**A LAND USE REGRESSION MODEL FOR
ULTRAFINE PARTICLES IN VANCOUVER, CANADA**

by

Rebecca Abernethy

B.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
(Occupational and Environmental Hygiene)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

March 2012

© Rebecca Abernethy, 2012

Abstract

Background and Aims:

Epidemiologic studies have associated adverse health outcomes with exposure to traffic-related air pollutants, principally NO₂, at levels below those showing effects in controlled exposure studies. This suggests the importance of related outdoor air contaminants, such as ultrafine particles (UFP) (<0.1 μm in diameter). Presently, no UFP monitoring exists in North America and little information is available regarding UFP spatial distributions.

We measured particle number concentrations (PNC) in Vancouver to develop a land use regression (LUR) model for use in epidemiologic studies and to identify important sources of UFP.

Methods:

During a two-week sampling period in spring 2010, PNC were measured with portable condensation particle counters (CPC) for 60-minutes at eighty locations used previously to characterize spatial variability in nitrogen oxides. Continuous PNC measuring occurred at four additional locations to assess temporal variation.

LUR modeling was conducted using 135 geographic predictors, including: road length, vehicle density, intersection and bus stop density, land use type, fast food restaurant density, population density and others, following previously developed methods. A novel buffer approach incorporated meteorologic data through wedge-shaped wind roses from measurements made during PNC sampling, in addition to circular buffers.

Results:

The range of measured (60-minute median) PNC across locations varied 70-fold (range: 1500 – 105 000 particles/cm³, mean [SD] = 18 200 [15 900] particles/cm³).

Correlations of PNC with concurrently measured two-week average NO₂, NO and NO_x concentrations at the same sites were 0.64, 0.65 and 0.70. A model ($R^2 = 0.48$, leave-one-out cross validation $R^2 = 0.32$) predicting PNC included length of truck routes within 50m, density of fast food locations within 200m and ln-distance to the nearest port. LUR models created with wind rose shaped buffers had lower predictive power than models with circular buffers ($R^2 = 0.29 - 0.34$).

Conclusions:

Measured PNC was highly variable across the Metro Vancouver region and correlated with nitrogen oxides. Geographic predictors explained a smaller proportion of variability in PNC than found previously for nitrogen oxides, suggesting some common sources and additional unknown factors influencing PNC spatial variability. This represents the first LUR model for UFP in North America.

Preface

The Metro Vancouver ultrafine particle sampling campaign, temporal data correction methods as well as summary statistics on measured PNC was presented orally at the 50th meeting of the Pacific Northwest International Section (PNWIS) of the Air and Waste Management Association in Missoula, Montana, November 3 – 5, 2010.

Sampling results as well as preliminary LUR models for ultrafine particles UFP were shared on a poster presentation at the British Columbia Particulate Matter Symposium in Vancouver, British Columbia, June 9, 2011.

Sampling results and LUR model results, including models incorporating meteorologic variables, were presented in a poster during the student symposium and in an oral presentation during the 23rd meeting of the International Society for Environmental Epidemiology (ISEE) conference in Barcelona, Spain, September 13 – 16, 2011.

Measurement data from the sampling campaign (section 3.2) were included in a journal article entitled: ‘Proximity of public elementary schools to major roads in Canadian urban areas.’ This paper was submitted to the International Journal of Health Geographics in September 2011, accepted in December 2011 and is now published online in its final form. My contribution to this paper was the Metro Vancouver ultrafine particle sampling campaign measurement data, including design and execution of the measurements and adjustments (temporal and instrument) to raw data. In addition I reviewed the manuscript and provided comments.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents	v
List of Tables	ix
List of Figures.....	xi
List of Abbreviations	xiii
Acknowledgements	xiv
Chapter 1: Introduction	1
1.1 Outdoor air pollution	1
1.1.1 Ultrafine particles	2
1.1.1.1 Health effects of UFP exposure	2
1.1.1.2 Sources of UFP and factors affecting concentrations	5
1.1.2 Methods of modeling PNC: LUR and others.....	8
1.1.2.1 Incorporating meteorology in LUR models.....	13
1.2 Rationale	13
1.3 Objectives and hypotheses.....	14
Chapter 2: Methods	16
2.1 UFP sampling	16
2.1.1 Pilot study	16
2.1.2 UFP sampling campaign study design.....	17
2.1.2.1 Site selection	17

2.1.2.2	Timing of sampling campaigns.....	20
2.1.2.3	Sampling equipment	20
2.1.2.4	Data collected during sampling	21
2.2	Data analysis	24
2.2.1	Descriptive statistics	24
2.2.2	Temporal corrections	24
2.2.3	Instrument correction.....	26
2.2.4	Comparisons with fall 2009 PNC measurements	26
2.3	LUR modeling	27
2.3.1	Geographic predictor variables.....	28
2.3.1.1	Circular buffers	29
2.3.1.2	Wind rose shaped buffers	30
2.3.1.3	Sources of geographic predictor variables.....	32
2.3.2	LUR model building procedure	39
2.3.3	Regression maps	41
2.4	Evaluation of measurements and models.....	41
2.4.1	Sixty-minute measurement evaluation.....	41
2.4.2	Model evaluation	42
Chapter 3:	Results	44
3.1	Sampling outcomes.....	44
3.1.1	Instrument correction factor.....	44
3.1.2	Temporal correction factors.....	44
3.1.3	Summary statistics	50

3.1.4	Comparing PNC with nitrogen oxides.....	52
3.1.5	Meteorologic data	54
3.1.6	Univariate correlations with PNC ₅₀ and geographic variables	56
3.2	PNC LUR models	57
3.2.1	Sensitivity analyses.....	60
3.2.2	Pollution surface images.....	61
3.3	Evaluation	66
3.3.1	Assessment of 60-minute samples.....	66
3.3.2	Comparing spring 2010 with fall 2009 measurements	68
3.3.3	Internal validation.....	69
3.3.4	External validation.....	70
Chapter 4:	Discussion.....	72
4.1	PNC measurements.....	72
4.2	LUR models.....	74
4.2.1	Comparison of Vancouver PNC LUR models with other LUR models.....	75
4.2.2	Comparison with other LUR models incorporating meteorology	82
4.2.3	Key findings of the Vancouver PNC LUR models.....	84
4.3	Strengths	84
4.4	Limitations.....	86
4.5	Applications.....	88
4.6	Future work.....	88
4.7	Conclusion	89
References.....		91

Appendices.....	99
Appendix A Metro Vancouver ultrafine particle database	99
Appendix B Ultrafine particle sampling campaign manual.....	103
Appendix C Ultrafine particle and nitrogen oxides comparisons.....	119
Appendix D Commands used in GIS variable extraction and R scripts.....	123
D.1 Geospatial modelling environment command examples	123
D.2 R script examples.....	124
Appendix E 2010 Ultrafine particle sampling campaign field notes.....	128
Appendix F Predictor variables for PNC LUR models (‘classic buffer’).....	129
Appendix G Predictor variables for PNC LUR model (based on measured wind rose data)	130
Appendix H Predictor variables for PNC LUR model (based on monitored wind rose data from a central monitoring station)	131

List of Tables

Table 2.1 Data collected during spring 2010 UFP sampling campaign	22
Table 2.2 Potential variables for UFP LUR modeling (adapted from (56))	37
Table 2.3 Summary of UFP LUR models created for Metro Vancouver	41
Table 3.1 Within-day variation correction factors applied to raw data	45
Table 3.2 Between-day variation correction factors applied to raw data	47
Table 3.3 Summary statistics for corrected one-hour PNC at 80 sites: median, 10 th and 90 th percentile.....	52
Table 3.4 Correlation coefficients between PNC and nitrogen oxides	52
Table 3.5 Examination of three sites with high residual values when comparing NO _x and UFP levels.....	54
Table 3.6 Summary statistics for site-specific and central station measured meteorology variables	54
Table 3.7 Range of wedge buffer lengths applied in measured meteorologic LUR model....	55
Table 3.8 Selected univariate correlations with PNC ₅₀	56
Table 3.9 LUR models developed for PNC in Metro Vancouver	58
Table 3.10 Selected univariate correlations between fast food and port variables and traffic- related variables	60
Table 3.11 Correction factors for comparing fall 2009 to spring 2010 PNC measurements .	68
Table 3.12 Predicting PNC at fall sites based on road length (PNC ₅₀) LUR model from spring sampling.....	69
Table 3.13 Predicting PNC levels at repeated and fixed sites by road length (PNC ₅₀) model	71

Table 4.1 Summary of LUR models published by other researchers for comparison with the Vancouver PNC LUR models.....	76
---	----

List of Figures

Figure 2.1 UFP sampling area in Metro Vancouver, showing fall locations (22 in orange), spring locations (58 in yellow plus 22 in orange, total 80), fixed sites (4), central monitoring site (1) with major road network.....	19
Figure 2.2 PNC sampling campaign equipment set-up, showing the CPC 3007 on the left and Kestrel 4500 on the right	21
Figure 2.3 Steps to creating a LUR model in Metro Vancouver (reprinted from (56)).....	28
Figure 2.4 500m and 1000m circular buffer around a point (sampling site) (adapted from (54))	30
Figure 2.5 Panel A: An example of a wind rose from a sampling site, used to make wind rose shaped buffers. Panel B: The windrose shapefile, used to extract values from geographic predictor variables. The blue, full extent of the wedges represents a 100% (maximum length) buffer while the yellow smaller wedges is a 50% buffer.....	32
Figure 3.1 Within-day variation correction factors by hour of day	45
Figure 3.2 Mean PNC by hour at four fixed sites over entire sampling period.....	46
Figure 3.3 Between-day variation correction factors by day of sampling.....	47
Figure 3.4 Mean PNC by day at four fixed sites over entire sampling period	49
Figure 3.5 Histograms of corrected PNC distributions (pt/cm^3). A: PNC_{50} . B: Ln-PNC_{50} . C: PNC_{10} . D: PNC_{90}	51
Figure 3.6 Residual plot of PNC_{50} (pt/cm^3) compared with average NO (ppb) from the same locations at the same sampling time	53

Figure 3.7 Pollution surface for road length, PNC ₅₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km), density of fast food restaurants within 200m.	62
Figure 3.8 Pollution surface for vehicle density, PNC ₅₀ LUR model. Variables in model: density of trucks within 25m, distance to port (ln km)	63
Figure 3.9 Pollution surface for road length, PNC ₁₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km). Note: different scale than previous figures.	64
Figure 3.10 Pollution surface for road length, PNC ₉₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km), density of fast food restaurants within 200m. Note: different scale than previous figures.	65
Figure 3.11 Random selection of 60-minute PNC measurements compared with two-week overall PNC at the same location. A, B, C, D: four fixed site locations.....	67

List of Abbreviations

CPC – condensation particle counter

DMTI – Desktop Mapping Technology Incorporated (DMTI Spatial is a Canadian publisher of spatial data)

GME – geospatial modeling environment

LOO – leave one out

LUR – land use regression

MV – Metro Vancouver

NO, NO₂, NO_x – nitric oxide, nitrogen dioxide, nitrogen oxides

PM_{2.5} – fine particulate matter (particles with an aerodynamic diameter <2.5µm)

PNC – particle number concentration or count, in particles per cubic centimeter

pt/cm³ – particles per cubic centimeter

RMSE – root mean square error

SD – standard deviation

SOEH – School of Environmental Health (now the Occupational and Environmental Health theme within the School of Population and Public Health)

SFU – Simon Fraser University

UBC – University of British Columbia

UFP – ultrafine particles (particles with a diameter <0.1µm)

Acknowledgements

First, I wish to thank my thesis committee for their academic guidance and expertise. My supervisor, Michael Brauer, has been available, encouraging and extremely helpful over the past two years through this project. Ryan Allen and Ian McKendry, my committee members, have also provided valuable assistance to this work.

I offer thanks to fellow students at the former School of Environmental Health (SOEH) and School of Population and Public Health (SPPH) for technical and peer support. In particular, Cris Barzan, Rongrong Wang and Perry Hystad provided a great deal of help with sampling equipment and computer programs. I am grateful for the work completed by my six research assistants, Rehema Ahmed, Sara Casallas, Mary Choi, Christie Cole, Brian Lee and Sarah St. John, during the field sampling campaign.

This study was made possible due to funding from Health Canada.

I wish to thank friends near and far for moral support and fun times while completing this degree. Finally, I am grateful to my parents for their unwavering support, always.

Chapter 1: Introduction

This chapter includes an overview of ultrafine particles (UFP), including the relationship with observed health impacts, sources and factors influencing concentrations, as well as information on land-use regression (LUR) modeling. A rationale for development of a UFP LUR in Metro Vancouver as well as study objectives and hypotheses are detailed.

1.1 Outdoor air pollution

An increasing base of evidence suggests chronic exposure to outdoor air pollution is associated with a variety of adverse respiratory, cardiovascular and reproductive health impacts. (1-7) The World Health Organization estimates that 1.4% of the global burden of disease can be attributed to outdoor air pollution. (8) With a significant portion of particulate matter (PM) and other pollutants in ambient air originating from motor vehicles, the direct health impacts of traffic-related air pollution (TRAP) and other sources are substantial. (8) Air pollution epidemiologic studies compare measured or modeled air pollutant exposure levels with health outcomes in a defined population. (9) Routine government-mandated air pollution monitoring typically used for exposure estimates lacks the station density required to capture small-scale spatial variations in concentrations of air contaminants. (9) In Metro Vancouver 22 monitoring stations are irregularly located throughout the region, placed to characterize regional background air quality and typically not located in close proximity to major roadways or other key sources of air pollution. (10) The shortest distance between two monitors in Metro Vancouver is 1.67km, with most monitors located an estimated five to seven kilometers apart. (10) The design of this network does not capture small-scale spatial

variability in pollutant concentrations, with major differences occurring over a distance of 300m or less in other studies. (15)

1.1.1 Ultrafine particles

UFP are a fraction of PM defined as having a diameter less than $0.1\mu\text{m}$. (11) UFP are often a direct product of combustion, though some secondary atmospheric formation occurs. (11,12) UFP are reactive and tend to coagulate (form accumulation mode particles) or disperse shortly after being emitted (on a scale of minutes to hours). (12) UFP are prevalent in urban areas and can account for over 90% of outdoor airborne particles by number. (11) $\text{PM}_{2.5}$ (diameter less than $2.5\mu\text{m}$) are regularly measured in British Columbia to evaluate observed levels relative to the BC Air Quality Objectives. (13) This measurement quantifies the mass of particles, of which UFP only constitute a small portion. UFP are one type of PM linked to adverse health outcomes and could be partially responsible for the effects observed in epidemiologic studies. (14,15) Given the importance of identifying specific components of the air pollution mixture that are harmful to humans, for the design of interventions and evidence-based air quality management strategies, understanding the spatial variability of specific outdoor air pollutants in addition to those routinely monitored is an important policy-relevant objective.

1.1.1.1 Health effects of UFP exposure

UFP are of particular interest from a health perspective, as these very small particles represent a large surface area which can contact cells in the body. UFP are defined by size and can exist in different shapes and may contain an array of contaminants, such as metals.

(11,12) Further, the small size of the particles make them respirable, permitting entry to the gas exchange area of the lung, as well as enabling entry into the circulatory system and penetrating cells via nonphagocytic means. (11) UFP interactions within cells generate reactive oxygen species, leading to oxidative stress (an imbalance of destructive reactive oxygen species and the body's ability to repair damage). (16) To date, a limited number of epidemiologic studies have been conducted for UFP. A summary of toxicology and short-term epidemiology studies on UFP and as well as expert opinions on health effects are provided below.

Animal studies have shown acute UFP toxicity effects and the ability of UFP to translocate within the body. One study using polytetrafluoroethylene (PTFE) particles with an approximate diameter of 18nm induced lethality in rats after fifteen minutes of inhalation at $50 \mu\text{g}/\text{m}^3$ (the author likened concentrations of 10 000 – 40 000 pt/cm³ to be about 2 $\mu\text{g}/\text{m}^3$, so $50 \mu\text{g}/\text{m}^3$ then represents 250 000 – 1 000 000 pt/cm³). (17) Rapid translocation of UFP to epithelial, interstitial and endothelial sites was observed. The same study found that aging PTFE fumes agglomerated into larger particles that reduced their toxicity. Another study again exposing rats to UFP in the form of TiO₂ particles found significantly more bronchoalveolar inflammation and interstitial and lymph node transfer of UFP compared with fine PM (PM_{2.5}) at 23 mg/m³ (equal doses of the contaminants by mass). (18) Further research on rats has shown that UFP can translocate into other organs in the body and remain at high levels up to seven days after exposure. (17,19) Animals were exposed to labelled ¹³C at $160 \mu\text{g}/\text{m}^3$ for six hours; afterward, the radioactive UFP were found in the liver and olfactory bulb in the brain. (19) Health effects of this translocation are unknown. The contrast in health effects from different exposure levels mentioned (lethal outcomes at 50

$\mu\text{g}/\text{m}^3$ for fifteen minutes versus survival at $160 \mu\text{g}/\text{m}^3$ for six hours) illustrates the variability of effects possible from UFP due to differing size and chemistry between particles. It also highlights the complex task of studying mixtures of UFP, in terms of particle size and composition, found in outdoor air.

A review paper (20) on human health effects of air pollution concluded that there were relatively few studies on the health impacts of UFP compared with fine and coarse PM, especially long-term epidemiologic studies. The review described some short-term studies linking UFP exposure to adverse cardiovascular outcomes, reduced pulmonary function, acute respiratory symptoms, autonomic nervous system problems, changes in blood pressure and impaired endothelial functions. Notable short-term studies found increases in UFP exposure to be correlated with increases in all-cause, cardiovascular and cardio-respiratory mortality. (21,22) A controlled exposure study in North Carolina found 24-hour standard deviation heart-beat intervals to decrease significantly while D-dimer plasma factor increased significantly (approximately 17%) with increased controlled exposure to UFP, which could suggest an increased tendency toward clotting and cardiac outcomes. (23) A study in Germany found that increased exposure to UFP was linked with increased plasma sCD40L levels, a marker for platelet activation. This can lead to coagulation and inflammation and decreased platelet count, which were hypothesized to be due to activation binding platelets to tissues. (24) Limited-term studies positively associated UFP exposure to all-cause morbidity, cardiovascular, pulmonary and hospital admission outcomes. (25)

One paper described an expert European panel (12 individuals including clinicians, toxicologists and epidemiologists) that followed systematic methods to determine the likelihood that UFP exposure contributes to health outcomes. (26) The panel generally

agreed that short-term exposure to high UFP concentrations was likely to contribute to all-cause mortality, cardiovascular hospital admissions, aggravation of symptoms in asthma patients and decreases in lung function. With long-term exposure, the experts tended to believe UFP are causally linked to all-cause mortality and cardiovascular morbidity. The panel was most confident that a pathway involving respiratory inflammation leading to a cardiac event through processes involving plaque rupture and thrombosis is most likely responsible for the outcome, when assuming UFP are causally related to cardiac events. The panel emphasized that spatially resolved UFP estimates are needed in order to conduct much-needed long-term epidemiologic studies.

While more research in this area is needed, overall, existing work suggests that UFP exposure may lead to a number of adverse health outcomes, especially cardiovascular events.

1.1.1.2 Sources of UFP and factors affecting concentrations

The primary source of UFP in outdoor air is motor vehicle exhaust, with more emissions generated from heavy-duty diesel than spark ignition engine vehicles. (27-31) Low-emission (of carbon dioxide) diesel engines tend to produce higher concentrations of UFP than conventional ones. (27,28) For gasoline-powered automobiles, UFP emissions are higher when the vehicle is older, accelerating or driving at high speeds. (28)

Stationary combustion sources, such as industrial parks, may be important sources of UFP emissions, (32) while residential wood combustion is also recognized as a major source, particularly at night and during cold seasons. (33,34)

There is indication from work in Amsterdam that marine vessels also significantly contribute to UFP levels onshore near major shipping areas. (35) The PM distribution emitted by ships is dominated by UFP. (36,37)

Restaurants, especially fast food and grill outlets that conduct a lot of frying, are also thought to contribute to UFP concentrations. For example, measurements from the California Children's Health Study reported high UFP levels (similar to those near major freeways) at a monitoring site next to a fast-food restaurant (38) while a study of UFP concentrations in Manchester, England suggested high restaurant density as a predictor of elevated UFP concentrations. (39) Indoor measurements of UFP indicate cooking as important source (32) and it is hypothesized that ventilation of indoor-generated UFP outside could increase outdoor concentrations also. (39) Additional work has specifically suggested meat cooking as a source of UFP. (33,40,41) A study in a single home in Reston, Virginia found particularly high levels ($>100\,000\text{ pt/cm}^3$) during cooking with gas or electric stoves or during use of electric toaster ovens. (42) Levels inside restaurants were also found to be consistently high during the course of eating a meal, with levels ranging from $50\,000 - 200\,000\text{ pt/cm}^3$. (42)

Concentrations of UFP are highly spatially variable. In a recent study conducted on cyclist exposures in Vancouver, UFP were heterogeneously distributed in space; areas with heavy traffic volumes had the highest concentrations. (30) Typically UFP measured by PNC exhibit much greater spatial variability within urban areas than measures of particle mass concentration. A publication from Augsburg, Germany found UFP to be 1.5 – 1.8 times higher at traffic versus background locations (15), although a study in The Netherlands found $\text{PM}_{2.5}$ to be only 17 – 18% higher at traffic sites compared with background sites. (43) A study in Amsterdam reported that median PNC levels were 80% higher at traffic locations

compared with urban background locations, while $PM_{2.5}$ at the same sites was only 9% higher at traffic sites. (44)

Several factors can explain UFP spatial patterns, including proximity to major roadways and meteorologic conditions. First, proximity to motor vehicle traffic, a major source, greatly affects UFP concentration. UFP levels have been shown to decrease exponentially with distance from major roadways. (15,28,45,46) Specifically, a 100m buffer around a traffic source found highest UFP concentrations (15,28) while over 300m away UFP levels were significantly reduced and similar to background concentrations. (46) There may be some temporal variation with UFP concentration, where a peak is reached during morning rush hour, which gradually declines over the course of the day, until a smaller peak is seen in the afternoon. (31,47) One study in the Los Angeles area measured UFP on and off major roadways and concluded that coupled with typical time allocation patterns of area residents, nearly 50% of most people's UFP exposure occurred on freeways or arterial roadways, underscoring the importance of proximity as well as time-of-day when considering UFP exposure. (48)

Second, meteorologic and topographic variables have an influence. Low and high wind velocities tend to lower UFP concentrations; at low speeds more time is provided for the particles to coagulate, while at high speeds rapid atmospheric dilution occurs. (12) Wind direction largely determines particle location. Long-range transport can increase UFP levels downwind of a primary emission site. (45) Downwind and medium wind speed conditions near roadways are associated with higher UFP levels. (48) Other research shows that unstable conditions disperse UFP, reducing concentrations near breathing height. (49) Relatively low temperature and high humidity are associated with slower atmospheric dispersion, suggesting

UFP levels are greater in the winter than summer; further, higher temperatures in summer favour increases in accumulation-mode ($PM_{2.5}$ range) particles. (15) Nocturnal surface temperature inversions pre-sunrise have also been associated with high concentrations of UFP further away from major roadways than observed during peak hours. (15,49) A study in Montreal found UFP concentrations to be inversely associated with wind speed and temperature. (50) Physical geographic factors, such as topography and altitude also dictate dispersion, where low-lying valleys collect PM and high elevations have greater atmospheric dispersion. (46,48,51)

Other land-use variables play a role in UFP concentration as well: industry operation, population density (which may simply be a surrogate for more sources in an area) and agricultural activity all increase concentrations principally within 500 – 1000m radii zones of the source. (14,15)

1.1.2 Methods of modeling PNC: LUR and others

Given the limited ability of routine monitoring networks to characterize spatial variability in air pollution, high-resolution exposure estimates are required for epidemiologic and risk assessment studies related to TRAP and other sources of outdoor air pollution. A study on NO_2 in Montreal noted that concentrations decreased logarithmically with increasing distance to nearest highway, with the most profound decreases occurring within 200m of the roadway. (52) Another study from New York reported that UFP concentration decreased 15 – 20% within 100m of a major roadway. (53) A study from Los Angeles reported PNC decreasing exponentially with distance to a major highway, with levels 300m away indistinguishable from background concentrations and the majority of the decrease in

concentrations occurring within the first 100m. (12) These examples highlight the spatial variability of outdoor air pollution, especially UFP, on scales of less than 300m. As the monitoring network typically consists of stations located kilometers apart and not close to major roads, these sources of data are insufficient to reasonably reflect spatial exposure patterns of these types of pollutants.

LUR models are a powerful, commonly used method for capturing this variability and estimating exposures for large study populations dispersed throughout large regions. (54,55) LUR models developed following specific procedures have traditionally been based on spatially dense measurements of TRAP indicator pollutant(s) (commonly NO₂ as it can be measured easily and affordably relative to other contaminants) and readily available geographic variables (such as land use designations, road types, elevation) surrounding the measurement locations. (55,56) The relationship between these site attributes and pollutant concentrations is determined empirically using standard multiple linear regression methods. (55,56) LUR models have been applied in a number of epidemiologic studies; for example, many have investigated TRAP and childhood respiratory disease outcomes. (4,57,58)

While there are a number of other modeling options, outlined below, for demonstrating spatial variability of air contaminant concentrations, LUR has been shown to offer a resource-effective, reliable approach for providing exposure estimates. (44,56,59) Three main types of alternatives are emission inventories, dispersion models and receptor methods. One major limitation of all alternatives, including LUR for UFP, is the lack of measurement data in many regions, including Vancouver.

Emission inventories estimate the amount of a pollutant released into the greater environment from specific sources. Data from emissions inventories can be geocoded and

used to create a map of emissions, highlighting areas of expected high concentrations. (9,56) However, emissions inventories are subject to large error and are not reported as concentrations, so supplemental field measurements also need to be completed. (56) Further, these models only incorporate emissions, excluding atmospheric dispersion and transport. Proximity models could be considered a type of emission inventory where the distance from the source is used to estimate exposure level. These types of studies are limited by considering the nearest location of a type of source (such as major road) to the exclusion of other source contributors that may also be located nearby and still contribute to pollutant concentrations. (9,56)

Dispersion models predict pollutant levels based on emissions and factors that may affect concentrations once released, such as meteorology and chemical and physical processes. (9,56) Dispersion models require detailed, localized input, which may be difficult to obtain and technically challenging to incorporate. Emissions data is needed for all pollutants from all sources and becomes complex when atmospheric chemical reactions occur and secondary pollutants (such as coagulation of UFP) are formed. (9,56) Eulerian Grid models are one type of dispersion model that are computationally complex and created at coarse resolutions (magnitude of kilometers). (56) Gaussian plume models are widely used to predict concentrations any distance from a source, but generally do not consider atmospheric chemistry. (9) LaGrangian or ‘puff’ models incorporate pockets of air pollution being moved by wind and are often regarded as more theoretically accurate than plume models. (56)

Receptor models begin with a pollutant measurement at specific sites and work backwards to determine contributions from sources, by incorporating chemical composition data and other information. (9,56)

Interpolation models, which could use network monitoring data, employ geospatial methods such as kriging, incorporating spatial dependence in measured levels to create a continuous pattern of pollution in space. This predicts concentrations at unmeasured locations in a study area. This option tends to homogenize exposure levels over an area and is particularly unreliable near edges of study zones. (9,56) Since UFP are not routinely monitored or regulated by governing bodies, (60) this method is particularly limiting for UFP.

Of the alternatives listed above, only emissions inventories would be practical for discerning spatial patterns of UFP in Vancouver, as dispersion models would have difficulty incorporating the reactive nature of UFP and virtually no data exist to develop receptor models. LUR then offers a high-resolution exposure estimate solution for UFP.

LUR models for Metro Vancouver and southern BC have been developed for NO/NO₂, PM_{2.5} and Black Carbon and woodsmoke (61,62), demonstrating the feasibility of this modeling approach in the region.

Most LUR modeling for TRAP exposure assessments have focused on NO and NO₂. (55) While exposure estimates from models based on NO₂ have been linked to adverse health outcomes, these concentrations are often below those that have demonstrated negative impacts in controlled exposure studies or in epidemiologic studies of indoor gas combustion sources. (21,63,64) It is possible that in studies of within-city air pollution gradients, associations observed between NO₂ and health impacts acts as a surrogate for other harmful

species in the traffic pollution mixture. (63,64) Correlation between UFP and NO/NO₂ levels has been reported in a limited number of studies, but further comparison of measurements is needed. (47,65,66) In a Los Angeles study, moderate correlation between PNC and NO/NO₂ levels were found ($r = 0.08 - 0.68$, depending on site) with high correlations found near traffic sites during morning rush hour ($r > 0.9$). (65) In Copenhagen, PNC-NO_x temporal and spatial correlations were also high ($r > 0.84$). (47)

A LUR model for UFP has been successfully developed for the city of Amsterdam and is at the time of writing the only published LUR model for UFP. (44) Measurements were successfully obtained at 46 sites outside homes, representing high traffic ($N = 20$) and urban background ($N = 26$) locations over a one-week period at each location, spread over two years. A single continuously monitoring background location provided data for temporal correction. Significant predictors of UFP concentrations included high traffic intensity (specifically traffic intensity and inverse distance to road squared), address density and industrial land use.

In a study conducted in Corpus Christi, Texas, UFP at the census block level were modeled with a modified California Line Source 4 (CALINE4) Gaussian dispersion model incorporating distance-scaling functions (which weight the receptor measurement according to distance from a major source: a major road). In New York, UFP levels were predicted with an autoregressive statistical model based on a dense ($N = 8225$) measurement campaign incorporating traffic count and distance from two major roadways, but not other land-use variables. (53) This study used smoothing-distance effects to monitor the percent decrease in UFP from two specific roadways and presented an explanatory equation showing how UFP changed in relation to proximity to the roadways. Neither of these models are both spatially

highly-resolved and cover large areas, two criteria that are important in epidemiologic studies of UFP.

1.1.2.1 Incorporating meteorology in LUR models

A number of studies have found that varying meteorology, especially regarding wind speed and direction, can substantially influence PNC and other pollutants. Most LUR models use a circular buffer around a sampling point to determine the influence of various geographic predictor variables on pollutant concentrations, under the assumption that over the long term, areas surrounding a point in all directions influence the pollutant in the centre equally. The absence of an approach to incorporate meteorologic phenomena in LUR models has been repeatedly noted as a limitation. (12,15,45,50,67)

1.2 Rationale

UFP represent the majority of PNC in urban air and are linked to a number of adverse health impacts. (14,15,68) UFP are different than more conventionally measured PM mass concentrations (PM_{2.5} and PM₁₀) in that their physical and chemical properties, largely due to their small size, may result in different toxicological impacts. Independently, measurement data on UFP in Vancouver adds to understanding concentrations and patterns of the pollutant in the region, due to its absence from regulatory monitoring. Comparing UFP with regulated contaminants, such as NO_x, provides insight to the degree contaminants share sources. A model that can predict individual exposures of residents based on residential location (56,61) may be useful to epidemiologic and risk assessment studies evaluating health effects of UFP. More directly, a LUR model for UFP has the potential to indicate the relative importance of

different sources in predicting pollutant concentrations. Both a predictive model of UFP and better understanding of UFP sources are relevant when considering policy implications, such as potential monitoring, exposure limits and air quality management.

1.3 Objectives and hypotheses

The specific objectives of this project are:

1. To conduct spatial measurements of PNC at 80 varied sampling sites located throughout the Metro Vancouver region.
2. To compare one-hour median PNC values and two-week average measurements of NO and NO₂ at the same sampling locations.
 - Hypothesis: PNC and NO or NO₂ concentrations are predicted to be highly, positively correlated, based on common sources and other studies showing correlation between TRAP components.
3. To assess the relationship between measured PNC and spatial predictors characterizing traffic flow, land use, topography and fast-food restaurant density/proximity in the study region.
 - Hypothesis: PNC is expected to be higher near high traffic areas, especially where diesel trucks travel, near fast-food locations and near industrial sources.
4. To use the PNC measurements in combination with the above spatial predictors and meteorologic data to develop a LUR model for PNC in Metro Vancouver.

- Hypothesis: Spatial patterns in PNC can be predicted using geographic predictors and meteorologic data. These models can explain a similar portion of variability as NO_x models.

Chapter 2: Methods

This chapter describes the methods used to develop LUR models for UFP in Vancouver, including the measurement campaign, data preparation after sample collection, creation of models and evaluation.

2.1 UFP sampling

Measurement of a pollutant in a spatially dense network is the first step in the development of a LUR model (and an informative exercise independently, especially for a contaminant such as UFP which is not routinely measured or regulated in North America).
(60)

2.1.1 Pilot study

Preliminary sampling was conducted at 22 sites in Metro Vancouver in October 2009 using a P-trak 8525 device (TSI®, Shoreview, MN, 2007) to record PNC values over one-hour periods. These pilot sites were selected from 116 sites used during the 2003 NO and NO₂ measurement period in the region by convenience from those representing a range of NO_x concentrations. Following examination of the collected data, including correlation between these UFP measurements and NO_x measurements made at the same locations during a 2003 (61) sampling campaign, a decision was made to proceed with a full PNC sampling campaign in conjunction with a 2009/2010 NO_x campaign.

2.1.2 UFP sampling campaign study design

Sampling for PNC was conducted at 80 sites (locations shown in Figure 2.1) across Metro Vancouver in April and May 2010. The 22 sites sampled in October 2009 are also shown. Four additional sites across the region were monitored continuously during the sampling period to provide data for temporal correction. All details of the sampling campaign are located in Appendix B, the ultrafine particle sampling campaign manual.

2.1.2.1 Site selection

The 80 sites where PNC was measured during the spring campaign were a subset of the 116 selected by the location-allocation method during the 2003 NO and NO₂ measurement period in the region. (61,69) Sampling locations were selected to characterize variability and to be geographically representative of the region. The study area was separated into four geographic quadrants, with 16 – 18 sites selected in each quadrant, including sites with low and high NO/NO₂ concentrations as measured in 2003 and those near fast food restaurants. The quadrants were not based on hard boundaries, but rather general portions of the region. They were: northwest (Vancouver), southwest (Richmond, Delta), northeast (east Burnaby, Coquitlam) and southeast (Surrey). A fifth grouping represented 12 geographic outlying sites in the region, including locations in the municipalities of North Vancouver, Tsawwassen and Abbotsford, to enhance the applicability of the model for the entire Lower Mainland region. All 22 sites from 2009 pilot sampling were included in spring sampling. Repeat sampling on different days occurred at 20 sites to validate the temporal adjustments applied to the measurements during the spring sampling campaign.

Further, four stationary sites measured PNC levels continuously to provide data for temporal (within day and between day) correction factors. The four fixed site locations were determined by attempting to represent different parts of the study area (including west and east parts of the region, those in high-traffic neighbourhoods and low-traffic neighbourhoods). Convenience also played a role in the fixed site selection, as the instruments had to be located at homes for security purposes and since the instruments required changing alcohol wicks on a daily basis. One monitor was located in eastern Vancouver (Kingsway, Site 1), one toward southern Vancouver (Cambie Village Site 2), one in and one in Port Coquitlam (Site 3) and one on the west side of Vancouver (Kitsilano, Site 4), covering three of the quadrants described above (excluding Surrey and outlying areas).

One central monitoring station (T18, South Burnaby) was selected to provide regional meteorologic data for the models due to its central location within the study region and is also shown in Figure 2.1.

Figure 2.1 UFP sampling area in Metro Vancouver, showing fall locations (22 in orange), spring locations (58 in yellow plus 22 in orange, total 80), fixed sites (4), central monitoring site (1) with major road network

2.1.2.2 Timing of sampling campaigns

The main sampling campaign occurred in April – May 2010 (April 19 – May 6) to correspond with NO_x sampling. As one of the study objectives was to compare PNC and NO_x values at the same locations during the same time period, the sampling campaigns were conducted at the same time of year (NO_x sampling in Metro Vancouver occurred in fall 2009 and spring 2010 as well). Measurements were taken Monday to Friday, between 8:00am – 6:00pm. The four continuous monitors recorded values seven days a week, twenty-four hours a day during the entire sampling period, to be sure that data covering all mobile sampling times was collected and to provide base data to determine a two-week period PNC average.

2.1.2.3 Sampling equipment

PNC were measured by Condensation Particle Counter (CPC) 3007 devices (TSI®, Shoreview, MN, 2007) in the spring sampling campaign. P-trak 8525 (TSI®, Shoreview, MN, 2007) were used in the fall pilot. P-trak 8525 and CPC 3007 devices were co-located during spring sampling to develop a correction factor between the two types of instruments, in order to compare the fall 2009 and spring 2010 measurements. Samples at all 80 sites were one-hour in length, with values recorded at one-minute intervals. The four fixed monitors recorded five-minute average UFP concentrations every fifteen minutes (the greatest frequency possible, given equipment operation and instrument memory limitations).

Additionally, at all sites, Kestrel® 4500 Pocket Weather Trackers (Nielsen-Kellerman, Boothwyn, PA, 2009) were used to record meteorologic variables, including wind speed, wind direction, temperature, humidity and pressure, as well as elevation. These variables were

recorded at five-minute intervals during the one-hour measurement period, the highest frequency option given the device's memory capacity.

A GPSMAP® 60CSx or Geko 301 (Garmin®, Olathe, KS, 2009) global positioning system (GPS) device was used to record latitude, longitude and elevation at each site. Waypoints were converted to Universal Transverse Mercator (UTM) zone 10 using an online calculator. (57) For sites with repeat sampling, the mean of measured coordinates was used to identify locations.

Figure 2.2 PNC sampling campaign equipment set-up, showing the CPC 3007 on the left and Kestrel 4500 on the right

2.1.2.4 Data collected during sampling

Table 2.1 outlines meteorology and field observations recorded at all 80 sampling sites and the four continuous monitoring locations.

Table 2.1 Data collected during spring 2010 UFP sampling campaign

	Measurement parameter	Method	N (# of sites)	Duration	Averaging time
80 Mobile sites	UFP concentration (PNC, pt/cm ³)	CPC 3007	80 (plus 20 repeats)	60 min sample	1 min
	Wind speed (m/s)	Kestrel 4500	80 (plus 20 repeats)	60 min sample	5 min
	Wind direction (°)	Kestrel 4500	80 (plus 20 repeats)	60 min sample	5 min
	Temperature (°C)	Kestrel 4500	80 (plus 20 repeats)	60 min sample	5 min
	X, Y coordinates (°)	GPS	80 (plus 20 repeats)		
	Elevation (m)	GPS	80 (plus 20 repeats)		
	Heavy-duty diesel vehicle counts	Visual observation of trucks and buses on road/intersection	80 (plus 20 repeats)	2 x 5 min sample	every 30 min
	Fast food restaurant counts	Visual observation of locations visible when walking 100m in all directions from site	80 (plus 20 repeats)		
	Field notes and photographs	Completion of Community Ultrafine Particle Sampling Field Log, Appendix B	80 (plus 20 repeats)		
4 Fixed sites	UFP concentration (PNC, pt/cm ³)	CPC 3007	4	18 days (entire sampling period)	5 min average every 15 min
	Wind speed (m/s)	Kestrel 4500	4	18 days (entire sampling period)	60 min
	Wind direction (°)	Kestrel 4500	4	18 days (entire sampling period)	60 min
	Temperature (°C)	Kestrel 4500	4	18 days (entire sampling period)	60 min

Four additional site-specific variables were created and outlined below, for use in the ‘best model.’

1. Manual heavy-duty vehicle counts

The mean of two, five-minute counts taken during each sample was calculated, divided by five (to obtain number of vehicles/minute) and multiplied by the number of minutes in the sample.

2. Wind direction

Research assistants estimated the overall wind direction during sampling based on visual observation (such as the direction trees were blowing). It is acknowledged this is a subjective assessment. Under conditions where no wind was discernable, the upwind designation was applied.

- 0 - upwind of potential major UFP source (major roads)
- 1 - downwind of potential major UFP source (major roads)

3. Fast food or other grill sites

- 0 - no locations visible within 1000m of the sampling site
- 1 - one location within 100m of the sampling site
- 2 – two locations within 100m of the sampling site
- 3 – three or more locations within 100m of the sampling site
- 4 - at least one location more than 100m but less than 1000m from the sampling site (or at least one location visible from site)

4. Construction activities visible within 500m of sampling site that were potential UFP emission sources

- 0 – no

- 1 – yes

2.2 Data analysis

2.2.1 Descriptive statistics

As many pollutants follow a lognormal distribution (70), the distributions of both median and log transformed (ln) median PNC values from each sample are displayed in Figure 3.5. The mean, median, 10th and 90th percentiles were all used for LUR modeling, the latter two assessed spatial variability in background levels of UFP (PNC₁₀) and locations with high peak levels (PNC₉₀).

2.2.2 Temporal corrections

Raw PNC measurements were adjusted for temporal variation, both within and between days, as well as for sampler variation to create a database of comparable values for LUR modeling. Data collected for temporal adjustments and calculations are in Appendix A.

Within-day variation adjustments were made using data from the four fixed locations according to methodology utilized previously (71) and the calculation is shown in equation 2.1.

Equation 2.1. Within-day correction factor (for each sampling hour)

$$CF_h = \frac{\left\{ \frac{\sum_{0,1}^{n,k} PNC_{50n,k}}{(n+1)k} \right\}}{\left\{ \frac{\sum_1^k PNC_{50h,k}}{k} \right\}}$$

CF = correction factor

h = hour for correction factor

n = hour of day (0 – 23)

k = fixed site number (1 – 4)

d = day for correction factor

m = day of sampling (1 – 19)

First, the mean PNC value for each sampling hour (median value at the four sites for that hour averaged) was calculated. Next, the overall mean value (median value of each hour at the four sites averaged) was determined. The overall mean was divided by the hourly mean value to determine the correction factor for a particular hour. Whole hours (i.e. from 09:00 to 10:00) were calculated and the correction factors were applied according to the hour the majority of the sample was in.

Between-day variations were calculated in a similar manner, shown in equation 2.2.

Equation 2.2 Between-day correction factor (for each sampling day)

$$CF_d = \frac{\left\{ \frac{\sum_{1,1}^{m,k} PNC_{50m,k}}{m k} \right\}}{\left\{ \frac{\sum_1^k PNC_{50d,k}}{k} \right\}}$$

*Symbols in equation 2.2 are the same as in equation 2.1

First, the mean PNC value for each sampling day (median value at four sites for that day averaged) was calculated. Next, the overall mean value (median value of each day at each of the four sites averaged) was determined. The overall mean value was divided by the daily mean value to determine the correction factor for a particular day.

2.2.3 Instrument correction

Two CPC instruments were used to measure PNC at the 80 spring sites. Based on collocation testing, one of the two instruments had consistently had lower readings than the other. One device was deemed most ‘correct’ based on its consistency with eight other instruments used during collocation testing (Appendix A), so a correction factor for the other instrument was developed. The correction factor was created by dividing raw measurements from one instrument by the other, taken at the same location at the same sampling minute, and averaged over all collocated minutes in three tests.

2.2.4 Comparisons with fall 2009 PNC measurements

As described above (section 2.1) PNC measurements were also made in fall 2009 at 22 of the 80 sites included in the spring sampling. For comparison with measured and modeled PNC in spring 2010, these measurements were first corrected for instrument differences. A correction factor was applied to convert measurements from the P-trak 8525 (fall) to the CPC 3007 (spring). This factor (1.51) was based on three collocated outdoor samples, described in Appendix A.

Next, predictions from the spring model were compared to unadjusted and adjusted fall measurements. In the unadjusted assessment, the spring model was used to directly predict measured PNC from the fall; R^2 and RMSE between unadjusted measurements and predictions

were calculated. In the adjusted assessment, temporal corrections for time of day (section 2.2.2, developed during spring sampling) were applied to the fall measurements. Two different adjustments for season were applied in separate comparisons: 1) based on the ratio of the overall mean (one-hour median values from all sites) PNC in the fall and in the spring (at the same sampling locations only) and 2) based on NO concentrations measured at all available Metro Vancouver monitoring stations (N = 20) during the two sampling periods. This second adjustment was conducted to determine the feasibility of using a regularly monitored pollutant to correct for seasonal variability.

To summarize, a total of three comparisons between fall and spring values were made: a comparison of predicted and unadjusted measurements, a comparison of predicted and adjusted measurements based on the ratio of spring:fall PNC measurements and a comparison of predicted and adjusted measurements based on the ratio of spring:fall NO measurements from Metro Vancouver monitoring stations.

2.3 LUR modeling

The Metro Vancouver UFP LUR models were developed using the same procedure as the original NO/NO₂ Vancouver models (61); the schematic in Figure 2.3 outlines the basic procedure in model creation. First, geographic predictor variables that may explain the variability in pollutant concentration, such as road networks, land use classifications, topographic features were calculated using geographic information systems (GIS) software for each measurement location. From this a regression model was created that predicts the level of the contaminant at any site in the region based on the geographic predictors. A pollution surface map, which is a visual representation of the regression model, can then be generated.

Figure 2.3 Steps to creating a LUR model in Metro Vancouver (reprinted from (56))

2.3.1 Geographic predictor variables

One hundred thirty-five geographic predictor variables were calculated in preparation for UFP LUR model development. Variable layers were developed using ArcGIS ArcInfo software, version 10.1 (ESRI®, Redlands, CA, 2011). The complete list of geographic predictors is outlined in Table 2.2. Selection of predictors was based on those used to develop previous LUR models in Vancouver, (61) but a number of additional variable types were included for UFP LUR model development, highlighted in yellow.

For each variable type, data were obtained in either vector (shapefile) or raster format (from sources outlined in Section 2.3.1.3) and various ArcGIS tools were used to prepare each variable type with the appropriate buffer size or distance. For example, all data sources were

clipped by a shapefile created previously (56,61) to limit the geographic extent of the study area (and projected to UTM Zone 10N). Using the select by features tool, new layers featuring only specific road types or locations of intersections were created. Shapefiles were then converted to a raster and the appropriate buffer size (if applicable) was generated. Then, commands using Geospatial Modelling Environment (GME) software (Spatial Ecology, 2011) generated values of a particular variable for each of the 80 sampling locations. Commands used included: intersect points in raster (isectpntrst), which calculated the raster value (for example, population density within 2500m) at each sampling location (point file) and sum line length in polygon (sumlinelengthspolys), which was used to sum the length of lines (such as highways) that intersected wind-rose shaped polygons from each sampling location. Appendix D provides a sample of GME commands used. Distance calculations of each variable at each sampling location were completed using the ‘near’ tool in ArcGIS.

2.3.1.1 Circular buffers

Most geographic predictor variables were derived for different buffer sizes, following two methods. The first is the ‘classic’ circular buffer, shown in Figure 2.4. These circular buffer sizes ranged from 25m to 2500m (buffer sizes for each variable type are listed in Table 2.2) depending on the variable class and likelihood that fine scale changes would impact UFP levels markedly, (e.g. highways, major roads, truck routes) based on previous research. (53) In the example shown in Figure 2.4, a 500m and 1000m buffer are drawn from a radius of that distance from the sampling location. Values for the variables were generated by summing the lengths of lines within the buffer in the case of a polyline variable (such as road length), summing the area

within the buffer in the case of a polygon variable (such as land use), or counting the number of points within the buffer in the case of a point variable (such as bus stops).

Figure 2.4 500m and 1000m circular buffer around a point (sampling site) (adapted from (54))

2.3.1.2 Wind rose shaped buffers

The second method for buffer generation was a novel approach incorporating measured wind speed and direction at time of pollutant sampling. In this procedure, a wind rose for each sampling location was created using the software WindRose PRO (Enviroware, Concorezzo IT, 2011); an example is shown in Figure 2.5. A wind rose includes the direction the wind was blowing, the length of time it was blowing in that direction (reflected by the magnitude of the wedge) and the wind speed (reflected by the colour of the wedges in Panel A). The shape of the wind rose for each site was exported from WindRose PRO as a shapefile, where the direction and time were preserved (Panel B).

Two types of wind roses were created. The first was developed from data measured at each location during the sampling campaign with the Kestrel 4500. The shapefile was assigned a

maximum length equal to the mean wind speed measured multiplied by the duration of the sample (minutes), to represent the maximum source area that could influence the sampling point. Buffers representing $1/8$, $1/4$, $1/2$, $3/4$ and 1 times the maximum length of the wind rose were also created, with the same angles and shapes as the maximum wind rose. Figure 2.5, Panel B shows an example of a 50% buffer in yellow and a 100% buffer in blue. Seventy-seven geographic predictor variables were derived for the meteorologic LUR models, shown in Table 2.2. Nine sites had a wind speed of <0.5 m/s measured (limit of detection of Kestrel 4500) during sampling. Since no wind rose could be produced for these sites, a circular buffer was used, assuming all directions affected the transport of UFP equally. In these instances, the circular buffer radius selected as half of the limit of measurement for the Kestrel 4500 was multiplied by the number of seconds in the sample. This value (0.25 m/s $\times$ ~ 3600 s = 900 m) represented the maximum distance the pollutant could have traveled in the sampling time. Circular buffers of $1/8$, $1/4$, $1/2$ and $3/4$ of this maximum distance were offered to models.

The second type of wind rose buffer used data from a central (T18 South Burnaby, Figure 2.2) Metro Vancouver monitoring site to create a regional rose that was applied to all sites. This wind rose was based on wind speed and direction data measured from 8:00am to 6:00pm for all days during the sampling campaign. These measurements were collected at 10m above ground, in contrast to the Kestrel 4500, which measured at about 1.5m above the ground, the same height as the CPC 3007 PNC sampler.

Figure 2.5 Panel A: An example of a wind rose from a sampling site, used to make wind rose shaped buffers. **Panel B:** The windrose shapefile, used to extract values from geographic predictor variables. The blue, full extent of the wedges represents a 100% (maximum length) buffer while the yellow smaller wedges is a 50% buffer.

2.3.1.3 Sources of geographic predictor variables

Road length variables were obtained from the CanMap 2010 Streetfile (Desktop Mapping Technologies (DMTI) Spatial Inc). Highways (RD1) were defined as carto classifications 1 and 2 (freeways and primary highways, respectively) while major roads (RD2) were carto classifications 3 and 4, (secondary highways and major roads, respectively). (72) The resolution of all road length variables was 10m. Intersections of highways with major roads or other highways were determined from the same data source using the script “Calculate Fnode Tnode 2.0” in ArcGIS 9.2 to generate nodes. Each node had a valence assigned to it and those with a

value ≥ 3 were defined as intersections. The point density tool was used to sum the number of intersections in a given buffer.

Truck route data was compiled manually by Dr. Sarah Henderson (54), using route data collected from municipalities and the CanMap 2003 Streetfile (DMTI Spatial Inc). Truck routes from the municipalities of Vancouver, West Vancouver, Burnaby, New Westminster, Coquitlam, Port Coquitlam, Surrey, Langley, Delta and White Rock were obtained from open source data on city websites or contact with planning departments. Other cities advise trucks to follow major arterial routes, RD1 and RD2, including North Vancouver, Richmond, Port Moody, Maple Ridge, Pitt Meadows and Abbotsford.

Traffic density variables were obtained from the Translink (regional transportation authority for Metro Vancouver) EMME/2 traffic model for 2008 provided by Clark Lim and Ken Tseng. EMME/2 is transportation planning software that estimates different types of traffic densities, speeds and times on road segments. Vehicle densities were based on morning peak hour volumes, from 7:30 – 8:30am. Data were received in a polyline file and auto volume was used for vehicle density and light truck and heavy truck volumes combined for truck density. The feature to point tool in ArcGIS was used to convert the file to points at a resolution of 1m and point density tool used to calculate the density of traffic volume points and associated values in the appropriate buffer size.

Bus stop location information was acquired as a point shapefile from Translink Transit GIS data, 2009. Data were accessed from Abascus on the UBC Library website. A search of “BC bus stops” yielded the file.

Land use classifications were based on data from the Metro Vancouver regional office (2006), as there were inconsistencies with land use data classifications between different years

from DMTI Spatial Inc. Data were available at a resolution of 10m. (12,17) Seventeen classifications were aggregated into five categories as follows: Industrial = Industrial-Extractive + Industrial + Transportation, Communication Utilities; Residential = Single family + Rural + Townhouses + Low-rise apartments + High-rise apartments; Open = Agricultural + Harvesting and research + Open and undeveloped + Regional watershed + Recreation and protected natural areas + Lakes and water bodies; Government = Institutional; Commercial = Commercial + Commercial – Residential/Mixed. The Metro Vancouver data did not include Abbotsford (and several sampling sites were located there) so DMTI Spatial Inc. land use data from 2006 was used outside of the Greater Vancouver Regional District.

The normalized difference vegetation index (NDVI), measuring the percentage of vegetation cover, was obtained from NASA Global Inventory Modeling and Mapping Studies (NASA GIMMS) group (2002). Dr. Jason Su provided the layer at a 25m resolution.

Population and dwelling density information was taken from Statistics Canada, Census of Canada using dissemination area level (2006). The population or number of dwellings was assigned to the centroid of its polygon and the calculate density (kernel) feature in ArcGIS ran at the buffers listed in Table 2.2. The resolution of these files was 5m.

Location variables were derived from their respective variable category (i.e. distance to nearest highway was calculated from the CanMap 2010 Streetfile). Port, airport and shoreline GIS data also came from CanMap 2010 Streetfile. (58)

Preparation of a restaurant location shapefile was completed manually. First, fast food chain restaurant locations that conduct cooking, grilling or deep-frying were located and their coordinates obtained using Google Earth. The selection of these restaurants was based on the assumption that these cooking methods could produce significant UFP emissions. These chains

included A&W, Burger King, Church's Chicken, Dairy Queen, Kentucky Fried Chicken, McDonald's and Wendy's. Next, we searched the study area in Google Earth for restaurants that contained the word "grill" "Indian" or "Chinese." For each potential restaurant that was identified, we assessed the name and description to determine if it was actively cooking food and included only those that were in the fast food restaurant database. For example, a location called "Benny's Grill" would be included in the final restaurant list whereas "Golden Dragon Cooking Supplies" would not. This approach assumed that Indian and Chinese restaurants (which are common in the study areas) employed cooking methods including a substantial amount of frying, which may be a source of UFP emissions. These point locations (a list of X,Y coordinates) were then imported into ArcGIS, converted to a point shapefile and used in further geographic analysis.

In addition to the geographic predictor variables considered for LUR modeling outlined in Table 2.2, additional variables combining traffic density and road length measures (such as $[\text{traffic density} \times 1/(\text{distance_to_highway})]$ or $[\text{traffic_density} \times 1/(\text{distance_to_highway})^2]$) were explored, as the LUR model developed for Amsterdam used similar compound variables. (44) These variables did not increase the R^2 of the models in preliminary analyses and were excluded from further model development in favour of road length, traffic density and distance to features used in the LUR models developed previously in Vancouver (56,61), which also included ln-distances to features.

The measured mean temperature during sampling was a variable offered to the road length, measured wind rose buffers (PNC₅₀) model (model 9 in Table 3.9). Temperature was included given its potential to influence PNC levels. After preliminary analysis, temperature was not found to be a significant predictor of PNC.

2.3. Best model

Four additional variables were created, based on counts and field notes collected during the UFP sampling campaign, to add to a ‘best model’ that incorporated site-specific field observations (Appendix E). This specific information, which may help describe determinants of PNC, would not be useful for estimating concentrations at unmonitored locations. The four additional variables were manual heavy-duty vehicle counts, wind direction, fast food sites and construction activities, detailed in section 2.1.2.4.

Table 2.2 Potential variables for UFP LUR modeling (adapted from (56))

Variable category (N)	Description	Sub-category	Circular buffer radii (m)	Wind rose buffer radii (portion of max. length)	Source
Road length (27)	Total length (km) of 3 road types	RD1 (highways) RD2 (major roadways) TRK (truck routes)	25, 50, 75, 100, 200, 300, 500, 750, 1000	0.125, 0.25, 0.5, 0.75, 1.0	CanMap Streetfile, DMTI Spatial Inc. (2010)
Vehicle density (18)	Density (vehicles/ha) of 2 types during morning peak	AD (automobiles) TD (trucks)	25, 50, 75, 100, 200, 300, 500, 750, 1000	Not used	Translink EMME Model (2008)
Bus stop density (15)	Density of bus stops	BUS	25, 50, 75, 100, 200, 300, 500, 750, 1000	0.125, 0.25, 0.5, 0.75, 1.0	Translink Transit GIS data, 2009
Intersection density (15)	Density of intersections of highways or major roads	INTER	25, 50, 75, 100, 200, 300, 500, 750, 1000	0.125, 0.25, 0.5, 0.75, 1.0	CanMap Streetfile, DMTI Spatial Inc. (2010)
Land use (20)	Total area (ha) of 5 land-use designations	RES (residential) COM (commercial) GOV (government) IND (industrial) OPN (open area)	300, 400, 500, 750	0.125, 0.25, 0.5, 0.75, 1.0	Greater Vancouver Regional District, 2001
Vegetation index (1)	Normalized difference index	VEG	N/A	N/A	NASA Global Inventory Modeling and Mapping Studies (2002)
Population density (6)	Density of population (people/ha)	POP (population)	750, 1000, 1250, 1500, 2000, 2500	Not used	Statistics Canada, Census of Canada (2006).

Table 2.2 Potential variables for UFP LUR modeling (adapted from (56))

Variable category (N)	Description	Sub-category	Circular buffer radii (m)	Wind rose buffer radii (portion of max. length)	Source
Dwelling density (6)	Density of dwellings (units/ha)	DWEL (dwelling)	750, 1000, 1250, 1500, 2000, 2500	Not used	Statistics Canada, Census of Canada (2006).
Location (19)	Variables describing attributes of site location (km from site)	ELEV (elevation) X (latitude) Y (longitude) DIST_HWY DIST_RD DIST_TRK DIST_AIR DIST_SHOR DIST_BUS DIST_INTER DIST_PORT (distance to highway, major road, truck route, airport, shoreline, bus stop, intersection, port)	N/A	N/A	CanMap Streetfile, DMTI Spatial Inc. (2010)
Fast food restaurants/grills (8)	Distance to nearest restaurant (km), restaurant density	FF (density of restaurants in buffer) FFD (distance to nearest restaurant)	50*, 75, 100, 200, 300, 500, 750, 1000	0.125, 0.25, 0.5, 0.75, 1.0	Developed for this project (2011)
Temperature	Measured (°C)			N/A	

*the nearest fast food restaurant/grill to any sites was 40m, so the 25m buffer was not used for this variable class

The highlighted portions of Table 2.2 indicate variables that were developed for UFP LUR modeling and were not considered in earlier (61) Vancouver LUR models of NO_x and PM_{2.5}. Buffers less than 100m are also novel. In total, 135 variables were used during LUR model development with circular buffers. The number of variables offered to the wind rose buffer models was 77.

2.3.2 LUR model building procedure

The LUR model development followed the same procedure as described previously and used in the development of previous Vancouver LUR models. (56,61)

The steps in creating a LUR were:

1. Rank all variables by absolute correlation with measured pollutant.
2. Identify the highest-ranking variable in each group (variables of the same type, but with different buffer radii).
3. Exclude variables highly correlated ($r > 0.6$) with the highest-ranking variable in each category.
4. Enter all remaining variables in a backward stepwise regression.
5. Remove variables that are inconsistent with *a priori* hypotheses (e.g. a negative coefficient for a traffic-related variable, as PNC is expected to increase with increasing length of roads within a buffer).
6. Remove variables that are insignificant ($p \geq 0.05$)
7. Repeat steps 4, 5 and 6 to converge on a final model. Remove any variable that contributes less than 1% to the model R^2 for a concise final model.

The final product is a multiple linear regression model of the form:

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

where α is the constant intercept, β_1 is the coefficient for variable X_1 (e.g. length of major roads in a 100m radius buffer) and e is the random error term.

Potential predictors (Table 2.2) were systematically evaluated by calculating correlations between each variable and the UFP measurement. PNC measurements used as outcome variables during modeling included PNC_{50} , $\ln\text{-PNC}_{50}$, PNC_{10} and PNC_{90} . One value for each site from the one-hour sample recorded at one-minute intervals was used to create the appropriate distribution. This method sought to achieve a final, parsimonious model with minimal collinearity in predictors. All model building was conducted using R software version 2.10.1 (R Development Core Team, Vienna AT, 2009). Code examples are located in Appendix D.

The two distinct types of models were developed: the road length model, which initially included all potential predictors except vehicle density variables, and the vehicle density model, which initially included all potential predictors except road length variables. As the models were created, it was apparent that the models built using road length variables and the median PNC values tended to explain more of the variation in measurements than others. As such, models that were developed subsequently with wind rose shaped buffers used an outcome of PNC_{50} and road length variables (excluding vehicle density variables). Table 2.3 below displays the models that were created and are shown in Chapter 3.

Spreadsheets containing PNC measurements along with extracted geographic variable values at each site, which were used to build the LUR models are contained in the appendices. Appendix F includes data used to build the models based on circular buffers and included variables incorporated into the best model. Appendix G contains data used to build the model

with wind rose shaped buffers, based on measured wind speed and direction. Appendix H contains data used to build the model with wind rose shaped buffers using data from a central monitoring location.

Table 2.3 Summary of UFP LUR models created for Metro Vancouver

Model	Road Length	Vehicle Density
PNC ₅₀ (pt/cm ³)	X	X
ln-PNC ₅₀ (ln pt/cm ³)	X	X
PNC ₁₀ (pt/cm ³)	X	X
PNC ₉₀ (pt/cm ³)	X	X
PNC ₅₀ (pt/cm ³), measured wind rose buffers	X	
PNC ₅₀ (pt/cm ³), central station wind rose buffers	X	
PNC ₅₀ (pt/cm ³), best model	X	

2.3.3 Regression maps

Following model development, the weighted sum tool in ArcGIS was used to create a pollution surface for each model by multiplying variable rasters by variable coefficients and adding a fixed value intercept raster. This yielded a map of the study area with predicted concentrations of the pollutant based on the model to be used for visual identification of potential ‘hotspots’ and other spatial patterns. Maps were truncated at the maximum PNC measured.

2.4 Evaluation of measurements and models

2.4.1 Sixty-minute measurement evaluation

In order to assess the representativeness of 60-minute UFP sample compared to longer-term averages, data from the four fixed sites were evaluated. For each of the four fixed sites, the overall mean (mean, median value of all hours) during the two-week sampling campaign was calculated. Then, 14 random one-hour mean PNC measures from each site were selected and

compared to the overall mean, to evaluate the representativeness of (the equivalent) one-hour sample measurements compared to the overall two-week period mean. One sample was selected for each sampling day (Monday – Friday) during sampling hours (8:00am – 6:00pm), to match the mobile sampling campaign protocol.

All three corrections were applied to the median PNC measurement from each of the 100 samples (80 sites plus 20 repeats), according to the time and day of the sample. The effectiveness of the adjustment factors was determined by comparing the absolute difference of the median PNC values as well as the root mean square error (RMSE) for raw and corrected data for the 20 sites that had repeat measurements taken.

2.4.2 Model evaluation

Initial model performance was assessed on the adjusted R^2 values of the models. Subsequent evaluations of model stability were based on the R^2 and root mean square error (RMSE) from a “leave one out” (LOO) cross-validation approach, in which models were re-generated by systematically leaving out one of the 80 sites and then using the remaining 79 sites to predict the concentration at the site that was left out. (56,61) R code for this evaluation is in Appendix D. The R^2 of the simple linear model between predictions from the LOO models and actual measurements and then the RMSE from these comparisons was calculated. A higher LOO- R^2 value and lower RMSE value (and standard deviation) were associated with more stable models. This LOO evaluation was completed on all eleven LUR models.

Finally, the model with the highest R^2 and lowest RMSE was compared to measured PNC at 20 sites where duplicate measurements were collected on separate sampling days and at the four fixed sites. For these four sites, comparisons were made for the measured PNC levels

both during the specific sampling hours (8:00am to 6:00pm, Monday to Friday) and during the entire study period (twenty-four hours a day, seven days a week). These comparisons were evaluated using RMSE.

Chapter 3: Results

This chapter will outline results of the UFP sampling campaign as well as LUR models and evaluations.

3.1 Sampling outcomes

3.1.1 Instrument correction factor

Two CPC 3007 instruments were used to measure PNC at the 80 locations around Metro Vancouver. Based on an outdoor collocation pretest and a sampling collocation measurement, an instrument correction factor of 1.29 (standard deviation = 0.19) was developed in order to make measurements from the two CPC devices comparable. (Correction factors of the individual collocation examples were 1.42 and 1.16.) Data used to create this correction are located in Appendix A.

3.1.2 Temporal correction factors

Temporal correction factors applied to the raw data to account for variability in time of day and day of sampling were normalized to the overall mean PNC from the hourly measurements across the four fixed sites.

The within-day variation corrections, shown in Table 3.1, range from 0.836 – 1.00, with a mean of 0.93 and standard deviation of 0.051. The trend of hourly adjustment factors throughout the day is shown in Figure 3.1.

Table 3.1 Within-day variation correction factors applied to raw data

Hour of day	Correction factor
08:00	0.95
09:00	0.96
10:00	0.92
11:00	0.96
12:00	0.87
13:00	0.84
14:00	0.90
15:00	0.92
16:00	1.00
17:00	1.00
18:00	0.96

Figure 3.1 Within-day variation correction factors by hour of day

The range of within-day correction factors was quite small, indicating there was little temporal variation in PNC from 8:00am to 6:00pm. From examining Figure 3.1, there appeared

to be a slight increasing trend in adjustment factors at later hours of the day, meaning that measurements made later in the day more closely reflect the overall mean PNC (twenty-four hours per day, seven days per week) and that measurements made earlier in the day are higher, which could be associated with morning rush hour. Figure 3.2 shows the mean hourly PNC values at the four fixed sites.

Figure 3.2 Mean PNC by hour at four fixed sites over entire sampling period

Adjustment factors to correct for the day of sample (between-day variation) are presented in Table 3.2. The range of correction factors applied was 0.613 – 3.82 with a mean of 1.52 and standard deviation of 1.03. The trend of daily adjustment factors throughout the sampling period is shown in Figure 3.2.

Table 3.2 Between-day variation correction factors applied to raw data

Day	Correction factor
Monday, 04/19/10	0.91
Tuesday, 04/20/10	0.68
Wednesday, 04/21/10	0.75
Thursday, 04/22/10	0.61
Friday, 04/23/10	0.96
Monday, 04/26/10	0.91
Tuesday, 04/27/10	1.1
Wednesday, 04/28/10	1.3
Thursday, 04/29/10	1.4
Friday, 04/30/10	0.90
Monday, 05/03/10	1.9
Tuesday, 05/04/10	2.9
Wednesday, 05/05/10	3.2
Thursday, 05/06/10	3.8

Figure 3.3 Between-day variation correction factors by day of sampling

From examining Figure 3.3, there is an upward trend of daily correction factors as sampling progressed. The last three days in particular (adjustments 2.9 – 3.8) have markedly higher correction factors than the first eleven days, which increases the slope of the line. Separate between-day correction factors were calculated based on only the two sites that were operated during the last three days of the campaign (the other two instruments had to be returned from loan and were no longer used). These new correction factors displayed a similar pattern as above, with correction factors on the last three days 50% higher than the campaign mean correction factor (overall mean of 1.22 versus last three days mean of 1.84), while the correction factors based on four sites were about 100% higher the last three days (overall mean of 1.52 versus last three days mean of 3.3). Thus, the trend of increasing correction factors on the last three days of the campaign (analogous to lower concentrations) appears to reflect atmospheric conditions and is not simply due to missing measurements.

Figure 3.4 shows the trend in mean PNC value by day at the four sites over the sampling period. It is clear that there is a general downward trend in PNC toward the end of sampling, which contributes to the elevated correction factors during the last three days of the campaign. Unique samples were collected at only four sites during the last three days of the campaign (the rest of the measurements on these days were repeats), so the number of sites included in LUR modeling that had large correction factors applied to them were minimal.

Figure 3.4 Mean PNC by day at four fixed sites over entire sampling period

All three corrections (instrument, within-day, between-day) were applied to the median PNC measurement from each of the 80 samples. Hereafter, all PNC measurements described are corrected, unless noted otherwise. For example, site 48, measured on April 19, 2010 from 9:40am to 10:39am had a raw median PNC count from the sample of 55 139 pt/cm³. An instrument factor of 1.29, a within-day adjustment of 0.92 (Table 3.1) and a between-day adjustment of 0.91 (Table 3.2) were applied to generate an adjusted median PNC concentration of $55139 \times (1.29) \times (0.92) \times (0.91) = 58\,871 \text{ pt/cm}^3$.

The root mean square error (RMSE) of the absolute differences between repeat measurements at 20 sites, before and after application of adjustment factors were 10 800 and 13 700 pt/cm³, respectively, an increase of 27%. However, the absolute median difference from uncorrected to corrected measurements decreased 11% (from 7500 to 6700 pt/cm³). Based on

these evaluations, it is unclear whether the corrections reduced temporal variability, but does suggest that factors other than time account for differences in PNC at the same locations. Notably, the corrections applied to the raw measurements were mostly relatively small, with the range of the three correction types ranging from 0.61 to 3.8.

3.1.3 Summary statistics

Histograms showing the distribution of adjusted measurements from the 80 sites sampled are displayed in Figure 3.5.

Figure 3.5 Histograms of corrected PNC distributions (pt/cm^3). A: PNC₅₀. B: Ln-PNC₅₀. C: PNC₁₀. D: PNC₉₀.

The distribution of PNC₅₀ appears lognormal in panel A, while the ln-transformed PNC₅₀ dataset in panel B appears to be approximately normally distributed. PNC₁₀ and PNC₉₀ in panel C and D both are skewed to the right, appearing lognormally distributed. Despite ln-PNC₅₀ being the mostly clearly normally distributed dataset, all were used in LUR modeling as it became

evident that using PNC_{50} as the outcome variable yielded models with systematically higher R^2 values than using $\ln-PNC_{50}$, (Table 3.9).

Summary statistics from the 80 sites sampled are reported as the median, 10th and 90th percentile of the 60 points recorded in a 60-minute sample, shown in Table 3.3.

Table 3.3 Summary statistics for corrected one-hour PNC at 80 sites: median, 10th and 90th percentile

Statistic	PNC₅₀ (pt/cm³)	Ln-PNC₅₀ (ln-pt/cm³)	PNC₁₀ (pt/cm³)	PNC₉₀ (pt/cm³)
Minimum	1500	7.30	1100	2300
10 th percentile	4200	8.47	3900	3700
First quartile	7900	8.97	6100	11 900
Mean	18 200	9.50	12 600	30 200
Median	13 600	9.54	10 000	20 447
Third quartile	21 900	9.99	15 400	41 500
90 th percentile	30 400	10.53	25 400	61 600
Maximum	105 000	11.6	72 300	169 000
Standard deviation	15 900	0.78	10 100	28 100

3.1.4 Comparing PNC with nitrogen oxides

Measurements of NO_x are compared with PNC in Table 3.4 and Figure 3.6.

Table 3.4 Correlation coefficients between PNC and nitrogen oxides

Correlated pollutants	Pearson correlation coefficient	Spearman correlation coefficient
PNC ₅₀ and NO	0.65	0.56
PNC ₅₀ and NO ₂	0.64	0.57
PNC ₅₀ and NO _x	0.70	0.58
ln-PNC ₅₀ and ln-NO	0.63	0.56
ln-PNC ₅₀ and ln-NO ₂	0.61	0.57
ln-PNC ₅₀ and ln-NO _x	0.60	0.58

Figure 3.6 Residual plot of PNC_{50} (pt/cm^3) compared with average NO (ppb) from the same locations at the same sampling time

PNC and NO_x were highly correlated ($r = 0.61 - 0.70$), suggesting the pollutants have some of the same sources. The residual plots between regression of PNC_{50} with NO_2 and NO_x displayed similar patterns to that of NO (Figure 3.6). Residuals were evenly distributed around zero, with three outlier locations evident (shown in red). Table 3.5 outlines some known geographic features of these sites that might explain some source differences between PNC and NO_x at those locations. Notably, all three sites had very high truck counts compared with other sites and were on or near truck routes. The site with the highest residual value, E7, was also very close to the port. This indicates that heavy-duty truck traffic and possibly port proximity may be more specific predictors of PNC than they are of NO_x . However, site 35 had a negative residual

and high truck counts, so there be other unaccounted for factors explaining the difference between PNC and NO_x.

Table 3.5 Examination of three sites with high residual values when comparing NO_x and UFP levels

Site	Site description	Residual (+/-)	Site features
E7	Clark Drive and Pender Avenue, Vancouver	+	-high truck count (378 in one hour) -on truck route -bus stop located nearby (within 20m) -high area of industrial land -fast food restaurant outlet nearby -very close to port (30m)
48	Knight Street and East 36 Avenue, Vancouver	+	-high truck count (366 in one hour) -on truck route
35	Copley Street and East 15 th Avenue, Vancouver	-	-high truck count (292 in one hour) -close to truck route -close to bus stop -close to intersection

3.1.5 Meteorologic data

Table 3.6 includes summary statistics for measured meteorology variables at the 80 sites sampled and those measured at the central Metro Vancouver (T18, Burnaby South) monitoring station data during the same time period.

Table 3.6 Summary statistics for site-specific and central station measured meteorology variables

Statistic	Wind speed (m/s)		Temperature (°C)	
	80 sites	Central (T18)	80 sites	Central (T18)
Minimum	0	1.1	8	5.3
First quartile	0.5	2.1	11.1	9.3
Mean	1.8	3.0	12.9	11.3
Median	1.2	2.9	12.3	10.8
Third quartile	2.2	3.7	14.2	12.5
Maximum	16.3	6.6	22.7	21.6
Standard deviation	2.3	1.2	2.8	3.0

The wind speed measured at the Burnaby South monitoring station was systematically higher (the mean was 1.7 and median 2.4 times greater) than the measured wind speeds from the 80 sampled sites. The central monitoring station had consistently lower temperatures than the 80 sites. These differences in meteorologic variable trends can likely be attributed to the height of the monitors. The central monitor, at 10m above ground, captured a more generalized pattern in the region whereas the Kestrel 4500® at each sampling location made recordings from 1.5m, the height the CPC 3007 devices were measuring (and approximate human breathing height).

The range of wedge buffer lengths used in the wind rose buffer methodology models is presented in Table 3.7.

Table 3.7 Range of wedge buffer lengths applied in measured meteorologic LUR model

Buffer size (portion of maximum wind rose length)	Minimum (m)	Maximum (m)	Mean (m)	Median (m)	Standard deviation (m)
0.125	113	3767	776	536	772
0.25	225	7534	1553	1073	1545
0.50	450	15067	3106	2145	3090
0.75	675	22601	4659	3218	4635
1.0	900	30134	6211	4290	6179

Notably, the total range of buffer sizes shown in Table 3.7 is from 113 – 30 134m, which represents a much larger maximum range than used in circular buffer variables (where 2500 m was the largest buffer size).

3.1.6 Univariate correlations with PNC₅₀ and geographic variables

Selected univariate correlations between geographic predictor variables and PNC₅₀ are shown in Table 3.8.

Table 3.8 Selected univariate correlations with PNC₅₀

Variable	Pearson correlation coefficient
ln-Distance to nearest intersection	-0.049
Mean temperature	0.03
Mean wind speed	0.12
Length of highways within 1000m	0.12
Population within 2500m	0.24
Number of intersections within 25m	0.27
Number of fast food sites within 200m	0.36
Area of industrial land within 750m	0.40
Number of bus stops within 100m	0.45
Length of major roads within 100m	0.49
Truck density within 25m	0.58
Automobile density within 25m	0.60
Truck density within 50m	0.63

Univariate correlations between predictor variables and PNC₅₀ measurements showed that the highest overall correlation was for truck routes within 50m ($r = 0.63$), followed by density of automobiles within 25m and the density of trucks within 25m. The highest distance-related correlation (ln-distance to the nearest intersection of major roads, $r = -0.049$) was quite low suggesting that simple road proximity measures are not very useful in the prediction of UFP spatial variability. The highest land use correlation was for industrial land within 750m at $r = 0.40$. With fast food and grill restaurants in a 250 m buffer, the correlation with PNC₅₀ was $r = 0.36$.

Mean sample temperature and wind speed, measured at the time of sampling, were weakly positively correlated with PNC₅₀ suggesting that these factors were not very important in predicting PNC.

3.2 PNC LUR models

A total of eleven LUR models for UFP were developed and are summarized in Table 3.9. Models described as “road length” had all qualifying variables initially offered into the model, except the vehicle density variables. Models described as “vehicle density” had all qualifying variables initially offered into the model, except the road length variables. The wind rose buffer based models (model 9 and 10 below) were only based on road length variables, as preliminary models had shown higher R^2 values for road length versus vehicle density models. A model for PNC₅₀ was also created using all possible variables (road length and vehicle density) and this final model was identical to the road length model.

Road length models consistently yielded higher R^2 values than their vehicle density counterparts by 0.03 to 0.10. PNC₅₀ models predicted 5 – 8% more variability than ln-PNC₅₀ models. PNC₉₀ models for both road length and vehicle density explained 1 – 5% more variability than PNC₅₀ models for the same model type, which in turn explained 9% more variability than the PNC₁₀ models. Models with wind rose shaped buffers had R^2 values 0.14 and 0.19 lower than the PNC₅₀ road length model.

The models with the highest R^2 , 0.53, were a road length model of PNC₉₀ and a model of PNC₅₀ including site observations. The model with the highest R^2 for median PNC levels excluding site observations was the road length model ($R^2 = 0.48$).

Table 3.9 LUR models developed for PNC in Metro Vancouver

Model type	Variables	Coefficients (SE) (pt/cm³ or ln-pt/cm³)	P values	Partial R² Model R²	LOO R² RMSE (SD)
1. Road length, PNC ₅₀	<ul style="list-style-type: none"> Truck route length (50m buffer) Distance to port (ln km) Fast food site density (200m buffer) Intercept 	150 800 (25 200) -3000 (900) 3700 (1700) 17 700 (2500)	<0.001 0.002 0.03	0.36 0.10 0.05 Model: 0.48	0.32 12 800 (23 100)
2. Vehicle density, PNC ₅₀	<ul style="list-style-type: none"> Truck density (25m buffer) Distance to port (ln km) Intercept 	88 (14) -3600 (990) 20 600 (2500)	<0.001 <0.001	0.32 0.11 Model: 0.42	0.29 13 100 (22 900)
3. Road length, ln-PNC ₅₀	<ul style="list-style-type: none"> Truck route length (50m buffer) Distance to fast food site (ln km) Distance to port (km) Latitude Area government land (300m buffer) Intercept 	6.6 (1.3) -0.29 (0.088) -0.041 (0.014) 3.4 x 10 ⁻⁵ (1.3 x 10 ⁻⁵) -0.085 -7.3 (6.1)	<0.001 <0.001 0.005 0.009 0.03	0.19 0.08 0.06 0.05 0.03 Model: 0.40	0.30 0.64 (0.78)
4. Vehicle density, ln-PNC ₅₀	<ul style="list-style-type: none"> Distance to bus stop (ln km) Truck density (25m buffer) Distance to intersection (ln km) Intercept 	-0.16 (0.060) 2.2 x 10 ⁻³ (8.6 x 10 ⁻⁴) -0.14 (0.058) 8.7 (0.16)	0.004 0.009 0.02	0.16 0.13 0.11 Model: 0.37	0.32 0.63 (0.78)
5. Road length, PNC ₉₀	<ul style="list-style-type: none"> Truck route length (50m buffer) Distance to port (ln km) Fast food site density (100m buffer) Intercept 	292 500 (40900) -4600 (1600) 17 700 (6600) 27 400 (4100)	<0.001 0.004 0.009	0.43 0.07 0.06 Model: 0.53	0.41 20 800 (34 700)
6. Vehicle density, PNC ₉₀	<ul style="list-style-type: none"> Vehicle density (25m buffer) Distance to port (ln km) Intercept 	8.0 (1) -5700 (1700) 30 800 (4400)	<0.001 0.001	0.37 0.08 Model: 0.43	0.31 22 400 (38 800)

Table 3.9 LUR models developed for PNC in Metro Vancouver

Model type	Variables	Coefficients (SE) (pt/cm ³ or ln-pt/cm ³)	P values	Partial R ² Model R ²	LOO R ² RMSE (SD)
7. Road length, PNC ₁₀	<ul style="list-style-type: none"> Truck route length (50m buffer) Distance to port (ln km) Intercept 	97 700 (17700) -2000 (700) 13 600 (1809)	<0.001 0.005	0.31 0.08 Model: 0.37	0.22 9100 (15600)
8. Vehicle density, PNC ₁₀	<ul style="list-style-type: none"> Truck density (25m buffer) Distance to port (ln km) Intercept 	57 (10) -2400 (700) 14 900 (1800)	<0.001 0.001	0.24 0.11 Model: 0.33	0.19 9300 (16600)
9. Road length, measured wind rose buffers (PNC ₅₀)	<ul style="list-style-type: none"> Distance to intersection (ln km) Distance to port (ln km) Intercept 	-4900 (1000) -3900 (1100) 14 300 (3600)	<0.001 <0.001	0.23 0.13 Model: 0.34	0.20 14 100 (24300)
10. Road length, central site wind rose buffers (PNC ₅₀)	<ul style="list-style-type: none"> Distance to port (ln km) Area residential land (100% buffer) Distance to major road (km) Distance to highway (ln km) Intercept 	-3400 (1100) -0.49 (0.19) -16 200 (5000) -2500 (1100) 36 300 (42300)	0.003 0.01 0.002 0.02	0.12 0.10 0.09 0.09 Model: 0.29	0.14 14 700 (24900)
11. Best model (PNC ₅₀)	<ul style="list-style-type: none"> Truck route, 50m Distance to port (ln km) Truck count Fast food site density, 200m Intercept 	103 100 (28100) -3100 (900) 38 (12) 3400 (1600) 16 800 (2400)	<0.001 <0.001 0.002 0.04	0.20 0.19 0.14 0.05 Model: 0.53	0.38 12 300 (21300)

SE = standard error; SD = standard deviation

Model R² are adjusted R² values

LOO = Leave one out cross validation, R² between predicted and measured values

RMSE = root mean square error

3.2.1 Sensitivity analyses

Additional univariate correlations between distance to the nearest fast food restaurant, fast food restaurant density within 200m, distance to the nearest port and traffic-related variables were also conducted, after the first variables appeared in a number of final models. Table 3.10 shows the results.

Table 3.10 Selected univariate correlations between fast food and port variables and traffic-related variables

Variable 1	Variable 2	Pearson correlation coefficient
Distance to nearest restaurant (km)	Distance to nearest intersection (km)	0.37
Number of restaurants in 200m	Length truck route within 50m	0.28
Number of restaurants in 200m	Length of major roads within 100m	0.39
Number of restaurants in 200m	Number of bus stops in 100m	0.55
Distance to port (km)	Length truck route within 50m	-0.19
Distance to port (km)	Density of trucks within 50m	-0.11
Distance to port (km)	Length of major roads within 100m	0.27
Distance to port (km)	Distance to nearest bus stop (km)	0.68

There is a strong correlation ($r = 0.68$) between distance to the port and distance to the nearest bus stop and a fairly strong correlation ($r = 0.55$) between number of restaurants within 200m and number of bus stops within 100m. All other comparisons show weak to moderate correlations, suggesting that restaurant and port variables independently contribute to observed UFP levels, or that other factors not considered are responsible for the PNC results obtained.

As a sensitivity analysis, a model was built with 79 sites, excluding the site with the highest median PNC (as it was an outlier). In this model, variables were the same as in the PNC₅₀ road length model and coefficients were comparable: length of truck routes in 50m (coefficient 105 000 pt/cm³), ln-distance to nearest port (coefficient -900 pt/cm³) and density of fast food restaurants in 200m (coefficient 5400 pt/cm³). The model R^2 was 0.45 and LOO- R^2 was 0.34.

The somewhat lower model R^2 and higher LOO- R^2 indicated that some of the instability of the road length PNC₅₀ model was due to this one site.

3.2.2 Pollution surface images

Maps representing the road length and vehicle density LUR models for PNC₅₀, and those for PNC₁₀ and PNC₉₀ are presented in Figures 3.7 – 3.10.

The vehicle density map overall predicted higher PNC than the road length map (Figure 3.8 versus 3.7) as the colour is darker in Figure 3.8 and there is not a gap of medium red over Delta (the region that appears as an island on the left side of the maps about two-thirds of the way down) as there is in Figure 3.7. It is evident that PNC was heterogeneous in this region. The road length model created a more heterogeneous surface than the vehicle density model, as the truck routes and nodes of restaurants were more evidently highlighted. The vehicle density model, with the density of trucks in a 25m buffer as a significant variable, appeared relatively homogeneously dispersed as the density of trucks is not restricted to official truck routes.

The PNC₁₀ model might best represent sites with background levels of UFP while the PNC₉₀ model could characterize spikes in concentrations or hotspots. As the two surfaces look similar and the models only differed by fast food locations within 100m variable (in the PNC₉₀ model), it seems plausible that fast food outlets may be an important source of peak UFP levels.

In all maps, the predicted level of PNC declined from west to east (generally as the distance to the nearest port increased) and away from the area more densely populated and developed. The truck route pattern was visible in all surfaces (three models have a truck route variable in them and the remaining model has density of trucks in 25m in it).

Figure 3.7 Pollution surface for road length, PNC₅₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km), density of fast food restaurants within 200m.

Figure 3.8 Pollution surface for vehicle density, PNC₅₀ LUR model. Variables in model: density of trucks within 25m, distance to port (ln km)

Figure 3.9 Pollution surface for road length, PNC₁₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km). Note: different scale than previous figures.

Figure 3.10 Pollution surface for road length, PNC₉₀ LUR model. Variables in model: length of truck route within 50m, distance to port (ln km), density of fast food restaurants within 200m. Note: different scale than previous figures.

3.3 Evaluation

3.3.1 Assessment of 60-minute samples

To evaluate the representativeness of 60-minute samples on longer-term PNC, a random selection of 14, 60-minute PNC₅₀ measurements (one from each day of the sampling campaign, during sampling hours) from each of the four fixed sites was plotted along with a line showing the overall mean value for that site during the entire two-week sampling campaign. Plots for the four sites are shown in Figure 3.11.

A

B

C

D

Figure 3.11 Random selection of 60-minute PNC measurements compared with two-week overall PNC at the same location. A, B, C, D: four fixed site locations.

The 14 random points are somewhat evenly distributed around the mean, with a tendency to over-estimate the mean (in panel D, this obvious pattern is enhanced by two outliers). This suggests that a 60-minute sample can characterize long-term averages of PNC (two-week averages in this case) reasonably well, though it may overestimate overall means. It should be noted that the 14 samples from each location were selected from sampling hours (8:00am to 6:00pm) while the overall mean value was calculated over 24 hours. As PNC is lower overnight than during the day (as observed in instrument logs) the positively skewed plots in Figure 3.11 are to be expected.

3.3.2 Comparing spring 2010 with fall 2009 measurements

The Pearson correlation between adjusted PNC_{50} measurements at 22 sampling locations in fall 2009 and spring 2010 was 0.58.

The road length (PNC_{50}) model developed with spring PNC measurements was used to predict the unadjusted and adjusted measurements collected in the previous fall campaign. Adjustment factors are displayed in Table 3.11 and an evaluation of the predicted versus measured concentrations is in Table 3.12.

Table 3.11 Correction factors for comparing fall 2009 to spring 2010 PNC measurements

Description	Correction factor
P-trak 8525 to CPC 3007 values	1.51
Fall to spring seasonal adjustment (based on mean PNC_{50} measures)	0.596
Spring to fall adjustment	1.68
Fall to spring seasonal adjustment (based on mean NO measures)	0.264
Spring to fall adjustment	3.79

Table 3.12 Predicting PNC at fall sites based on road length (PNC₅₀) LUR model from spring sampling

Method of comparison	RMSE (pt/cm ³)	R ² (adjusted)
Uncorrected	23 500 (31500)	0.29
Temporally corrected (hour of day, season) by UFP	12 100 (14000)	0.28
Temporally corrected (hour of day, season) by NO	13 900 (15800)	0.26

The R² of the seasonal validations were slightly lower than the LOO method (0.26 – 0.29 versus 0.32). The RMSE was approximately halved when fall measurements were temporally adjusted. RMSE values for adjusted fall data were large at over 12 000 pt/cm³, representing 72-82% of the mean (fall PNC corrected mean: 16 900 pt/cm³). Temporal adjustment by UFP measurements versus by NO measurements did not greatly impact the RMSE of the predicted PNC, as those using the other pollutant to adjust had only 14% greater error.

3.3.3 Internal validation

The LOO cross-validation on the eleven LUR models yielded R² between predicted and measured values from 0.14 and 0.41 (central site wind rose buffers for PNC₅₀ and road length for PNC₉₀ models respectively) and RMSE between 9100 and 22 400 (road length for PNC₁₀ and vehicle density for PNC₉₀ models). This is shown in Table 3.9.

LOO values shared a similar pattern to model R², with decreases from the model R² ranging from 14% (ln-PNC₅₀ by vehicle density, model 4) to 52% (PNC₅₀, road length model based on central wind-rose shaped buffers, model 10). The PNC₉₀ road length model had a higher LOO R² than the PNC₅₀ model with site observations. RMSE values were high, ranging from 9300 to 22 400 pt/cm³. Considering mean values were 12 600, 18 200 and 30 200 pt/cm³ for PNC₁₀, PNC₅₀ and PNC₉₀, RMSE values were 72 – 74% of the mean for the relevant distribution. RMSE values were generally higher for vehicle density models, PNC₉₀ models (and

lower than PNC_{50} for the PNC_{10} models) and wind-rose shaped buffer models compared with the PNC_{50} road length model.

3.3.4 External validation

Using the road length PNC_{50} model to predict replicate measurements at 20 locations, (i.e. measurements not included in building the model, but at sites that were included) yielded a larger validation R^2 (0.63) and smaller RMSE (10 600 pt/cm^3) than the LOO methods, shown in Table 3.13. The model explained more variability in these repeat measures and they were subject to less error than the LOO approach. This is logical as these same 20 sites were included in the model building (but with different measurements taken at a different time). This outcome increased confidence in the ability of the model (PNC_{50} , road length) to predict the spatial pattern of UFP, however, the absolute error (RMSE 58% of the mean) is still quite large.

Similarly, the PNC_{50} road length model was also used to predict PNC at the four fixed sites, both during sampling hours (8:00am to 6:00pm, Monday to Friday) and during all hours during the two-week sampling period. Results are in Table 3.13. The RMSE values increased to about 15 000 pt/cm^3 , which is larger than the LOO results and represents about 82% of the mean. This was a high error level, suggesting predictions at these sites were not very accurate. Interestingly, the RMSE for the four sites was nearly the same for sampling hours versus all hours, indicating that temporal variation is not that important when predicting PNC. The R^2 was not calculated, as it is not a valid statistic for four sites.

Table 3.13 Predicting PNC levels at repeated and fixed sites by road length (PNC₅₀) model

	RMSE (SD) (pt/cm³)	R² (adjusted)
Twenty repeat sites	10 600 (13 400)	0.63
Four fixed sites: sampling hours	15 000 (16 600)	N/A
Four fixed sites: all hours	15 200 (16 400)	N/A

Overall, these external validation results increase confidence in the predictability of the PNC₅₀ road length model, but highlight its limitations due to large RMSE values.

Chapter 4: Discussion

This chapter compares the Vancouver PNC measurements and LUR models with other LUR models to contextualize this work. The key findings of this study, as well as strengths, limitations, applications, future work and success of the modeling process are described.

4.1 PNC measurements

Variation in PNC was dictated more by spatial factors than temporal factors. Fairly small temporal correction factors (most between 0.6 and 1.5) were applied to raw measurement data. Further, a LUR model developed from spring measurements was able to predict measurements of PNC during the fall nearly as well as in the spring, with similar validation R^2 values (0.32 versus 0.28 – 0.29) and comparable RMSE values (12 800 versus 12 100 – 13 900 pt/cm^3). Temporal adjustments for the hour of the day and season reduced the RMSE markedly (from 23 500 to 12 100 and 13 900 pt/cm^3). Interestingly, the R^2 and RMSE values for models with temporal adjustment based on PNC measurements and models adjusted for temporal trends using NO measurements from the Metro Vancouver monitoring sites were quite similar (0.28 and 12 000 pt/cm^3 versus 0.26 and 13 900 pt/cm^3). This suggests that NO (or likely NO₂ or NO_x), a routinely monitored, related pollutant in the TRAP mixture, could be used to seasonally adjust PNC data. This could be useful in an epidemiologic application of the UFP LUR model to times of the year when continuous measurements of PNC are not available.

Conversely, spatial distribution of PNC varied widely across the region, with a seventy-fold increase from minimum to maximum site median measurements observed (1500 – 105 000

pt/cm³). There was a three- and seven-fold increase from 25th to 75th percentile and 10th to 90th percentile respectively.

The last three days of between-day temporal correction factors (2.9 – 3.8) were about twice as large as the corrections applied (0.61 – 1.9) in earlier days of sampling. This increase could be caused by the lack of measurements from two fixed sites with higher PNC which were stopped a few days earlier as instruments had to be returned. The relatively large daily correction factors in the last three days of sampling compared to the other days might have biased the results by increasing the median values for measurements collected on those days. This would be a positive bias, that could have influenced models to over-predict PNC levels. However, as a number of the sites visited during the last days of sampling were repeat sites (and were not included in model development, only evaluation) the potential impact of this is likely to be minimal. Only four sites sampled in the last three days were first-time samples, with measurements used to make the LUR models. Additionally, review of the trend of PNC over the last three days of sampling showed a real decrease in concentrations, which is further supported by correction factors developed from the two sites that were operated during the last three days showing similar patterns as when all four sites were included in calculation of correction factors. Between-day corrections were greater than within-day corrections (ranges), with the former similar in magnitude to spatial variability. This trend could be due to varying activity patterns by day of the week. It is unlikely that temperature, which varies more within a day than between days usually and was not an important predictor of PNC, can predict this difference.

Examination of Figure 3.11 shows that the 14 randomly selected one-hour samples at three sites are approximately evenly distributed around the individual site mean during the two weeks of continuous measurements, while site four's samples are skewed above the mean. From

this, one-hour measurements appear to reflect at least two-week periods reasonably well. Further, a correlation of 0.58 between one-hour measurements in fall 2009 and spring 2010 also suggests that one-hour samples can capture a fair amount of long-term spatial variability in PNC.

4.2 LUR models

Eleven LUR models for UFP in Vancouver were produced. The amount of variability in measured PNC explained by the models (R^2) ranged from 0.29 – 0.53. The most representative model from this suite was the road length (PNC₅₀) model, with an R^2 of 0.48, as it had the highest R^2 for a model predicting a median or ln-median PNC value and lowest RMSE. The LOO cross-validation R^2 was 0.32 and RMSE was 12 800 pt/cm³, representing 70% of the mean.

The LOO cross-validation approach to assess the R^2 and RMSE of predictions versus measurements yielded a range of 0.14 – 0.41 and 9100 – 22 400 (pt/cm³), which suggest modest performance. The R^2 between predictions and measurements using the road length (PNC₅₀) model and 20 repeated sites was 0.63, indicating the model performed quite well in predicting PNC at those locations. The RMSE at repeated sites was slightly lower than for the model by the LOO method, at 10 600 (pt/cm³).

Using the road length (PNC₅₀) model to predict PNC at the four fixed sites yielded RMSE values around 15 000 pt/cm³, quite high considering fixed site mean values were 8000 – 14 000 pt/cm³. However, there was little difference between the RMSE of the sampling hours and that of all hours of the two-week period (15 000 versus 15 200 pt/cm³), suggesting that this model can predict PNC mean levels over a longer term (at least two-week periods).

Creation of regression maps for some of the PNC models (Figures 3.7 – 3.10) showed that the road length models typically predicted lower levels of PNC than vehicle density models,

but had prediction values more heterogeneously distributed. In all surfaces, PNC tended to decline from west to east, suggesting that some factors in the western part of the region (ports, increased population density and associated motor vehicle activity) accounted for increased levels of PNC.

4.2.1 Comparison of Vancouver PNC LUR models with other LUR models

Table 4.1 lists a selection of previously published LUR models with which to compare the Vancouver PNC LUR model results. It is important to note that this listing of models is not comprehensive; a review on LUR models in 2008 (55) identified 25 published models and there are even more available now. These comparisons focus on models that examined the same pollutant (PNC) (44), were located in the same region (Vancouver) (61), or provided examples of the greater North American and world context. A discussion comparing the model in Table 4.1 to the PNC models in Vancouver follows.

Table 4.1 Summary of LUR models published by other researchers for comparison with the Vancouver PNC LUR models

Location (N)	Investigator (year)	Pollutant	Variables in LUR models	LUR R² (validation R²)	RMSE (SD)	Mean or median measurements (SD)
Vancouver (80)	Abernethy <i>et al.</i> (2012)	PNC	-TRK.50, LN-DIST.PORT, FFD.200 -removal of site observations	0.53 (0.38) 0.48 (0.32)	12800 (23100)	13600 pt/cm ³ ^a
Vancouver (114)	Henderson <i>et al.</i> (2007) (61)	logNO	-RD.100, RD.1000, RD2.100, POP.2500, ELEV, X, Y; AD.100, TD.1000, ELEV, X, Y	0.57-0.62 (0.36-0.65)	1.65-1.9 (12.7-13.1)	31.3 (21.6) ppb ^b
(114)		NO ₂	-RD.100, RD.1000, RD2.100, POP.2500, COM.750, ELEV, X; AD.100, TD.200, TD.1000, COM.750, POP.2500, ELEV	0.56-0.60 (0.31-0.79)	0 (2.61-2.65)	16.2 (5.6) ppb ^b
(25)		PM _{2.5}	-ELEV, COM.300, RES.750, IND.300, AD.100, ELEV, COM.300, RES.750	0.52 (0.14)	0 (1.5-1.53)	4.08 (1.98) µg/m ³ ^b
Vancouver (39)	Larson <i>et al.</i> (2009) (71)	Black carbon	-RD1.500, RD2.750, ELEV, X, RD2.300, TD.300	0.51-0.72 (0.40-0.63)	0.4-5.08	9.72 (4.85) σ _{ap} ^a
Amsterdam (46)	Hoek <i>et al.</i> (2011) (44)	PNC	-traffic intensity x inverse DIST.RD2, DWELL.300, PORT.300 -removal of site observations	0.65 (0.57) 0.44	Not published	22400-40400 pt/cm ³ ^a
		PM _{2.5}	-traffic intensity x inverse DIST.RD2, OPEN.300, AD.100	0.54 (0.50)		22-24 µg/m ³ ^a
Netherlands (40)	Brauer <i>et al.</i> (2003) (74)	PM _{2.5}	-AD.250, address 300m, traffic site, DIST.RD2, region, TRAF.50, TRAF.50-250, street canyon, POP.1000-5000	0.73-0.78 0.81-0.90	1.1 – 1.6 µg/m ³ for PM _{2.5} 0.22 – 0.31 10 ⁻⁵ /m for absorbance	17.5 (µg/m ³) ^b 1.64 (10 ⁻⁵ /m) ^b
Munich (40)		PM _{2.5} Absorbance		0.56-0.76 0.67-0.83		13.6 (µg/m ³) ^b 1.84 (10 ⁻⁵ /m) ^b
Stockholm County (42)		PM _{2.5} Absorbance		0.50-0.63 0.66-0.76		10.3 (µg/m ³) ^b 1.28 (10 ⁻⁵ /m) ^b
Amsterdam (80)	Briggs <i>et al.</i> (1997) (59)	NO ₂	-RD1.50, RD2.50, RD1.200, RD2.500, RD3.50, RD3.200, Built_land.100, DIST.RD1	0.62 (0.79 – 10 sites)	4.45 µg/m ³	36.4-51.8 (6.4-12.2) µg/m ³ ^b

Table 4.1 Summary of LUR models published by other researchers for comparison with the Vancouver PNC LUR models

Location (N)	Investigator (year)	Pollutant	Variables in LUR models	LUR R² (validation R²)	RMSE (SD)	Mean or median measurements (SD)
New York (36-62)	Ross <i>et al.</i> (2007) (75)	PM _{2.5}	-traffic counts 500m, POP.1000, IND.300, OPEN.1000	0.61-0.64	0.77-0.93	14.3-15.3 (1.4-1.8) (µg/m ³) ^b
North Rhine-Westphalia, Germany (40)	Hochadel <i>et al.</i> (2006) (76)	PM _{2.5}	-TD.250-10000, AD.250, DIST.RD2, building count 2500x2500	0.07	2.16	22.4 (µg/m ³) ^b
		PM _{2.5} Absorbance	-TD.100-10000, AD.100, DIST.RD1	0.81	0.144	1.71 (10 ⁻⁵ /m) ^b
Montreal (57)	Gilbert <i>et al.</i> (2005) (77)	NO ₂	-DIST.HWY, traffic counts RD1, RD1.100, RD2.100, RD3.500, OPEN.100, POP.2000	0.54 (0.52)	Not published	11.6 (3.0) ppb ^b
Toronto (94)	Jerrett <i>et al.</i> (2007) (78)	NO ₂	- RD1.200, RD2.50, IND.750, DWEL.2000, X, DWIND.1500, TRAF.500	0.67	0.163 ln ppb (1.2 ppb)	32.2 (9.2) ppb ^b
Los Angeles (23)	Moore <i>et al.</i> (2006) (79)	PM _{2.5}	-TRAF.300, IND.5000, GOV.5000	0.69 (0.63)	Not published	18.4 (6.01) µg/m ³ ^b

^a median

^b mean

* Note: variables presented in this table are denoted using the same coding as variables used in Vancouver model development, shown in Table 2.2 for those that are common. Other variables are written out fully.

Land use regression models are geographically specific, due to different geographic and meteorologic factors in different regions as well as varying emissions and chemical features of pollutants. It is difficult to transfer models developed in one region to another and to compare models created in other parts of the world. (80,81) The transfer of the Vancouver LUR model for NO₂ (61) to Victoria and Seattle, locations geographically close and similar to Vancouver, yielded reduced explanation of variation compared with the original model (R^2 of 0.33 and 0.51, compared with 0.62). (80) Similar results were obtained when interchanging NO and NO₂ models between Edmonton and Winnipeg. (81) As such, it is important to recognize that the exercise of comparing the Vancouver UFP LUR model to other models is more qualitative than quantitative in meaning.

The PNC model developed for Amsterdam is similar in several ways to the models developed in Vancouver. In both cases, relatively few variables (usually about three), including one related to ports, were in the final model. This is strong evidence that something linked to ports is contributing to PNC. The PM distribution emitted by ships is known to be dominated by UFP (36,37) and has been shown to be a likely important source of UFP in a separate study in Amsterdam. (35) The Amsterdam model also included a combined traffic intensity and inverse distance to nearest major road variable, while the principal Vancouver model included length of truck route within 50m. Both cases suggest vehicle traffic as an important source of UFP, while the Vancouver model more specifically suggests heavy-duty diesel engine trucks to be the primary traffic-related source. This finding is consistent with other studies that have shown exponential decreases in PNC with increased distance from heavy-duty vehicles. (12,15,28) The third variable in the Amsterdam model was address density within 300m and in Vancouver it was fast food locations in 200m. This difference could indicate distinct sources of UFP in the

two locations, although the Amsterdam model did not offer restaurant related variables to the models. Notably, the Vancouver models did have population and dwelling density (address density) variables offered to them, but none of the final models contained any of those variables (in contrast with some of the other LUR models shown in Table 4.1). This could be because population or address density type variables are surrogates for other variables, such as restaurant density or traffic density, which may be the actual drivers of PNC levels. Another possible explanation is that Vancouver uses less fossil fuel (electricity is a major heating source) for residential heating than other areas. (33,34) Vancouver also has a relatively mild climate, so lower fossil fuel use due to less heating in winter could be another reason for this observation.

The R^2 for the Amsterdam PNC model including site observations was 0.65 (validation: 0.57), which is analogous to the 0.53 (validation: 0.38) found in Vancouver ‘best’ model (model 11, Table 3.9). In this case the Amsterdam model explained more of the variability in PNC than the Vancouver model and the LOO cross-validation R^2 remained quite a bit higher. However, once site-specific observations were excluded, the Amsterdam model R^2 dropped to 0.44 (no validation R^2 reported), while the main Vancouver model was 0.48 (LOO $R^2 = 0.32$). The main site observations that contributed to the more predictive model in Amsterdam were distance measurements to the edge of the nearest road. The authors suggested the discrepancy could be due to GIS measurements going to the middle of a street, small GIS errors having an amplified effect in a compact city and contributions of mopeds and motorcycles traveling at the edge of the road. In Vancouver, manual truck counts increased the R^2 , providing the most accurate heavy-duty traffic data for a specific site at a specific time. As only a model where variable information is available at any site in the area of interest is applicable in an epidemiologic setting, these two models from different parts of the world seem quite similar in their predictive power.

The median PNC was higher in Amsterdam than Vancouver, 22 400 – 40 400 pt/cm³ versus 18 200 pt/cm³. This range is due to the stratification of traffic and background sites. The Amsterdam model also included fewer sites than Vancouver (46 versus 80), which may partially account for the higher R² of the ‘best model’ as LUR models with fewer sites tend to generate higher R² (they may over-estimate the explanatory ability of the model). (55,56)

The Vancouver PNC LUR model explained a bit less variability than the Vancouver NO_x models, which had R² from 0.56 – 0.62. The NO_x models contained more variables than the PNC model, including multiple road length or vehicle density variables, a population density variable and a geographic location variable. The Vancouver PM_{2.5} model, based on only 25 sampling locations, had an R² of 0.52, which is comparable to the PNC₅₀ and PNC₉₀ road length models of 0.48 and 0.53. The validation R² for NO_x showed similar patterns as that of PNC, ranging from 0.31 – 0.79. This represents a fairly large (up to 0.25) decrease in R² values from model to validation, which was observed to a generally lesser degree in the PNC models. External validation with repeat and monitoring sites showed high R² values, up to 0.79, which is similar to the increase to 0.63 shown in validation of 20 repeat sites with PNC. The RMSE values for NO modeling were only about 6% of the mean, whereas they were over 50% for PNC, indicating one has much more confidence in the predicted levels of NO than of PNC. The Vancouver Black Carbon model performed similarly to the NO_x models, with a higher model R² than the PNC model (0.51 – 0.72) and comparatively high validation R² (0.40 – 0.63) values. However, the RMSE values represented 4 – 52% of the mean, which is closer to the scale of the PNC RMSE values. This higher error (relative to NO_x models) may be due to the smaller number of sites (39).

Similarly, the Amsterdam PNC LUR model ($R^2 = 0.44 - 0.65$) explained a bit less variability than the LUR model for $PM_{2.5}$ done at the same time ($R^2 = 0.54$ without site observations), the LUR models for $PM_{2.5}$ and Absorbance across the Netherlands in a previous study ($R^2 = 0.73 - 0.90$) and the original NO_2 model for Amsterdam ($R^2 = 0.62$). (44,59,74) There is insufficient published validation numbers to compare within the Netherlands as some studies reported LOO R^2 , some sub-sample R^2 , and some RMSE. Generally, LOO R^2 values were slightly lower than model R^2 and RMSE numbers were less than 20% of the mean of measurements. As with the Vancouver models, a number of the variables in the Dutch models were similar, pertaining to length of roads (especially major roads within 500m), vehicle density and building density. Also as in Vancouver, the PNC model remained the only one to include a variable related to the port, underscoring the possibility that port-related emissions are an important source of UFP specifically.

As with the NO_x models in Vancouver discussed earlier, the other models in Table 4.1 tended to have some type of population variable and geographic location variable in addition to traffic or road variables. R^2 for the other models ranged from 0.5 – 0.9, placing the Vancouver PNC models at the lower end of this range of explanation offered by models. RMSE values for the other models ranged from 4 – 24% of the reported measurement mean, which is markedly less than the >50% RMSE numbers obtained from the Vancouver PNC models. However, many of the models in Table 4.1 were based on fewer sites than the Vancouver PNC model, which could contribute to those models having higher R^2 values.

Overall, the development of LUR models for UFP in Vancouver was successful. The amount of variability the models explained ($R^2 = 0.29 - 0.53$) is on the lower end of the spectrum of related published LUR models, but within the range of other studies.

As the UFP LUR models highlighted the sharp spatial gradient of that pollutant with inclusion of variables with buffer sizes <100m, one reason these models tend to have lower R^2 values could be due to the difficulty in modeling something that varies so much over small distances and the limitation of accurate coordinates. Perhaps the most limiting aspect of the Vancouver PNC LUR models developed are the very high RMSE values associated with the models, often exceeding 50% of the mean measurements for a particular dataset. However, internal and external validation methods showed R^2 reasonably close to model R^2 values, suggesting models are at least reliable in predicting overall trends in concentration on a fine spatial scale.

4.2.2 Comparison with other LUR models incorporating meteorology

Other studies have found that varying meteorology, especially regarding wind speed and direction, can influence PNC substantially. (12,15,45,50) Several studies have incorporated wind speed and direction into LUR models for other pollutants.

One LUR model for NO_2 in Portland incorporated wind direction using the wind rose for the entire region during the sampling time (mean wind direction and two standard deviations on either side for the range) and applied this wedge-shaped buffer to the geographic predictor variables (in concert with circular buffers). (51) Inclusion of wind direction improved model predictive power by 0.15. Another study in Hamilton-Toronto on NO_2 interpolated observed wind speed and direction from weather stations and found this to improve models slightly (R^2 increased 0.01 – 0.04). (82) However, a coarse surface (10km grid) of wind patterns did not improve models in this case. A third study from Vancouver incorporated wind speed, wind

direction and cloud cover in a source area LUR model for NO and NO₂ and found that due to interpolating meteorology variables, this approach was not more helpful than LUR using circular buffers. (73) A fourth paper modeling PM_{2.5}, elemental carbon and NO₂ in Boston incorporated wind direction using meteorology data from one central monitoring location. (67) Variables such as percent of hours a site was downwind of a major road, mean wind speed during sampling time and percent of daytime with still winds were developed. These variables appeared in final models and the authors recommended further studies incorporating more site-specific meteorology variables.

In this project, site-specific meteorology data as well as central station monitoring data were used to create two models (model 9 and 10 in Table 3.9) using wind rose shaped buffers instead of circular buffers. Use of these buffers did not improve predictive abilities of the models, as R² values dropped from 0.48 to 0.34 and 0.29. It is unclear why the meteorology models did not improve R² or maintain the R² previously achieved. Perhaps due to the tendency of UFP to be short-lived and reactive (rapidly coagulating in the atmosphere), transport by wind over 60 minutes is not as important in determining concentrations as sources directly around a point (equidistant from the sampling location in a circular buffer). As most of the wind speeds were quite low (interquartile range for measured speeds was 0.5 – 2.2 m/s), perhaps circular buffers were a sufficient approximation for source contribution. One additional way of testing meteorologic impacts on PNC is to conduct measurements in areas where strong wind directions are highly predictable, such as near the ports, where winds are predominantly onshore or offshore. From this experience, the additional effort of measuring meteorology and conducting additional analyses and model development was not worthwhile.

4.2.3 Key findings of the Vancouver PNC LUR models

Variables that were consistently significant in final models included: those related to heavy-duty trucks (length of truck route or density of trucks), distance to the port, density or proximity to fast food restaurants or grills, distance to the nearest intersection of major roads. This suggests these variables as important sources of UFP in Metro Vancouver. The heavy-duty diesel truck related variables in the final models were for small buffer sizes (truck route length within 50m and truck density within 25m), suggesting that marked changes in UFP concentration occur on a very small scale from a source. Further, the distance to port variable, which had a negative coefficient, frequently showed up in the form $\ln$ -distance to port in the models, as PNC decreased logarithmically from the port. Fast food restaurant variables in final models were for 100m or 200m buffers, underscoring the fine spatial scale of UFP distribution. The smallest buffer investigated for fast food restaurants was 50m, as the closest restaurant to a sampling site was 40m (so a 25m buffer was uninformative).

4.3 Strengths

This design of this study benefitted from previous LUR model development in Vancouver for other pollutants and associated methodology developed therein. The number of sites and length of measurements, presence of four fixed site monitors, quantity and type of geographic predictor variables and inclusion of meteorologic considerations enhance the robustness of the methods employed to develop UFP LUR models for Vancouver.

Mobile sampling occurred at 80 sites for 60-minute periods, which is at the high end of number of sites needed to create a LUR, as 40 – 80 locations have been suggested by others (55) as the minimum number needed to provide enough data to resolve spatial patterns in pollutant

concentration. Further, 60-minute samples appeared to adequately represent long-term averages of PNC (two-week and seasonal). Overall, the number and length of samples optimized the quantity of data and resource expenditure.

Another major strength of this study was the inclusion of four fixed sites, which continually monitored PNC during sampling and were used to temporally adjust data. These four locations were situated to capture suspected high and low PNC locations (high traffic and low traffic areas), to be representative of the region. Other studies used a single central monitoring site, often measuring a different pollutant to correct data, or had no corrections at all. (50,67,71) From the temporal adjustments, it can be concluded that temporal variation (factor of three variability) appears to be less important than spatial variation (factor of seventy variability from minimum to maximum, factor of seven from 10th to 90th percentiles) in determining UFP levels.

A total of 135 geographic predictor variables were prepared for consideration in the LUR models. This represents nearly a doubling of the number of variables offered to NO_x LUR models created previously in the region (61) and included many new types of variables, highlighted in Table 2.2. Many more factors were considered when attempting to predict PNC levels than had previously been done. Inclusion of new variables was valuable in this study, as some of the new variables were included in the final LUR models and evidently helped to explain sources and factors affecting PNC. They were especially relevant as many of the final models only included a small number of variables (two to five) and contributed substantial partial R² values: 0.05 – 0.24 for fast food to distance to nearest intersection respectively. This is the first LUR model to include fast food restaurant variables as potential predictors. Notably, these were significant positive predictors of PNC in the two most predictive models (PNC₅₀, road

length and PNC₉₀, road length), suggesting that there should be more emphasis on restaurants as potential UFP sources in urban areas.

A novel aspect of this LUR model development was the incorporation of meteorologic variables, including wind speed, wind direction and temperature measured at the sampling site. While inclusion of these data did not improve the predictive power of the models in this instance, failure to consider meteorology is often cited as a limitation in other LUR models predicting air pollution species. (45,50,67)

4.4 Limitations

There are some limitations associated with this study. First, the LUR models were based on measurements made in the spring and may not represent the entire year. While a subset of PNC measurements at 22 sites completed in the fall were correlated with spring measurements, values were 1.68 – 3.79 times higher. Using the spring model to predict fall measurements yielded validation results (RMSE) comparable to spring measurements, suggesting this model is as applicable in spring as it is to fall seasons. Further, while fall and spring were selected as periods of the year to best represent the annual average of NO_x, the annual pattern of PNC in Metro Vancouver is not well known. A paper from Stockholm, Sweden used a combination of measurements each month of the year over a number of years and dispersion modeling to estimate monthly PNC concentrations in that city. (83) The researchers found a 2.5-fold change between the month with the highest levels (December) and lowest (August). Concentrations tended to be lowest during the summer, June-September, and highest during fall and winter, November to February. The annual average PNC levels are well represented in the months of

April and October in Stockholm, the same months when Vancouver sampling was completed, providing support that samples during these times reflects the annual average.

The UFP LUR models exhibited similar performance to the Amsterdam UFP LUR model (44) and relatively poor performance compared to LUR models for other pollutants in this region (Table 4.1). Model R^2 values were not higher than 0.53, LOO R^2 were not higher than 0.41 and RMSE values for all models were quite high, representing 68 – 81% of the mean. These observations all suggest that other unaccounted factors also contribute to UFP concentration variability or that the short-term presence of UFP in air is difficult to model. Application of these models in epidemiologic studies should consider these limitations. However, the most predictive models do account for nearly half of the variability in PNC levels and due to the pollutant's highly variable spatial distribution, likely can offer a better estimate than a coarsely applied estimate over a large area.

Two predominant variables in final LUR models were distance to the nearest port and density of fast food outlets. However, it is unclear what the specific source of UFP is relating to these variables, as emissions may be from the port and fast food sites themselves or those variables may be a surrogate for sources at or near those locations. Contributions to PNC are likely from a combination of both scenarios. Correlations between fast food and port variables with traffic-related variables showed some moderate relationships between these two types of variables. Based on Table 3.10, it seems possible that bus stops, major roads and intersections contribute to PNC near fast food restaurants. Somewhat surprisingly, distance to the nearest bus stop was highly correlated with distance to the port ($r = 0.68$) while other traffic-related variables were weakly correlated with the port. This suggests that bus stops may play a role in observed high levels of PNC near the port.

4.5 Applications

The UFP LUR models developed in this study have the capacity to be used in environmental epidemiology studies in Metro Vancouver, linking levels of pollutants to health outcomes in an exposed population. While the LUR models do not have R^2 values as high as other models and RMSE values are large, these models are fairly temporally stable (60-minute samples represent long-term levels fairly well and can be used to predict concentrations in other seasons) and can provide predictions of relatively high or low PNC, capturing trends and contrasts in levels. This represents an important addition to information currently available to assess long-term, population-level health effects of UFP.

The understanding of temporal trends of UFP (through temporal correction) as well as spatial variation and distribution of UFP in the Metro Vancouver is useful to know when developing future sampling campaigns which can focus on teasing out specific UFP sources and considering seasons for sampling. The regression maps created from the LUR models inform policymakers where UFP hotspots are located, which could be used to inform air quality management.

Finally, the commencement of a database of UFP measurements in Metro Vancouver offers a launching point for future sampling campaigns and for possible future development of a monitoring network and regulatory limits in the area.

4.6 Future work

Additional work on refining UFP LUR models for Metro Vancouver and for better understanding the sources of UFP and factors influencing concentrations can be completed in

order to better understand spatial-temporal patterns of UFP in the area and to increase confidence in applying models to epidemiologic studies. Additional sampling around the ports and fast food locations can be completed to determine whether the port itself (ships) and restaurants themselves (cooking processes) contribute to UFP levels or whether related traffic is responsible.

Further assessment of the validity of 60-minute samples to accurately reflect long-term concentrations of UFP can be completed, as well as determination of whether shorter time periods can capture the same trends. Sampling could also be conducted over different times of the year, in order to gain a better understanding of annual PNC patterns. This should be a lower priority than other sampling as comparison between fall 2009 and spring 2010 measurements yielded seasonal differences (adjustment factor 1.68 – 3.79) but the temporal variation was not found to be as important as spatial variation.

4.7 Conclusion

All project objectives outlined in the introduction have been completed and land use regression models for ultrafine particles in Metro Vancouver have been successfully developed. Sixty-minute measurements of PNC using CPC 3007 devices at 80 locations, a subset of sites previously used to develop LUR models for other pollutants in Vancouver, measured spatially variable concentrations with median sample values ranging from 1500 – 105 000 pt/cm³. Temporal adjustments to raw data based on four continuous monitors were generally small, ranging from 0.613 – 3.82, suggesting UFP temporal variation over a two-week period is lower than the spatial variability. Comparisons to measurements collected in a different season indicated that seasonally corrected PNC using NO measurements from central monitoring sites is

nearly as effective as using PNC data to temporally correct measurements. In addition, one-hour samples seem to represent longer-term (two-week period) PNC averages fairly well.

Eleven land use regression models, drawing from 135 potential predictor variables, were created. The most representative model created, based on road length variables predicting PNC_{50} , included the variables: length of truck routes within 50m, density of fast food locations in 200m and ln-distance to the nearest port. This model explained 48% of the variation in PNC; a LOO cross-validation exercise yielded an R^2 of 0.32 and RMSE of 12800 pt/cm³ (70% of the mean), suggesting the model is subject to a large degree of error. A new method to incorporate measured and monitored meteorologic variables (wind speed and wind direction) into LUR models was developed and applied but did not improve the predictive power of the models compared to those using circular buffers.

The established UFP LUR models can be used in epidemiologic applications in Metro Vancouver. Additional monitoring around ports and fast food sites will help determine the contribution of these features to PNC in the region. This project had laid a foundation for PNC data collection and modeling in Metro Vancouver from which additional studies and perhaps eventually policy-related monitoring protocol can be developed.

References

1. Health Effects Institute: Panel on the Health Effects of Traffic-Related Air Pollution. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects [Internet]. Boston, MA: Health Effects Institute; 2010. Report No.: Special Report 17. [cited 2011 Dec. 13]; Available from: <http://pubs.healtheffects.org/getfile.php?u=553>
2. Hoffmann B, Moebus S, Dragano N, Möhlenkamp S, Memmesheimer M, Erbel R, et al. Residential traffic exposure and coronary heart disease: results from the Heinz Nixdorf Recall Study. *Biomarkers*. 2007;14(s1):74–8.
3. Hoffmann B, Moebus S, Stang A, Beck E-M, Dragano N, Möhlenkamp S, et al. Residence close to high traffic and prevalence of coronary heart disease. *European Heart Journal*. 2006;27(22):2696–702.
4. Brauer M, Hoek G, Smit HA, de Jongste JC, Gerritsen J, Postma DS, et al. Air pollution and development of asthma, allergy and infections in a birth cohort. *Eur. Respir. J*. 2007;29(5):879–88.
5. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ. Health Perspect*. 2008;116(5):680–6.
6. Clark NA, Demers PA, Karr CJ, Koehoorn M, Lencar C, Tamburic L, et al. Effect of Early Life Exposure to Air Pollution on Development of Childhood Asthma. *Environ Health Perspect*. 2010;118(2):284–90.
7. Karr CJ, Demers PA, Koehoorn MW, Lencar CC, Tamburic L, Brauer M. Influence of Ambient Air Pollutant Sources on Clinical Encounters for Infant Bronchiolitis. *Am. J. Respir. Crit. Care Med*. 2009;180(10):995–1001.
8. World Health Organization. The world health report 2002: reducing risks, promoting healthy life [Internet]. Geneva, CH: World Health Organization; 2002. [cited 2010 June 1]; Available from: http://www.who.int/whr/2002/en/whr02_en.pdf
9. Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, et al. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol*. 2005;15(2):185–204.
10. Metro Vancouver - Air Quality Index by Station [Internet]. 2010. [cited 2010 Jun 28]; Available from: <http://apps2.metrovancouver.org/aqi/>
11. Geiser M, Rothen-Rutishauser B, Kapp N, Schürch S, Kreyling W, Schulz H, et al. Ultrafine Particles Cross Cellular Membranes by Nonphagocytic Mechanisms in Lungs and in Cultured Cells. *Environ Health Perspect*. 2005;113(11):1555–60.

12. Zhu Y, Hinds WC, Kim S, Shen S, Sioutas C. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. *Atmospheric Environment*. 2002;36(27):4323–35.
13. Government of British Columbia. New Ambient Air Quality Criteria for PM_{2.5} [Internet]. BC Air Quality. 2009 [cited 2010 Jun 1]; Available from: <http://www.env.gov.bc.ca/epd/bcairquality/regulatory/pm25-objective.html>
14. Sioutas C, Delfino RJ, Singh M. Exposure Assessment for Atmospheric Ultrafine Particles (UFPs) and Implications in Epidemiologic Research. *Environ Health Perspect*. 2005;113(8):947–55.
15. Cyrys J, Pitz M, Heinrich J, Wichmann H-E, Peters A. Spatial and temporal variation of particle number concentration in Augsburg, Germany. *Sci. Total Environ*. 2008;401(1-3):168–75.
16. Xia T, Li N, Nel AE. Potential Health Impact of Nanoparticles. *Annual Review of Public Health*. 2009;30(1):137–50.
17. Oberdörster G. Toxicology of ultrafine particles: in vivo studies. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*. 2000;358(1775):2719 –2740.
18. Ferin J, Oberdörster G, Penney J. Pulmonary retention of ultrafine and fine particles in rats. *American journal of respiratory cell and molecular biology*. 1992;6(5):535.
19. Oberdörster G, Sharp Z, Atudorei V, Elder A, Gelein R, Kreyling W, et al. Translocation of Inhaled Ultrafine Particles to the Brain, *Inhalation Toxicology, Informa Healthcare. Inhal. Toxicol*. 2004;16(6-7):437–45.
20. Rückerl R, Schneider A, Breitner S, Cyrys J, Peters A. Health effects of particulate air pollution: A review of epidemiological evidence. *Inhal Toxicol*. 2011;23(10):555–92.
21. Wichmann HE, Spix C, Tuch T, Wolke G, Peters A, Heinrich J, et al. Daily mortality and fine and ultrafine particles in Erfurt, Germany part I: role of particle number and particle mass. Research report (Health Effects Institute). 2000;(98):5–86.
22. Stolzel M, Breitner S, Cyrys J, Pitz M, Wolke G, Kreyling W, et al. Daily mortality and particulate matter in different size classes in Erfurt, Germany. *J Expos Sci Environ Epidemiol*. 2007;17(5):458–67.
23. Samet J, Graff D, Berntsen J, Ghio A, Huang Y, Devlin R. A Comparison of Studies on the Effects of Controlled Exposure to Fine, Coarse and Ultrafine Ambient Particulate Matter from a Single Location, *Inhalation Toxicology, Informa Healthcare. Inhal. Toxicol*. 2007;19(Suppl. 1):29–32.

24. Rückerl R, Phipps RP, Schneider A, Frampton M, Cyrus J, Oberdörster G, et al. Ultrafine particles and platelet activation in patients with coronary heart disease – results from a prospective panel study. *Part Fibre Toxicol.* 2007;4:1.
25. Hoek G, Boogaard H, Knol A, de Hartog J, Slottje P, Ayres JG, et al. Concentration Response Functions for Ultrafine Particles and All-Cause Mortality and Hospital Admissions: Results of a European Expert Panel Elicitation. *Environ. Sci. Technol.* 2010;44(1):476–82.
26. Knol A, de Hartog J, Boogaard H, Slottje P, van der Sluijs J, Lebre E, et al. Expert elicitation on ultrafine particles: likelihood of health effects and causal pathways. *Part. Fibre Toxicol.* 2009;6(19):1–16.
27. Gertler AW, Gillies JA, Pierson WR, Rogers CF, Sagebiel JC, Abu-Allaban M, et al. Real-world particulate matter and gaseous emissions from motor vehicles in a highway tunnel. *Res Rep Health Eff Inst.* 2002;(107):5–56; discussion 79–92.
28. Kittelson DB. Engines and nanoparticles. *Journal of Aerosol Science.* 1998;29(5-6):575–88.
29. Buonocore JJ, Lee HJ, Levy JI. The influence of traffic on air quality in an urban neighborhood: a community-university partnership. *Am J Public Health.* 2009;99 Suppl 3:S629–635.
30. Thai A, McKendry I, Brauer M. Particulate matter exposure along designated bicycle routes in Vancouver, British Columbia. *Science of The Total Environment.* 2008;405(1–3):26–35.
31. Shi JP, Evans DE, Khan A., Harrison RM. Sources and concentration of nanoparticles (<10nm diameter) in the urban atmosphere. *Atmospheric Environment.* 2001;35(7):1193–202.
32. Lurmann F. Where are people exposed to ultrafine particles? HEI Conference, Portland, Oregon; 2009.
33. Kleeman MJ, Riddle SG, Robert MA, Jakober CA, Fine PM, Hays MD, et al. Source Apportionment of Fine (PM_{1.8}) and Ultrafine (PM_{0.1}) Airborne Particulate Matter during a Severe Winter Pollution Episode. *Environ. Sci. Technol.* 2009;43(2):272–9.
34. Martin CL, Longley ID, Dorsey JR, Thomas RM, Gallagher MW, Nemitz E. Ultrafine particle fluxes above four major European cities. *Atmospheric Environment.* 2009;43(31):4714–21.
35. Van der Zee S, Van Der Laan J, Hoek G. The contribution of ships to measured NO_x and ultrafine particle concentrations along the waterways in Amsterdam. ISEE Conference, Barcelona, Spain; 2011.

36. Saxe H, Larsen T. Air pollution from ships in three Danish ports. *Atmospheric Environment*. 2004;38(24):4057–67.
37. Healy RM, O'Connor IP, Hellebust S, Allanic A, Sodeau JR, Wenger JC. Characterisation of single particles from in-port ship emissions. *Atmospheric Environment*. 2009;43(40):6408–14.
38. Kim S, Shen S, Sioutas C. Size distribution and diurnal and seasonal trends of ultrafine particles in source and receptor sites of the Los Angeles basin. *Journal of the Air & Waste Management Association* (1995). 2002;52(3):297.
39. Dennekamp M, Howarth S, Dick C, Cherrie J, Donaldson K, Seaton A. Ultrafine particles and nitrogen oxides generated by gas and electric cooking. *Occup Environ Med*. 2001;58(8):511–6.
40. Cass GR, Hughes LA, Bhavsar P, Kleeman MJ, Allen JO, Salmon LG. The chemical composition of atmospheric ultrafine particles. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*. 2000;358(1775):2581–2592.
41. Kleeman M, Robert M, Riddle S, Fine P, Hays M, Schauer J, et al. Size distribution of trace organic species emitted from biomass combustion and meat charbroiling. *Atmospheric Environment*. 2008;42(13):3059–75.
42. Wallace L, Ott W. Personal exposure to ultrafine particles. *J Expos Sci Environ Epidemiol*. 2011;21(1):20–30.
43. Hoek G, Meliefste K, Cyrys J, Lewné M, Bellander T, Brauer M, et al. Spatial variability of fine particle concentrations in three European areas. *Atmospheric Environment*. 2002;36(25):4077–88.
44. Hoek G, Beelen R, Kos G, Dijkema M, Van der Zee SC, Fischer PH, et al. Land Use Regression Model for Ultrafine Particles in Amsterdam. *Environmental science & technology*. 2011;45(2):622–8.
45. Zhou Y, Levy JI. Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis. *BMC Public Health*. 2007;7(1):89.
46. Hagler GSW, Thoma ED, Baldauf RW. High-Resolution Mobile Monitoring of Carbon Monoxide and Ultrafine Particle Concentrations in a Near-Road Environment. *Journal of the Air & Waste Management Association*. 2010;60(3):328–36.
47. Ketzel M, Wählén P, Berkowicz R, Palmgren F. Particle and trace gas emission factors under urban driving conditions in Copenhagen based on street and roof-level observations. *Atmospheric Environment*. 2003;37(20):2735–49.

48. Fruin S, Westerdahl D, Sax T, Sioutas C, Fine PM. Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles. *Atmospheric Environment*. 2008;42(2):207–19.
49. Hanna SR, Britter R, Franzese P. A baseline urban dispersion model evaluated with Salt Lake City and Los Angeles tracer data. *Atmospheric Environment*. 2003;37(36):5069–82.
50. Weichenthal S, Dufresne A, Infante-Rivard C, Joseph L. Determinants of ultrafine particle exposures in transportation environments: findings of an 8-month survey conducted in Montreal, Canada. *J Expos Sci Environ Epidemiol*. 2008;18(6):551–63.
51. Mavko ME, Tang B, George LA. A sub-neighborhood scale land use regression model for predicting NO₂. *Science of The Total Environment*. 2008;398(1–3):68–75.
52. Gilbert NL, Woodhouse S, Stieb DM, Brook JR. Ambient nitrogen dioxide and distance from a major highway. *Science of The Total Environment*. 2003;312(1-3):43–6.
53. Zwack LM, Paciorek CJ, Spengler JD, Levy JI. Modeling Spatial Patterns of Traffic-Related Air Pollutants in Complex Urban Terrain. *Environ Health Perspect*. 2011;119(6):852–9.
54. Henderson S, Brauer M. Measurement and modeling of traffic-related air pollution in the British Columbia Lower Mainland for use in health risk assessment and epidemiological analysis [Internet]. Vancouver, BC: University of British Columbia; 2005. [cited 2010 Sept. 1]; Available from: <http://www.cher.ubc.ca/UBCBAQS/Traffic%20Report/Traffic%20Final%20Report.pdf>
55. Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*. 2008;42(33):7561–78.
56. Brauer M, Henderson S, Marshall J. A land use regression road map for the Burrard Inlet area local air quality study. [Internet]. 2006. [cited 2010 Sept. 15]; Available from: http://www.cher.ubc.ca/PDFs/GVRD_BI_Report.pdf
57. Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer PH, Wijga A, et al. Air Pollution from Traffic and the Development of Respiratory Infections and Asthmatic and Allergic Symptoms in Children. *American Journal of Respiratory and Critical Care Medicine*. 2002;166(8):1092 –1098.
58. Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S, et al. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. *Occupational and Environmental Medicine*. 2007;64(1):8 –16.
59. Briggs D, Collins S, Elliott P, Fischer P, Kingham S, Lebret E, et al. Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science*. 1997;11(7):699–718.

60. Government of Canada HC. Regulations Related to Health and Air Quality [Internet]. 2003 [cited 2010 Jun 1]. Available from: <http://www.hc-sc.gc.ca/ewh-semt/air/out-ext/reg-eng.php>
61. Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter. *Environ. Sci. Technol.* 2007;41(7):2422–8.
62. Larson T, Su J, Baribeau A-M, Buzzelli M, Setton E, Brauer M. A Spatial Model of Urban Winter Woodsmoke Concentrations. *Environ. Sci. Technol.* 2007;41(7):2429–36.
63. Seaton A, Dennekamp M. Hypothesis: Ill health associated with low concentrations of nitrogen dioxide--an effect of ultrafine particles? *Thorax.* 2003;58(12):1012–5.
64. World Health Organization. Air quality guidelines, global update: particulate matter, ozone, nitrogen dioxide and sulfur dioxide [Internet]. Geneva, CH: World Health Organization; 2005. [cited 2011 Feb. 15]; Available from: http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_eng.pdf
65. Delfino RJ, Sioutas C, Malik S. Potential Role of Ultrafine Particles in Associations between Airborne Particle Mass and Cardiovascular Health. *Environ Health Perspect.* 2005;113(8):934–46.
66. Sardar SB, Fine PM, Yoon H, Sioutas C. Associations between particle number and gaseous co-pollutant concentrations in the Los Angeles Basin. *Journal of the Air & Waste Management Association.* 2004;54(8):992–1005.
67. Clougherty JE, Wright RJ, Baxter LK, Levy JI. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. *Environ Health.* 2008;7:17.
68. Vinzents PS, Møller P, Sørensen M, Knudsen LE, Hertel O, Jensen FP, et al. Personal Exposure to Ultrafine Particles and Oxidative DNA Damage. *Environ Health Perspect.* 2005;113(11):1485–90.
69. Kanaroglou PS, Jerrett M, Morrison J, Beckerman B, Arain MA, Gilbert NL, et al. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment.* 2005;39(13):2399–409.
70. Hornung RW, Reed LD. Estimation of Average Concentration in the Presence of Nondetectable Values. *Applied Occupational and Environmental Hygiene.* 1990;5(1):46–51.
71. Larson T, Henderson SB, Brauer M. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. *Environ. Sci. Technol.* 2009;43(13):4672–8.

72. DMTI Spatial Inc. DMTI Spatial Inc. CanMap® Streetfiles: User manual v2010.3 [Internet]. 2010. [cited 2011 Jan 5]; Available from: <http://abacus.library.ubc.ca/jspui/handle/10573/42371>
73. Su JG, Brauer M, Ainslie B, Steyn D, Larson T, Buzzelli M. An innovative land use regression model incorporating meteorology for exposure analysis. *Science of The Total Environment*. 2008;390(2-3):520–9.
74. Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U, et al. Estimating Long-Term Average Particulate Air Pollution Concentrations: Application of Traffic Indicators and Geographic Information Systems. *Epidemiology*. 2003;14(2):228–39.
75. Ross Z, Jerrett M, Ito K, Tempalski B, Thurston GD. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*. 2007;41(11):2255–69.
76. Hochadel M, Heinrich J, Gehring U, Morgenstern V, Kuhlbusch T, Link E, et al. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmospheric Environment*. 2006;40(3):542–53.
77. Gilbert NL, Goldberg MS, Beckerman B, Brook JR, Jerrett M. Assessing Spatial Variability of Ambient Nitrogen Dioxide in Montreal, Canada, with a Land-use Regression Model. *J. Air & Waste Manage. Assoc.* 2005;55:1059–63.
78. Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert NL, et al. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health A*. 2007;70(3-4):200–12.
79. Moore DK, Jerrett M, Mack WJ, Künzli N. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *J. Environ. Monit.* 2007;9(3):246–52.
80. Poplawski K, Gould T, Setton E, Allen R, Su J, Larson T, et al. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. *J Expos Sci Environ Epidemiol*. 2008;19(1):107–17.
81. Allen RW, Amram O, Wheeler AJ, Brauer M. The transferability of NO and NO₂ land use regression models between cities and pollutants. *Atmospheric Environment*. 2011;45(2):369–78.
82. Arain MA, Blair R, Finkelstein N, Brook JR, Sahsuvargolu T, Beckerman B, et al. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment*. 2007;41(16):3453–64.

83. Johansson C, Norman M, Gidhagen L. Spatial & temporal variations of PM10 and particle number concentrations in urban air. *Environmental Monitoring and Assessment*. 2006;127(1-3):477–87.

Appendices

Note: Appendices A, C and E – H are electronic appendices. These appendices are spreadsheets with project data. A brief written description of each file is provided below.

Appendix A Metro Vancouver ultrafine particle database

This appendix is located in the file “Metro_Vancouver_UFP_database_2010.xls” This file contains all of the measurements made during the spring 2010 UFP sampling campaign.

The first tab, “1. Sample UFP Data” includes the spatial UFP measurements from the 80 sites (plus 20 repeats). This includes the date, site identification, measured coordinates of the site, description of the site, instrument used, start and end time of the sample, raw measurements during each minute of sampling, trucks in sample, wind direction during sample, fast food locations, construction code, as well as statistics on data from each sample (minimum concentration, 5, 10, 25, 50, 75, 90, 95%ile, mean and standard deviation) in addition to the three correction factors (time of day, day of sample, instrument) individually and compositely applied.

The second tab, “2. Sample meteorologic data” includes direction (of wind), wind speed, temperature, relative humidity and altitude measured during sampling. Calculations of mean sample temperature (used as a variable in LUR modeling) as well as maximum wind rose length (and 0.125, 0.25, 0.50, 0.75 fractions) for each site are also included.

The third tab, “Fixed site UFP data” includes all raw data from the 4 fixed sites as well as calculation of the within (hour of day) and between day (day of sample) correction factors, used to temporally adjust mobile data.

The fourth tab, “Instrument 4 versus 2” contains collocation data for the two CPC 3007 instruments used in mobile monitoring, which was used to develop the instrument correction factor.

The fifth tab, “CPC and P-trak comparisons” contains collocation data for a CPC 3007 instrument and P-trak 8525, which was used to develop the between-instrument correction factor.

Table A.1 Sampling site locations and site PNC measurements (N=80)

Site ID	X	Y	PNC ₅₀ (pt/cm ³)	Ln-PNC ₅₀ (pt/cm ³)	PNC ₁₀ (pt/cm ³)	PNC ₉₀ (pt/cm ³)
2	490416	5463965	19144	9.9	17118	20160
3	494066	5463943	7652	8.9	5239	15888
4	496650	5463946	15886	9.7	13248	18731
6	494505	5462440	12291	9.4	6637	19757
7	496478	5461852	16280	9.7	13932	18481
8	512888	5460432	6066	8.7	4484	7838
10	495601	5459427	13861	9.5	9784	26406
11	498971	5459429	18726	9.8	6407	49653
12	514925	5459392	7772	9.0	6803	9782
13	490002	5458928	9285	9.1	7145	13316
15	496980	5458455	21999	10.0	19120	40573
16	502058	5458479	4510	8.4	3423	6525
17	510886	5458415	7400	8.9	5063	14509
18	518562	5458438	2655	7.9	1770	4226
19	490985	5457925	22599	10.0	8038	31956
20	485015	5457380	20706	9.9	12601	77282
22	489495	5457456	11473	9.3	9180	14981
23	495003	5457439	45636	10.7	28995	78963
25	503489	5457513	10980	9.3	9291	12078
27	514011	5457519	4297	8.4	2937	6140
28	516571	5460031	20098	9.9	16260	24557
29	498027	5457006	7702	8.9	6210	10310
31	491990	5456397	44891	10.7	36031	65570
32	493505	5456489	26278	10.2	20042	44603
33	485899	5455956	7088	8.9	5054	8078
35	495871	5456049	13274	9.5	9890	20347
37	497530	5455468	29661	10.3	21811	58043
38	506961	5455418	12910	9.5	11952	14643
39	509533	5455456	14651	9.6	9410	36423

Table A.1 Sampling site locations and site PNC measurements (N=80)

Site ID	X	Y	PNC₅₀ (pt/cm³)	Ln-PNC₅₀ (pt/cm³)	PNC₁₀ (pt/cm³)	PNC₉₀ (pt/cm³)
40	491940	5454909	7947	9.0	4818	10716
41	520994	5454686	22773	10.0	15609	48105
42	486515	5454517	7873	9.0	5409	11933
44	499443	5454408	9709	9.2	7517	11649
46	516003	5454428	14673	9.6	13328	16773
47	488544	5453962	17518	9.8	14565	25425
48	494428	5453972	58871	11.0	25298	120766
49	497414	5453989	4812	8.5	16043	21843
50	490439	5453478	6858	8.8	2431	29290
52	508537	5453234	42352	10.7	29199	61448
55	494471	5452403	21121	10.0	16089	32338
56	500404	5452408	15767	9.7	13719	22016
59	503042	5451946	14120	9.6	11268	17631
60	491594	5451418	7206	8.9	5516	10731
61	498071	5451478	19167	9.9	10768	26590
62	530425	5451493	16912	9.7	13610	22296
63	493482	5450945	23353	10.1	17794	34305
68	512555	5450387	19504	9.9	16474	21142
69	514584	5449456	21845	10.0	11295	44402
70	509839	5448950	9309	9.1	6595	14559
71	511994	5447958	29607	10.3	26008	35317
72	507525	5447362	13975	9.5	11548	15731
74	490437	5446388	35895	10.5	23700	57387
76	516532	5445999	11311	9.3	10188	15116
79	490056	5445458	24276	10.1	14923	46466
80	512051	5445340	17012	9.7	13968	23955
82	509839	5448950	14931	9.6	11823	21176
83	486897	5443953	6323	8.8	5422	9331
85	490374	5442931	5516	8.6	4553	8000
86	491999	5442965	1473	7.3	1096	2317
87	506984	5442442	12658	9.4	11362	14012
88	522552	5442364	24972	10.1	15297	51353
89	486452	5441941	2310	7.7	1054	4080
90	509425	5441892	10893	9.3	7638	20548
92	524491	5439415	18110	9.8	13710	30417
93	531368	5437889	5727	8.7	4384	7095
96	514954	5431440	11971	9.4	10722	13808
98	512486	5430950	10814	9.3	8647	19206
99	514970	5429927	8048	9.0	7611	10410
100	494177	5428967	7880	9.0	6467	9244

Table A.1 Sampling site locations and site PNC measurements (N=80)

Site ID	X	Y	PNC ₅₀ (pt/cm ³)	Ln-PNC ₅₀ (pt/cm ³)	PNC ₁₀ (pt/cm ³)	PNC ₉₀ (pt/cm ³)
101	519675	5439863	13480	9.5	9438	26356
102	489749	5450709	43150	10.7	30004	62277
103	489930	5456767	43073	10.7	29695	61336
104	491451	5450831	47856	10.8	33632	81058
105	491555	5453533	37137	10.5	18216	67853
106	494406	5451604	22964	10.0	15200	35107
107	494407	5458937	105000	11.6	72288	169032
108	516809	5450585	4602	8.4	3884	6905
110	507518	5449502	8864	9.1	4698	16903
113	547943	5434160	4455	8.4	3729	9190
114	552095	5435382	4297	8.4	3831	12832

Table A.2 Repeat PNC measurements (N=20)

Site ID	PNC ₅₀ (pt/cm ³)
3	7514.450802
11	10557.70521
17	12724.34099
22	13787.84694
23	35677.55335
31	10744.69571
31	21500.85245
33	9335.833256
38	11873.76788
40	8144.88406
49	18378.06365
50	26138.24295
59	10377.92845
60	6400.095252
70	18972.10104
79	4349.820324
85	9428.899416
103	34498.77198
103	43390.61669

Appendix B Ultrafine particle sampling campaign manual

Metro Vancouver Ultrafine Particulate Sampling Campaign Manual Spring 2010

Prepared by: Rebecca Abernethy
School of Environmental Health
University of British Columbia

April 9, 2010

Thank you for agreeing to be a part of the Spring 2010 Ultrafine Particulate Matter Sampling Campaign across Metro Vancouver. This is a student project based out of the University of British Columbia.

This study will provide important new research to the region in terms of the sources of a particular type of air pollution, ultrafine particles, and their exposure to humans. Briefly, the objective of this project is to measure concentrations of ultrafine particles across Metro Vancouver and to develop a predictive model to capture fine spatial variation in ambient levels.

Please familiarize yourself with all items in this manual and ensure you are confident with all procedures and pieces of equipment. Follow the protocol and schedule strictly; I must approve ANY changes. However, providing feedback throughout the sampling period on methods is encouraged.

There is a timesheet in this package for you to record the hours worked. While the schedule has been made and it is estimated that work hours will be from 9am – 5pm each day, it is important that actual hours are recorded and submitted, as that is what your compensation will be based on. You will be provided with an honourarium once the sampling period is complete to thank you for your contribution.

Your commitment to this project is appreciated. Please let me know if you have any concerns or feedback, otherwise enjoy your time in the field!

Sincerely,

Rebecca Abernethy
MSc Student
School of Environmental Health
University of British Columbia

rebecca.abernethy@gmail.com
604-999-0510

Overall Mobile Sampling Protocol

Please complete all of these steps at each site. More detailed operating instructions for the TSI 3007 Condensation Particle Counter, Kestrel 4500 Pocket Weather Tracker and GPS devices follow.

Equipment Checklist

Please ensure you have all items before leaving for the day.

Sampling Equipment

- CPC 3007
 - alcohol cartridge
 - isopropyl alcohol in small container
 - (there is a possibility each team will get a spare 3007 part way through sampling)
- 3007 tripod mounting case
- 3007 tripod
- Zero filter for 3007
- Kestrel 4500 (and spinning device and weather vane)
- 4500 tripod

Miscellaneous

- GPS
- Digital camera
- Extra AA batteries
- Extra AAA batteries
- Stopwatch
- 2 traffic counters
- Binder
 - sampling manual
 - log sheets
 - site satellite photos (Google Earth images)
 - schedule with sites
 - list of site addresses/intersections
- Clipboard
- Pens
- Sharpie marker
- CPC 3007 instrument manual
- Kestrel 4500 instrument manual

- 2 High visibility vest
- Cell phone
- Rebecca's cell number: (604) 999-0510

Field Sampling Procedure

The primary data collection period for this project will occur between 9:00 am and 5:00 pm, weekdays. Please plan your days accordingly.

Please double check with Rebecca BEFORE making changes to the timing of sampling or anything else to do with the schedule..

1. Put on your high visibility vests.
2. Identify the next site you will be visiting based on the schedule. Refer to the list of sites for reference. **Please plan out your route the night before**, in the order outlined on the schedule (use Google Maps, GPS navigation etc to help you). Drive to the site.
3. Place the alcohol wick in and turn the 3007 on when you estimate you are about 10 minutes away from the site, so that it is warmed up and ready to go upon arrival. (See 3007 details for how to do this)
4. Pull up to the site. Look for the Ogawa sampler which will already be in place at some sites. It will be on the pole identified in the site description about 2.5m above ground in a little white cup (the cup is about 10cm high, so it's fairly small). See photo.
 - a. NOTE: It's a good idea to park so that you have a good view of the road on which you will be counting traffic.

Installed Ogawa sampler.

5. Record the Site ID on the field log.
6. Prepare for sampling.
 - a. See operating instructions for the CPC 3007, attached.
 - b. See operating instructions for the Kestrel 4500, attached.

- i. Ensure the CPC 3007 and Kestrel 4500 are facing traffic and facing the same direction. Indicate the direction the instruments are facing on the Google Earth image for that site.
7. Start the 3007 logging, by selecting Log Mode 1 and shut the case. Ensure the 4500 is also recording by double-checking that Auto Store under Memory Options is on (see 4500 instructions). Record the sample start time on the field log. Allow the devices to run for one hour.
8. Complete the log sheet for the site:
 - a. Record notes on any interesting activity or nearby buildings/facilities. **Take photos in all directions of the surrounding site as well as sampling equipment.** Label the template roads and indicate where the sample site is on the sketch.
 - b. Record the direction the wind is coming from and whether the sampler is upwind or downwind of the road.
 - i. Using a sharpie marker, indicate the direction of the wind on the zoomed in Google Earth image for that site.
 - c. Record the name, location (ie “intersection of Main & Maple”), and approximate distance to any fast food restaurants.
 - i. Using a sharpie marker, indicate the location of any nearby fast food restaurants on the zoomed in Google Earth image for that site.
 - d. Record any necessary drawings on the Google Earth images. If things are significantly different from the Google Earth image, draw nearby building and activities on the sketch template, on the log sheet.
 - e. Use the hand-held clickers to count diesel truck and bus traffic travelling in both directions:
 - i. Use the stopwatch to make sure you count for 5 minutes. Count the trucks/buses on the busiest road adjacent to the samplers (ie the busiest road near the site that is not blocked by buildings) and mark this road on the sketch.
 - ii. Record the number of trucks/buses and the counting duration (this should be 5 minutes under normal circumstances).
 - iii. Repeat the 5-min diesel truck/bus count a second time. The two 5-minute counts should be conducted at least 30 minutes apart (to best capture average conditions over the hour of ultrafine particle sampling).
9. After 1 hour stop the 3007 logging session.
10. Turn off the 3007 and 4500.
 - a. Remove the wick from the 3007 and immerse it in the alcohol container during travel. Place the “for storage only” cap in the instrument for travel.
11. Return all equipment to the car.
12. Go to the next site.
13. Repeat above.
14. **At the end of the day, drivers: download CPC 3007 data, 4500 data and photos and send to Rebecca.**

- a. For the 3007:
 - i. Install the software on a PC. Connect device to the computer, turn it on and open the Aerosol Instrument Manager Program.
 - ii. Click 'New File' and save it in the following format: "mmddyyUBC#-siteID#" For example, a file collected April 20, 2010 with instrument UBC #5 at site ID 15 would be saved as 042010UBC5-15.
 - iii. Then go to Run -> Receive Logged Data, Select All and each file for each site will come up. You can scroll along the
 - iv. Go to File -> Export and select .xls and Time Stamp under data point time format. Save the.xls files as "April 19, UBC #5, Site 18" as there will be one file per site. Email all files to Rebecca.
 - v. There is a computer manual that should help with any problems, or ask Rebecca.
- b. For the 4500: download the software from <http://www.nkhome.com/kestrel/kestrel-4500/more.php> (under software). Install the program and follow the manual attached to download data.
- c. Please label all photos the same way (ie 042010UBC5-15).
Please record all file names for the 3007, 4500 and photos on the log sheet for the appropriate site.

15. Ensure batteries are charged for the next day and that all equipment is in working condition.

16. Arrange to meet with your teammate for the next day and if necessary, pass the equipment on to the next driver.

- a. A complete contact list is included in this manual and is also on the schedule.

***If community members ask you questions, please try to answer them if you feel able. If you do not know an answer, please do not make something up - just direct them to call Rebecca (604-999-0510).

***Please be respectful of people in the community you are in and activities going on. If you are asked to relocate, please do so. If you feel unsafe at any time then you should also relocate. The sampler can be stopped (make a note of when on the logsheet) and restarted in a different area. Take another GPS waypoint as well.

***I would much rather have too much information and notes than too little – you will have a lot of time while the instrument is sampling. Please be thorough in your site descriptions, photos and notes.

Operation of the TSI 3007 Condensation Particle Counter

**Note: this is a summary. For details on instrument operation refer to the Operation and Service Manual or call Rebecca!*

1. Place the alcohol wick in the 3007.

- a. The 3007 should be OFF.
- b. Open the clear alcohol capsule and set the alcohol wick on a clean surface with the end standing up.
- c. Squeeze alcohol into the fill capsule until the liquid level is even with the fill line near the base.
- d. *After making sure the alcohol cartridge is clean*, insert the alcohol cartridge into the fill capsule and lock it into place.
- e. Wait a few minutes while the wick inside the cartridge soaks up alcohol.
- f. Remove the alcohol cartridge from the fill capsule and gently shake it to allow excess alcohol to drain back into the capsule.
- g. Insert the cartridge into the 3007 and lock into place.

The 3007 with the “for storage only” cap in place, to be used during transit, and wick soaking up alcohol.

- 2. Turn the device on and allow it to warm up (10 min). It will count down from 600s and you will not be able to do anything until this countdown is complete.**
- 3. Ensure enough battery life remains for sampling.**
 - a. Ensure the spare set is charged; charge drained batteries in vehicle while sampling.
- 4. Zero the instrument.**
 - a. Attach the zero filter assembly to the inlet.
 - b. The particle concentration should go to zero in approximately 5-10 seconds. Leave the zero filter attached to the instrument for 30 seconds to make sure the zero reading is stable.
 - c. Record the 30s zero reading on the log sheet for the first site of the day.
 - d. Remove the zero filter. The instrument is now ready for operation.
 - e. If there are problems, refer to the troubleshooting page in the manual (31).
 - i. If the device does not go to zero, add an additional zero assembly. Readings of up to 10 are ok, but ensure you record the reading on the logsheet. Anything above this, please call Rebecca.

3007 with zero filter assembly attached.

5. Prepare for sampling.

- a. Place the 3007 in its case on the tripod where sample will be collected from, pointing the aerosol inlet toward the road.
- b. The 3007 should be approximately 4' off the ground (i.e. the tripod should be fully extended).

3007 in case on fully extended tripod.

- c. 3007 should be placed on the edge of the road (on a sidewalk/median strip if possible) out of the way of oncoming vehicles and pedestrians.
- d. Press down arrow to 'setup.'
 - i. Select log interval. Ensure it is at 1:00 minute.

6. Start sampling.

- a. Press down arrow to 'log mode 1' on main menu and hit enter (bottom bent arrow key).

“Log mode 1” – select to start logging with enter key. Press enter key again to stop logging after 1 hr.

“enter” key – used to select items

b. Sample for one hour. At the end of sampling, press the enter key to stop.

Note: Always keep the 3007 level during sampling

1. You can review the sample information by selecting 'sample' in the main menu, then 'statistics' in the next menu. Select 'logged test stats' and hit enter. Use the up and down arrow keys to scroll through saved tests.
2. Hit enter to stop logging.
3. Press the power button to turn off.
Remove alcohol cartridge and replace with the storage cap when moving to the next site/ending for the day.

Operation of the Kestrel 4500 Pocket Weather Tracker

**Note: this is a summary. For details on instrument operation refer to the Operation and Service Manual or call Rebecca!*

1. Turn the device on (the circle button with red on it).

“on” and selection button

2. Ensure the device has sufficient battery power (>80% at the start of the day).
3. Calibrate the compass.
 - a. Place the 4500 upright in its box and rotate around three times, slowly (each rotation should be 10s). Do not calibrate on metal surfaces and ensure the device is straight the whole time.
 - b. The device will tell you once calibration is complete; if it is unsuccessful, refer to the manual (page 5-7) for troubleshooting.
 - c. The device must be calibrated every time the battery compartment is opened or batteries replaced. Otherwise, once at the beginning of the day is sufficient.

Place device upright in case on a flat surface and rotate around three times to calibrate.

4. Verify that the device is logging and that the time and date are correct.
 - a. Press the 'on' key for the menu.
 - b. Select 'Memory Options' with the centre, oval button and ensure 'Auto Store' is on and that the 'Store Rate' is 5 minutes.

(top) Menu showing Memory Options. (bottom) Under Memory Options, ensure Auto Store is on.

5. Affix the 4500 to the tripod casing, ensuring the device is fully upright and facing into the traffic (same direction as the inlet on the 3007).
6. After sampling, turn the device off.
 - a. Note: the Kestrel will turn its screen off automatically after about 10 minutes. As long as it says “Auto Store” on at the start and end of sampling, it is recording.

Operation of the GPS

1. Turn GPS on.
2. Ensure it has adequate battery power.
3. At the sampling location, save a way-point. Note the point number and latitude/longitude down on the logsheet.

Community Ultrafine Particle Sampling Field Log

Date: _____ CPC 3007 instrument ID: UBC # _____ Kestrel 4500 ID: UBC # _____ GPS ID: _____
 Site ID: _____ Site Description _____
 CPC 3007 file name: _____ Kestrel 4500 file name: _____ CPC 3007 zeroing reading: _____
 Photo file names: _____

3007 Start Time	3007 Stop Time	TRAFFIC COUNTS					WIND		NEARBY FAST FOOD RESTAURANTS			GPS Point # and Latitude- Longitude
		Road Name(s)	Count 1 Trucks	Count 1 Minutes	Count 2 Trucks	Count 2 Minutes	Direction wind is coming <i>from</i> [N,NE, E,...]	Sampler upwind or downwind of road? [U, DW]	Name(s)	Location(s)	Approx. Distance(s) (m)	
:	:			:		:						
If the sampler is located near the intersection of two roads you should count the trucks on both roads.							Also record the wind direction on the Google Earth image.		Codes: McDonald's = McD; Burger King = BK; KFC; Tim Horton's = TH; A&W = AW; Church's Chicken = CC, Dairy Queen = DQ, Wendy's = WE; Other = O Also record the location(s) of nearby fast food restaurant(s) on the Google Earth image.			

Site Notes (and sketch if different from Google Earth image):

Vehicle Use Tracking Log

Please keep accurate record of the kilometers driven per day for ultrafine sampling purposes. At the end of the sampling period, submit this form to be compensated for vehicle use (\$0.41/km), per UBC vehicle use policy.

***Zipcar use will be reimbursed. Please keep receipts for each day but also indicate on this sheet days used.

Vehicle owner/driver name: _____ Date submitted:

Date driven	Odometer reading start (km)	Odometer reading end (km)	Kilometers driven (km)	Description of route	Notes

Time Sheet

Please keep accurate record of the number of hours worked per day. You will be paid an honourarium at the end of the sampling period according to number of hours worked. Note that a standard sampling workday will be from 9am – 5pm. If your day will be different than that, particularly if it will be longer, please call Rebecca THAT DAY (604-999-0510) to discuss why.

Name: _____ Date submitted: _____

Date worked	Start time	End time	Hours worked	Description of route	Notes

Team Members

Name	Phone Number	Email Address
Rebecca		
Christie		
Rehema		
Sara C		
Brian		
Sarah		
Mary		

*phone numbers and emails were shared amongst team members to facilitate sampling

Appendix C Ultrafine particle and nitrogen oxides comparisons

This appendix is located in the file “UFP_NOx_comparisons.xls” This file contains the average NO, NO₂ and NO_x values sampled at one hundred fourteen sites during the same time period as UFP sampling was occurring at eighty of the same sites. UFP median values at the same sites are included in the database.

Table C.1 UFP and NOx comparison data

Site ID	Notes	PNC ₅₀ (pt/cm ³)	NO ₂ (ppb)	NO (ppb)	NO _x (ppb)
1			4.60	4.29	8.89
2		19144	9.00	6.15	15.15
3		7652	6.95	6.11	13.06
4		15886	6.37	5.50	11.87
5			10.03	7.72	17.75
6		12291	12.41	13.66	26.07
7		16280	9.43	9.82	19.25
8		6066	4.46	5.36	9.81
9			15.10	10.19	25.29
10		13861	11.47	16.00	27.47
11		18726	9.23	7.91	17.13
12		7772	8.01	5.53	13.55
13		9285	14.03	10.13	24.16
14			14.24	23.67	37.91
15		21999	10.70	7.60	18.30
16		4510	7.07	4.90	11.97
17		7400	13.00	18.65	31.65
18		2655	4.62	2.59	7.22
19		22599	15.55	10.45	25.99
20		20706	8.84	22.11	30.95
21			10.09	7.29	17.38
22		11473	9.70	14.62	24.32
23		45636	18.96	37.12	56.08
24			9.36	4.44	13.80
25		10980	8.65	3.20	11.85
26			6.40	4.10	10.50
27		4297	5.38	8.83	14.22
28		20098	6.87	8.88	15.75

Table C.1 UFP and NOx comparison data

Site ID	Notes	PNC ₅₀ (pt/cm ³)	NO2 (ppb)	NO (ppb)	NOx (ppb)
29		7702	14.49	12.65	27.14
30			7.56	9.24	16.80
31		44891	13.09	13.04	26.12
32		26278	12.31	24.58	36.89
33		7088	7.48	6.39	13.86
34			6.56	7.75	14.31
35		13274	14.15	16.76	30.91
36			6.96	10.66	17.62
37		29661	12.98	17.10	30.08
38		12910	10.28	20.64	30.92
39		14651	8.96	14.93	23.89
40		7947	7.31	8.35	15.67
41		22773	5.40	9.15	14.55
42		7873	6.06	11.84	17.90
43	missing				
44		9709	11.96	7.62	19.58
45			5.67	7.60	13.27
46		14673	6.26	8.25	14.51
47		17518	6.14	10.03	16.18
48		58871	13.61	16.14	29.75
49		4812	8.41	6.93	15.34
50		6858	8.36	14.20	22.56
51			6.69	8.20	14.89
52		42352	16.62	33.27	49.89
53			7.35	8.98	16.33
54			8.04	9.50	17.54
55		21121	8.64	12.48	21.13
56		15767	5.86	23.75	29.60
57			8.77	13.26	22.03
58			5.78	7.32	13.10
59		14120	10.04	8.05	18.09
60		7206	6.10	11.08	17.18
61		19167	7.72	5.04	12.77
62		16912	4.16	6.21	10.37
63		23353	14.10	37.94	52.04
64			9.67	9.61	19.28
65			8.05	12.46	20.51
66			7.47	12.19	19.67
67			8.60	8.78	17.38
68		19504	5.97	7.51	13.49

Table C.1 UFP and NOx comparison data

Site ID	Notes	PNC ₅₀ (pt/cm ³)	NO2 (ppb)	NO (ppb)	NOx (ppb)
69		21845	13.84	24.79	38.63
70		9309	7.60	7.40	14.99
71		29607	7.79	14.75	22.54
72		13975	7.17	6.67	13.84
73			5.84	7.44	13.28
74		35895	12.06	18.49	30.54
75			7.02	7.12	14.15
76	missing	11311			
77			4.51	6.34	10.84
78			3.65	8.12	11.77
79		24276	8.42	11.57	19.99
80		17012	13.06	25.98	39.04
81			7.23	6.54	13.77
82		14931	6.62	11.69	18.31
83		6323	2.90	11.51	14.42
84			5.63	9.01	14.65
85		5516	4.26	5.74	10.00
86		1473	3.92	7.24	11.16
87		12658	4.24	8.08	12.32
88		24972	8.46	20.86	29.32
89		2310	4.09	7.81	11.89
90		10893	5.25	7.44	12.69
91			6.97	8.89	15.85
92		18110	8.46	20.40	28.86
93		5727	3.12	4.17	7.29
94			6.21	17.37	23.58
95			3.54	5.21	8.75
96		11971	4.24	4.70	8.94
97			4.59	6.94	11.53
98		10814	4.32	9.26	13.58
99		8048	2.35	5.28	7.64
100		7880	4.26	4.30	8.56
101		13480	7.58	18.69	26.27
102		43150	8.86	15.23	24.08
103		43073	10.31	31.82	42.12
104		47856	13.06	31.37	44.43
105		37137	7.25	22.89	30.15
106		22964	11.03	37.47	48.49
107		105000	20.31	43.22	63.53
108		4602	5.16	5.35	10.51
109			10.06	7.50	17.56
110		8864	9.99	17.80	27.79

Table C.1 UFP and NOx comparison data

Site ID	Notes	PNC₅₀ (pt/cm³)	NO2 (ppb)	NO (ppb)	NOx (ppb)
111			7.40	21.92	29.32
112			6.10	15.21	21.31
113		4455	5.12	6.31	11.43
114		4297	5.15	8.59	13.75
115			4.65	4.43	9.08
116			9.89	8.41	18.29

Appendix D Commands used in GIS variable extraction and R scripts

D.1 Geospatial modelling environment command examples

GME from Spatial Ecology was used to extract values for the geographic predictor variables (shown in Table 2.2 and 2.3) used in LUR model development. Below are examples of commands used to derive this information.

Using the traditional buffer, the 'isectpntrst' command was used to intersect the sampling location point shapefile with the raster (with appropriate buffer). The example below is for determining the length of highways 25m from a site.

```
isectpntrst(in="C:\Users\rabernethy\Desktop\Thesis\3. GIS  
data\ufp_final_waypoints\80points.shp", raster="C:\Users\rabernethy\Desktop\Thesis\3. GIS  
data\Road\rd1_25_f", field="rd1_25");
```

For the wind rose buffers, a polygon shapefile of the wind roses from each site, with site attribute information attached, interacted with polyline files to sum the line length within the polygon, with point files to count the points within the polygon and with polygon files to calculate the overlapping area. Examples of the use of these three commands include the length of highways in a 0.125 maximum wind rose buffer, the number of bus stops in a 0.25 maximum wind rose buffer and the area of commercial land in a 1.0 maximum wind rose buffer.

```
sumlinelengthsinpolys(line="C:\Users\rabernethy\Desktop\Thesis\3. GIS data\Road\RD1.shp",  
poly="C:\Users\rabernethy\Desktop\Thesis\3. GIS data\Windroses\0125\0125_merge_j80.shp",  
field="rd1_0125w");
```

```
countpntsinpolys(poly="C:\Users\rabernethy\Desktop\Thesis\3. GIS
data\Windroses\025\025_merge_j80.shp", pnt="C:\Users\rabernethy\Desktop\Thesis\3. GIS
data\bus_stops_march09\BusStops_09March2010_Clip.shp", field="bus_wr_025");
```

```
isectpolypoly(in="C:\Users\rabernethy\Desktop\Thesis\3. GIS
data\Windroses\100\100_joinatt.shp", poly="C:\Users\rabernethy\Desktop\Thesis\3. GIS
data\Landuse\landuse_GVRD\comwr.shp", field="CATNUM", prefix="cw100",
thematic=TRUE);
```

D.2 R script examples

R was used to prepare and validate (through LOO cross validation) the LUR models.

The following script is for the development of a LUR model, including ranking the correlation of the variables with PNC, removing variables correlated >0.6 with the highest ranking variable in their category and creating a linear model.

```
# Function to return a table of rankings between input Y and the
# potentially predictive covariates.
# Function variable group should be of form "rd1."
# Function output is a list of character class, names of variables
# y = character value ("pnc50", "lnpnc50", "pnc10" etc.)
# ranks = data frame with three columns (variable name, univariate r2
# with Y, and variable type)
```

```
get_ranks_table = function(y, modeltype){
  if (modeltype == "length"){
    covariates = names(ourdata[7:48])
  }
  if (modeltype == "density"){
    covariates = names(cbind(ourdata[7],ourdata[20:60]))
  }
  r2 = numeric()
  for (variable in covariates){
    model = lm(get(y)~get(variable), na.rm = T)
    r2 = c(r2, as.double(summary(model)$r.squared))}
  ranks = as.data.frame(covariates)
  ranks$r2 = r2
```

```

ranks$covariates = as.character(ranks$covariates)
ranks$vartype = strsplit(ranks$covariates,"\\.")
ranks$vartype = sapply(ranks$vartype, '[[', 1)
ranks$vartype = sapply(ranks$vartype, paste, ".", sep='')
return(ranks)
}

```

```

# Function to return the names of variables within the same group
# correlated by less than 0.6
# Function variable group should be of form "rd1."
# Function output is a list of character class, names of variables

```

```

include_in_stepwise = function(ranks){
  vargroups = as.character(unique(ranks$vartype))
  stepwiselist = character()
  for (group in vargroups){
    maxr2 = max(ranks$r2[ranks$vartype == group])
    maxvar = ranks$covariates[ranks$r2 == maxr2]
    subdata = as.data.frame(ourdata[,grep(group, names(ourdata))])
    if (dim(subdata)[2] > 1){
      varcor = cor(subdata)[maxvar,]
      valid = c(maxvar, names(varcor)[varcor < 0.6])
    }
    else {valid = c(maxvar)}
    stepwiselist = c(stepwiselist, valid)
  }
  return(stepwiselist)
}

```

```

# Function to return a step-wise selected model from a set of eligible
# variables
# pollutant = pollutant type
# modeltype = "length" or "density"
# stepmodel = step-wise selected model of class lm.

```

```

get_stepwise_model = function(y, modeltype){
  ranks = get_ranks_table(y, modeltype)
  stepwiselist = include_in_stepwise(ranks)
  variablelist = c(y, stepwiselist)
  newdata = ourdata[,variablelist]
  basemodel = lm(newdata, na.action = na.omit)
  stepmodel = step(basemodel, trace = FALSE)
}

```

```

return(stepmodel)
}

```

The stepmodel would be updated to only include variables with coefficients in an a priori direction and with a p value < 0.05.

```

stepmodel = update(stepmodel, .~-dist.portln)
stepmodel = step(basemodel, trace = FALSE)
summary(stepmodel)

```

The LOO cross validation entailed scripting R to re-create the model at each site excluding its information from the site and then predicting the concentration of the pollutant at that site from the new model. The script is shown below.

```

# Leave-one-out

setwd("/Users/rebecca/Desktop/ufp")
ufp = read.table("pn_final.csv", header=TRUE, sep=",")

store<-matrix(nrow=78,ncol=3)

for (i in 1:78){
  ourdata=as.matrix(ufp)
  ourdata=ourdata[-i,]
  ourdata=as.data.frame(ourdata)
  attach(ourdata)

  #stepmodel=lm(pnc.50~dist.port.ln+newff.200+trk.50) # pnc_50 (length)

  store[i,1]=predict(stepmodel, newdata=ufp[i,])
  store[i,2]=ufp[i,"pnc.50"]
  store[i,3]=store[i,1]-store[i,2]

  detach(ourdata)
}

store=as.data.frame(store)

```

```
names(store)=c('predicts','measurements','diff')  
  
write.csv(store, file=paste('L00meas.csv',sep=''))  
  
mean(store$diff, na.rm=T)  
sd(store$diff, na.rm=T)  
cor(store$predicts,store$measurements)
```

Appendix E 2010 Ultrafine particle sampling campaign field notes

This appendix is located in the file “UFP_sampling_campaign_field_notes.xls” This spreadsheet contains all of the notes collected during the spring 2010 UFP sampling campaign. The tab “1. Field notes” simply contains all field notes received from research assistants. The tab “2. New variables” is where data from the notes was compiled to create four new variables (number of trucks counted, upwind or downwind, number of visible fast food sites nearby and presence of construction or abnormal activity), which were used in the best LUR model.

Appendix F Predictor variables for PNC LUR models ('classic buffer')

This appendix is located in the file “pnc_final.csv” This file contains the extracted geographic predictor variable values for all of the sites, as well as PNC_{50} , $\ln-PNC_{50}$, PNC_{10} and PNC_{90} values used to develop LUR models. The four variables added into the ‘best model’ (based on site observations) are included also. Category abbreviations are identified in Table 2.2.

Appendix G Predictor variables for PNC LUR model (based on measured wind rose data)

This appendix is located in the file “pnc_measure_wind.csv” This file contains the geographic predictor variable values for all sites (this is a subset of all variables; those included are listed in Table 2.3) extracted by wind rose shaped buffers based on wind data measured at the time of each sample.

Appendix H Predictor variables for PNC LUR model (based on monitored wind rose data from a central monitoring station)

This appendix is located in the file “pnc_mv_wind.csv” This file contains the geographic predictor variable values for all sites (this is a subset of all variables; those included are listed in Table 2.3) extracted by wind rose shaped buffers based on average wind data from a central monitoring station (T-18, South Burnaby) during the entire sampling period.