models. The provincial standardization of data made it possible to upscale models for the GVRD to the Georgia Air Basin, and the estimates for the city of Victoria are being evaluated with secondary set of 42

measurements. Preliminary results suggest that estimates from the GVRD model were systematically higher than the Victoria measurements, but the relative difference between sites was well-predicted *(48)*. Given that LUR estimates are generally categorized for epidemiological analyses, this is a promising result within the context of public health.

Table 4. Summary of previous LUR studies

Investigator, Year	Study Location	Domain Size (km^2)	Site Selection Method	Mean NO_2 (SD) in ppb	R^2 for NO_2 (N sites)	Mean ABS (SD) in $10^{-5}m^{-1}$	R^2 for ABS (N sites)
Henderson, 2006 Larson, 2006	Vancouver, BC	2200	Location-allocation Subset (mobile)	16.2 (5.6) -	0.56-0.60 (114) -	0.84 (0.47) 1.28 (0.83)	0.39-0.41 (25) 0.56-0.65 (39)
Sahsuvaroglu, 2006 *(27)*	Hamilton, ON	1400	Location-allocation	16.4 (3.7)	0.76 (101)	-	-
Jerrett, 2006 *(44)*	Toronto, ON	900	Location-allocation	32.7 (10.5)	0.69 (95)	-	-
Gilbert, 2004 *(16)*	Montreal, QC	1200	Location-allocation	11.6 (3.0)	0.54 (67)	-	-
Ryan, 2007 *(45)*	Cincinnati, OH	1600	Manual selection based on proximity to sources, population, etc.	-	-	0.67 (0.29)*	0.75 (24)
Ross, 2005 *(19)*	San Diego, CA	2100	Public buildings, stratified by expected concentration	14.8 (5.7)	0.77 (39)	-	-
Gonzales, 2005 *(17)*	El Paso, TX	800	Elementary schools, no stratification specified	20.6 (7.1)	0.81 (20)	-	-
Hochadel, 2006 *(20)*	Western Germany	3300	Study domain, stratified by urbanization and traffic density	13.7	0.89 (40)	1.71	0.81 (40)
Hoek, 2002 *(23)* Brauer, 2003 *(34)* (TRAPCA)	Netherlands Rotterdam Stockholm Munich	38000 200 150 80	Study domain, stratified by urbanization and traffic density	15.4 (4.9) 17.5 (3.9) 10.1 (4.0) 15.2 (4.1)	0.85 (40) 0.79 (18) 0.73 (42) 0.62 (40)	1.64 (0.58) 1.79 (0.56) 1.29 (0.35) 1.84 (0.43)	0.81 (40) 0.77 (18) 0.66 (42) 0.67 (40)
Briggs, 1997 *(15)* (SAVIAH)	Amsterdam Huddersfield Prague	30 300 50	Study domain, stratified by urbanization and traffic density	20.1-28.6 (3.4-6.7) 14.1-26.3(5.2-7.8) 12.3-21.9(5.7-9.9)	0.63 (80) 0.61 (80) 0.72 (80)	-	-

*Study measured elemental carbon (EC). Values estimated from GVRD relationship where ABS = 0.091 + 1.196*EC *(32)*

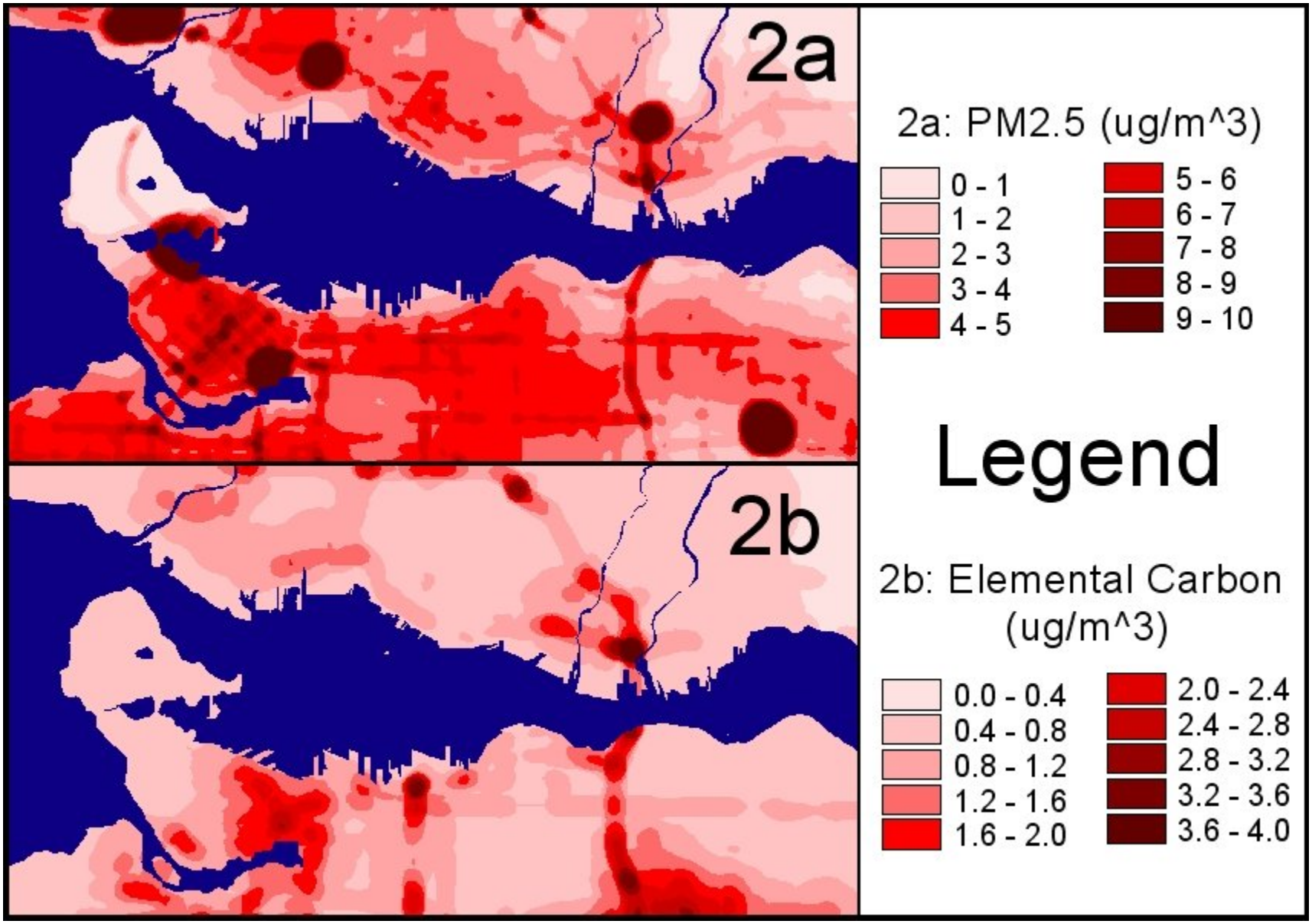


Figure 2. Land use regression estimates of $PM_{2.5}$ (2a) and elemental carbon (2b, as approximated from the absorbance coefficient) in the Burrard Inlet. Note that maximum estimated values were truncated to 120% of the maximum measured values.

Chapter 2 Conducting LUR in the GVRD

Air Pollutant Measurements

Although it is preferable to make study-specific measurements for LUR analyses, pre-existing data can inform the study design.

Regulatory Network

The GVRD has a comprehensive network of 18 regulatory monitoring stations, 9 of which are clustered around the Burrard Inlet as shown in Figure 3. Pollutants monitored at these stations are summarized in Table 5. Although it is possible to use fewer than 10 sites for LUR, measurements from regulatory networks tend not to capture much spatial variability because most monitors are located to reflect background concentrations. As such, there is little co-variation between measurements and geographic predictors, which results in an artificially narrow range of concentration estimates. However, data from these stations may be valuable when selecting sites for LUR monitoring, either manually or with location-allocation-type models.

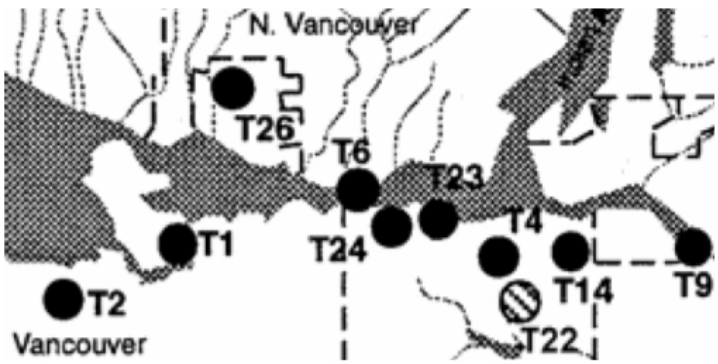


Figure 3. Regulatory air quality monitoring stations around the Burrard Inlet

Table 5. Pollutants* measured at AQ sites in the proposed study area.

ID	Name	SO_2	TRS	NO_2	CO	O_3	PM_{10}	$PM_{2.5}$	THC	VOC
T1	Downtown Vancouver	✓		✓	✓	✓				✓
T2	Kitsilano	✓		✓	✓	✓	✓	✓		
T4	Kensington Park	✓	✓	✓	✓	✓	✓	✓		
T6	Second Narrows	✓		✓	✓	✓				
T9	Rocky Point Park	✓	✓	✓	✓	✓	✓	✓		✓
T14	Burnaby Mountain			✓		✓				
T22**	Burmount		✓						✓	✓
T23	Capitol Hill	✓	✓							
T24	Burnaby North	✓	✓						✓	✓
T26	Mahon Park	✓		✓	✓	✓	✓			

*SO_2 = sulfur dioxide; TRS = total reduced sulfur; NO_2 = nitrogen dioxide; CO = carbon monoxide; 0_3 = ozone; PM_{10} = particulate matter less than 10 microns; $PM_{2.5}$ = particulate matter less than 2.5 microns; THC = total hydrocarbons; VOC = volatile organic compounds.
**T22 was a temporary station.

Other Existing Measurements

Five years of NO_2 measurements from the regulatory network were entered into a location-allocation model to help identify sampling sites for LUR studies already conducted in the GVRD *(21)*. Sites near to the Burrard Inlet are shown in Figure 4. Data from these locations could easily be reused for hypothesis-generating in further LUR analyses with new and/or different geographic predictor variables. Though few of the sites are ideally located to capture the influence of marine sources, the data could be augmented with supplemental measurements sometimes made by the GVRD.

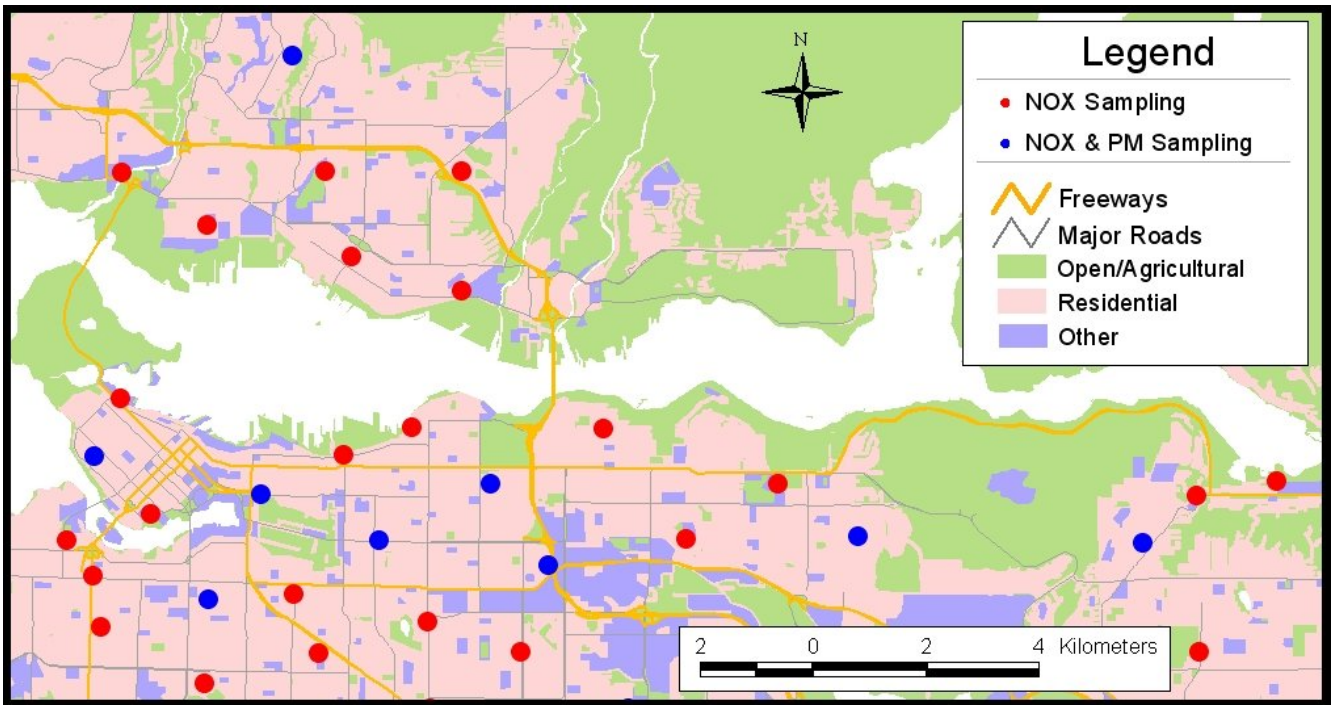


Figure 4. Sampling sites from previous LUR studies that fall around the Burrard Inlet.

Data Sources for Geographical Predictor Variables

Land Traffic

Municipalities within the GVRD have different regimens for enumerating automobile and truck traffic. Disparities between these methods make it challenging to aggregate all data into a single GIS file for LUR variable generation. While the magnitude of this problem would be diminished around the Burrard Inlet, other ready-made approaches can provide similar information.

Several LUR studies have used third-party road classifications as a proxy for traffic volume *(15, 16, 44)*. DMTI Spatial (Markham, ON) distributes a street network file that categorizes each road segment as a freeway, principal highway, secondary highway, major road or local road based on its accessibility and volume capacity. As such, the volume of traffic that impacts a point in the domain may be estimated by summing the length of each road type within near-, mid- and far-radius neighborhoods, as described in Table 2. The DMTI file was used to generate the Road Length variables discussed in Section Chapter 1224220. Similarly, the Digital Road Atlas (DRA) for British classifies each segment as a freeway, highway, arterial, collector or local road, and further classifies the highway, arterial and collector roads into major and minor sub-categories. In the GVRD LUR, substituting DRA for DMTI values resulted in a negligible impact on models.

The Vehicle Density variables discussed in Section Chapter 1224220 were derived from TransLink's EMME/2 transportation planning model (INRO, Montreal, QC). This tool simulates rush hour traffic volume based on vehicle count data, sociodemographic patterns and travel-demand surveys *(49)*. Each road segment receives a vehicles/hour value for its volume of heavy-duty trucks, light-duty trucks and automobiles. One processing methodology is described in Table 6, though other approaches could be used.

While classified street networks and simulated traffic volumes may be good proxy measures of actual traffic impact, some error is inherent in both methods. To further investigate this Setton et al. *(50)* compared actual traffic counts at 215 GVRD locations to the DMTI classification, DRA classification, and EMME/2 estimates for those sites. They concluded that actual volumes were best differentiated by the 8-category DRA classification scheme, but that caution should be taken when using any proxy measure of traffic impact.

Although not included in any LUR models to date, researchers have begun to investigate the association between measured pollutant concentrations and the density of traffic intersections. With some data cleaning, GIS could be used to identify non-residential street nodes and to calculate their distance from and density around any location.

Marine Traffic

In its year 2000 emissions inventory, the GVRD estimated that marine vessels (primarily the ocean-going fleet) are responsible for 14% of the $PM_{2.5}$ emitted in the Lower Mainland. No published studies have generated high-quality LUR variables to account for this source, but some possibilities exist within the GVRD.

As ocean-going vessels enter and exit the Vancouver harbour their movements are tracked by the Vancouver Marine Communications and Traffic Services Center (MCTS) – a branch of the Canadian Coast Guard. Radar is generally used prior to harbour entry. Ship positions are manually recorded on a digital board ever 4-6 minutes, and the records are retained indefinitely. By plotting all coordinates recorded during one year (for example) and weighting them by ship size one could generate variables describing the annual density of shipping emissions at any location. The MCTS readily admits that these data are imperfect (because manual updates may not be accurate), but they probably capture long-term trends in ship movement quite well.

The MCTS also maintains detailed records of ships in port. Information on each ship's arrival time, departure time, port and berth is recorded in a database. Although most ships shut down their main engines while in port, they continue to use diesel-powered generators to meet their energy needs. The impact of in-port emissions might be captured by multiplying ship size by stay duration, and summing the total for each port. Spatial interpolation of these values could provide valuable LUR variables.

The Chamber of Shipping of British Columbia (CSBC) is currently working to capture these mobile and stationary emissions in a province-wide Marine Vessel Emissions Inventory (MVEI). The MVEI used spatial data from the MCTS along with survey information about ship sizes, engine types, fuel types and cargo loads to estimate ship-specific emissions at five-minute intervals for every vessel in the harbour *(51)*. Final results of this study are expected in January 2007, though questions about data access for external research have yet to be addressed *(52)*. It is difficult to make recommendations about how MVEI data could be used in LUR without knowing their format, but this would certainly negate the need for separate variables reflecting ships underway and anchored at port.

Inclusion of any MCTS or MVEI data in LUR analyses will be time-consuming and computationally intensive. If resources are scare the simplest way to integrate a shipping variable is to characterize ports within the modeling domain. The Coast Guard has a comprehensive map of ports and terminals

and the CSBC Ports Handbook could be used to assign several informative attributes to each (number of berths, transport connections, etc.). Several LUR variables could be generated by measuring the distance to ports of various types, the sum distance to nearest ports, the number of ports in a specific direction, etc. as described in Table 6.

Industrial Point Sources

The Canada West Emission Inventory (CWEI) identified thousands of point-source polluters in BC and assigned estimates for annual CO, NO_X, VO_C, NH_3, SO_X, PM_{10} and $PM_{2.5}$ emissions to each. By linking these values to 2630 point source locations in and around the GVRD, Marshall *(53)* estimated relative exposure to a composite of industrial NO_X, SO_X, $PM_{2.5}$ and VOCs for 91,314 postal codes in the area (see Appendix 3). First, emissions for each pollutant at all point sources were converted from a tons/day rate into their percentile ranking. For example, a source whose NO_X emissions are at the 85th percentile amongst the 2630 emitters was assigned a NO_X value of 0.85; a source with no NO_X emissions was assigned a value of 0. Second, percentile scores for the four pollutants were summed to yield the relative emissions each point source. The largest relative emission rate was 3.96 for a point source that was in the 99th percentile for all four pollutants. The lowest rate was zero for a source with no emissions for these pollutants (N=221).

Within search radii of 10 and 40 km, Marshall used MatLab (The MathWorks, Natick, MA) to assign source counts and relative emission exposure values to each postal code. Similar density analyses within a GIS framework could produce LUR variable layers for composite emissions or individual pollutants. Other variables reflecting distance to nearest and/or upwind source(s) and tons/day emissions at those sources could also be derived.

Land Use

Assuming that air emissions in a given area are associated with its zoning, information about land use may capture some variability in measured concentrations that other source-specific variables do not reflect. Both DMTI and the GVRD have polygon files that classify the region into 7 and 8 categories, respectively. For the studies described in Section Chapter 1224220 the DMTI classifications were collapsed into those categories shown in Table 2 and processed as described in Table 6. Although visual inspection reveals considerable discrepancy between the DMTI and GVRD files, negligible differences were found when conducting analyses with both data sets.

The wood smoke study described in Section Chapter 1224220 used high-resolution spatial property assessment data (SPAD) instead of the DMTI/GVRD land use polygons. The BC Assessment Authority maintains a database of attributes for each land parcel in the province while regional authorities maintain their own cadastral (survey) polygons. The two data sources can be linked by individual parcel identifiers. Hystad et al. *(54)* describe how these data can be used to improve the quality of LUR variables by increasing their spatial accuracy and variability. Where the variables generated from the DMTI and GVRD polygons can reflect land use areas, those generated from SPAD can reflect land use densities. This distinction is analogous to the difference between the Road Length and Vehicle Density variables described in Section Chapter 1224220, with the exception that the SPAD are more accurate than the vehicle counts generated by EMME/2.

Population

While SPAD calculations can estimate residential unit density, other data sources may give more accurate information about the density of persons in the GVRD. At the dissemination area (DA) level Statistics Canada provides boundary polygons that can be linked to age-categorized population tables. For previous studies we generated population density variables by assigning the population count in

each DA polygon to its centroid and calculating densities within several radii. This method can provide good estimates where DAs are small, but the centroids of larger DAs may not reflect the actual distribution of the population. Because DAs are larger when they include areas of commercial and industrial zoning, another approach is likely more appropriate for LUR analyses specific to the Burrard Inlet.

Statistics Canada also produces population data at the block-face level for most of the GVRD. Coordinates for the center of each side of every block in the district can be linked to a population database by individual identifiers. Blocks with no residential population have 0 values and blocks with large apartment complexes have values in the hundreds. Unlike the DA population data, the block-face counts are not stratified by age. There are 58000 block face points in the GVRD and approximately 10000 around the Burrard Inlet. Population density calculated from these data would provide a high degree of spatial accuracy.

Geographic Effects

Because elevation is associated with ambient concentrations several pollutants it is an important variable to include in LUR analyses specific to the GVRD. In previous LUR work in the GVRD elevation was a significant predictor in the NO, NO_2 and $PM_{2.5}$ models. Although direct elevation has been used previously, it may be useful to further characterize the topography within an LUR domain. Variables reflecting the standard deviation or range of elevations within certain radii are simple to generate and may help to clarify the conditions under which the variable has real influence. In the wood smoke LUR, hydrological catchment basins were used to incorporate the impact of topography on nighttime pollutant drainage flows. Other geographic variables like latitude and longitude may simply capture some variability in measured concentrations not explained by other variables. They should not be significant predictors if other influences have been well-characterized, but they may serve to reflect meteorological trends that are difficult to account for within LUR analyses.

Meteorology

Experimentation with complex source area buffers to is ongoing by atmospheric scientists (see Section Chapter 1224220), but some simpler approaches may be worth pursuing in upcoming research. Wind speed and direction data are available from seventeen stations in the GVRD, and spatial interpolation can provide estimates of wind speed and direction for sites throughout the district. Wind vectors from weather forecasting models could likely be used in conjunction with topographical information to improve the interpolation results. By including new variables that reflect the average wind speed (1) overall, (2) from the eight principal directions, and (3) from the predominant direction at each site, LUR analyses might be better adjusted for the elusive influence of meteorology. Although such variables would be somewhat less accurate than those discussed above, their significance would suggest that more rigorous research in this area is warranted.

Notes on Geoprocessing

Several suggestions are made in Table 6 about the geoprocessing of predictor variables, but multiple other approaches could be taken with most of these data sets. Although LUR requires little specific expertise to apply, an experienced and creative GIS analyst is an important asset.

Table 6. Data sources and processing techniques for generation of potentially predictive LUR variables in the GVRD.

Data, Source and Format	Possible LUR Variables	Suggested Geoprocessing
• **Data:** Classified street network • **Source:** DMTI, DRA • **Format:** Line file with all necessary attributes	1. Cumulative road length within buffer radii of interest 2. Distance to specific road types 3. Distance to intersections 4. Density of intersections	1. Convert each road classification into a raster file with cell values representing pixel length (no greater than 0.001 km). Use Focal Statistics* to sum the length of roads within radii of interest. 2. Use *Euclidean Distance** for as-the-crow-flies interpretation. Could also use *Cost Distance** with a Digital Elevation Model as the cost to further account for topography. 3. Aggregate all segments by their street name. Use a node extractor from the ESRI Support Center to identify intersections. Complete visual check of the results. Use *Euclidean Distance** to find nearest node. 4. Use *Kernel* or *Point Density** to estimate intersection density.
• **Data:** Rush hour traffic volume • **Source:** TransLink • **Format:** Line file with all necessary attributes	1. Vehicle density within buffer radii of interest 2. Traffic count on nearest, 2nd nearest etc. road segment.	1. Convert line file to points spaced at one meter. Assign each point the automobiles/m and trucks/m values of the line segment from which it was derived. Use *Kernel* or *Point Density** to estimate the density values within search radius of interest. Alternately, use *Line Density**. 2. Use *Euclidean Distance** in ArcGIS for the nearest segment. The Hawths Tools *Distance Between Points* function calculates point-to-point distances matrices (with optional attributes) for feature classes.
• **Data:** Ship Tracking • **Source:** MCTS • **Format:** Table with lat/long locators and attributes	1. Number of points recorded within buffer radii of interest 2. Size-weighted density of ship traffic within radii of interest.	1. Import ASCII table and convert to event data. Use *Focal Statistics** to sum the number of events within in radii of interest. 2. Use *Kernel* or *Point Density** to reflect total tonnage of ship traffic within radii of interest.
• **Data:** Port Locations • **Source:** MCTS • **Format:** Point file with some necessary attributes	1. Distance to ports and/or terminals 2. Traffic at nearest, 2nd nearest etc. ports and terminals.	1. Use *Euclidean Distance** and/or the *Distance Between Points* tool in the Hawths Tools extension. 2. Add attribute data from MCTS port usage records and CSBC port handbook to reflect relative or absolute shipping-related traffic (marine or land-based) at ports and terminals.
• **Data:** MVEI • **Source:** CSBC • **Format:** Unknown	Relative or absolute density of marine emissions within radii of interest.	Assuming that MVEI output are points with associated emissions/five minute period, use *Kernel* or *Point Density** on absolute or percentage of emissions values.

• **Data:** Industrial Point Sources • **Source:** RWDI Air • **Format:** Table with lat/long locators and attributes	1. Distance to nearest source(s) 2. Relative or absolute emissions at nearest source(s) 3. Density of sources and/or emissions within radii of interest	1. Import the ASCII file and convert to event data using the lat/long coordinates. Save as point file. Use *Euclidean* and/or *Cost Distance** and/or *Distance Between Points* in the Hawths Tools extension to get distances. 2. The *Distance Between Points* tool can provide both distance and attributes of nearest neighbors. 3. *Point* or *Kernel Density*.*
• **Data:** Land Use • **Source:** DMTI, GVRD • **Format:** Polygon file with all necessary attributes	Total area of each land use type within radii of interest.	Convert each category of land use polygons to a raster file with pixel representing the pixel area (no more than 0.01 hectares). Use *Focal Statistics** to sum the total number of hectares within radii of interest.
• **Data:** SPAD • **Source:** BC and regional Assessment Authorities • **Format:** Tabulated attributes linked to cadastral polygons	Density of residentially-, commercially- and industrially-zoned buildings and/or parcels within radii of interest.	Link attribute data to cadastral polygons and derive polygon centroids. Calculate *Point* or *Kernel Density** for parcels with different zonings.
• **Data:** DA population • **Source:** StatsCan • **Format:** Polygon file with all necessary attributes	Density of persons within radii of interest.	Convert census polygons to centroids. Assign each centroid the population count (total or age-specific) of the centroid from which it was derived. Use *Kernel Density** to estimate values for each search radius.
• **Data:** Block Face Population • **Source:** StatsCan • **Format:** Table with lat/long locators and attributes	Density of persons within radii of interest.	Import ASCII file and convert to event data using lat/long coordinates. Save as point file. Use *Kernel Density** to estimate values for each search radius.
• **Data:** Elevation • **Source:** Census Package • **Format:** Digital Elevation Model with 30m Resolution	1. Absolute elevation 2. Range or standard deviation of elevation within radii of interest	1. None! 2. *Focal Statistics** within each radius using the Range and Standard Deviation functions
• **Data:** Wind Speed/Direction • **Source:** GVRD • **Format:** ASCII table with lat/long locators and attributes	1. Average wind speed 2. Wind speed from principal directions 3. Wind speed from predominant direction	1. Average wind speed at each site over the relevant duration. Import the ASCII file and convert to event data using the lat/long coordinates. Save as point file. Use appropriate interpolation tool (*IDW**, *Kriging**, etc.) 2. Average wind speed for 8 principal directions. Repeat #1 for each. 3. Extract maximum of 8 values from #2 and repeat #1.

*Tools in the ESRI Spatial Analyst toolbox

Chapter 3 Road Map for LUR in the Burrard Inlet

In this section we propose a two-phase plan for using LUR to study air quality in the Burrard Inlet, as summarized in Table 7. Please refer to Appendix 3 to review alternate modeling approaches that could be used to address the objectives stated here.

Table 7. Outline of objectives and methods for Phase I and II LUR in the Burrard Inlet

Objective	Method	Measurements
Phase I: Concentration Maps to Highlight Potential Hotspots* for Criteria Pollutants	Measure concentrations throughout the study area. Use all available geographic predictor variables to identify the best regression equation. Apply this equation throughout the study area to generate maps of pollutant concentrations.	Any of the Criteria Pollutants can be measured, using fixed-site, rotating-site, or mobile monitors. Pollutants with health-related thresholds are preferable for regulatory purposes.
Phase IIa: Tracer Maps to Highlight the Spatial Impact of Specific Sources	Measure concentrations of a source-specific tracer species throughout the study area. Develop a regression equation with predictor variables that are specific to the source of interest. Map concentrations of the tracer as above.	Any well-described tracer species, using fixed-site, rotating-site, or mobile monitors. Filter-based measurements of particulate matter are the most versatile option.
Phase IIb: Maps of Criteria Pollutant Concentrations Attributable to Specific Sources	Complete Phase IIa. Develop a regression model for the ratio between the tracer and the pollutant of interest (e.g. PM). Map ratio as above, and multiply by Phase IIa to get PM concentrations.	Same as above. If the ratio between the tracer and other pollutant is not well-known, then source testing would be required to make estimates.

*See definition at beginning of Section Chapter 1224220.

Phase I: Characterization of Local Air Quality

Objective: Concentration Maps to Highlight Potential Hotspots for Criteria Pollutants

Mapping the results of LUR models will show *hotspots* of elevated pollutant concentrations where multiple predictors overlap or where a single predictor is densified. The definition of a hotspot is context specific, and can be based on absolute concentrations (e.g. air quality guidelines) or relative values within a modeling domain (e.g. concentrations in the top five percentile of the overall distribution). Examples of LUR traffic-related hotspots can be seen in both Figure 2a and 2b (page 17). While the magnitude of the estimated concentrations of $PM_{2.5}$ and elemental carbon may not be accurate for these locations, their relative positioning is intuitively correct. Both the Lions Gate and Ironworkers Memorial bridges show entrance/exit hotspots for $PM_{2.5}$, but only the latter, which is open to truck traffic, has the same hotspots for elemental carbon. A classical LUR study specific to the Burrard Inlet could help air quality managers to better (1) map and understand this small-scale variability and (2) explain how this variability is associated with several geographic predictor variables.

Pollutant Measurement

For Phase I LUR in the Burrard Inlet we recommend measuring pollutants with established, health-based concentration thresholds. Table 8 summarizes the most recent air quality guidelines suggested by the Worth Health Organization (WHO), the GVRD Air Quality Objectives and the air quality standards mandated by the California Environmental Protection Agency. Rows for $PM_{2.5}$ and NO_2 are highlighted because these pollutants have long-term (annual) standards and have traditionally been measured for LUR. We recommend that filter-based measurements of $PM_{2.5}$ (time-averaged or, ideally, combined with continuously-logged measurements) would provide the most versatile measurements for addressing Phase I and II objectives in the Burrard Inlet. A single field campaign would provide (1) mass concentrations for conducting traditional LUR analyses, and (2) filter particle samples to be stored indefinitely and later fanalyzed for source-specific LUR if deemed necessary. Other pollutants of particular interested could be concurrently measured with inexpensive passive samplers.

Table 8. Guidelines and standards for criteria air pollutants.

Pollutant	Averaging Period	GVRD Objectives ($\mu g/m^3$)	WHO Guidelines ($\mu g/m^3$)	California EPA Standards ($\mu g/m^3$)
$PM_{2.5}$	Annual	25	25	35
	24-hour	12	10	12
PM_{10}	Annual	50	50	50
	24-hour	20	20	20
NO_2	Annual	40	40	100
	1-hour	200	200	470
O_3	8-hour	126	100	137
	1-hour	-	-	180
CO	8-hour	10,000	10,000	10,000
	1-hour	30,000	30,000	23,000
SO_2	24-hour	125	20	105
	1-hour	450	-	655
	10-minutes	-	500	-

Site Selection

We recommend that no fewer than 40 sites should be sampled to accurately characterize the distribution of any measured pollutant. Beyond the (1) time savings there and (2) simplified adjustment for temporal variability is no demonstrable advantage to using a single array of simultaneously fixed samplers over multiple, smaller arrays of rotating fixed samplers. In either case samplers should be located such that they are expected to optimize the distribution of concentration while maximizing the inter-sampler distance during each measurement period. This can be achieved with complex algorithms like location-allocation modeling, or with simpler geoprocessing methods that rely on some *a priori* assumptions. For example, the $PM_{2.5}$ concentration map shown in Figure 2a could be reformed into deciles, and four sites could be picked randomly from each. Weighting the selection for population density would ensure that residential areas were preferentially sampled.

Analysis

The ideal LUR model is parsimonious, with fewer than eight variables in total, and with all variables explaining at least one percent of the variability in measured concentrations. However, this stipulation should not limit the number of potentially predictive variables considered in the analyses. Any variable that is (1) theoretically associated with the measured pollutant and (2) reasonably approximated in GIS should be generated and included in the preliminary data analysis. Even when a variable is not selected for a final LUR model, the strength of its association with the measured outcome can provide valuable information about source-specific relationships. For example, we generated two variables describing truck route intersections in the GVRD and they are strongly associated (correlation coefficient > 0.6) with ABS measurements (see Section Chapter 1224220), but collinearity with other, more general variables (i.e. vehicle density) prevented their inclusion in the most predictive LUR model. If the truck-specific fraction of ABS could be separated from the general ABS, these truck-specific variables would likely become more significant than those reflecting generalized traffic. Note that a model may also include predictor variables that relate to different sources of interest, such as truck traffic and marine emissions, and statistical modeling could be used to estimate the relative impact of different source-related predictors.

Phase II: Assessing Source-Specific Contributions

Objective IIa: Tracer Maps to Highlight the Spatial Impact of Specific Sources

Although there are no published examples of such, we propose that the spatial impact of emissions from a specific source can be mapped by conducting LUR on measurements of a source-specific tracer. As stated above, LUR models for generalized pollutants will include variables that describe a variety of emission sources. We expect that models for source-specific pollutants will include (or could be statistically constrained to include) only variables specific to that source. For example, vanadium (V) and nickel (Ni) are found in the bunker fuel that most ocean-going vessels use when in transit. Thus, we expect that a LUR model for V or Ni would include (or could be constrained to include) marine-related variables as geographic predictors of concentrations.

Both species have shown a strong correlation with NO emissions from ships in Sweden *(55)*. Concentrations of V and Ni could be non-destructively measured from the filter-based PM collected in Phase I, and the results could be used to developing a shipping-specific LUR model with the shipping-specific variables described in Section Chapter 1224220. Similarly, field measurements at Slocan Park in Vancouver suggested that sulfate particles measuring 60 – 250 nanometers ($PM_{0.25}$) could also be used to trace marine emissions *(56)*, though sulfur is also found in diesel emissions from truck traffic. Table 9 provides a list of emission sources likely to be important in the Burrard Inlet and their potential tracer species.

Objective IIb: Maps of Criteria Pollutant Concentrations Attributable to Specific Sources

If the emission ratio between a tracer species and a criteria pollutant (e.g. $PM_{2.5}$) is known, the tracer-specific LUR maps generated in Phase IIb could be multiplied against that ratio to produce maps of criteria pollutant concentrations attributable to a specific source. For example, Cooper and Gustafsson *(57)* report 93 mg/kg of vanadium in residual-oil fuel and 1.7 mg/kg in marine distillates (a less common type of marine fuel). For characteristic operating conditions (slow- or medium-speed engine operation) using residual-oil fuel they report $PM_{2.5}$ emission factors of 11 – 27 g/kg. These estimates suggest that multiplying vanadium concentrations by 120 (11/0.093) – 290 (27/0.093) would yield the $PM_{2.5}$ concentration attributable to ship emissions. Analysis of Vancouver-specific ship operating conditions (as described previously) and ship fuel would provide a more accurate estimate. Several

other articles report useful information about emissions factors from ships *(58-61)* including some in Washington State *(62)*.

A more sophisticated source-specific approach was recently described by Ryan et al. *(45)*, and was discussed briefly in Section Chapter 1224220. In this application, researchers used a multivariate receptor model to estimate the percent contribution of traffic to elemental carbon (EC) measurements made at 24 sites. Next they multiplied the measured EC concentrations against the traffic-related fraction to estimate the fraction of elemental carbon attributable to traffic. Then LUR was used to predict the EC concentrations attributable to traffic as a more specific indicator of diesel combustion. In theory, similar approaches could be applied to determine spatial patterns of other sources (e.g. marine emissions) that can be resolved by source apportionment methods.

Table 9. Emissions sources in the Burrard Inlet and their potential pollutant tracers

Source	**Tracer**	**Notes**
Marine sources	Vanadium, nickel, sulfate particles in specific size range	Refer to Section Chapter 1224220.
Diesel engines	Elemental carbon, lubricating oils	Diesel engines include on-road (e.g. trucks) and off-road (e.g. construction equipment, electrical generators, marine vehicles) sources. Elemental carbon is associated with diesel emission, but not specific to this source *(63)*. Lubricating oils (organic molecules) are source-specific, but expensive to measure *(64)*.
Fossil fuel combustion	Elemental carbon, NO_X, CO	All three species listed are emitted by any internal combustion engine, but emission concentrations vary among fuels *(65)*.
Oil refinery	Lanthanum/samarium ratio	Both elements are found in catalysts for cracking crude oil. Their ratio has been strongly associated with particulate matter from this point source *(66, 67)*.
Cement plant	Calcium	Several elements (especially heavy metals) are found in cement dust, but calcium (in the form of CaO) concentrations are relatively consistent from region to region *(68)*.
Sewage treatment plant	Aldehydes, keytones	Incomplete combustion in sewage incinerators can lead to the formation of odorous (and difficult to measure) partially-oxidized hydrocarbons *(69, 70)*.
Wood smoke	Levoglucosan	A well-characterized and widely-used tracer for emissions from biomass burning *(71, 72)*.

References

(1) Roosli M, Braun-Fahrlander C, Kanzli N, Oglesby L, Theis G, Camenzind M, Mathys P and Staehelin J. Spatial Variability,of Different Fractions of Particulate Matter within an Urban Environment and between Urban and Rural Sites. *Journal of the Air & Waste Management Association (1995)* **2000,** 50, (7), 1115-1124.

(2) Janssen NAH, Van Mansom DFM, Van Der Jagt K, Harssema H and Hoek G. Mass concentration and elemental composition of airborne particulate matter at street and background locations. *Atmospheric Environment* **1997,** 31, (8), 1185-1193.

(3) Monn C, Fuchs A, Hogger D, Junker M, Kogelschatz D, Roth N and Wanner H-U. Particulate matter less than 10 microns (PM10) and fine particles less than 2.5 microns (PM2.5): relationships between indoor, outdoor and personal concentrations. *Science of The Total Environment* **1997,** 208, (1-2), 15-21.

(4) Roemer WH and van Wijnen JH. Differences among Black Smoke, PM10 and PM1 Levels at Urban Measurement Sites. *Environmental Health Perspectives* **2001,** 109, (2), 151.

(5) Cyrys J, Heinrich J, Brauer M and Wichmann H. Spatial Variability of Acidic Aerosols, Sulfate and PM1 in Erfurt, Eastern Germany. *Journal of Exposure Analysis and Environmental Epidemiology* **1998,** 8, (4), 447-464.

(6) van Vliet P, Knape M, de Hartog J, Janssen N, Harssema H and Brunekreef B. Motor Vehicle Exhaust and Chronic Respiratory Symptoms in Children Living near Freeways. *Environmental Research* **1997,** 74, (2), 122-132.

(7) Weiland S, Mundt K, Ruckmann A and Keil U. Self-reported wheezing and allergic rhinitis in children and traffic density on street of residence. *Annals of Epidemiology* **1994,** 4, (3), 243-247.

(8) Wjst M, Reitmeir P, Dold S, Wulff A, Nicolai T, von Loeffelholz-Colberg EF and von Mutius E. Road traffic and adverse effects on respiratory health in children. *BMJ: British Medical Journal* **1993,** 307, (6904), 596.

(9) Oosterlee A, Drijver M, Lebret E and Brunekreef B. Chronic respiratory symptoms in children and adults living along streets with high traffic density. *Occupational and Environmental Medicine* **1996,** 53, 241-247.

(10) Nitta H, Sato T, Nakai S, Maeda K, Aoki S and Ono M. Respiratory Health Associated with Exposure to Automobile Exhaust. I. Results of Cross-Sectional Studies in 1979, 1982, and 1983. *Archives of Environmental Health* **1993,** 48, (1).

(11) Duhme H, Weiland SK, Keil U, Kraemer B, Schmid M, Stender M and Chambless L. The Association between Self-Reported Symptoms of Asthma and Allergic Rhinitis and Self-Reported Traffic Density on Street of Residence in Adolescents. *Epidemiology* **1996,** 7, (6), 578-582.

(12) Ciccone G, Forastiere F, Agabiti N, Biggeri A, Bisanti L, Chellini E, Corbo G, Dell'Orco V, Dalmasso P, Volante T, Galassi C, Piffer S, Renzoni E, Rusconi F, Sestini P and Viegi G. Road traffic and adverse respiratory effects in children. SIDRIA Collaborative Group. *Occupational and Environmental Medicine* **1998,** 55, (11), 771-778.

(13) Roosli M, Theis G, Kunzli N, Staehelin J, Mathys P, Oglesby L, Camenzind M and Braun-Fahrlander C. Temporal and spatial variation of the chemical composition of PM10 at urban and rural sites in the Basel area, Switzerland. *Atmospheric Environment* **2001,** 35, (21), 3701-3713.

(14) Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J and Giovis C. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Environ. Epidemiol.* **2005,** 15, (2), 185-204.

(15) Briggs DJ, Collins S, Elliot P, Fischer P, Kingham S, Lebret E, Pryl K, van Reeuwijk H, Smallbone K and van der Veen A. Mapping urban air pollution using GIS: a regression-based approach. *Int. J. Geogr. Inf. Sci.* **1997,** 11, (7), 699-718.

(16) Gilbert NL, Goldberg MS, Beckerman B, Brook JR and Jerrett M. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *J. Air Waste Manage. Assoc.* **2005,** 55, 1059–1063.

(17) Gonzales M, Qualls C, Hudgens E and Neas L. Characterization of a spatial gradient of nitrogen dioxide across a United States-Mexico border city during winter. *Sci. Total Environ.* **2005,** 337, (1-3), 163-173.

(18) Hoek G, Meliefste K, Brauer M, van Vliet P, Brunekreef B and Fischer P. *Risk assessment of exposure to traffic-related air pollution for the development of inhalant allergy, asthma and other chronic respiratory conditions in children (TRAPCA). Final Report*; IRAS, University: Utrecht, 2001.

(19) Ross Z, English PB, Scalf R, Gunier R, Smorodinsky S, Wall S and Jerrett M. Nitrogen dioxide prediction in Southern California using land use regression modeling: Potential for environmental health analyses. *J. Expo. Anal. Environ. Epidemiol.* **2005**.

(20) Hochadel M, Heinrich J, Gehring U, Morgenstern V, Kuhlbusch T, Link E, Wichmann H-E and Kramer U. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmos. Environ.* **2006,** 40, (3), 542-553.

(21) Henderson S and Brauer M. *Measurement and modeling of traffic-related air pollution in the British Columbia Lower Mainland for use in health risk assessment and epidemiogical analysis*; Report submitted to Health Canada: Vancouver, BC, 2005; p 40 pages.

(22) Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer PH, Wijga A, Koopman LP, Neijens HJ, Gerritsen J, Kerkhof M, Heinrich J, Bellander T and Brunekreef B. Air Pollution from Traffic and the Development of Respiratory Infections and Asthmatic and Allergic Symptoms in Children. *Am. J. Respir. Crit. Care Med.* **2002,** 166, (8), 1092-1098.

(23) Hoek G, Meliefste K, Cyrys J, Lewne M, Bellander T, Brauer M, Fischer P, Gehring U, Heinrich J, van Vliet P and Brunekreef B. Spatial variability of fine particle concentrations in three European areas. *Atmos. Environ.* **2002,** 36, (25), 4077-4088.

(24) Briggs DJ, de Hoogh C, Gulliver J, Wills J, Elliott P, Kingham S and Smallbone K. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *Sci. Total Environ.* **2000,** 253, (1-3), 151-167.

(25) Lebret E, Briggs D, van Reeuwijk H, Fischer P, Smallbone K, Harssema H, Kriz B, Gorynski P and Elliott P. Small area variations in ambient NO2 concentrations in four European areas. *Atmospheric Environment* **2000,** 34, (2), 177-185.

(26) Larson T, Su J, Baribeau A-M, Buzzelli M, Setton E and Brauer M. A Spatial Model of Urban Winter Woodsmoke Concentrations. *Environmental Science and Technology* **Submitted June 12, 2006.**

(27) Sahsuvaroglu T, Arain A, Beckerman B, Kanaroglou P, Brook JR, Finkelstein N, Finkelstein MM, Gilbert NL, Newbold B and Jerrett M. A land-use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Canada. *J. Air Waste Manage. Assoc.* **2006,** 56, (8), 1059-1069.

(28) Cyrys J, Hochadel M, Gehring U, Hoek G, Diegmann V, Brunekreef B and Heinrich J. GIS-Based Estimation of Exposure to Particulate Matter and NO2 in an Urban Area: Stochastic versus Dispersion Modeling. *Environmental Health Perspectives* **2005,** 113, (8), 987-992.

(29) Briggs D, de Hoogh K and Gulliver J. *Matching the metric to need: modelling exposures to traffic-related air pollution for policy support.* In Presentation at NERAM V. Vancouver, BC; November 2006.

(30) Kanaroglou PS, Jerrett M, Morrison J, Beckerman B, Arain MA, Gilbert NL and Brook JR. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmos. Environ.* **2005,** 39, (13), 2399-2409.

(31) Noullett M, Jackson P and Brauer M. Winter measurements of children's personal exposure and ambient fine particle mass, sulphate and light-absorbing components in a northern community. *Atmospheric Environment* **2006,** 40, 1971-1990.

(32) Rich K. Air pollution and patients with implanted cardiac defibrillators: An epidemiological analysis and assessment of exposure. MSc, University of British Columbia, Vancouver, 2003.

(33) Cyrys J, Heinrich J, Hoek G, Meliefste K, Lewné M, Gehring U, Bellander T, Fischer T, van Vliet P, Brauer M, Wichmann E and Brunekreef B. Comparison between different traffic-related particle indicators: Elemental carbon (EC), PM2.5 mass, and absorbance. *Journal of Exposure Analysis and Environmental Epidemiology* **2003,** 13, (2), 134-143.

(34) Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer P, Gehring U, Heinrich J, Cyrys J, Bellander T, Lewne M and Brunekreef B. Prediction of long term average particulate air pollution concentrations by traffic indicators for epidemiological studies. *Epidemiology.* **2003,** 14, 228-239.

(35) Larson T, Garcia N, Covert D and Brauer M. *Mobile Monitoring of Particulate Black Carbon Concentrations in an Urban Area.* In ISEE/ISEA International Conference on Environmental Epidemiology and Exposure. Paris, France; 2006.

(36) Larson TV and Koenig JQ. Wood Smoke: Emissions and Noncancer Respiratory Effects. *Annual Review of Public Health* **1994,** 15, (1), 133-156.

(37) Rudra C, Koenig J and Williams M. *Estimation of Monthly Average Ambient Air Pollutant Concentrations Using Geographic Information Systems.* In Poster presentation at ISEE/ISEA. Paris, France; 2006.

(38) Palmes E and Gunnison A. Personal monitoring device for gaseous contaminants. *Am Ind Hyg Assoc J* **1973,** 34, (2), 78-81.

(39) NO, NO2, NOx and SO2 sampling protocol using the Ogawa sampler. In 4th Edition ed.; 1998.

(40) Westerdahl D, Fruin S, Sax T, Fine P and Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment* **2005,** 39, 3597-3610.

(41) Ainslie B and Steyn DG. *Source Area Model for estimating population exposure.* In PNWIS 2005, Air and Waste Management Association. "International Perspectives on Environmental Management". Blaine, Washington; November 8-11, 2005. http://www.cher.ubc.ca/UBCBAQS/Presentations/PDF%20versions/steyn_baqs_nov9.pdf.

(42) Buzzelli M, Su J, Ainslie B, Steyn DG, Brauer M and T L. *A GIS Spatio-Temporal Model of Ambient Air Pollution Exposure.* In ISEE/ISEA International Conference on Environmental Epidemiology and Exposure. Paris, France; 2006.

(43) Brauer M and Henderson SB. *Diesel exhaust particles and related air pollution from traffic sources in the Lower Mainland*; Report submitted to Health Canada: Vancouver, BC, 2003; p 23 pages.

(44) Jerrett M, Arain M, Kanaroglou P, Beckerman B, Crouse D, Gilbert N, Brook J, Finkelstein N and Finkelstein M. Modelling the intra-urban variability of ambient traffic pollution in Toronto, Canada. *J. Toxicol. Environ. Health A* **in press**.

(45) Ryan PH, LeMasters GK, Biswas P, Levin L, Hu S, Lindsey M, Bernstein DI, Lockey J, Villareal M, Khurana-Hershey GK and Grinshpun SA. A Comparison of Proximity and Land Use Regression Traffic Exposure Models and Wheezing in Infants. *Environ Health Perspect.* **in press,** Check DOI# 10.1289/ehp.9480 at http://dx.doi.org/.

(46) Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G and Jarup L. Using Geographic Information Systems To Assess Individual Historical Exposure to Air Pollution from Traffic and House Heating in Stockholm. *Environmental Health Perspectives* **2001,** 109, (6), 633.

(47) Rosenlund M, Berglind N, Pershagen G, Hallqvist J, Jonson T and Bellander T. Long-term exposure to urban air pollution and myocardial infarction. *Epidemiology* **2006,** 17, (4), 383-390.

(48) Poplawski K. *Evaluating the Transferability of a Land-Use Regression Model.* In Pacific Northwest International Section of the Air & Waste Management Association. November 8-10. Victoria, BC; 2006.

(49) Yu L, Yue P and Teng H. Comparative study of EMME/2 and QRS II for modeling a small community. *Transport. Res. Rec.* **2003,** 1858, 103-111.

(50) Setton EM, Hystad PW and Keller PC. *Road Classification Schemes – Good Indicators of Traffic Volume?*; Spatial Sciences Laboratories, Department of Geography, University of Victoria: 2005; p 11.

(51) British Columbia Chamber of Shipping Fact Sheet on the BC Marine Vessel Emission Inventory. http://www.chamber-of-shipping.com/index/cms-filesystem-action?file=/EIFactsheet.pdf

(52) Bryant R. *Telephone conversation about the BC Marine Vessel Emission Inventory*. With Henderson, S. Vancouver, BC. November 17th.

(53) Marshall J. *Memo on point-source exposure surface for the Border Air Quality Study.* With Brauer, M. Vancouver, BC.

(54) Hystad PW, Setton E and Keller PC. *Uses of Spatial Property Assessment Data for Air Pollution Exposure Assessment and Epidemiological Analyses*; Spatial Sciences Laboratories, Department of Geography, University of Victoria: 2005; p 9.

(55) Isakson J, Persson TA and Lindgren ES. Identification and assessment of shipnext term emissions and their effects in the harbour of Göteborg, Sweden. *Atmospheric Environment* **2001,** 35, (21), 3659-3666.

(56) Huang L, Brook JR, Zhang W, Li SM, Graham L, Ernst D and Chivulescu A. Identification and characterization of inland ship plumes over Vancouver, BC. *Atmos Environ* **2006,** 40, (15), 2767-2782.

(57) Cooper DA and Gustafsson T. *Methodology for calculating emissions from ships: Two Emission factors for 2004 reporting*; Swedish Methodology for Environmental Data: Stockholm, Sweden, 2004.

(58) Sinha P, Hobbs PV, Yokelson RJ, Christian TJ, Kirchstetter TW and Bruintjes R. HCB, PCB, PCDD and PCDF emissions from ships. *Atmospheric Environment* **2005,** 39, (27), 4901-4912.

(59) Cooper DA. Exhaust emissions from ships at berth. *Atmospheric Environment* **2003,** 37, (27), 3817-3830.

(60) Cooper DA. HCB, PCB, PCDD and PCDF emissions from ships. *Atmospheric Environment* **2005,** 39, (27), 4901-4912.

(61) Endresen O, Bakke J, Sorgard E, Flatlandsmo Berglen P and Holmvang P. Improved modelling of ship SO2 emissions -- a fuel-based approach. *Atmospheric Environment* **2005,** 39, (20), 3621-3628.

(62) Corbett JJ. Emissions from Ships in the Northwestern United States. *Environmental Science and Technology* **2002,** 36, 1299-1306.

(63) Schauer JJ. Evaluation of elemental carbon as a marker for diesel particulate matter. *Journal of Exposure Analysis & Environmental Epidemiology* **2003,** 13, (6), 443-453.

(64) Shah SD, Cocker DR, Miller JW and Norbeck JM. Emission Rates of Particulate Matter and Elemental and Organic Carbon from In-Use Diesel Engines. *Environmental Science and Technology* **2004,** 38, 2544-2550.

(65) Westerholm R and Egeback KE. Exhaust Emissions from Light- and Heavy-Duty Vehicles: Chemical Composition, Impact of Exhaust after Treatment, and Fuel Parameters. *Environ Health Perspect.* **1994,** 102, (Supp. 4), 12-23.

(66) Kulkarni P, Chellam S and Fraser M, P. Lanthanum and lanthanides in atmospheric fine particles and their apportionment to refinery and petrochemical operations in Houston, TX. *Atmospheric Environment* **2006,** 40, (3), 508-520.

(67) Kitto ME, Anderson EL, Gordon GE and Olmez I. Rare earth distributions in catalysts and airborne particles. *Environmental Science and Technology* **1992,** 26, (7), 1368-1375.

(68) Adejumo JA, Obioh IB, Ogunsola OJ, Akeredolu FA, Olaniyi HB, Asubiojo OI, Oluwole AF, Akanle OA and Spyrou NM. The Atmospheric Deposition of Major, Minor and Trace Elements Within and Around Three Cement Factories. *Journal of Radioanalytical and Nuclear Chemistry* **1994,** 179, (2), 195-204.

(69) Dewling RT. Sewage-Sludge Incineration Raises Air Pollution Concerns. *Water and Sewage Works* **1980,** 127, (10), 26-29.

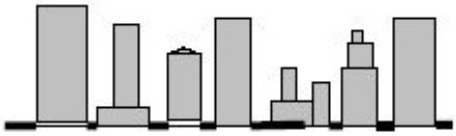
(70) *Locating and Estimating Air Emissions from Sources of Methyl Ethyl Keytone*; US Environmental Protection Agency, Office of Air Quality Planning and Standars: Research Triangle Park, NC, 1994; p 154.

(71) Simoneit BRT, Schauer JJ, Nolte CG, Oros DR, Elias VO, Fraser MP, Rogge WF and Cass GR. Levoglucosan, a tracer for cellulose in biomass burning and atmospheric particles. *Atmospheric Environment* **1999,** 33, (2), 173-182.

(72) Fraser MP and Lakshmanan K. Using Levoglucosan as a Molecular Marker for the Long-Range Transport of Biomass Combustion Aerosols. *Environmental Science & Technology* **2000,** 34, (21), 4560-4564.

Appendix 1: Urban Climate Zone Classification

This classification system (Oke 2004, unpublished) has been used to categorize LUR sites in the GVRD and Seattle.

Urban Climate Zone, UCZ[1]	*Image*	*Rough-ness class*[2]	*Aspect ratio*[3]	*% Built (imperm-eable)*[4]
1. *Intensely developed urban with detached close-set high-rise buildings with cladding, e.g. downtown towers*		8	> 2	> 90
2. *Intensely developed high density urban with 2 – 5 storey, attached or very close-set buildings often of brick or stone, e.g. old city core*		7	1.0 – 2.5	> 85
3. *Highly developed, medium density urban with row or detached but close-set houses, stores & apartments e.g. urban housing*		7	0.5 – 1.5	70 - 85
4. *Highly developed, low or medium density urban with large low buildings & paved parking, e.g. shopping mall, warehouses*		5	0.05 – 0.2	70 - 95
5. *Medium development, low density suburban with 1 or 2 storey houses, e.g. suburban housing*		6	0.2 – 0.6, up to >1 with trees	35 - 65
6. *Mixed use with large buildings in open landscape, e.g. institutions such as hospital, university, airport*		5	0.1 – 0.5, depends on trees	< 40
7. *Semi-rural development, scattered houses in natural or agricultural area, e.g. farms, estates*		4	> 0.05, depends on trees	< 10

Key to image symbols: buildings; vegetation; impervious ground; pervious ground

[1] A simplified set of classes that includes aspects of the schemes of Auer (1978) and Ellefsen (1990/91) plus physical measures relating to wind, thermal and moisture controls (columns at right). Approximate correspondence between UCZ and Ellefsen's urban terrain zones is: 1(Dc1, Dc8), 2 (A1-A4, Dc2), 3 (A5, Dc3-5, Do2), 4 (Do1, Do4, Do5), 5 (Do3), 6 (Do6), 7 (none).

[2] Effective terrain roughness according to the Davenport classification (Davenport *et al.*, 2000); see Table 2.

[3] Aspect ratio = z_H/W is average height of the main roughness elements (buildings, trees) divided by their average spacing, in the city centre this is the street canyon height/width. This measure is known to be related to flow regime types (Oke 1987) and thermal controls (solar shading and longwave screening) (Oke, 1981). Tall trees increase this measure significantly.

[4] Average proportion of ground plan covered by built features (buildings, roads, paved and other impervious areas) the rest of the area is occupied by pervious cover (green space, water and other natural surfaces). Permeability affects the moisture status of the ground and hence humidification and evaporative cooling potential.

Appendix 2: Technical Memo on Point Source Pollution

From: Julian Marshall
To: Mike Brauer
Date: October 4, 2006
Re: Point-source exposure surface for BAQS

Summary

This memo documents the approach for generating the point-source exposure surface for the BAQS epidemiology study.

Background and introduction

The Border Air Quality Study (BAQS) is a large epidemiology study of air pollution in the Georgia Air Basin, a region that includes Vancouver, Victoria, and Seattle. Other investigators have generated exposure metrics for traffic emissions, such as Sarah Henderson's land-use regression (LUR) model. Here, I describe a metric of exposure for point-source (i.e., industrial) emissions in Vancouver.

The exposure metric described here can be calculated for any arbitrary location within the study region. At present, it has only been calculated at the locations of PC centroids in Vancouver.

The metric is a proxy for at-home exposure to industrial point-source emissions. It is only a proxy. No actual concentrations are estimated. The approach employed here is modeled in part after the work of Yu *et al*. (2006), "Residential Exposure to Petrochemicals and the Risk of Leukemia: Using Geographic Information System Tools to Estimate Individual-Level Residential Exposure" (Am J Epidemiol 164, 200–207). One difference, however, is that Yu and colleagues employed wedges (shape: pie-piece) that explicitly account for wind direction. Here, uniform circles are used, thus ignoring wind direction. The study location for the Yu *et al.* article (Kaohsiung, Taiwan) includes only four point sources, all major petrochemical complexes. In contrast, Vancouver includes hundreds of industrial point sources.

Approach

The point-source exposure metric is a proximity-weighted summation of relative emissions within a given radius of each PC centroid. The metric incorporates three main inputs:

1. Point-source emissions and locations

Point-source emissions and locations were supplied by Thomas Nipen (tnnipen@interchange.ubc.ca), a UBC graduate student working with Professor Roland Stull. Professor Stull uses these data as input to the CMAQ air dispersion model. Emission locations are throughout Southwestern Canada and Northwestern United States. Emissions (tons per day) are given for the following 8 pollutants: CO, NOX, VOC, NH3, SO2, PM10, PM2_5, and PMC (coarse PM; i.e., PM10 minus PM2_5).

Emission inventories can be notoriously inaccurate, in part because (1) they are labor-intensive to generate, and (2) important information is often difficult or impossible to obtain and/or verify. Prof. Stull's group has attempted to fix obvious errors in the official inventory (and thus there may be minor differences between the dataset employed here and the official inventory). Nevertheless, the data are not perfect. Here is an example of a (minor) inconsistency I identified in the data: for 38 out of 9,458

point sources[9], reported emissions are greater for PM2.5 than for PM10 (which is not possible). In all but a few cases, the discrepancy is modest (30% or less). For this work, I have employed the dataset as it was provided to me, without modification. To evaluate or confirm the data would be a significant task, and is beyond the scope of this investigation.

2. PC locations

Latitudes and longitudes of PC centroids were supplied to me by Cornel Lencar (a SOEH/BAQS staff member; clencar@interchange.ubc.ca). The dataset contains 93,716 PCs, with lat/lon values for 91,348 PCs. (The remaining 2,368 PCs are no longer in use or are otherwise missing location information.) In addition, 34 of the 91,348 PCs are not considered here because they are located outside the Georgia Air Basin. Thus, results below are reported for the 91,314 PC with lat/lon values inside the GAB.

3. Cut-off threshold distance

If the distance between a PC and an emission source is greater than a certain value, that emission source is ignored when evaluating that PC. Given the nature of air dispersion, the variability among sources in stack height and plume rise, and the complex topology and meteorology of Vancouver, it is not possible to identify a single value as the "correct" cut-off threshold distance. For elevated emission sources (unlike for ground-level sources), maximum concentrations are typically some distance down-wind of the emission location, owing to the presence of a stack and plume rise.

Two cut-off threshold distance values are separately employed here: 10km and 40km. These values are intended to represent the approximate distance needed for a point-source plume to fully mix throughout the atmospheric mixing height. As mentioned below, the point source surfaces separately generated by these two cut-off threshold values are highly correlated with each other (r=0.91).

These two values (10km and 40km) were derived from the so-called Pasquill-Gifford curves, which are available on-line and in standard air pollution texts (e.g., *Atmospheric Chemistry and Physics, 2nd ed.*, by Seinfeld and Pandis; John Wiley, 2006; see p. 865). This approach offers an order-of-magnitude estimate of the "impact zone" for a point source, though the true size of an impact zone will vary widely over time and among sources, based on parameters such as emissions, stack height, exit velocity, meteorology, and topography. A small point source may have a localized impact, similar to a roadway (i.e., less than 1 km), while a large point source with a tall stack and significant plume rise may impact 100's of kilometers (and also may have little impact immediately next to the stack itself). The latter type of large point source (i.e., high stack and significant plume rise) may have little impact on an urban area if a stagnant layer aloft prevents the elevated plume from mixing down to ground level. In conclusion, there is not a single "correct" value for the cut-off threshold; the two values employed here (10km and 40km) are approximate and span a reasonable range for this parameter.

The point-source exposure metric is calculated using the following formula:

$$W_i = \sum_{j;\{d_{ij}>x\}} E_j \Big/ d_{ij} \; .$$

[9] Here are the Source IDs for these 38 sources: 4669, 5504, 7271, 8080, 8081, 8082, 8085, 8086, 8087, 8088, 8089, 8090, 8091, 8928, 8949, 9020, 9182, 9235, 10457, 10489, 10491, 10565, 10566, 10635, 10701, 10786, 10808, 10831, 11017, 11423, 11443, 11479, 11841, 11971, 11994, 12064, 12281, 12362.

Here, W_i is the point-source exposure metric for postal code i, E_j is the relative emissions for point source j, and d_{ij} is the distance between postal code i and point source j. The summation is carried out for all point sources within distance x of the PC. This approach mirrors the work by Yu *et al*. (2006).

Relative emissions for point source j (E_j) are calculated as follows. First, emissions for each point source are converted from a raw emission rate (tons per day) into the percentile of that source among all emitting point sources. This step is repeated for each of four pollutants (PM2.5, SOx, NOx, and VOCs). For example, a point source that does not emit SOx is assigned a percentile of zero; a source whose SOx emissions are at the 85th percentile (i.e., 85% of the SOx-emitting point sources have an emission rate that is less than this source, and 15% of the SOx-emitting sources have greater emissions than this source) is assigned a SOx value of 0.85. Next, the percentile scores for the four pollutants are summed to yield the relative emissions for a specific point source. The largest relative emission rate is 3.96, which is for a specific point source that is in the 99th emission percentile for all four pollutants. The lowest relative emission rate is zero, representing sources with no emissions of the four pollutants.

Further details regarding the calculations, output, and specific files employed are in Appendix 2.1. Appendix 2.2 provides the full MatLab code employed in this investigation.

Results

Each postal code has an average of 173 point sources within 10km and 753 point sources within 40km. Mean (st dev) values for the point-source exposure metric are 21.6 (21.8) for x=10 km, and 41.5 (27.6) for x=40km. Geometric means (GSD) are 12.7 (3.5) for x=10 km, and 30.3 (2.5) for x=40 km. The correlation between the two exposure metrics is very high (r = 0.91), suggesting that the results from using either surface in an epidemiology study should be similar.

Appendix 2.1: Description of the Calculations

Input: point-source emissions data

Emissions data were supplied to me by Thomas Nipen in two files: Emission_data.rpt, and Location_data.csv. Here is the relevant note from Thomas Nipen (email date: Thursday 3/16/2006 6:54 PM) explaining the content of these files:

```
Julian,

I have prepared the emission inventory.

There are two files attached (both ASCII files):
      - Emission_data contains the emissions of each pollutant for each
SourceID. It also contains a facility name to each source. It is a
semicolon separated values file.
      - Location_data contains the lat/long of each SourceID. It is a
comma seperated values file.

The sourceID is only an ID that lets you link the sources in the
Emission_data file to the same source in the Location_data file. The
IDs do not have any significance otherwise. Note that both files
contain the Stack height, stack diameter, stack temperature, stack
exit velocity, except Emission_data has the data to two decimal
digits, where as Location_data has them to higher accuracy. The
columns should be self explanatory.
      - Latitude / longitude values are in degrees,
```

```
      - Stack height in meters
      - Stack diameter in meters
      - Stack temperature in Kelvins
      - Stack exit velocity in m/s.
```

In the Emission_data file, the SourceID values range from 1 to 12395 but there are only 9458 rows of data (i.e., several SourceIDs are unused). As mentioned above, the point sources include a very large area throughout British Columbia, Washington State, and Idaho. Based on the lat/lon location of the point sources, I excluded point-sources located far from Vancouver. (A list of sources included are not – a 1 for include; a 0 for not – are in the STATA file location-include_or_not.dta). The remaining 2630 point-sources (i.e., the sources that are in or near Vancouver), along with the lat/lon location and the relative emissions, are given in the file pt_src_of_interest.xls. The data in this file is used as input to the MatLab program described below.

Calculation: relative emissions

Relative emissions are calculated in the spreadsheet Emission_data2.xls. As discussed above, the relative emission calculation involves four pollutants. As a sensitivity analysis, I compared this 4-pollutant composite score against the NOx-only score and the PM2_5-only score. The correlation between the composite and NOx-only scores is good (r=0.76); correlation between the composite and PM2_5-only scores is moderate (r=0.48); correlation between the NOx-only and PM2_5-only scores is nearly zero (r=0.02). The Emission_data2.xls spreadsheet (specifically, the bottom of 'calculate relative emission' sheet) contains scatter plots of the composite versus NOx-only metrics, and of the composite versus PM2_5-only metrics. Slopes of the best-fit lines are 0.31 for the first plot (composite versus NOx-only) and 0.20 for the second plot (composite versus PM2_5-only). Of note here is that (1) the slopes are positive, indicating the same directional trend among the three metrics (i.e., point-sources ranked high in one metric tend to rank highly using the other two metrics also), and (2) the slope magnitude is consistent with expectations and suggests that on average each of the four pollutants contributes roughly equally to the composite score. (Since there are four pollutants, and the composite score is the sum of the scores for each pollutant, then if all pollutants contributed the exact same amount as each other, then the slopes would be roughly 25%, or 0.25. The actual slopes are 0.31 and 0.20, which are close in value to the exactly-even value of 0.25.) Taken together, these two pieces of evidence – the slope of the best-fit line when plotted against the composite score is greater for NOx than for PM2_5; and, the correlation with the composite score is better for NOx than for PM2_5 – indicate that the composite score is largely a marker for combustion-related industrial emissions such as NOx, but it also includes moderate influence from the three other pollutants as well.

Calculation: point-source exposure metric

The exposure metric is calculated using MatLab (Version 5.3.1.29215a (R11.1), September 28,1999). Briefly, the three sets of input data are (1) lat/lon for each postal code, (2) lat/lon for each point-source (abbreviated as "PC" and "SRC", respectively, in the MatLab code), and (3) a weighting associated with each point-source (i.e., the relative emissions; abbreviated as "weight" in the code). The main output is the point-source exposure metric for each PC (abbreviated as "score" in the code). The code consists of a pair of nested loops: the outer loop includes all PCs; the inner loop covers all point-sources. The portion of the exposure metric score that is attributable to each pairing of a point source and a PC (abbreviated as "wt" in the code) is initially set to zero, and only changed (employing the equation above) if the distance between that PC/SRC pair is less than 40km (or, 10km).

Before running the MatLab code, lat/lon values for each of the PCs and the point-sources are loaded as 1-dimentional arrays with the names lat_pc, lon_pc, lat_src, and lon_src. Calculation of the distance between two points requires converting the lat/lon values from degrees to radians.

Here is the core of the calculation, in MatLab code.

```
score = zeros(length(lat_pc),1);
for pcs = 1:length(lat_pc);        %outside loop;
   wt = zeros(length(lat_src),1);
   for srcs = 1:length(lat_src);   %inside loop;
      f1 = lat_pc(pcs) * (pi/180);
      l1 = lon_pc(pcs) * (pi/180);
      f2 = lat_src(srcs) * (pi/180);
      l2 = lon_src(srcs) * (pi/180);
      d = 6372*acos(sin(f1)*sin(f2)+cos(f1)*cos(f2)*cos(l2-l1));
      if d<40
         wt(srcs)=weights(srcs)/d;
      end;
   end;
   score(pcs) = sum(wt);
end;
```

The full MatLab code as actually run is in Appendix 2.2. The full code includes the following additional attributes. It repeats the same calculations listed above twice (once with a cut-off of 40 km, once with 10 km). It also includes output statements along the way (e.g., "whos" provides a list of variables and their array size.) It checks to see if there are errors, and counts the number of PCs with an exposure metric value of zero. It provides a count of the number of point sources within the cut-off distance of each PC. Finally, the data are output to files.

(As mentioned below, the only calculation "error" from running the code is that one of the PCs has the same lat/lon coordinates as one of the point sources. In this case, the distance between the two points is zero. When the MatLab code tries to divide by zero, it ends up generating the "NaN" error code, meaning "Not a Number", for the exposure metric for that PC. The code in Appendix 2.2 converts this "NaN" value to –1.)

Output

Full results are in the Stata file ptsrc_surface.dta. Each row represents a single PC. The headers for the columns, and what they stand for are as follows:

PCode	The postal code (e.g., "V6T1Z3")
Lat	Latitude of the PC centroid
Lon	Longitude of the PC centroid
Ptsrc10	Point-source exposure surface, using a radius of 10km
Ptsrc40	Point-source exposure surface, using a radius of 40km
Ptsrc10_count	The number of point sources within 10km
Ptsrc40_count	The number of point sources within 40km

The ptsrc_surface.dta file combines the following information. The point-source exposure surface is given in the files pt_src_weights_results_40km.txt and pt_src_weights_results_10km.txt. The counts data – i.e., the number of point-sources that are used in calculated the point-source exposure surface for each PC – is given in the files pt_src_weights_results-counts_10km.txt and pt_src_weights_results-counts_40km.txt. All four of these text files simply contain a list of numbers (i.e., a 1-dimentional array, with each value corresponding to a single PC). Values are in the same order as the input PCs.

For the three PCs listed in the following table, the calculated point source exposure metric is an outlier value.

Dataset Row	Postal Code	PC lat	PC lon	Calculated point-source weighting (x=10 km)	Calculated point-source weighting (x=40 km)
43145	V5M4J9	49.2658	123.0275	-1	-1
3086	V2S2E5	49.0478	122.2886	666.458	677.577
18488	V3M6H9	49.1759	-122.9315	2798.61	2826.25

These outlier values are postal codes that are extremely close to a point source. The first PC above (V5M4J9) is co-located with a point source. Thus, the separation distance is zero, and the distance-weighted relative emission, which includes dividing by the distance (see the equation above), is infinite. In the dataset, the "infinity" value (which generates a "NaN" error in MatLab) has been replaced with the value -1. For use in an epi study, it is encouraged that the scores for these three PCs (for both x=10km and x=40 km) be re-assigned to the value 400, which is approximately the maximum value other than these three outliers.

At present, the text files with the raw output (pt_src_weights_results_40km.txt and pt_src_weights_results_10km.txt) do NOT have the values changed for these three outliers, but the Stata summary output file (ptsrc_surface.dta) DOES have these three outlier values changed to 400, for both the 40km and the 10km point-source surfaces. If the user prefers to change them back, values are given in the above table.

The *results* section of the memo (above) mentions that the correlation between the two point-source exposure metrics (i.e., employing x=10km and x=40km) is very high (r = 0.91). With the three outlier values changed to 400, the correlation is reduced slightly, to r=0.90.

The point-source surface values (which, of course, incorporate both proximity and emissions) are reasonably well-correlated with straightforward counts of the number of point sources within the radius. Specifically, the correlation between the point-source surface and the number of emission sources within the radius is 0.83 for 10km and 0.77 for 40km. For 787 of the 91,314 PCs (i.e., ~1% of the PCs), there are zero point sources within 10km. None of the PCs have zero point sources within 40km.

Appendix 2.2: MatLab Code

```
score_40 = zeros(length(lat_pc),1);
score_10 = zeros(length(lat_pc),1);
count_40 = zeros(length(lat_pc),1);
count_10 = zeros(length(lat_pc),1);

for pcs = 1:length(lat_pc);
   wt_40 = zeros(length(lat_src),1);
   wt_10 = zeros(length(lat_src),1);
   for srcs = 1:length(lat_src);
```

```
      f1 = lat_pc(pcs) * (pi/180);
      l1 = lon_pc(pcs) * (pi/180);
      f2 = lat_src(srcs) * (pi/180);
      l2 = lon_src(srcs) * (pi/180);
      d = 6372*acos(sin(f1)*sin(f2)+cos(f1)*cos(f2)*cos(l2-l1));
      if d<40    wt_40(srcs) = weights(srcs)/d; count_40(pcs)=count_40(pcs)+1;
end;
      if d<10    wt_10(srcs) = weights(srcs)/d; count_10(pcs)=count_10(pcs)+1;
end;

   end;
   score_40(pcs) = sum(wt_40);
   score_10(pcs) = sum(wt_10);
end;

whos

sum(score_40)
sum(score_10)
sum(count_40)/length(lat_pc)
sum(count_10)/length(lat_pc)



junk_40=0;
count_junk_40 = 0;
count_zero_40 = 0;
for i = 1:length(score_40);
   if score_40(i) == 0 count_zero_40 = count_zero_40 + 1;
   elseif score_40(i)/score_40(i) == 1 junk_40 = junk_40;
   else
      score_40(i) = -1;
      count_junk_40 = count_junk_40 + 1;
   end;
end;


junk_10=0;
count_junk_10 = 0;
count_zero_10 = 0;
for i = 1:length(score_10);
   if score_10(i) == 0 count_zero_10 = count_zero_10 + 1;
   elseif score_10(i)/score_10(i) == 1 junk_10 = junk_10;
   else
      score_10(i) = -1;
      count_junk_10 = count_junk_10 + 1;
   end;
end;




fid = fopen('C:\Documents and Settings\Julian Marshall\My
Documents\research\UBC\point source
emissions\matlab\pt_src_weights_results_40km.txt','w');
fprintf(fid,'%10.3f\n',score_40);
fclose(fid);
```

```
fid = fopen('C:\Documents and Settings\Julian Marshall\My
Documents\research\UBC\point source emissions\matlab\pt_src_weights_results-
counts_40km.txt','w');
fprintf(fid,'%10.3f\n',count_40);
fclose(fid);

fid = fopen('C:\Documents and Settings\Julian Marshall\My
Documents\research\UBC\point source
emissions\matlab\pt_src_weights_results_10km.txt','w');
fprintf(fid,'%10.3f\n',score_10);
fclose(fid);

fid = fopen('C:\Documents and Settings\Julian Marshall\My
Documents\research\UBC\point source emissions\matlab\pt_src_weights_results-
counts_10km.txt','w');
fprintf(fid,'%10.3f\n',count_10);
fclose(fid);



whos
```

Appendix 3: Review of Alternatives to LUR

Land use regression is one approach among many for generating information useful to air quality management. Broadly, there are three other types of approaches that may be used to understand source-specific contributions to ambient air pollutant concentrations. *Emission inventories* provide estimates of the pollutant mass released into the environment from specific sources. *Dispersion models* simulate the fate and transport of pollutants based on meteorology and mechanistic understanding of chemistry and physics – these models start with emissions to predict concentrations. *Receptor methods* begin with concentration measures at specific locations, and work backwards to estimate source contributions based on chemical composition, particle sizes, ratios of specific compounds, etc.

For primary (i.e., directly emitted) pollutants, all three approaches offer useful information. For secondary pollutants, such as ozone or secondary PM, emission inventories by definition provide little information; to date, receptor methods have limited ability to distinguish among sources for secondary pollutants. This appendix focuses on primary pollutants, for which all three methods have a demonstrated utility.

Emission Inventories

Given that an inventory exists for the GVRD, it would be worthwhile to evaluate, extend, and improve the inventory for sources in the study region, for example, incorporating the marine traffic. As part of that process, one could spatially reference the inventory, and then generate maps of emissions within the study area. This type of map would provide a crude indication of where concentration hotspots are likely to be. As an example of what this map might look like, the California Air Resources Board has developed an interactive online mapping tool, whereby users can look at air pollution sources and emission levels. The tool, called CHAPIS (Community Health Air Pollution Information System) and is available online[1]. However, given that emission inventories tend have significant uncertainties and that emissions are not concentrations, field measurements would be required to evaluate the utility of an emissions map for estimating the spatial extent of concentrations.

Dispersion Models

A wide variety of dispersion models exist, with a range of abilities and costs. Three model types are described below: Eulerian (grid-cell), Gaussian plume, and LaGrangian (trajectory) models. In general, these models use emission rates and meteorological information (wind speed and direction) as inputs, and generate concentration estimates at specific locations as output.

Eulerian Grid Cell

These models are generally considered by air quality modelers to be the most sophisticated and most labor-intensive approach available. Such models are computationally intensive, and offer several advantages such as the ability to simulate ozone chemistry and the ability to simulate "what if" scenarios, such as estimating the impact of reducing emissions from a specific source. A major disadvantage of these models is the size grid cells, which are generally a few kilometers on each side. Concentrations are modeled as uniform within each grid cell; within-grid concentration variability is generally not quantified (with some exceptions: for major sources, a plume-in-grid approach is possible). Thus, the model is not well-designed to identify concentration hot-spots. A grid cell model has been set up for the Vancouver region by Professor Roland Stull at University of British Columbia,

[1] www.arb.ca.gov/ch/chapis1/chapis1.htm.

using 4 km × 4 km grid cells. By running this model with typical and then modified emissions, it would be possible to estimate the impact on grid-cell-average concentrations of specific emission reductions.

Gaussian

These are widely used for regulatory applications, simulate concentrations downwind of a specific emission source. The theory behind plume models – that meteorology and air dispersion is uniformly at steady state during each one-hour time step – is not compelling, but it offers a straightforward modeling approach that is less computationally intensive than Eulerian models. A major advantage of these models is that they can predict concentrations at any arbitrary distance from a source, and thus are well-designed to estimate hot-spot concentrations associated with a specific source. A major disadvantage of these models – that they cannot simulate the complex chemistry involved in ozone production – is unlike to be important for the cases considered here (Burrard Inlet Study).

LaGrangian

These models, which include so-called puff models (e.g., CalPuff), allow parcels of air to meander with prevailing wind conditions. This approach offers a more theoretically compelling model formulation than plume models, while still offering the advantage of predicting concentration hot-spots. After decades of relying on plume models, the US EPA recently switched to puff models as the required approach for many regulatory situations.

Receptor Methods

Data Analysis

The following methods are often performed prior to implementing more formal receptor modeling, with the goals of generating hypotheses and focusing measurement efforts. These methods can often be completed with existing monitoring data.

(A) Time series plots of hourly or shorter-duration measurements

Diurnal profiles during weekend and weekdays may show systematic peaks that can potentially be attributed to major sources such as traffic rush-hour peaks and combustion of wood or fossil fuels for home-heating. Evening peaks are often weaker in summer than in winter because of the deeper atmospheric mixing layer.

(B) Averaging by wind speed and/or wind direction

Averaging pollutant concentrations by wind-speed is useful for identifying dust suspension, such as from soil and from construction activities. Averaging concentrations by wind direction highlights potential sources and source directions. Local sources can be straightforward to identify using a simple pollution rose. Identifying regional sources may require estimating air-mass back-trajectories.

(C) Comparisons of concentrations and chemical compositions among locations

Spatial differences in concentrations and chemical compositions can highlight influence zones of specific sources. For example, one might estimate the importance of urban-scale versus regional sources by comparing concentrations in an urban area versus upwind of that area. Similarly, because several trace elements (iron, lead, copper, zinc) are associated with industrial emissions, the presence of these elements indicates industry contributions.

Models

Inputs to these models are usually the chemical composition of air pollution, which the models then associate with sources. There are several examples of elements and sources that have been associated. Soil and dust are often associated with Al, Si, K, Ca, and Fe. The metals Cu and Zn are generally associated with specific motor vehicle components (brake and clutch linings; vehicle construction). Coal power station exhaust contains similar elements, but in different proportions, as fugitive dust. Elements found in gasoline exhaust vary significantly from one vehicle to the next. Organic carbon and black carbon are major components of gasoline vehicle exhaust and of wood smoke. Most K in wood smoke is water-soluble, in contrast to the insoluble K in fugitive dust. Recent work in Vancouver and Toronto *(1)* employed hopanes and steranes as markers for petroleum emissions.

The following elements and processes were generally associated in the past: V and Ni for residual oil combustion and refinery catalyst crackers; Pb, Cl, and Br for leaded-gasoline combustion; Cu, Fe, Mn, Cr, Zn, Zr, and As for metal smelting and refinery processes. Reformulation of fuels and processes has generally offered environmental improvements, but with the side effect of reducing or removing the utility of these elements as precise tracers. A possible exception to this general trend is ship fuel: because bunker fuel and residual oil have not been environmentally reformulated like many other fuels, two elements – V and Ni (and their ratio) – may offer a useful tracer of ship emissions.

In addition to chemical composition, particle size distributions and particle number concentrations can also provide information about sources. Combustion sources generally omit a large number of small particles, which are removed from the environment as they migrate downwind. Thus, the presence of large particle number concentrations indicates proximity to fresh combustion emissions. Elemental isotopes have also been employed, including in Vancouver *(2, 3)*.

(A) Chemical Mass Balance Models

The CMB approach combines chemical compositions for (1) air pollution concentration measurements and (2) emissions from specific sources or source categories. For example, one might measure the abundance of specific elements (sulfur, potassium, magnesium, silicon) and/or organic compounds (hopane, syringol, cholesterol, isoeugenol) in ambient air, and also in the exhaust gas for sources such as a coal boiler, a gasoline motor vehicle, a ship, and wood smoke. A mathematical optimization (error-minimization) estimates the strength of each source or source category. This approach works best when one or more compounds can be identified as tracer species for a source. Alternatively, during fieldwork one can supplement ambient measurements with the intentional release of a tracer from a specific source.

(B) Enrichment Factors Models

EF models are a scaled-down version of CMB. EF models employ a small number of measured compounds, rather than a large suite of compounds as with CMB, and EF models usually identify the source contributions from one specific source, rather than the quantitative source apportionment in the CMB. EF models work well when there are tracer species for sources of interest, such as silicon for soil and calcium for cement manufacture. Vanadium or the nickel-vanadium ratio may be a useful for ship emissions.

(C) Multiple Linear Regression; Eigenvector Analysis Models

These two approaches (MLR and EA) employ statistical analyses of a large number of samples (50-100 or more samples). By identifying spatial or temporal correlations among species concentration or properties such as light extinction coefficient, these methods identify pseudo-sources likely responsible for the concentrations. In fact, these methods do not identify true sources. Rather, they identify

mathematical “factors” or “principal components” based on correlations among species, and the user can attempt to match these principal components profiles with specific sources.

Application of Receptor Methods to the Burrard Inlet Study

For the Burrard Inlet study, two of the receptor methods seem especially promising. (1) If ambient monitoring data are available at 1-hour or shorter averaging times, the data analysis techniques mentioned above can potentially offer insights and hypothesis-generation are relatively low cost. For example, one could compare concentration measurements at the ambient monitoring stations during periods when the wind is and is not coming from the study area. (2) The chemical mass balance (CMB) approach involves sampling chemical species in ambient air, and matching these measurements to chemical “fingerprints” from source categories. Often, a CMB study would involve new measurements from emission sources to ensure locally-accurate chemical fingerprints; however, libraries of emission species for specific source categories are available.

One previous study has conducted receptor modeling in this region. The Regional Visibility Experimental Assessment in the Lower Fraser Valley studies (REVEAL I & II) were conducted in 1993 and 1994-95. Results were generated for Chilliwack, Clearbrook, and Vancouver. Results from this investigation are summarized online, at the website http://www.indiana.edu/~geog/reveal, and in the journal articles and reports cited there. The REVEAL studies did not, for example, distinguish between truck and ship emissions.

References

(1) Brook JR, Graham L, Charland JP, Cheng Y, Fan X, Lu G, Li SM, Lillyman C, MacDonald P, Caravaggio G and MacPhee JA. Investigation of the motor vehicle exhaust contribution to primary fine particle organic carbon in urban air. *Atmospheric Environment* **in press**.

(2) Huang L, Brook JR, Zhang W, Li SM, Graham L, Ernst D and Chivulescu A. Identification and characterization of inland ship plumes over Vancouver, BC. *Atmos Environ* **2006,** 40, (15), 2767-2782.

(3) Norman AL, Anlauf K, Hayden K, Thompson B, Brook JR, Li SM and Bottenheim J. Aerosol sulphate and its oxidation on the Pacific NW coast: S and O isotopes in PM2.5. *Atmospheric Environment* **2006,** 40, (15), 2676-2689.