ADVANCING THE USE OF MOBILE MONITORING DATA FOR AIR POLLUTION MODELLING

ADVANCING THE USE OF MOBILE MONITORING DATA FOR AIR POLLUTION MODELLING

By Matthew D. Adams, HBESc, MES

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2015) McMaster University

(School of Geography and Earth Sciences) Hamilton, Ontario

TITLE: Advancing The Use Of Mobile Monitoring Data For Air Pollution Modelling

AUTHOR: Matthew David Adams

HBESc. (Lakehead University, 2009)

MES (Lakehead University, 2011)

SUPERVISOR: Professor Pavlos S. Kanaroglou

NUMBER OF PAGES: xii, pp. 128

Abstract

Air pollution is highly variable in both space and time, which presents many challenges to researchers when they wish to model concentrations. The modelling of air pollution is necessary for a number of reasons, which include the determination of human health effects, providing warning of health risks, and to understand general ecosystem health. In this thesis, modelling of air pollution through both space and time has been explored, with a focus on improving models that can be used to assign air pollution exposure. The techniques presented in this thesis have leveraged the ability of mobile monitoring units to collect air pollution concentration data multiple locations throughout a study period. First, we explore the use of combining mobile air pollution monitoring data with traditional fixed location monitoring. We find that the mobile data is able to provide insight into changes in spatial pattern between two temporal periods that could not be identified solely with the fixed location monitors, which demonstrates value in this monitoring approach that can be built upon with refinement of techniques. Second, we present a method to determine the amount of classical error that will be introduced when a long-term mean concentration is calculated from a discontinuous time-series dataset, which are the type of datasets collected by mobile air pollution monitoring. Third, we merge mobile and stationary air pollution monitoring data, along with meteorological, transportation, and land use information to model the hourly air pollution field using neural network models. The models developed allowed for the assignment of air pollution exposure incorporating human activity patterns. Also, they can be used to provide a spatially refined air quality health index. Lastly, we demonstrate exposure assignment that incorporates human activity patterns to calculate the dose exposure for students during their trips to school.

This work commences with a demonstration of the basic utility of mobile air pollution monitoring data, which is to increase the number of monitored locations. Building on that utility of mobile technology, a technique was developed to estimate the error when mobile units are used for long-term estimates, similar to stationary monitoring units; and we were able to provide guiding principles for mobile monitoring data collection. Furthering our objective, to better understand the value of mobile data in a fully integrated monitoring network, we utilized both mobile and stationary data collection techniques together, in a single model, to produce accurate estimates of an air pollution field on an hourly basis. Finally, the research culminates with the demonstration of how mobile monitoring can be used for activity based air pollution exposure estimates, which was shown with a case-study of students' trips between home and school. Overall, the chapters in this thesis work toward a better understanding of how to incorporate mobile monitoring data into air pollution assessment studies.

Acknowledgements

I would like to acknowledge my supervisor, Dr. Pavlos Kanaroglou for his support, guidance and mentorship that allowed me to accomplish this work. I am thankful to my thesis committee members Drs. Paulin Coulibaly, Antonio Páez and Niko Yiannakoulias, they have both challenged me intellectually and provided critical input during my PhD studies.

I would be remiss to not mention the member of the Centre for Spatial Analysis through the years, particularly Chris Higgins who I went through the PhD experience with, Laura Labate for providing endless administrative support, and Pat De Luca for providing technical support and great conversations about GIS.

To all the founding members of Geographers Without Borders, particularly Charles Burke and Justin Hall, hopefully, we have developed a lasting organization.

Lastly, I am forever thankful to my parents Judi and David Adams for their unwavering lifelong support in all my endeavours. And to Kerry Foley, my partner, for her relentless support during this four-year undertaking.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AQHI	Air Quality Health Index
AQI	Air Quality Index
CO	Carbon Monoxide
CT	Census Tract
EPA	Environmental Protection Agency
GIS	Geographic Information System
GTHA	Greater Toronto and Hamilton Area
IDW	Inverse distance squared weighting
MET	Metabolic Equivalent of Task
MOE	Ontario Ministry of the Environment and Climate Change
MVR	Minute Ventilation Rate
NO	Nitrogen Oxide
NO_2	Nitrogen Dioxide
NO_x	Nitrogen Oxides
O_3	Ozone
PM_{10}	Particulate matter 10 microns or smaller in aerodynamic diameter
$PM_{2.5}$	Particulate matter 2.5 microns or smaller in aerodynamic diameter
ppb	Parts Per Billion
RMSE	Root Mean Square Error
SO_2	Sulphur Dioxide

Preface

This thesis is presented as a composite of four substantive chapters, that are either accepted (Chapter 2 and 3), submitted (Chapter 4) or in preparation for publication (Chapter 5). The basis of each chapter including the hypothesis, experimental design and execution were developed and conducted by the lead author. The substantive chapters in their individual forms are:

Chapter 2: Adams, M.D., De Luca, P., Corr, D., and Kanaroglou, P. (2012). Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005-2010. Social Indicators Research 108(2): 351-364

Dr. Corr and Mr. De Luca provided input and support for data acquisition and preparation.

Chapter 3: Adams, M.D., and Kanaroglou, P.S. A Method for Reducing Classical Error in Long-Term Average Air Pollution. Editors: Kanaroglou, P.S., Delmelle, E., and Paez, A. *Edited Volume: Spatial Analysis in Health Geography*, Ashgate Publishing (In Press).

Chapter 4: Adams, M.D., and Kanaroglou, P.S. Development of real-time air pollution models with neural networks in a land use regression framework: Combining mobile and stationary air pollution monitoring to better understand the spatial variation of air pollution. (Under Review)

Chapter 5: Adams, M.D., Yiannakoulias, N., Kanaroglou, P.S. Are children who walk between school and home in lower socioeconomic status neighbourhoods inhaling higher doses of air pollution. (Prepared for Submission)

Dr. Yiannakoulias provided data on students' routes to school in Hamilton and also provided review and editorial advice for this manuscript.

Chapter 1: Introduction

1.1 Introduction

Air pollution is a global issue, with most regions of the globe affected by concentrations that are known to have negative health outcomes. For example, global estimates of particulate matter 2.5 microns or smaller in aerodynamic diameter ($PM_{2.5}$) suggest a population-weighted geometric mean concentration of 20 μg/m^3 (van Donkelaar et al., 2010). The anthropogenic component of $PM_{2.5}$ is estimated to account for 8, 12.8 and 9.4 percent of global mortality for cardiopulmonary disease, lung cancer and ischemic heart disease, respectively (Evans et al., 2013). The debate of air pollution as a carcinogen also ended during the preparation of this thesis, with the International Agency for Research on Cancer (World Health Organization) classifying air pollution a human carcinogen (International Agency for Research on Cancer, 2013).

Many government organizations (e.g. Environmental Protection Agency, USA; Environment Canada, Canada; World Health Organization, International) have set standards for air quality. These standards are continually refined as knowledge about the health effects improves, such as understanding the effects associated with short-term exposure (Brook et al., 2010). The refinement of these standards and the improved knowledge of the health effects is driven by improved exposure estimates both spatially and temporally. This need to refine exposure estimates is called for in the epidemiological literature, in particular the need to move away from central site monitors for exposure assignment (Baxter et al., 2013b; Ozkaynak et al., 2013).

Recently mobile monitoring technologies have become widely available. They include both units that can be set-up in different locations (semi-stationary), and truly mobile technologies that can monitor air pollution concentrations as they travel. Even autonomous mobile air pollution monitors have been developed (Reggente et al., 2010). Mobile monitoring of air pollution is promising to overcome some of the deficiencies of stationary monitoring units, such as the lack of spatial coverage; however, the data these mobile systems produce is different from the spatially static long-term time-series data that is generally produced from air pollution monitoring. This chapter will introduce the works that identified the importance of monitoring air pollution in space, move into an overview of how stationary monitoring networks are developed, define the objectives of this research, and then conclude with a description of the thesis contents.

1.2 Air Pollution Spatial Variability

Goldstein and Landovitz (1977a, 1977b), in a pair of papers, analyzed the air pollution field for New York City, NY, using the monitoring network, which consisted of 40 stationary monitoring units. Their initial research was to determine if a single monitor would be able to capture the spatial-temporal field across New York City. They assessed hourly SO_2 and bi-hourly smoke shade concentrations. The pair-wise correlation statistics between stations revealed a broad range of values from r = -0.7 to 1.0. Further analysis identified that high pair-wise correlations in one time period (season) did not result in high correlations for other periods, often even the sign of the correlation statistic would change (Goldstein and Landovitz, 1977a). These initial findings indicated that the then current practice of utilizing a single station in epidemiological air pollution research was unsuitable, and those results should be re-examined.

Building on these findings in their second paper, (Goldstein and Landovitz, 1977b) were interested to see if Tobler's first law of geography held; if stations nearer to each other demonstrated a higher correlation than those further apart. Simply put, their results were not what they expected, and stations that were nearer to each other did not demonstrate a stronger correlation than those further apart. As well, by comparing relative downwind stations and those stations with relative direction out of the wind path, the downwind stations did not possess stronger correlation than the stations that were not aligned. The two papers by Goldstein and Landovitz set the stage for increased methodological development for the siting of air pollution monitors. This knowledge has led us to the use of land use regression modelling in air pollution where distance between locations is not the critical factor, but that the surrounding land uses are the critical factor (D. J. J. Briggs et al., 1997).

Handscombe and Elsom (1982) analyzed 24-h concentrations from the Greater London Area Pollution Monitoring Network, applying correlation techniques similar to Goldstein and Landovitz (1977a, 1977b). In contrast to the American's findings, their data demonstrated a high correlation between monitors. It was suggested that the American monitors capture a greater proportion of larger particles (10 μm and greater) compared to the British monitors, which primarily capture finer particles, and these larger particles are more spatially heterogeneous. Their spatial heterogeneity occurs because they settle out of the atmosphere faster due to their increased weight. To effectively capture their spatial distribution a denser network would be required compared to finer particles, suggesting the

American network was not dense enough. Applying the correlation analysis between sites, they deemed many stations redundant. With a required correlation of $r = 0.9$, 19% of the stations could be removed from the network.

These three papers laid the foundation for the need to investigate and understand the particular location, pollutant and objectives of the study when an air pollution assessment is conducted.

1.3 Locating Air Pollution Monitors

Shortly following the publications of Goldstein and Landovitz; Lee et al. (1978) published an approach on how to locate air pollution monitors effectively. Citing that the "EPA has provided only subjective and often conflicting guidelines to local agencies. These local agencies must subjectively weigh such location criteria as pollutant concentration, source location and meteorological conditions, population density, growth projections and geographic coverage in determining instrument location, and must also resolve the question of sampling frequencies." The approach relied on an atmospheric simulation model to estimate the air pollution field for the study domain. The model's input data are annual point source data. The model predicted probabilities of violating a standard at any location on a grid. They took from the field of facility location problems and applied a discrete space model. This process partitioned the region into a n x n grid, potential monitors were located at the centroid of each cell. Applying a hierarchical interpreted rendering of the EPA guidelines, they determined the first monitoring priority is that monitors are to be located in areas with the highest air pollution concentrations; the second priority was to locate monitors near areas of high population density or growth. They formulated a mathematical programming problem that would identify the locations that maximize the identification of the number of expected standard violations. A second formula, which when maximized, indicated the monitor locations that would maximize the population that would be covered when monitors identified standard violations. They added a constraint for proximity to ensure all monitors were not located coincidentally. Their procedure maximized the first problem constrained by resources (number of available monitors) and the proximity constraint. Then they maximized the second problem with both the resource and proximity constraints along with a detection criterion constraint to ensure a similar solution to the first priority. The use of a dispersion model is common in air pollution monitoring network design to obtain potential realizations of the field from which monitors can be located (Mazzeo and Venegas, 2007; Zwack et al., 2011)

The identification of pollution standard violations is an effective objective when acute air pollution effects are of interest. Noll and Mitsutomi (1983) proposed a design approach to optimize monitor location based on dose. They applied a dispersion model to output hourly concentrations across a grid of receptor cells. At each 1-hour step in the model, they identified clusters of receptors that were contiguous and estimated to be above a prescribed concentration. Each cluster was then assigned a dose based on all receptor cells in the cluster; sum of area dosage. Grid receptors that occurred in a high number of clusters became potential monitoring locations. Each monitoring location was given an efficiency score based on the proportion of the total study areas' dose accounted for at each location. To determine which monitoring locations were chosen, the station with the highest efficiency was selected first. All the clusters it was associated with were removed, and efficiency scores were recalculated for all other potential monitoring locations. This algorithm was continued until the number of available monitors was reached. Their approach of focusing on dose may provide a monitoring network that is better optimized to estimate the health effects of long-term elevated air pollution concentrations across a study area, assuming the monitors end up being located in areas with sufficient population. Their technique did not include a population control. As well, areas with estimated low dosage would be avoided by this approach. However, areas of low air pollution concentrations are important for epidemiology to conduct statistical analysis with high power (Le and Zidek, 2006). The approach proposed by Noll and Mitsutomi (1983) is a greedy optimization technique. Greedy techniques are characterized by making the locally optimal choice at each iteration of the technique, which may not result in the global optima.

Another greedy optimization technique was proposed by Modak and Lohani in a set of three papers (Modak and Lohani, 1985a, 1985b, 1985c). In their first paper, they applied a spatial correlation approach, which first required a correlation cut-off used to indicate the degree of correlation that stations would be considered uncorrelated. The technique is based on a pattern score, which is the number of stations that are correlated to the station above the correlation cut-off. The algorithm they propose operates as:

(1) Select the station with the highest pattern score,
(2) Select the next station with the minimum overlap to the previous station based on pattern score,
(3) Select the next station with the minimum overlap to the previously selected stations,
(4) Stop if the entire domain is covered or if the number of available monitors is reached.

Temporal variation must be removed prior to this technique, it is also only applicable for locating monitors to a regular grid with no significant topological characteristics.

Modak and Lohani (1985b) extended their methodology to allow for multiple objectives. They retained pattern score as an objective and included a violation score, which was based on the number of air pollution occurrences above a set of standards. Each level in the standard was given a weight, with higher weights given to standards of higher concentrations. The violation score was the weighted sum of the violations. A utility score was derived based on the product of the pattern and violation score to use for the optimization.

Most air pollution monitoring networks have multiple objectives because often the monitoring network is developed to full-fill a government organization's mandate. There are also multiple end-users with different needs and a multitude of pollutants. Generally these multi-objective or multi-pollutant networks require a compromise between individual optimal solutions for pollutants or objectives (Chen and Xu, 2012). This compromise can be applied through weights applied to pollutants or objectives when solving the objective function (Chen et al., 2006), goal-programming (Chen and Xu, 2012), or providing a set of solutions (Pareto front) to choose from by adjusting the weights (Boix et al., 2011). All the approaches to multi-objective network design require a decision maker with the expertise to provide some input regarding the weighting of objectives.

Locating monitors using geostatistical techniques is very common. These monitoring networks formulate the objective function to locate monitors that will maximize data collection in areas of the domain with high-kriging variance (Baalousha, 2010; Nunes et al., 2006; Trujillo-Ventura and Ellis, 1991; Zimmerman, 2006). This approach can be applied prior to any monitors being located if a dispersion model is available or if historic data from a network can be used as input for kriging. By designing a network to minimize kriging variance, spatial interpolation will be improved. However, the areas of high kriging-variance may not exhibit the highest concentration, which is important if health effects are to be identified.

Kanaroglou et al. (2005) adopted a location-allocation model for the optimal location of a dense network of pollution monitors, which was to be used to assess exposure. They first developed a demand surface, which was based on an air pollution model calibrated from a network of monitors covering Ontario and then applied the model to the study area (Toronto, Ontario). The demand surface represented the spatial variability of pollutants, these values were calculated with a formula inspired by the semi-variogram equation. High values of the demand surface required increased monitoring. To this point, the method would be very similar to approaches that apply kriging variance. They modified this surface by increasing values where higher populations occurred. The location-allocation technique ensured they were able to identify locations that represent differences in land use and transportation network, as part of the initial model; and the distribution of the population. This approach can be modified to include any set of particular characteristics that are of concern. Combining population into the design of pollution monitoring is a common technique to ensure the network covers the population of the domain (Langstaff et al., 1987).

This overview of the approaches for locating air pollution monitors presents a number of different techniques with varying objectives. This review does not judge any particular technique as superior, only that with the number of approaches available no single approach can satisfy all current or future objectives.

1.4 Objectives of this Research

Recently, there has been a call for new techniques to model ambient air pollution concentrations from the epidemiological literature (Baxter et al., 2013b; Buonanno et al., 2013). In particular, the models should be able to estimate exposure with finer resolutions and be applicable for use with activity analysis data. For example, instead of assigning a long-term value at a single location, the model should be able to assign different exposures based on a person's location and activity throughout the day. One technique that has recently been adopted for the monitoring of air pollution concentrations, that may be able to help satisfy the need for improved modelling, is mobile air pollution modelling. Mobile monitoring systems are beginning to become widespread in air pollution studies (Adams et al., 2012; DeLuca et al., 2012; Ferri et al., 2010; Kanaroglou et al., 2013; Xu et al., 2007; Zwack et al., 2011). These mobile systems are able to collect air pollution concentrations at a variety of locations, which is unique compared to traditional stationary monitoring units.

The value of this mobile data has yet to be fully exposed because of the differences in the data collection processes, the monitoring techniques (a single mobile monitoring unit typically collects data with variation in both space and time) and the reduction in long-term single location data from these monitors. It is this author's hope that this research can contribute to the better understanding of air pollution modelling and exposure assignment through the inclusion of mobile monitoring techniques. This thesis considers the main objective through four lenses that comprise sub-objectives aligned to the main goal.

1.4.1 Objective One

The first objective is to determine if mobile air pollution monitoring in Hamilton, Ontario, Canada is able to identify changes in air pollution exposure that are not captured by the stationary monitoring network. This research will explore the use of mobile air pollution data with a standardization technique currently applied in the literature to determine where ambient air pollution concentrations are changing, identifying both increases and decreases in ambient air pollution using mobile air pollution monitoring data.

1.4.2 Objective Two

The second objective build on the techniques applied in the first objective to better understand how the adjustment of short-term mobile air pollution concentrations using fixed location stationary monitoring units leads to long-term estimates. This objective identifies the errors introduced when a discontinuous time-series of measurements are used to estimate a long-term mean concentration, which may be applied as the estimate of ambient exposure to air pollutants.

1.4.3 Objective Three

The third objective is to leverage both mobile and stationary air pollution monitoring to develop air pollution models that can operate in real-time to model air pollution surfaces with across the City of Hamilton, Ontario that include the spatial variation in air pollution concentrations. These models can be applied for real-time mapping of air pollution risk, as well since they model air pollution on an hourly basis they can be applied for estimating air pollution exposure with an activity based analysis.

1.4.4 Objective Four

The final objective is to conduct an activity-based air pollution exposure study in the Hamilton, Ontario region to demonstrate the utility of activity-based air pollution exposure modelling. Specifically looking at the dose of particulate matter air pollution for students during their trips to and from school when they use active transportation. This analysis relies on the air pollution model developed in objective three to assign activity and location specific air pollution concentrations and the resulting dose exposure.

1.5 Thesis contents

The thesis contains five chapters in addition to this introduction, comprised of four substantive chapters (either published or prepared for publication) and a concluding chapter exploring the contributions of this research and identifying future research. The four substantive chapters each align to one of the objectives in section 1.4 and form a body work that is coherent and advances the understanding of the utility of mobile air pollution monitoring data.

Chapter two examines the change in air pollutant concentrations between 2005 and 2010 occurring in the City of Hamilton, Ontario, Canada. After analysis of stationary air pollutant concentration data, we analyze mobile air pollutant concentration data. Air pollutants included in the analysis are CO, $PM_{2.5}$, SO_2, NO, NO_2, and NO_X. Stationary monitoring indicates a continuous reduction in air pollutant concentrations. Stationary monitors only cover a small spatial extent of Hamilton. Mobile monitoring of air pollutant concentrations, averaged over census tract boundaries, indicates both improvement and decline in air quality. These improvements and declines in air quality are spatially clustered throughout Hamilton. Mobile data indicated a significant decline in median pollutant concentration for CO, SO_2, $PM_{2.5}$, and NO_2; but a significant increase in NO and the resulting total NO_X concentrations. Air quality change in Hamilton is spatially heterogeneous and is not captured based on the current stationary monitoring network. Coupling of mobile and stationary air pollutant concentration monitoring provides a more accurate spatial assessment of local air quality.

Chapter three explores how air pollution observational epidemiologic studies that require long-term mean concentrations assigned to the subjects may be affected by both Berkson and classical error when using discontinuous time-series data. Researchers have begun to use mobile monitoring techniques that allow for rapid relocation of monitors to attempt to reduce Berkson, but in our study we demonstrate this approach creates classical error due to the incomplete dataset for the period of interest. A data adjustment method is presented that harnesses a fixed-location monitor's observations of the entire period of interest to adjust each datum in the incomplete dataset to reduce the classical error in the calculation of the observation's mean. A computer simulation is applied to test the adjustment method, where complete datasets of time-series observations from the monitoring network in Paris, France, are sampled to simulate incomplete collection. The classical error in the determination of the long-term mean is based on the actual long-term mean (using all observations); the error in calculating the mean concentration is determined for both the adjusted and unadjusted incomplete datasets. Three periods of interest are examined, which include one-week, one-season, and one-year. The computer simulation to generate incomplete samples varies the number of repeat observations during the period of interest and the length of each observation. We find the adjustment approach is beneficial when the incomplete dataset is comprised of less than one-quarter of the entire time-series of potential observations, and as the incomplete dataset approaches one-third of all potential observations in the period of interest, the adjustment approach may introduce additional error. As well, we find that it is best to balance the total number of sampling hours between the number of repeated observations and the length of each observation to minimize classical error. We conclude the article with suggestions to help aid researchers in designing mobile monitoring campaigns with minimized classical error.

Chapter four presents two models that capture the real-time spatial variation of air pollution in Hamilton, Ontario, Canada. We applied neural network models within a framework that is inspired by land use regression. Mobile air pollution monitoring campaigns were conducted across Hamilton from 2005 to 2013. These mobile air pollution data were modeled with a number of predictor variables that included information on the surrounding land use characteristics, the meteorological conditions, air pollution concentrations from fixed location monitors, and traffic information during the time of collection. The two pollutants that were modelled included Fine Particulate Matter and Nitrogen Dioxide. During the model fitting process, we reserved twenty percent of the data to determine the prediction validation. The models' performances were measured with a coefficient of determination at 0.78 and 0.34 for $PM_{2.5}$ and NO_2, respectively. We apply a relative importance measure to identify the importance of each variable in the neural network to partially overcome the black box issues of neural network models.

Chapter five presents an analysis of the $PM_{2.5}$ exposure of children in Hamilton, Ontario, Canada during their trips to and from school. Air pollution exposure was estimated with a neural network model using a land use regression approach, with a prediction accuracy of R^2 = 0.78. This model was built with a combination of air pollution data collected by both mobile and stationary monitoring units. The doses were calculated for 250 different students' routes to and from school for both cycling and walking as the mode of travel. During morning cycling trips the average dose was 2.17 μg (range 0.085 – 5.67 μg), afternoon trips were higher with a mean of 2.19 μg (range of 0.097 – 5.61 μg). Walking tips were higher with a mean dose of 3.19 μg (range 0.126 – 8.327) in the morning and 3.23 μg (range of 0.14 – 8.26 μg). Students' average household ambient $PM_{2.5}$ concentrations were 15.7 μg/m^3 and 15.6 μg/m^3 in the morning and afternoon respectively. The school concentrations were higher at 18.3 μg/m^3 and 19.3 μg/m^3 for the morning and afternoon respectively. Students living in lower income neighbourhoods were not found to have higher $PM_{2.5}$ doses during their trips to or from school with either mode of transportation. The primary policy implication from this work is for programs that encourage active transportation to and from school. If these programs can encourage more students to cycle versus walk they can reduce the students' exposure to air pollution.

Chapter six concludes the thesis by detailing the contributions of the work to the field of air pollution monitoring. This chapter also identifies future avenues for exploration that were identified within this document or were identified by the author during the completion of this work.

1.6 References

Adams, M. D., De Luca, P. F., Corr, D., & Kanaroglou, P. S. (2012). Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005–2010. *Social Indicators Research*, 108(2), 351–364. doi:10.1007/s11205-012-0061-5

Baalousha, H. (2010). Assessment of a groundwater quality monitoring network using vulnerability mapping and geostatistics: A case study from Heretaunga Plains, New Zealand. *Agricultural Water Management*, 97(2), 240–246. doi:10.1016/j.agwat.2009.09.013

Baxter, L. K., Dionisio, K. L., Burke, J., Ebelt Sarnat, S., Sarnat, J. A., Hodas, N., et al. (2013). Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *Journal of Exposure Science & Environmental Epidemiology,* 23(6), 654–9. doi:10.1038/jes.2013.62

Boix, M., Montastruc, L., Pibouleau, L., Azzaro-Pantel, C., & Domenech, S. (2011). A multiobjective optimization framework for multicontaminant industrial water network design. *Journal of Environmental Management*, 92(7), 1802–8. doi:10.1016/j.jenvman.2011.02.016

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., et al. (1997). Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science*. doi:10.1080/136588197242158

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V, et al. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation*, 121(21), 2331–78. doi:10.1161/CIR.0b013e3181dbece1

Buonanno, G., Stabile, L., & Morawska, L. (2013). Personal exposure to ultrafine particles: The influence of time-activity patterns. *The Science of the Total Environment*, 468-469C, 903–907. doi:10.1016/j.scitotenv.2013.09.016

Chen, A., & Xu, X. (2012). Goal programming approach to solving network design problem with multiple objectives and demand uncertainty. *Expert Systems with Applications*, 39(4), 4160–4170. doi:10.1016/j.eswa.2011.09.118

Chen, C.-H., Liu, W.-L., & Chen, C.-H. (2006). Development of a multiple objective planning theory and system for sustainable air quality monitoring networks. *The Science of the Total Environment*, 354(1), 1–19. doi:10.1016/j.scitotenv.2005.08.018

DeLuca, P. F., Corr, D., Wallace, J., & Kanaroglou, P. (2012). Effective mitigation efforts to reduce road dust near industrial sites: assessment by mobile pollution surveys. *Journal of Environmental Management*, 98, 112–8. doi:10.1016/j.jenvman.2011.12.023

Evans, J., van Donkelaar, A., Martin, R. V., Burnett, R., Rainham, D. G., Birkett, N. J., & Krewski, D. (2013). Estimates of global mortality attributable to particulate air pollution using satellite imagery. *Environmental Research*, 120, 33–42. doi:10.1016/j.envres.2012.08.005

Ferri, G., Mondini, A., Manzi, A., & Mazzolai, B. (2010). DustCart, a mobile robot for urban environments: Experiments of pollution monitoring and mapping during autonomous navigation in urban scenarios. *Environments*, 3(2).

Goldstein, I., & Landovitz, L. (1977a). Analysis of air pollution patterns in New York City—I. Can one station represent the large metropolitan area? *Atmospheric Environment*, 11(1), 47–52. doi:10.1016/0004-6981(77)90205-0

Goldstein, I., & Landovitz, L. (1977b). Analysis of air pollution patterns in New York City—II. Can one aerometric station represent the area surrounding it? *Atmospheric Environment*, 11(1), 53–57. doi:10.1016/0004-6981(77)90206-2

Handscombe, C., & Elsom, D. (1982). Rationalisation of the national survey of air pollution monitoring network of the United Kingdom using spatial correlation analysis: a case-study of the greater London. *Atmospheric Environment*, 16(5), 1061–1070.

International Agency for Research on Cancer. (2013). IARC: Outdoor air pollution a leading environmental cause of cancer deaths. *World Health Organization*, No. 221.

Kanaroglou, P., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R. (2005). Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment*, 39(13), 2399–2409. doi:10.1016/j.atmosenv.2004.06.049

Kanaroglou, P. S., Adams, M. D., De Luca, P. F., Corr, D., & Sohel, N. (2013). Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model. *Atmospheric Environment*, 79, 421–427. doi:10.1016/j.atmosenv.2013.07.014

Langstaff, J., Seigneur, C., Mei-Kao, L., Behar, J., & McElroy, J. L. (1987). Design of an optimum air monitoring network for exposure assessments. *Atmospheric Environment - Part A General Topics*, 21(6), 1393–1410. doi:10.1016/0004-6981(67)90087-X

Lee, T., Graves, R., & McGinnis, L. (1978). A procedure for air monitoring instrumentation location. *Management Science*, 24(14), 1451–1461.

Mazzeo, N. A., & Venegas, L. E. (2007). Design of an Air-Quality Surveillance System for Buenos Aires City Integrated by a NO x Monitoring Network and Atmospheric Dispersion Models. *Environmental Modeling & Assessment*, 13(3), 349–356. doi:10.1007/s10666-007-9101-y

Modak, P. M., & Lohani, B. N. (1985a). Optimization of ambient air quality monitoring networks (Part I). *Environmental Monitoring and Assessment*, 5(1), 1–19. doi:10.1007/BF00396391

Modak, P. M., & Lohani, B. N. (1985b). Optimization of ambient air quality monitoring networks (Part II). *Environmental Monitoring and Assessment*, 5(1), 21–38. doi:10.1007/BF00396392

Modak, P. M., & Lohani, B. N. (1985c). Optimization of ambient Air Quality Monitoring Networks (Part III). *Environmental Monitoring and Assessment*, 5(1), 39–53. doi:10.1007/BF00396393

Noll, K. E., & Mitsutomi, S. (1983). Design methodology for optimum dosage air monitoring site selection. *Atmospheric Environment* (1967), 17(12), 2583–2590. doi:10.1016/0004-6981(83)90086-0

Nunes, L. M., Caeiro, S., Cunha, M. C., & Ribeiro, L. (2006). Optimal estuarine sediment monitoring network design with simulated annealing. *Journal of Environmental Management*, 78(3), 294–304. doi:10.1016/j.jenvman.2005.04.024

Ozkaynak, H., Baxter, L. K., Dionisio, K. L., & Burke, J. (2013). Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science & Environmental Epidemiology*, 23(6), 566–72. doi:10.1038/jes.2013.15

Reggente, M., Mondini, A., Ferri, G., & Mazzolai, B. (2010). The dustbot system: Using mobile robots to monitor pollution in pedestrian area. *Proc. of NOSE*, 23, 273–278. doi:10.3303/CET1023046

Trujillo-Ventura, A., & Ellis, J. H. (1991). Multiobjective air pollution monitoring network design. *Atmospheric Environment. Part A.*, 25(2).

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, 118(6), 847–855. doi:10.1289/ehp.0901623

Xu, X., Brook, J. R., & Guo, Y. (2007, November). A statistical assessment of saturation and mobile sampling strategies to estimate long-term average concentrations across urban areas. *Journal of the Air & Waste Management Association* (1995).

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17(6), 635–652. doi:10.1002/env.769

Zwack, L. M., Hanna, S. R., Spengler, J. D., & Levy, J. I. (2011). Using advanced dispersion models and mobile monitoring to characterize spatial patterns of ultrafine particles in an urban area. *Atmospheric Environment*, 45(28), 4822–4829. doi:10.1016/j.atmosenv.2011.06.019

Chapter 2: Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005–2010

2.1 Introduction

Numerous air pollution studies demonstrate the detrimental effects of various air pollutants on human health; effects which result in an adverse impact on life expectancy and quality of life (Neupane et al., 2010; Sanhueza et al., 2010; Pope et al., 2009; Medina-Ramon et al., 2006; Kan et al., 2010; and Chiusolo et al., 2011). Pope et al., (2009), for example found that a decrease by10 $\mu g/m^3$ in the mean annual concentration of particulate matter under 2.5 microns in aerodynamic diameter ($PM_{2.5}$) is associated with an increase in life expectancy of 0.77 years. Similar health impacts from increased air pollution were found in the Netherlands (Brunekreef 1997), Finland (Nevalainen and Pekkanen 1998), and Canada (Coyle et al., 2003). Thus, reducing air pollution in urban areas has become an important initiative of governments globally including those of Canada (Environment Canada 2011a).

Air pollutant concentrations are typically measured with stationary air quality monitors, often few in number due to high initial and continued maintenance costs, which results in a sparse spatial coverage of monitors. Ontario, Canada has a stationary, ambient air quality monitoring network with 37 stations in southern Ontario, operated by the Ontario Ministry of the Environment and Climate Change (MOE). Three of those stations are located in the City of Hamilton; these stations are invaluable for their long-term record of pollution. Unfortunately, the three Hamilton monitors are clustered near the core of the city, and they do not capture the spatial contrasts of the urban-suburban or industrial-commercial-residential environments. Vardoulakis et al., (2005) have demonstrated that the concentrations recorded by stationary monitors may not reflect the values of surrounding areas, and, therefore, may not be adequate in assessing population exposure.

Mobile monitoring techniques can evaluate air quality while in transit, using specialized vehicles. Air is collected during transit and analyzed onboard for pollutant concentrations. Mobile monitoring techniques are a powerful addition to air quality data obtained from stationary monitoring networks, because of the ability to move to many collection locations. The complementarities between mobile and stationary monitors allow for a more detailed picture of air pollution impacts to be constructed. Data collection mobility is particularly important because short-term peak exposures to air pollution can have serious detrimental health impacts (Atkinson et al., 2006; Dominici et al., 2006; Pope et al., 2006). For example, the mobile unit can roam city-wide (Wallace et al., 2009); or focus on specific locations of concern such as areas with high amounts of road dust (DeLuca et al., 2012) and areas undergoing temperature inversions (Wallace et al., 2010).

Given the links between air pollution, health, and overall quality of life, we investigate the change in air quality occurring in the City of Hamilton over a six-year period, from 2005 to 2010. Specifically, we demonstrate that stationary monitors, while sufficient for examining long-term trends, are insufficient in capturing all spatial variability throughout the city. Further, we demonstrate how information from stationary monitors can be combined with data from mobile monitors to provide a more detailed assessment of the variability of several pollutants. In terms of specific pollutants, we examine the concentrations of sulphur dioxide (SO_2), which is related primarily to heavy industry;

nitrogen oxide (NO), nitrogen dioxide (NO_2), total nitrogen oxides (NO_X) and carbon monoxide (CO), which is linked primarily to traffic; and $PM_{2.5}$ which is tied to both industry and traffic but is largely attributed to open sources; approximately 72% of all emissions. Open sources include road dust resuspension, construction, and farming (Environment Canada 2011b). Each of the pollutants examined have both acute and chronic health effects (Kampa and Castanas 2008; Pope and Dockery 2006).

2.2 Study Area

The City of Hamilton (Hamilton), with a population of 520 000 (Statistics Canada 2012) is situated at the western end of Lake Ontario (43.3^{o}N, 79.9^{o}W), and is separated into an upper and lower city by the Niagara Escarpment (average height ~90 m). Several suburban satellite villages are incorporated in the larger urban area, suburban sprawl. Four major expressways encapsulate most of the city proper (Figure 2.1). Transportation corridors include highway and railroad traffic between the Greater Toronto Area and the United States of America; and shipping flows, which initiate and terminate at Hamilton Harbour.

Between 2005 and 2010, Hamilton has witnessed the downsizing of its primary metals industry; a shutdown of one of its major steel manufacturers; and the closure of several other its industries, which were located in the traditional industrial zone of the city (City of Hamilton, 2010). Suburbanization and diversification of employment occurred with the establishment of several new industrial zones throughout the city. These zones are defined as Bayfront, East Hamilton, Stoney Creek, Red Hill, Ancaster, Flamborough, West Innovation District, and the Airport/Airport Employment Growth District zones (City of

Hamilton, 2011). Zones are indicated on a map of Hamilton in Figure 2.1. These newly developed zones increase the potential for spatial variation in pollutant concentrations. Variability which may be attributed to increased and new traffic between zones, construction-related pollutants, and the redevelopment of traditionally rural land to an urban land form.

Hamilton's air quality is of interest due to the known past spatial variability of air pollutant concentrations (Wallace et al., 2010; Wallace et al., 2009; Sahsuvaroglu et al., 2006; Jerrett et al., 2001; Finkelstein et al., 2004; Finkelstein et al., 2003; Buzzelli et al., 2003). This variability is partially attributed to: (1) Meteorological conditions including northeast lake effect winds passing over the traditional industrial core and depositing pollution in the city, and atmospheric inversion conditions resulting in pollutant buildups, particularly in the lower city (Wallace et al., 2010); (2) Pollutant releases from 143 facilities spread throughout the city (Environment Canada 2011c); and (3) Emissions from vehicular traffic.

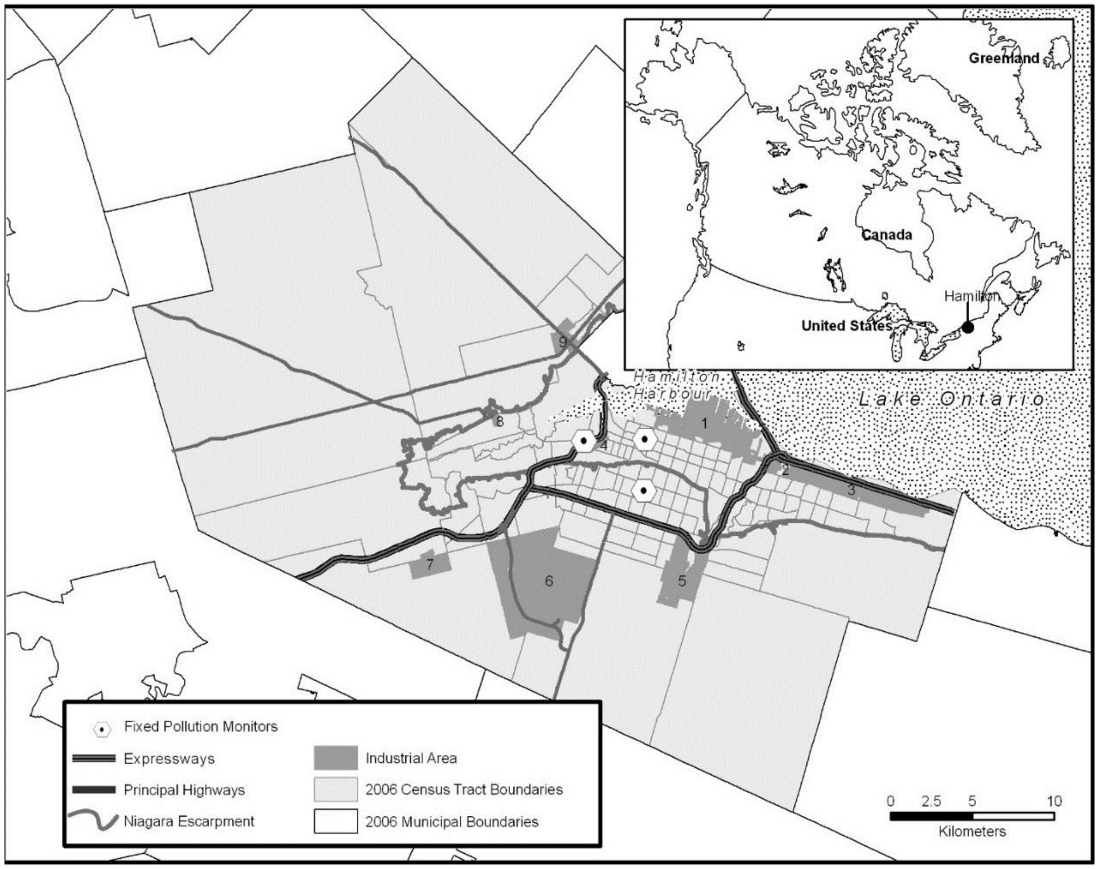

Figure 2.1 The City of Hamilton study area map. Industrial areas of the city are numbered: 1 - Bayfront, 2 – East Hamilton, 3 – Stoney Creek, 4 – West Innovation District, 5 – Red Hill, 6 – Airport Employment Growth District, 7 – Ancaster, 8 – Dundas, 9 – Flamborough.

2.3 Methods

2.3.1 Mobile Monitoring

Mobile monitoring methods are fully described in Wallace et al., (2009). An industrial van equipped to measure CO, SO_2, $PM_{2.5}$, NO, NO_2 and NO_X, traversed Hamilton collecting air samples through a roof-mounted air intake (~3 m above ground level). Air samples were analyzed onboard with three different air quality monitoring instruments. The make and model, the principle of operation, operating range, and precision of the continuous pollution monitoring instrumentation are presented in Table 2.1. Two global positioning system (GPS) units collected positional information: (1) A roof-mounted Garmin GPS16-HVS detector; and (2) A windshield mounted Garmin 18 GPS. Pollutant and positional data were simultaneously logged with a Campbell 23X data logger and stored in an integrated database with 1-second temporal resolution. Mobile data collection occurred between speeds of 5 and 25 km per hour.

Table 2.1 Mobile air quality monitoring instrumentation

Pollutant	Instrument Type	Principle of Operation	Operating Range	Precision
NO, NO_2, NO_X	Thermo Scientific Model 42i	Chemi-luminescence	0 – 1000 ppb	0.4 ppb
CO	Thermo Scientific Model 48	Gas Filter Correlation	0 – 50 ppm	0.1 ppm
SO_2	Monitor Labs 8850	Fluorescence	0 – 10000 ppb	5 ppb
$PM_{2.5}$	Grimm Model 1.107	Laser Optical	0 – 6,500 μg/m^3	1 μg/m^3

2.3.2 Air Pollutant Data from Stationary Monitors

Six years, January 1st, 2005 to December 31st, 2010, of hourly averaged air pollutant concentration data from stationary monitors were obtained from the Ontario Ministry of the Environment's Historical Air Pollution Data Base (OMOE 2011). These stationary monitors are active automated monitors reporting average air quality on an hourly basis. Pollutant concentrations included: CO (ppm), SO_2 (ppb), $PM_{2.5}$ (μg/m^3), NO (ppb), NO_2 (ppb), and NO_X (ppb). Three stationary monitors were located in Hamilton: Hamilton Downtown, Hamilton Mountain, and Hamilton West (see Figure 2.1 for locations). Hamilton Downtown was the only station with a complete dataset for all pollutants. $PM_{2.5}$

data were available for all stations. SO_2, NO, NO_2, and NO_X data were available for Hamilton Mountain beginning January 2007.

All three Hamilton Air quality monitors are located in close proximity; the maximum distance between any two monitors is less than 5 km. Both the Hamilton Downtown and Hamilton West monitors are located in the lower city, and the Hamilton Mountain monitor is located in the upper city. The Hamilton Mountain monitor is located to the direct south of the Hamilton Downtown monitor. And the Hamilton West monitor is located to the direct west of the Hamilton Downtown monitor. During north-east wind conditions the monitors would receive winds passing over the traditionally industrial zone of Hamilton. North-east winds are the secondary winds in Hamilton; none of the monitors are located on the path to receive pollutant concentrations of a north-east wind passing through the new industrial zones. South-west winds are the primary winds.

2.3.3 Air Pollution Data from Mobile Surveys

Mobile pollution survey campaigns occurred during 16 days in 4 months for 2005-2006 and 21 days in 5 months for 2009-2010, between November and April for both biennial groups. No data collection occurred in March or April during 2005-2006 surveys, or January during 2009-2010 surveys. Surveys for both time periods included a combination of city-wide and targeted area scans. Targeted areas included industrial sectors, the downtown core, specific neighborhood studies, and along major traffic corridors.

All mobile air pollutant concentration data collected in each biennial grouping were used in producing models. To reduce variation in data due to seasonality effects, data were standardized. Data standardization followed a similar approach to Larson et al., (2009), to remove seasonal variation. The formula for standardization is presented in equation 1.

$$MS_t = MO_t \times \frac{S2Y_t}{SD_t} \tag{1}$$

where,
MS_t = standardized mobile value at time t,
MO_t = original mobile value at time t,
$S2Y_t$ = biennial period mean from the MOE Downtown Stationary monitor that time t falls in,
SD_t = daily mean from the MOE Downtown stationary monitor that time t falls in.

This standardization was applied to increase/decrease values relative to the two-year mean collected under conditions of generally lower/higher pollutant concentrations. Standardization was based only on the Hamilton Downtown stationary monitor values because it was the only stationary monitor with a complete dataset.

2.4 Statistical Methods

Descriptive statistics are presented separately for both the mobile and stationary data. Mobile and stationary data were each grouped into two biennial groups, which were 2005-2006 and 2009-2010. Both mobile and both stationary biennial groups were distributed log-normal, as expected for air pollution data. We report median values as the central tendency when comparing between biennial groups because it is an appropriate measure of central tendency for a log-normal distribution.

For both mobile and stationary data, we determined if significant increases or decreases occurred between the first and second biennial periods for Hamilton's air pollutant concentrations, without spatial considerations, with the Wilcoxon rank sum two-sample unpaired test. This test determines if a significant difference in the median values between biennial groups occur. The Wilcoxon test was chosen over the t-test because of the log-normal distributions and zero data; zero data resulted from either no air pollutants occurring in the air, or values below the detection limit of the monitoring instrumentation. Zero data do not allow for a natural logarithm to be taken to transform the data to approximate a normal distribution.

In understanding if significant differences, in stationary monitoring data, occurring between biennial periods were discrete or continuous phenomena, linear models were applied to determine the amount of explanatory power time has to changes in concentration measured at the Hamilton Downtown stationary monitor. If differences were discrete, no downward or upward trend in air pollutant concentrations would be significant. Stationary monitor pollutant concentrations were averaged with a 30-day running mean to capture long-term trends without averaging out seasonal or yearly variability. These running means were calculated using the running mean function from the package igraph (Csardi and Nepusz 2006) for the statistical software environment R, which was used for all non-spatial statistics (R Development Core Team 2011). The running averages were plotted against time, and linear regression models were estimated based on the Hamilton Downtown station's air pollution data. Linear models included time, measured in hours, as the independent variable and the natural logarithm of the pollutant's running mean as the dependent variable. The natural logarithm was used to transform data; this could be taken because the running mean removed all zero values.

Inverse distance squared weighting (IDW) interpolation was used to estimate air pollutant surfaces from the mobile data for each pollutant, in both biennial periods. IDW was applied because these values are irregularly spaced, and air pollution is a continuous phenomenon. IDW is appropriate because we previously minimize seasonality concerns with the standardization, and samples are spaced relatively near. We limit our evaluation of the interpreted surface to the areas which initially contained monitoring points; i.e. no extrapolation. Model accuracy was assessed as the agreement (correlation) between the mobile monitored location's value and the corresponding interpolated surface value. Spearman's rho was used to assess correlation between estimated and actual values because data were log-normal distributed with many zero values present.

Mean air pollutant concentration within each census tract (CT) was computed from the surfaces. Only CTs with mobile air pollution data for both biennial periods were included in further analysis. For each CT, the difference between biennial periods for air pollutant concentrations were calculated, either as an increase or decrease in concentration to the latter period. To determine if global clustering of differences in air pollutant concentrations were occurring, we applied Moran's I to the CTs positive and negative values of air quality change. Local Moran's I was applied *post hoc* used to identify specific significant spatial clustering in CT values. Local significant spatial clustering is an indication of spatial variability of air pollutant concentrations throughout Hamilton.

2.5 Results

Stationary monitor (Hamilton Downtown) air pollutant concentrations, decreased for all pollutants during the study period. Biennial periods' median values for all pollutants were statistically significantly lower in the latter period. Biennial periods' median values for stationary monitors are presented in Table 2.2.

Table 2.2 Downtown Hamilton air pollution monitor descriptive statistics

Pollutant	Years	Mean	S.D.	Min	Max	Median*
CO	05-06	0.31	0.20	0	2.83	0.27
(ppm)	09-10	0.21	0.18	0	5.02	0.18
SO2	05-06	5.03	7.99	0	93	2
(ppb)	09-10	3.32	6.50	0	136	1
PM2.5	05-06	9.76	9.3	0	82	7
(µg/m3)	09-10	7.48	6.72	0	64	5
NO	05-06	9.16	20.57	0	393	3
(ppb)	09-10	5.66	11.9	0	179	2
NO2	05-06	18.36	10.75	1	77	16
(ppb)	09-10	13.41	8.93	0	59	11
NOX	05-06	27.72	28.30	2	470	19
(ppb)	09-10	18.85	18.47	0	231	13

*All median values between biennial groupings were statistically significant using the Wilcoxon test with p < 0.001, rejecting the null hypothesis of the distributions of both groups being equal.

Linear regression models, with mean air pollutant concentration at the Hamilton Downtown monitor as the dependent variable and time as the independent variable, were all statistically significant ($p < .05$), with explanatory abilities between 10% and 35% of pollutant concentration variation. The explanatory abilities of the models were based on a continual decline during the period; indicated by all negative coefficients for time (hour). Air pollution decline is continuous during the study period at the Hamilton Downtown stationary monitor. Adjusted R^2 and coefficient values are presented with running mean plots in Figure 2.2.

Table 2.3 Descriptive statistics for normalized mobile monitoring data. Normalized by Hamilton Downtown daily station average for CO, SO2, PM2.5, NO, NO2 and NOX.

Pollutant	Years	N	Mean	S.D.	Median*
CO	05-06	2097	1.24	1.57	0.96
(ppm)	09-10	6974	1.40	5.43	0.71
SO2	05-06	2178	24.87	33.18	13.71
(ppb)	09-10	2178	8.73	11.00	4.43
PM2.5	05-06	1187	32.67	43.23	23.96
(μg/m3)	09-10	2384	18.46	14.83	14.96
NO	05-06	2197	39.18	58.26	21.2
(ppb)	09-10	7309	76.75	91.49	42.62
NO2	05-06	1913	20.86	16.99	17.06
(ppb)	09-10	5017	15.96	15.45	12.46
NOX	05-06	2036	52.27	48.74	39.34
(ppb)	09-10	8048	71.84	70.95	44.94

*All Median Values between biennial groupings were statistically significant using the Wilcoxon test with $p < 0.001$, rejecting the null hypothesis of distributions equality.

Mobile sampled air pollutant concentration data were statistically significantly increased and decreased when comparing the first to the latter biennial period, depending on which air pollutant concentration was evaluated. Air pollutant concentrations median values are presented in Table 2.3. CO, SO_2, $PM_{2.5}$, and NO_2 were all significantly lower ($\alpha = 0.001$) in the latter biennial period. NO and NO_X were both significantly higher ($\alpha = 0.001$) in the later biennial period.

Mobile data based IDW estimated pollutant surfaces' accuracy were high. Spearman's rho values were between 0.646 – 0.844, indicating a strong agreement between the monitored mobile values and those estimated at each monitored location from the IDW pollutant surfaces.

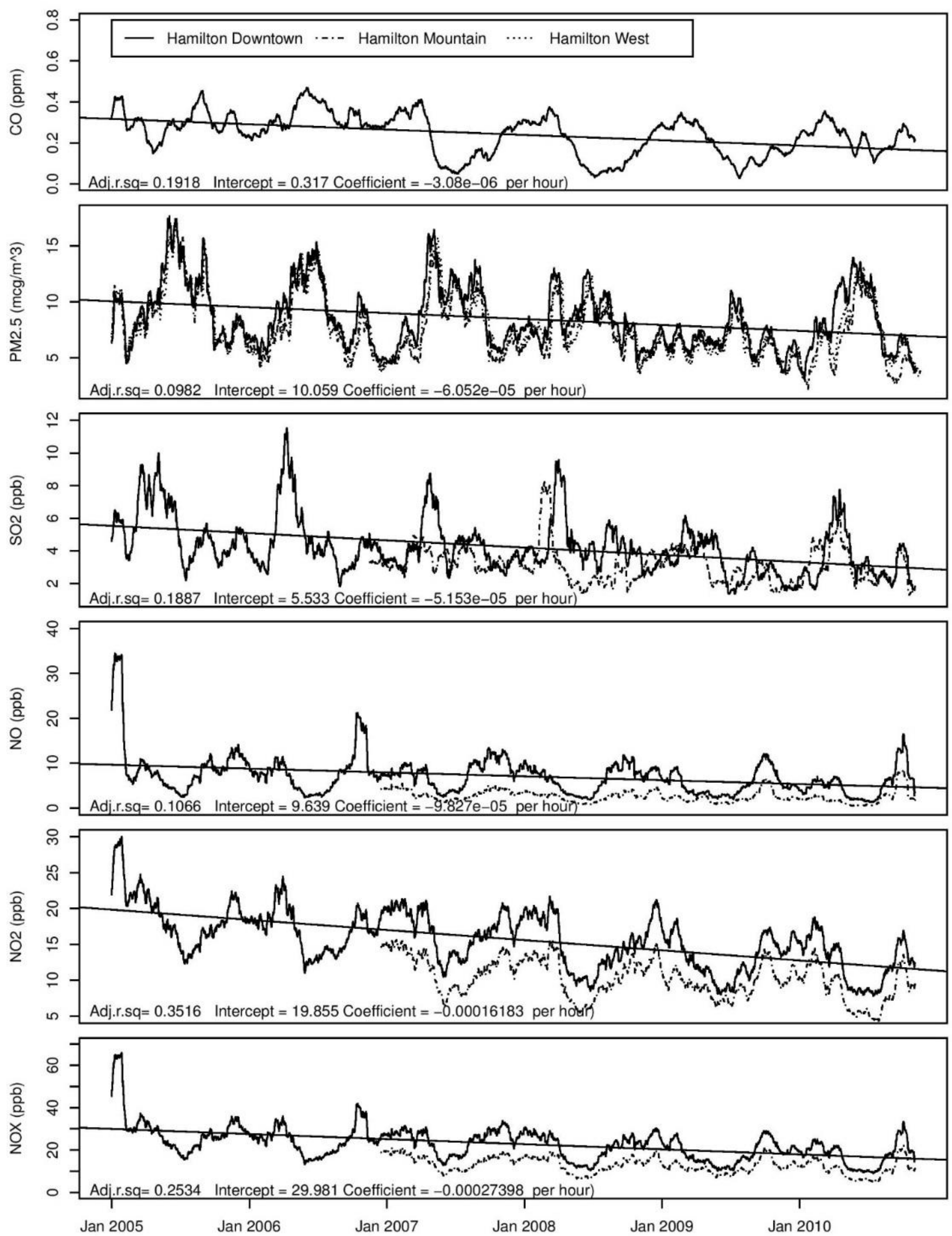

Figure 2.2 Hamilton's air pollution 30-day running averages. SO_2, $PM_{2.5}$, CO, NO, NO_2 and NO_X concentrations from the stationary monitors in Hamilton.

Census tracts' mean air pollutant concentrations, estimated with mobile data, were heterogeneous in decline and improvement between 2005-2006 and 2009-2010. CTs overall, decline significantly in median pollutant concentration for CO, SO_2, $PM_{2.5}$, and NO_2; but increase for NO and NO_X. The numbers of CTs which increase and decrease in mean air pollutant concentrations, in the latter biennial period, are presented in Table 2.4. These locations of CT air quality improvement and decline are mapped for Hamilton in Figure 2.3.

Table 2.4 Descriptive statistics for change in mean/median pollution level comparing 2005-2006 to 2009-2010 for the census tract level of geography.

Pollutant	N	Improved	Declined	Mean	Median	Moran's I
CO	67	45	22	-0.637	-0.374	0.373*
SO_2	67	53	14	-4.878	-4.77	0.786*
$PM_{2.5}$	50	43	7	-7.573	-8.635	0.442*
NO	89	18	71	26. 25	19.45	0.511*
NO_2	87	53	34	-1.166	-2.052	0.402*
NO_X	64	31	33	4.686	1.702	0.503*

*$p < 0.001$

Global Moran's I indicates statistically significant spatial clustering of the CTs that increase or decrease in the latter biennial period for all pollutants. Moran's I statistics are presented in Table 2.4. Applications of local Moran's I indicate statistically significant local clustering of increased and decreased air pollutant concentrations at the CT geographic level. CTs that are significantly locally clustered for each pollutant, based on local Moran's I, are mapped for Hamilton in Figure 2.4.

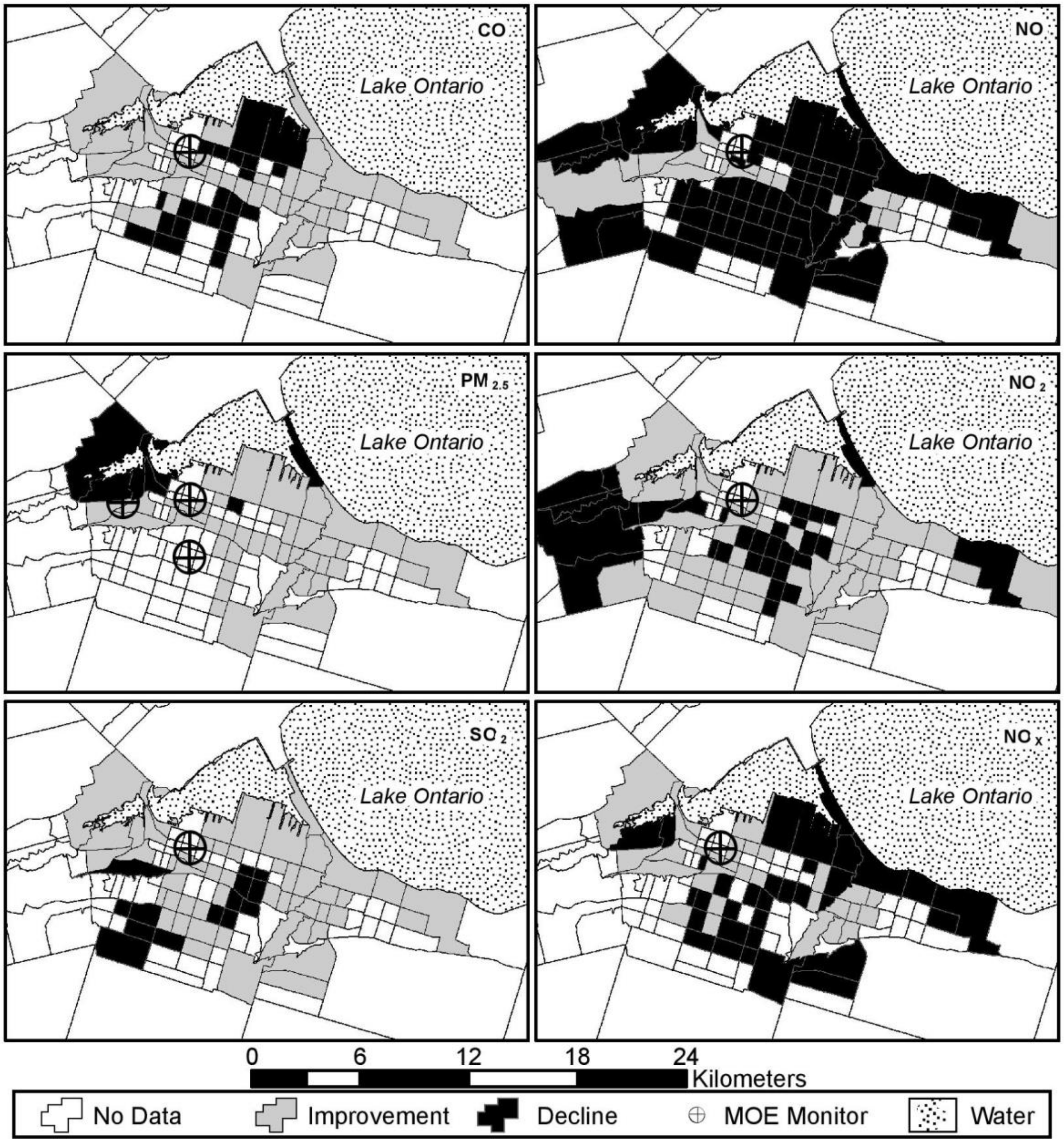

Figure 2.3 Locations of air quality improvement and decline between 2005-2006 and 2009 - 2010 at the census tract level of geography. MOE stationary monitors are included for each map, only those monitors with full data sets for the period between January 2005 and December 2010 are included.

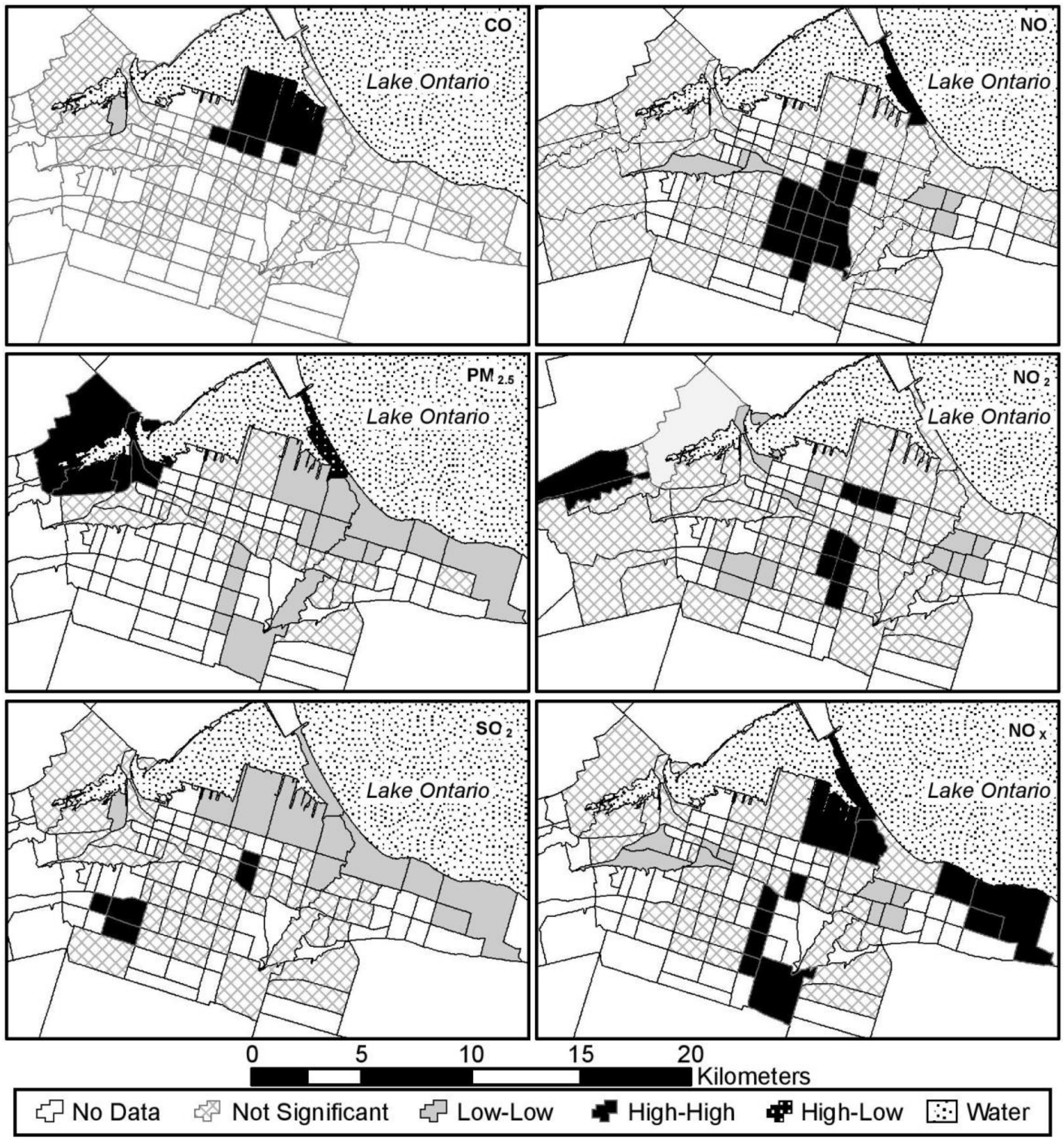

Figure 2.4 Spatial clustering of air quality improvement and decline between 2005-2006 and 2009 - 2010 at the census tract level of geography. Significant clusters were determined with the local Moran's I.

Epidemiological, environmental justice, and spatial variability studies require a much larger area to be interpolated than is available for Hamilton, an area that generally covers the exposed population. To help with this issue in Hamilton, an NO_2 study employed short-term passive monitors (n = 100). These passive monitors were distributed throughout the city and left in place for a period of time and then analyzed to determine local pollutant accumulation. Unfortunately passive monitors are subject to damage/vandalism, which occurred in about 5% of the monitors initially deployed for the Hamilton NO_2 study (Sahsuvaroglu et al., 2006). Mobile monitoring allowed a more extensive area of Hamilton to be validly interpolated, with convex hulls generated by the sample points covering areas between 96 and 207 km^2, depending on the pollutant. Additionally, spatially diverse sample collection occurred covering the complexities of the urban terrain of Hamilton, similar to Zwack et al., (2011).

Stationary monitoring indicated air quality improvement from 2005 - 2010, whereas mobile monitoring indicated areas that experience both improvement and decline thereby corroborating other results on the spatial heterogeneity of air pollutant concentrations in Hamilton (Wallace et al., 2010; Wallace et al., 2009; Sahsuvaroglu et al., 2006; Finkelstein et al., 2004; Buzzelli et al., 2003; Finkelstein et al., 2003; Jerrett et al., 2001). Local Moran's I tests indicate significant local clustering in all air pollutant concentrations evaluated at the CT level. City trends for CO, SO_2, $PM_{2.5}$, and NO_2 air quality indicate improvement, but 33%, 21%, 14%, and 39% of the census tracts respectively show a decline in air quality. NO and NO_X city-wide trends indicate a decline in air quality, with 20% and 48% of the census tracts improving in air quality respectively. Hamilton's census tracts with declining air quality would not have been quantified based solely on the current stationary monitors, which indicated air quality improvement over the time period.

CO and NO_X pollutant concentration increases were significantly clustered in Hamilton's Bayfront industrial area, see Figure 2.4. CO pollutant releases, from industrial sources in this area, declined by one-third between 2005 and 2010, which is in opposition to the increased concentrations; NO_2 releases from industrial sources do not change significantly (Environment Canada 2011c). Seven of the days monitored within 2009 – 2010, experienced offshore north-east winds, winds which would pass over the industrial area pushing pollutants towards the city. These elevated concentrations are likely due to CO and NO_X being brought towards the city during these north-east wind conditions.

Elevated NO_X levels occurred in the eastern most part of the city. This area is a recent suburban satellite community with continued housing development during the study period. These elevated NO_X levels may be derived from the increasing vehicular traffic that couples with the increased population. Elevated NO, NO_2, and NO_X concentration clusters occur to the south and south-west of the Bayfront industrial region. This may be partially attributable to wind patterns, and similar to the eastern part of the city, some of these CTs have increased in population from 2006 – 2011(Statistics Canada 2012). The increased population will increase local traffic and traffic congestion resulting in increased NO, NO_2, and NO_X.

Significant reductions of SO_2 and $PM_{2.5}$ occur throughout most of the waterfront area, which can be attributed to economic decline and localized industrial departure. Two pockets indicate an overall increase in SO_2, both located to the south-west of the industrial waterfront. In each pocket, the sampling was more intense in the second time period, with seven of the twenty-one days of monitoring occurring with north-east winds passing over the industrial area. Elevated $PM_{2.5}$ concentrations clustered in Dundas Valley, to the west of the Bayfront industrial area. This area consists of a growing community with significant housing construction and is surrounded by other growing communities to the north, north-east, south and west. Additionally, a major highway runs along the eastern and southern parts of the cluster of census tracts. Dundas valley also often suffers from increased air pollutant concentrations due to temperature inversions (Wallace et al., 2010).

2.6 Conclusions

Mobile monitoring coupled with stationary monitoring provided a more detailed spatial distribution of Hamilton's air quality. Values from various mobile monitoring scans of the city, standardized to the Hamilton Downtown stationary monitor, allow for the use of multi-season data to be utilized, with minimal variation due to seasonal fluctuations. These data may be used for epidemiological, environmental justice, and local spatial variation studies. Mobile data collection allows for air pollutant concentrations to be obtained with a higher spatial distribution and density than is possible with stationary or passive monitors. Lastly, the comprehensive collection of mobile air pollutant data allowed for the estimation of varying concentrations within Hamilton. This more comprehensive collection resulted in the determination that the current stationary monitors are inadequate in distribution to quantify the change in air quality for the entire city of Hamilton, Ontario, Canada.

2.7 References

Atkinson, R., Cohen, A.J., Carrington, J.C., & Anderson, H.R. (2006). Worldwide Spatial Variation in Daily Mortality Relative Rates Due to Short-Term Exposure to Air Pollution: A Meta-Analysis. *Epidemiology*, 17(6), pS19.

Buzzelli, M., Jerrett, M., Burnett, R., & Finklestein, N. (2003). Spatiotemporal perspectives on air pollution and environmental justice in Hamilton, Canada, 1985–1996. *Annals of the Association of American Geographers*, 93(3), 557–573.

Brunekreef, B. (1997). Air pollution and life expectancy: is there a relation? *Occupational And Environmental Medicine*, 54, 781-784.

Chiusolo, M., Cadum, E., Stafoggia, M., Galassi, C., Berti, G., Faustini, A., Bisanti, L., Vigotti, M.A., Dessi, M.P., Cernigliaro, A., Mallone, S., Pacelli, B., Minerba, S., Simonato, L., & Forastiere, F. (2011). Short term effects of nitrogen dioxide on mortality and susceptibility factors in 10 Italian cities: the EpiAir study. *Environmental Health Perspectives*, 119, 1233-1238.

City of Hamilton (2010). Hamilton Economic Development Strategy, 2010 – 2015. http://www.investinhamilton.ca/wp-content/uploads/2011/06/Hamilton-EcDev-Strategy2010.pdf. Accessed 2012.

City of Hamilton (2011). One Billion – Economic Development Review 2010. http://www.investinhamilton.ca/onebillion/. Accessed 2012.

Coyle, D., Stieb, D., Burnett, R.T, DeCivitab, P., Krewskic, D., Chen, Y., & Thun, M.J. (2003). Impact of particulate air pollution on quality-adjusted life expectancy in Canada. *Journal of Toxicology and Environmental Health, Part A: Current Issues*, 66(16-19), 1847- 1863.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research, *InterJournal Complex Systems* 1695. http://igraph.sf.net Accessed 2012.

DeLuca, P.F., Corr, D., Wallace, J., & Kanaroglou, P. (2012). Effective mitigation efforts to reduce road dust near industrial sites: assessment by mobile pollution surveys. *Journal of Environmental Management* 98,112-118.

Dominici. F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L., Samet, J.M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, 295(10),1127–1134.

Environment Canada (2011a). Acts, Regulations and Agreements. http://www.ec.gc.ca/default.asp?lang=En&n=48D356C1-1 Accessed September 2011.

Environment Canada (2011b). Criteria Air Contaminant Emission Summaries, http://ec.gc.ca/pdb/websol/emissions/ap/ap_query_e.cfm Accessed Jan 2012.

Environment Canada. (2011c). National Pollutant Release Inventory (NPRI) Facility Location Table. http://www.ec.gc.ca/inrp-npri/default.asp?lang=En&n=20DE1DC2-1 Accessed Feb 2012.

Finkelstein, M.M., Jerrett, M., & Sears, M.R. (2004). Traffic air pollution and mortality rate advancement periods. *American Journal of Epidemiology*,160, 173-177.

Finkelstein, M.M., Jerrett, M., DeLuca, P., Finkelstein, N., Verma, D.K., Chapman, K., & Sears, M.R. (2003). Relation between income, air pollution and mortality: a cohort study. *Canadian Medical Association Journal*, 169(5), 397-402.

Hamilton Air Monitoring Network 2010. Annual Air Quality Report 2010. http://www.hamnair.ca/uploads/HAMN-2010-Annual-Report.pdf Accessed January 2012.

Jerrett, M., Burnett, R.T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., & Brook J.R. (2001). A GIS - environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, 33, 955–973.

Jerrett, M., Burnett, R.T., Brook, J., Kanaroglou, P., Giovis, C., Finkelstein, N., & Hutchison, B. (2004). Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, Canada. *Journal of Epidemiology and Community Health*, 58, 31–40.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, 151, 362-367.

Kan, H., Wong, C.M., Vichit-Vadakan, N., & Quan, Z. (2010). Short-term association between sulfur dioxide and daily mortality: The Public Health and Air Pollution in Asia (PAPA) study. *Environmental Research*, 110, 258-264

Larson, T., Henderson, S.B., & Brauer, M. (2009). Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environmental Science & Technology*, 43, 4672-4678.

Medina-Ramón, M., Zanobetti, A., & Schwartz, J. (2006). The effect of ozone and PM10 on hospital admissions for pneumonia and chronic obstructive pulmonary disease: a national multicity study. *American Journal of Epidemiology*, 163, 579–588.

Neupane, B., Jerrett, M., Burnett, R.T., Marrie, T., Arain, A., & Loeb, M. (2010). Long-term exposure to ambient air pollution and risk of hospitalization with community-acquired pneumonia in older adults. *American Journal of Respiratory and Critical Care Medicine*, 181, 47-53.

Nevalainen, J., & Pekkanen, J. (1998). The effect of particulate air pollution on life expectancy. *Science of the Total Environment*, 217(1-2), 137-141

OMOE. Ontario Ministry of the Environment, 2011. Historical Air Quality Pollutant Data. http://www.airqualityontario.com/science/data_sets.php Accessed December 2011.

Pope III, C.A., Ezzati, M., & Dockery, D.W. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine*, 360. 376-386.

Pope III, C.A., & Dockery, D.W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, 56, 709-742.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Sahsuvaroglu, T., Arain, A., Kanaroglou, P., Finkelstein, N., Newbold, B., Jerrett, M., Beckerman, B., Brook, J., Finkelstein, M., & Gilbert, N.L. (2006). A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario, Canada. *Journal of the Air & Waste Management Association*, 56, 1059–69

Sanhueza, P., Pizarro, J., Vargas, C., Torreblanca, M., & Passalacqua, M. (2010). Health risk estimation due to carbon monoxide pollution at different spatial levels in Santiago, Chile. *Environmental Monitoring and Assessment*, 167, 165-173.

Statistics Canada. (2012). Population and dwelling counts, for Canada and census subdivisions (municipalities), 2011 and 2006 censuses. http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=301&S=3&O=D Accessed 2012.

Wallace, J., Corr, D., DeLuca, P., Kanaroglou, P., & McCarry, B. (2009). Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring*, 11(5), 998-1003.

Wallace, J., Corr, D., & Kanaroglou, P. (2010). Topographic and spatial impacts of temperature inversions on air quality using mobile air pollution surveys. *Science of the Total Environment*, 21, 5086-5098.

Vardoulakis, S., Gonzalez-Flesca, N., Fisher, B.E.A., & Pericleous, K. (2005). Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. *Atmospheric Environment*, 39(15), 2725-2763.

Zwack, L.M., Paciorek, C.J., Spengler, J.D., & Levy, J.I. (2011). Modeling Spatial Patterns of Traffic-Related Air Pollutants in Complex Urban Terrain. *Environmental Health Perspectives*, 119(6), 852-859.

Chapter 3: A Method for Reducing Classical Error in Long-Term Average Air Pollution Concentrations from Incomplete Time-Series Data

3.1 Introduction

Air pollution exposure negatively affects human health including reduced cognitive function (Hutter et al., 2013), reduced lung function (Wallner et al., 2012), early childhood cancer (Ghosh et al., 2013), increased low and underweight births (Padula et al., 2012), and mortality and morbidity due to cardiovascular and respiratory diseases (Hoek et al., 2013). Observational epidemiologic studies that use ambient air pollution concentrations assigned to research subjects are the primary method of determining the association between long-term air pollution exposure and health effects. Hoek et al., (2013) reviewed long-term air pollution exposure studies of health effects and cardio-respiratory mortality, including such research studies as the Harvard Six Cities, American Cancer, German Cohort, California Teachers and the Nurses' Health Study. These studies were fundamental in identifying possible associations and health effect outcomes to varying levels of air pollution.

Current air pollution studies assess exposure over space, because air pollution in cities is long-known to vary spatially and that a single value or monitor is not representative (Goldstein & Landovitz, 1977a, 1977b); however, the number of air pollution monitoring stations is typically limited, often due to their high-cost, and concentration data must be spatially interpolated to provide values for each individual's location. Kriging or land-use regression modelling are commonly employed to spatially interpolate air pollution concentrations at unmonitored locations (Jerrett et al., 2005; Kanaroglou et al., 2005; Kumar, 2009). The effectiveness of these spatially refined estimates is dependent on a well-designed spatial sampling approach, which, in general, should maximize the probability of capturing the spatial variability (Delmelle, 2014). For a primer on spatial sampling see Delmelle (2014), who introduces two-dimensional sampling, geostatistical sampling, and second-phase sampling. More in-depth information on spatial sampling for a spatially correlated phenomena can be found in (Griffith, 2005).

Two types of error, Berkson and classical, affect observational epidemiologic studies (Heid et al., 2004). Classical error occurs when multiple measurements, commonly in time-series data, do not represent the true value of interest because of the type of monitoring strategy. In other cases, this occurs because of missing values in the time series. These effects may combine to bias the estimated effect measure upwards or downwards (Armstrong 1998). Berkson error reduces the study's power, which results in increased confidence intervals on the coefficients. This error occurs when subjects are assigned a group-average exposure (Armstrong 1998), for example by assigning all residents within 2 km of a pollution monitor the same value.

Attempts to reduce classical error (in estimating a long-term mean) caused by incomplete datasets have included adjusting data values based on a fixed-location continuous monitor (Larson et al., 2009; Adams et al., 2012; Kanaroglou et al., 2013; Hoek et al., 2002). This practice entails determining if air pollution conditions in the study area are above or below average by comparing the current concentration at a fixed-location continuous monitor to its long-term mean. The incomplete data are then adjusted to account for the above or below average conditions. After the adjustment, all observations in the incomplete dataset should be closer to the true long-term mean, which when averaged together should reduce classical error in estimating the long-term mean concentration. The adjustment method requires that regional phenomena affect air pollution concentrations uniformly in the study area. If this occurs, when one monitor is observing higher than average concentrations, all monitors should be observing higher than average concentrations. Thus, if monitoring times were biased, in that, data were collected mainly during above or below average conditions, this bias would be reduced.

Researchers increasingly use mobile monitoring to collect data on air pollution concentrations (Adams et al., 2012; Kanaroglou et al., 2013; Larson et al., 2009; Reggente et al., 2010); this concentration data is often collected to supplement existing monitoring networks that have few air pollution monitors. Mobile monitoring is different from traditional fixed-location continuous air pollution monitoring because the monitors are designed for rapid relocation, which researchers take advantage of in an attempt to reduce Berkson error by monitoring at various locations. This strategy produces incomplete time-series datasets that are prone to give rise to classical error.

Our study evaluates an adjustment formula, which is designed to reduce classical error when one uses incomplete time-series datasets to estimate a long-term mean concentration. Evaluation is based on a set of incomplete datasets derived through sampling from a database of observations obtained with continuous fixed-location monitors. The error is determined by calculating the actual long-term mean, which is estimated from the entire time-series, with both the adjusted and unadjusted incomplete datasets. It is critical that researchers understand the effect that mobile monitoring may have on the assignment of ambient air pollution concentrations to subjects in their research and identify any approaches to reducing Berkson and classical error. Understanding the effect of the adjustment method will help ensure appropriate study designs.

3.2 Methods

3.2.1 Observed Air Pollution Data

The Paris, France, air pollution monitoring network was selected for study, which consists of 67 air pollution monitors. We focus on observations from 2012, which was a year that Paris' air pollution concentrations often exceeded guidelines for particulate matter, nitrogen dioxide, ozone, and benzene; along with highly variable weather conditions. In January to March, meteorological conditions were conducive to escalated air pollution episodes. Pollution concentrations reduced in the next months because of cool and wet weather, which continued through the autumn (Airparif, 2013). The temporal variability of Paris' air pollution concentrations in 2012 provides us with a suitable dataset for testing the proposed adjustment method, because, without temporal variability, even incomplete datasets would adequately represent the long-term mean, and no adjustment would be necessary.

This study examined the effect of the adjustment method of particulate matter 10 microns or smaller in aerodynamic diameter (PM_{10}) concentrations, which were observed at 24 locations with fixed-location continuous monitors and reported as hourly averages. We present a map of the locations in Figure 3.1. Monitors were located in the following four land-use types: urban, peri-urban, rural, and transportation focused locations. All monitors in the network are within 60 km of the central monitor (PA04C), located in the city centre, which will be used to determine any adjustments to the data.

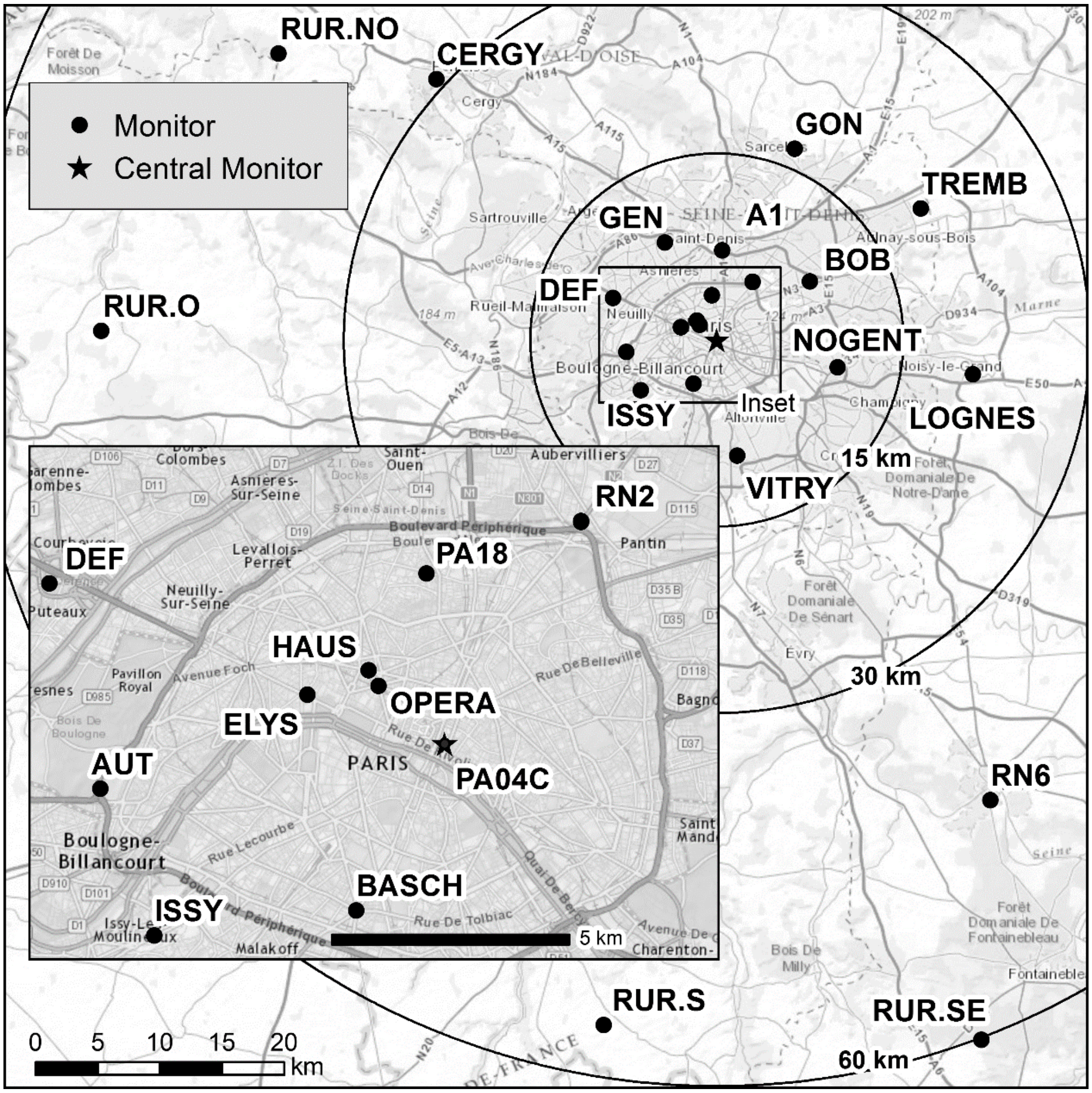

Figure 3.1 Paris, France study area and air pollution monitors. Monitors identified by their ID. Circles with radii of 15, 30 and 60 km are included and centre on the Paris Centre monitor. An inset is included of the downtown region.

During 2012, each fixed monitor recorded 8,784 hourly-observations. Missing, erroneous, or incomplete data in these time-series ranged between 1.6 percent and 7.4 percent (mean = 3.3 percent). The network's central monitor's missing data were filled by down filling with the previous hours' values. Down-filling was chosen over spatial interpolation for this monitor to not increase its correlation with the other monitors; increased correlation would increase the datasets similarity and artificially increase the effectiveness of the adjustment method. When a datum in the time-series was missing for the central monitor, the down filling would replace the missing value with the previous hour's record. The other stations' missing data were filled by spatially interpolating a value with the other monitoring stations' data, excluding the central location. The spatial interpolation method was inverse distance squared interpolation.

3.2.2 Time-Series Correlation

The adjustment method requires that monitors in the study area be temporally correlated. Without correlation between monitors, the application of this or a similar adjustment method would be in vain. We identify the level of correlation with Pearson's r correlation coefficient, which is calculated between each monitor and the adjustment monitor (Paris Centre monitor). Natural logarithms of the data were used because the data were distributed log-normal. To investigate if the correlation between monitors were associated with the distance between monitors or the station type, we regressed the correlation coefficients against distance to the Paris Centre monitor while controlling for the different monitoring land use types. If the distance in the model was significant, it would indicate the adjustment approach is biased based on distance to the central monitor.

3.2.3 Adjustment method

In our adjustment method, we term the fixed-location continuous monitor used to adjust the incomplete datasets as the adjustment monitor. Our adjustment method is a linear adjustment defined with by equation 2.

$$O_A = O_R \div \frac{log_e(S_h + e)}{log_e(S_L + e)} \tag{2}$$

Where, O_R is the air pollution observation to be adjusted, O_A is the adjusted air pollution observation, S_H is the concurrent hourly observation at the adjustment monitor, and S_L is the long-term arithmetic mean at the adjustment monitor. We add the base of the natural logarithms (e) as a constant to all values in the adjustment monitor's dataset to ensure the lowest value after the logarithm is taken is not less than one. Zero values in S_h would be indivisible, and the adjustment formula would fail. Paris Centre is used as the adjustment monitor because of its central location in the city. We limit the influence of extreme values when adjusting data by using log-values of the adjustment monitor's data.

3.2.4 Computer Simulation to Generate Incomplete Observations

We test our adjustment approach with a computer simulation to generate incomplete observations of air pollution time-series data. These incomplete data were obtained by sampling the Paris, France, time-series data. We analyzed the adjustment method for three periods of interest, which included one-week, one-season, and one year; one-season and one-year are general periods of interest in epidemiological studies, and one-week was chosen to explore the method with a shorter period. The geometric mean was chosen to represent the long-term mean for a period of interest, because, the monitoring stations' data were distributed log-normal, and it is a better expected value of the data than the arithmetic mean.

Mobile monitoring that collects incomplete time-series data consists of two parameters that detail how sampling is conducted at a single location, which include the number of repeated observations (the number of times a monitor is set-up at a location) and the length of each observation. Regular relocation is common in mobile monitoring (Adams et al., 2012; Kanaroglou et al., 2013). Our computer simulation generated the incomplete observations varying those two parameters. We first stipulated that the total number of sampling hours (total sample hours) be less than one-third of the total hours in the period of interest, which ensures that the monitor would be able to observe three locations during the period of interest. For the one-week simulation, samples lengths included 1, 2, 4, 8, 16, and 32 hours. For the one-season (2,184 hours) simulation, sample lengths included additional sample lengths of 64, 128, 256 and 512 hours. The one-year (8,784 hours) simulation included all the sample lengths used for one-week and one-season with the addition of 1,024 and 2,048 hours. Sample counts began at one and were doubled until the total number of hours sampled would be greater than one-third of the period of interest.

The process of the computer simulation for the selection of incomplete datasets follows:

(1) Define the simulation parameters:

   a. Sample Length (*SL*)
   b. Sample Count (*SC*)
   c. Period of Interest (*POI*)

(2) Randomly choose a time-period equal to the length of the *POI*. For the one-year simulation, this step is skipped as one-year cannot be varied. The one-week and one-season periods of interest are selected by a time-period with the correct number of continuous hours, in that, they could begin at any hour within the year with sufficient hours remaining.

(3) Choose one monitor at random and select data for the time-period, excluding the Paris Centre monitor.

(4) Calculate the long-term mean from the entire dataset obtained in 3.

(5) Select *SC* samples of *SL* length without repetition from the dataset obtained in step 3 to generate the incomplete sample.

(6) Apply the adjustment method to the incomplete sample.

(7) Determine the percent error in estimating the long-term mean for both the adjusted and unadjusted incomplete sample.

For each combination of sample length, sample count, and period of interest, we repeated the simulation 50,000 times.

3.2.5 Statistical Analysis

We determined if the adjustment method reduced classical error using statistical analysis. The statistical significance comparisons were conducted with the student's *t*-test comparing the adjusted error and unadjusted error calculated for each of the 50,000 simulations. This was conducted for each combination of the period of interest, sample length, and sample count; $\alpha = 0.05$. Air pollution monitoring data and the error data were distributed log-normal; appropriate transformations were used to satisfy the assumptions of the statistical tests. Throughout the results, we refer to the percent error in calculating the long-term mean concentration from the unadjusted incomplete dataset or the adjusted incomplete dataset as the unadjusted error and adjusted error, respectively. All statistical analysis and simulations were conducted in R (R Core Team, 2013).

3.3 Results

3.3.1 Monitor Correlation

The minimum correlation between any monitor with the Paris Centre monitor was $r = 0.7$, with a maximum of $r = 0.91$, and a mean $r = 0.83$ (s.d. = 0.05). Table 3.1 presents the pairwise correlations between all monitors with the adjustment monitor. It also includes the Euclidean distance between each monitor and the adjustment monitor, and each monitor's minimum, maximum, and geometric mean air pollution concentrations.

A linear regression model with the dependent variable of correlation between each monitor and the adjustment monitor, using the predictor variables of (1) distance between the monitors and (2) dummy variables for each of the land-use types, identified only one significant variable, which was rural land use; rural monitors' correlation with the Paris Centre monitor were significantly lower ($p < 0.05$) than the other monitors. The distance between any monitor to the Paris Centre monitor was not a significant factor in the linear regression model. Further analysis excluded the rural monitors because of their significantly lower correlations to the Paris Centre monitor, and research suggests that rural and urban ambient air pollution should be examined separately because of different causal factors and resulting air pollution conditions (Pedersen et al., 2013).

The one-week period of interest simulations demonstrated, for all combinations of sample lengths and counts, significantly reduced the classical error for calculating the long-term mean from incomplete datasets. Table 3.2 presents the unadjusted classical error and the amount of reduction in the classical error by applying the adjustment method.

Table 3.1 Descriptive statistics of Paris' air pollution monitors. Correlation and distance to the Paris Central Monitor, the monitor's land-use type, and the minimum, maximum and geometric mean values for their time-series of data.

	Pearson's *r*	Distance to Central Paris Monitor (m)	Type	Concentration Min	Max	Geo-Mean
PA18	0.91**	3,611	Urban	1	166	21.7
VITRY	0.88**	9,437	Urban	1	172	21.5
ISSY	0.88**	7,326	Urban	1	128	20.4
OPERA	0.87**	1,851	Transportation	1	282	29.7
DEF	0.87**	8,984	Urban	0	172	21.9
BASCH	0.87**	3,972	Transportation	4	154	35.8
NOGENT	0.86**	9,968	Urban	0	108	17.9
GON	0.86**	16,609	Urban	0	132	21
GEN	0.86**	8,880	Peri-Urban	0	180	21.5
BOB	0.86**	8,850	Urban	1	144	20.7
TREMB	0.85**	19,528	Peri-Urban	2	122	21.1
HAUS	0.84**	2,229	Transportation	0	165	28.9
LOGNES	0.83**	20,782	Urban	1	134	19.3
ELYS	0.83**	3,072	Transportation	3	174	35.4
CERGY	0.83**	30,790	Urban	0	126	20.1
RN2	0.82**	5,496	Transportation	3	220	35.6
RN6	0.8**	43,046	Transportation	0	245	26.6
AUT	0.8**	7,316	Transportation	6	423	43.2
RUR.O	0.79**	49,491	Rural	1	101	18.5
RUR.SE	0.76**	60,163	Rural	0	105	16.8
RUR.NO	0.74**	42,115	Rural	0	95	13.9
RUR_S	0.73**	55,823	Rural	1	107	17.3
A1	0.7**	7,236	Transportation	7	282	49.4
PA04C	N/A	0	Urban	2	154	23.7

$p < 0.01$ **

The one-season and one-year periods of interests' statistical evaluations identified that the adjustment method did not always significantly reduce the mean, in some cases additional error was introduced. Results for the one-season period of interest are found in Table 3.3, and the one-year results are presented in Table 3.4.

Table 3.2 Error results from the computer simulation for the one-week period of interest. The classical error when estimating the long-term mean from the unadjusted incomplete dataset, and the reduction in classical error from the adjustment method is included in parenthesis. All reductions were statistically significant ($p < .05$).

		Sample Length					
		1	2	4	8	16	32
Sample Count	1	24.76 (7.77)	24.22 (8.21)	23.29 (8.93)	21.14 (8.54)	18.64 (8.02)	15.25 (6.07)
	2	18.76 (5.36)	18.29 (5.7)	17.28 (5.92)	15.3 (5.15)	12.55 (3.84)	
	4	15.9 (4.72)	15.19 (4.72)	13.83 (4.28)	11.68 (3.07)		
	8	16.63 (5.95)	15.14 (5.53)	12.65 (3.92)			
	16	17.45 (7.24)	14.61 (5.78)				
	32	15.69 (6.54)					

Table 3.3 Error results from the computer simulation for the one-season period of interest. The classical error for estimating the long-term mean from the unadjusted incomplete dataset is presented, and the reduction in classical error from the adjustment method is included in parenthesis. Cells with grey backgrounds are not statistically significantly different, and cells with black backgrounds are when the adjustment method significantly increased the error.

Sample Count \ Sample Length	1	2	4	8	16	32	64	128	256	512
1	66 (17)	64 (18)	63 (20)	60 (20)	55 (19)	51 (17)	45 (12)	37 (6)	31 (2)	**21 (-4)**
2	50 (10)	49 (11)	47 (11)	44 (10)	41 (9)	38 (6)	32 (2)	**26 (-2)**	**19 (-6)**	
4	39 (5)	39 (5)	37 (5)	35 (5)	32 (2)	28* (0)	**23 (-4)**	**17 (-6)**		
8	38 (6)	37 (6)	36 (5)	32 (3)	27* (0)	**23 (-3)**	**16 (-6)**			
16	42 (11)	40 (9)	36 (7)	31 (2)	**25 (-1)**	**17 (-5)**				
32	44 (13)	39 (9)	33 (4)	27* (0)	**17 (-5)**					
64	41 (11)	34 (5)	27* (1)	**18 (-5)**						
128	34 (5)	28 (1)	**18 (-5)**							
256	28 (1)	**18 (-5)**								
512	**18 (-5)**									

$p > 0.05$*

Table 3.4 Error results from the computer simulation for the one-year period of interest. The classical error for estimating the long-term mean from the unadjusted incomplete dataset is presented, and the reduction in classical error from the adjustment method is included in parenthesis. Cells with grey backgrounds are not statistically significantly different, and cells with black backgrounds are when the adjustment method significantly increased the error.

Sample Count \ Sample Length	1	2	4	8	16	32	64	128	256	512	1024	2048
1	333 (61)	332 (67)	340 (71)	311 (70)	309 (74)	285 (63)	270 (50)	235 (31)	210 (24)	160* (-5)	124* (-5)	87 (-7)
2	260 (32)	255 (33)	252 (38)	251 (40)	229 (32)	214 (26)	196 (15)	170* (-2)	136 (-15)	96 (-22)	69 (-17)	
4	215 (31)	210 (33)	211 (36)	203 (35)	193 (29)	175 (21)	154 (10)	121* (-6)	92 (-23)	55 (-24)		
8	243 (104)	235 (103)	228 (101)	230 (99)	216 (91)	179 (76)	150 (49)	111* (5)	62 (-18)			
16	306 (191)	289 (175)	269 (162)	273 (154)	218 (127)	179 (89)	136 (26)	67 (-13)				
32	303 (196)	296 (192)	286 (163)	225 (134)	202 (104)	151 (34)	67 (-12)					
64	314 (206)	298 (173)	231 (138)	201 (103)	154 (38)	70 (-9)						
128	306 (183)	235 (139)	196 (92)	159 (38)	71 (4)							
256	230 (135)	192 (85)	161 (38)	73 (5)								
512	194 (86)	160 (38)	70* (2)									
1024	159 (36)	69* (-1)										
2048	69* (-2)											

p > 0.05*

3.4 Discussion

A city's diverse urban structure and local meteorology create spatially varying air pollution concentrations, which is one causal factor for different levels of ambient air pollution exposure across the population. Another factor is a person's movement throughout the city in a day. This issue of spatial variability has been known for a while, and that research concluded that a single monitor cannot represent an entire city (Goldstein

& Landovitz 1977a; Goldstein & Landovitz 1977b). Currently, many monitoring locations are established in a city to capture the variability of air pollution exposure. The air pollution data obtained from the Paris, France, monitoring network were spatially variable with yearly-mean concentrations ranging from 14 $\mu g/m^3$ to 50 $\mu g/m^3$, which allowed us an effective study of the adjustment method. We feel these results are generalizable because of the similarity in concentration to many of areas, such as Greece (Sfetsos & Vlachogiannis 2010; Grivas & Chaloulakou 2006), Italy (Badaloni et al., 2013), Germany (Liu et al., 2013), in general Western Europe (Vienneau et al., 2013), Canada (Brook et al., 1997) and the United States (Samet et al., 2000).

The Paris, France monitors were, for the most part, temporally correlated, satisfying the main requirement of the adjustment method. Four primary land use types exist in this monitoring network, which included rural, peri-urban, urban, and transportation; rural monitors were significantly lower in their correlation with the Paris Centre monitor. Transportation related pollution does not affect these areas significantly compared to urban areas, because of the low population and traffic density in rural regions. Our removal of these locations aligns to the thought that in epidemiological studies rural and urban areas should be assessed independently because of differing air pollution concentrations and respiratory health and exposure factors (Pedersen et al., 2013).

Recently, mobile monitoring technologies have been used to study the variability of air pollution in a city (Kanaroglou et al., 2013; Larson et al., 2009). The incomplete datasets that are observed with mobile monitoring may introduce classical error; data adjustment methods have been applied with the purpose to reduce classical error (Larson et al., 2009; Adams et al., 2012; Kanaroglou et al., 2013; Hoek et al., 2002). We identify in our research that many different sampling parameters affect the amount of error for incomplete datasets, which include the number of samples obtained, and the length of the samples. Every simulation indicated that incomplete time-series datasets exhibit classical error, which occurred when either the unadjusted or adjusted observations were used to estimate the long-term mean concentration. When we compare a particular combination of sample count and sample length across all three periods of interest, the amount of error is positively correlated with the length of the period of interest, for example, an incomplete observation of two samples of two continuous hours resulted in 18 percent, 49 percent, and 255 percent average error for the unadjusted data for one-week, one-season, and one-year periods of interest respectively. The increased variability in meteorology that occurs with a longer period of interest is the probable cause for the increase in classical error; our findings are in agreement with the rationale of controlling for seasonality when modelling air pollution (Chen et al., 2010; Pandey et al., 2014).

The total number of sampling hours consists of the sample count multiplied by the sample length. When we examine all the combinations of sample count and sample length that are a multiple of a particular total number of sampling hours, within a particular period of interest, we find that the least error occurs when the sample count and length are in the middle of the range of values tested. If we examine all the combinations that consist of thirty-two total sampling hours for the one-week period of interest, the lowest error occurs with four samples of eight hours each, the second lowest error occurs with eight samples of four hours each, and the highest error occurs with thirty-two samples of one hour each.

This result has an implication for the design of monitoring programs that collect incomplete time-series data samples, which is that monitoring should occur with a balance between the number of observations and the length of each observation. Neither the adjusted or unadjusted data deviate from this finding.

When considering the use of an adjustment method, the total sampling time is important. If the total number of hours is a small portion of the period of interest, we find the adjustment method beneficial; however, as the portion sampled increases towards one-third of the period of interest, the utility of the adjustment method diminishes and the adjustment method may increase the classical error. Based on our findings the research that has incorporated an adjustment method would have benefitted (Larson et al., 2009; Adams et al., 2012; Kanaroglou et al., 2013; Hoek et al., 2002).

3.5 Conclusions

It is apparent from our results that the optimal method for minimized classical error is to obtain the entire time-series with a fixed-location monitor. We understand this is not always possible and incomplete datasets may be the only option, to reduce classical error with these circumstances, we suggest the following guidelines:

(A) The total monitoring time should be equally divided by the number of samples and the sample length.

(B) When less than one-quarter of all possible observations are obtained, it is likely useful to employ an adjustment method.

(C) When possible, an evaluation similar to this study should be conducted on more than two monitors' historical data to provide an estimate of the classical error.

(D) If no data from historic monitoring are available, locate adjustment stations in the different primary land uses, e.g. urban and rural.

The classical error will not be eliminated by following our guidelines; however, it should be reduced. Our findings indicate that researchers who are using incomplete datasets have challenging decisions for sampling design that extend beyond the choice of locations for mobile monitoring.

3.6 References

Adams, M. D., De Luca, P. F., Corr, D., & Kanaroglou, P. S. (2012). Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005–2010. Social Indicators Research, 108(2), 351–364. doi:10.1007/s11205-012-0061-5

Airparif (2013). Air quality assessment network in the Paris Region. Accessed April 2013 from: http://www.airparif.asso.fr/_pdf/publications/bilan-2012-anglais.pdf

Armstrong, BG 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational and environmental medicine, 55(10), 651–6.

Badaloni, C., Ranucci, A., Cesaroni, G., Zanini, G., Vienneau, D., Al-Aidrous, F., De Hoogh, K., Magnani, C., & Forastiere, F. (2013). Air pollution and childhood leukaemia: a nationwide case-control study in Italy. *Occupational and Environmental Medicine*, 70(12), 876–83.

Brook, J., Dann, T., & Burnett, R. (1997). The relationship among TSP, PM10, PM2.5, and inorganic constituents of atmospheric particulate matter at multiple Canadian locations. *Journal of the Air & Waste Management*, 47, 2–19.

Chen, L., Bai, Z., Kong, S., Han, B., You, Y., Ding, X., Du, S., & Lie, A. (2010). A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. *Journal of Environmental Sciences*, 22(9), 1364–1373.

Delmelle, E. M. (2014). Spatial Sampling. In M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 1385–1399). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ghosh, J. K. C. , Heck, J. E. , Cockburn, M., Su, J., Jerrett, M., & Ritz, B. (2013). Prenatal exposure to traffic-related air pollution and risk of early childhood cancers. *American Journal of Epidemiology*, 178(8), pp.1233–9.

Goldstein, I., & Landovitz, L. (1977a). Analysis of air pollution patterns in New York City—I. Can one station represent the large metropolitan area? *Atmospheric Environment*, 11(1), 47–52. doi:10.1016/0004-6981(77)90205-0

Goldstein, I., & Landovitz, L. (1977b). Analysis of air pollution patterns in New York City—II. Can one aerometric station represent the area surrounding it? *Atmospheric Environment*, 11(1), 53–57. doi:10.1016/0004-6981(77)90206-2

Griffith, D. (2005). Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95(4), 740–760.

Grivas, G., & Chaloulakou, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment*, 40(7), 1216–1229.

Heid, I. M., Küchenhoff, H., Miles, J., Kreienbrock, L., & Wichmann, H.E. (2004). Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology*, 14(5), 365–77.

Hoek, G., Krishnan, R. M., Beelan, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environmental Health*, 12(1), 43.

Hoek, G., Meliefste, K., Cyrys, J., Lewné, M., Bellander, T., Brauer, M., Fischer, P., Gehring, U., Heinrich, J., van Vliet, P., & Brunekreef, B. (2002). Spatial variability of fine particle concentrations in three European areas. *Atmospheric Environment*, 36(25), 4077–4088.

Hutter, H. P., Haluza, D., Piegler, K., Hohenblum, P., Fröhlich, M., Scharf, S., Uhl, M., Damberger, B., Tappler, P., Kundi, M., Wallner, P., & Moshammer, H. (2013). Semivolatile compounds in schools and their influence on cognitive performance of children. *International Journal of Occupational Medicine and Environmental Health*, 26(4).

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., & Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15(2), 185–204.

Kanaroglou, P., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R. (2005). Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment*, 39(13), 2399–2409.

Kanaroglou, P. S., Adams, M. D., De Luca, P. F., Corr, D., & Sohel, N. (2013). Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model. *Atmospheric Environment*, 79, 421–427. doi:10.1016/j.atmosenv.2013.07.014

Kumar, N. (2009). An Optimal Spatial Sampling Design for Intra-Urban Population Exposure Assessment. *Atmospheric Environment*, 43(5), 1153.

Larson, T., Henderson, S.B., & Brauer, M. (2009). Mobile monitoring of particle light absorption coefficient in an urban area as a basis for land use regression. *Environmental Science & Technology*, 43(13), 4672–8.

Liu, C., Flexeder, C., Fuertes, E., Cyrys, J., Bauer, C-P., Koletzko, S., Hoffmann, B., von Berg, A., & Heinrich, J. (2013). Effects of air pollution on exhaled nitric oxide in children: Results from the GINIplus and LISAplus studies. *International Journal of Hygiene and Environmental Health*. DOI: 10.1016/j.ijheh.2013.09.006

Padula, A. M., Mortimer, K., Hubbard, A., Lurmann, F., Jerrett, M., & Tager, I. B. (2012). Exposure to traffic-related air pollution during pregnancy and term low birth weight: estimation of causal associations in a semiparametric model. *American Journal of Epidemiology*, 176(9), 815–24.

Pandey, B., Agrawal, M., & Singh, S. (2014). Assessment of air pollution around coal mining area: Emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis. *Atmospheric Pollution Research*, 5.

Pedersen, M., Siroux, V., Pin, I., Charles, M. A., Forhan, A., Hulin, A., Galineau, J., Lepeule, J., Giorgis-Allemand, L., Sunyer, J., Annesi-Maesano, I., & Slama, R. (2013). Does consideration of larger study areas yield more accurate estimates of air pollution health effects? An illustration of the bias-variance trade-off in air pollution epidemiology. *Environment International*, 60, 23–30.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Reggente, M., Mondini, A., Ferri, G., & Mazzolai, B. (2010). The dustbot system: Using mobile robots to monitor pollution in pedestrian area. *Proc. of NOSE*, 23, 273–278. doi:10.3303/CET1023046

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., & Zeger, S. L. (2000). Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *New England Journal of Medicine*, 343(24), 1742–1749.

Sfetsos, A., & Vlachogiannis, D. (2010). A new approach to discovering the causal relationship between meteorological patterns and PM10 exceedances. *Atmospheric Research*, 98(2-4), 500–511.

Vienneau, D., de Hoogh, K., Bechle, M. J., Beelen, R., van Donkelaar, A., Martin, R., Millet, D., Hoek, G., & Marshall, J. D. (2013). Western European land use regression incorporating satellite- and ground-based measurements of NO2 and PM10. *Environmental Science & Technology*. 47, 13555−13564.

Wallner, P., Kundi, M., Moshammer, H., Piegler, K., Hohenblum, P., Scharf, S., Fröhlich, M., Damberger, B., Tappler, P., & Hutter, H-P. (2012). Indoor air in schools and lung function of Austrian school children. *Journal of Environmental Monitoring*, 14(7), 1976–1982.

Chapter 4: Development of real-time air pollution models with neural networks in a land use regression framework: Combining mobile and stationary air pollution monitoring to better understand the spatial variation of air pollution.

4.1 Introduction

The City of Hamilton, Ontario, Canada is located at the western tip of Lake Ontario, which is a body of water that bridges Canada and the United States of America. Hamilton's economic prosperity relied on its steel industry, which began in the early 20th century. Access to Lake Ontario benefitted the industry from low-cost transportation and the use of water as the cooling medium. World War II brought an economic boom to Hamilton's steel sector; however, the jolt to the economy resulted in air, water and soil contamination. During steel production air pollution is a result of the processing of steel that begins with creation of pig iron, which consists of iron ore, coke (residue after the distillation of bituminous coal), and limestone. This processing releases criteria air contaminants including Nitrogen Oxides, Carbon Monoxide, Sulphur Dioxide, Particulate Matter, and Volatile Organic Compounds, which are monitored in Canada under the National Pollutant Release Inventory. Currently, Hamilton has a population of 520,000 people, with a diversified economy and reduced reliance on the steel industry. Air pollution issues are still prevalent in Hamilton but are now due to a diverse set of sources (Adams et al., 2012).

Nations around the world have established that it is necessary to inform citizens about the risks of air pollution in their region. The main tool that is used to provide this information is an index of air quality, which at the most basic definition translates a cocktail of air pollutants onto a scale that can be interpreted by the public. These air quality indexes

(AQI) vary in their approach for translating pollutant concentrations onto the scale. The scales are determined by regional policies (Plaia and Ruggieri, 2010). Canada, has developed and adopted an Air Quality Health Index (AQHI), which is an 11-point scale that is a non-linear combination of particulate matter 2.5 microns or smaller in aerodynamic diameter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) (Steib et al., 2008). AQHI values in Canada are presented to the public through the national and local news outlets and allow citizens to make informed decisions about their activity level for the day based on their personal health risk. Users in the high-risk category, such as the elderly, may rely on this information to plan their activity for the day. The information the public is presented with is single representative values for the entire city where they reside. Unfortunately, these representative values do not account for any spatial variation that may occur within the city.

Urban centres, particularly in North America, have developed with sprawled patterns requiring significant computing, they have isolated pockets of industry and commercial activity, and sometimes are characterized by highly variable meteorological conditions. All of these factors drive space-time variation in urban air pollution (Johnson et al., 2010; Tang et al., 2013). Modellers expend significant effort in designing techniques for the identification of these factors. Geographic Information System (GIS) tools play a prominent role in enhancing efforts in space-time modelling (Briggs, 2006). Data for the modelling is commonly provided through a network of monitoring instruments that are installed within the region and record concentrations over a time interval that is defined by the operator. Optimal location theory can be applied when locating the instruments to ensure spatial variability is observed (Ainslie et al., 2009). Epidemiological studies identify that space-time variation must be accounted for to elicit accurate estimates (Ozkaynak et al., 2013). It is undeniable that space-time variation will be a focus for the continued understanding of air pollution exposure.

In this paper, we present a model for predicting in real-time spatially resolved air quality health index maps. The technique applies neural network models for the spatial prediction of air pollutants across Hamilton, Ontario, Canada. The prediction of air quality across space is based on a number of data including mobile and stationary air pollution monitoring data, meteorological data, land use characteristics, and traffic information. These models can improve the health risk information that is provided to the general public in a region.

4.2 Methods

4.2.1 Study Area

Hamilton, Ontario, Canada, situated at the western tip of Lake Ontario (43.3^{o}N, 79.9^{o}W) is Canada's 9th and Ontario's 3rd largest city with a population of 520,000 (Statistics Canada 2012). Between the census years of 2006 and 2011 Hamilton's population increased by 3%. The city is divided into an upper and lower city by a 90-meter escarpment. Hamilton has traditionally been an industrial city, focused on steel production. Air pollution concerns have led to multiple air pollution studies in Hamilton (Adams et al., 2012; Jerrett et al., 2001; Kanaroglou et al., 2013; Wallace and Kanaroglou, 2008; Wallace et al., 2009). Two air quality monitoring networks operate in Hamilton. The first has four monitors near and within Hamilton and is operated by the provincial ministry of the environment; the second network has 14 monitors and is operated in partnership with the local industries, focusing on the northern portion of the city. In Figure 4.1, we present a map of industrial lands within Hamilton along with the fixed monitors of the provincial network, which record the pollutants used in the AQHI. Hamilton was chosen for this study due to the variation identified across the city by mobile air pollution monitoring (Adams et al., 2012).

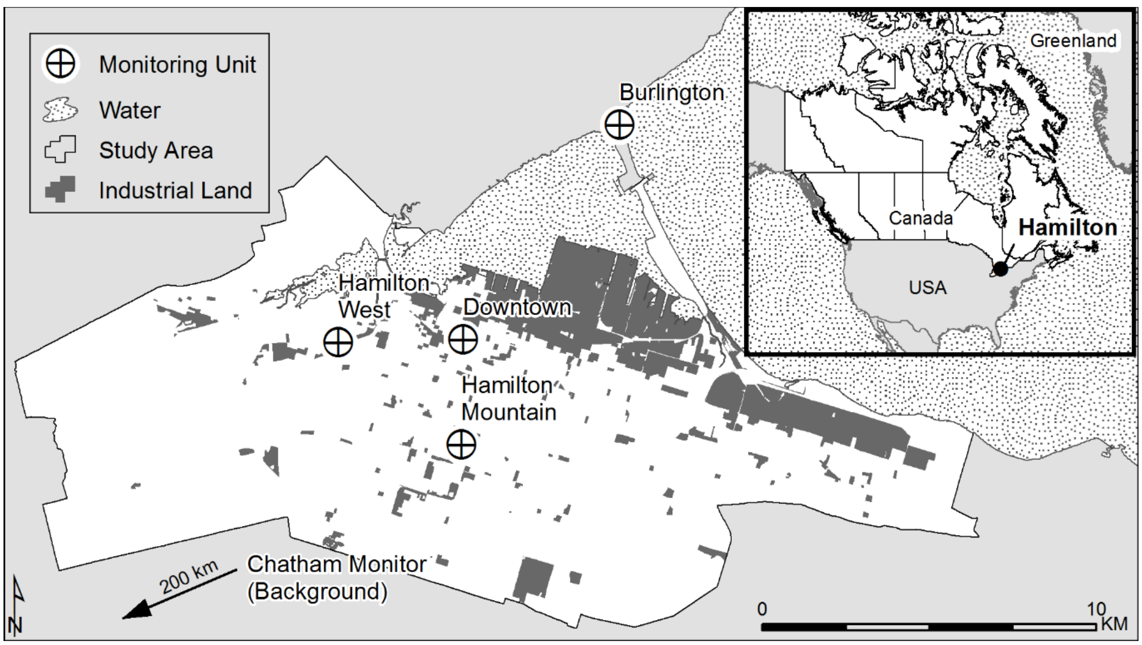

Figure 4.1 Hamilton stationary monitor locations and industrial land use.

4.3 Modelling

Air pollution health risk is typically presented to the public as individual values for a city or region, excluding any spatial variation. In Figure 4.2, we present two air quality health index maps for Alberta and British Columbia in Canada, both of which present air quality health risk at single geographical points. This approach provides a general overview of the air pollution, but it can be more informative if the spatial variation within the cities became available. The challenge for many locations is the limited number of monitoring sites that renders spatial interpolation inappropriate. For example, Hamilton has four monitoring stations that could be used to determine the AQHI.

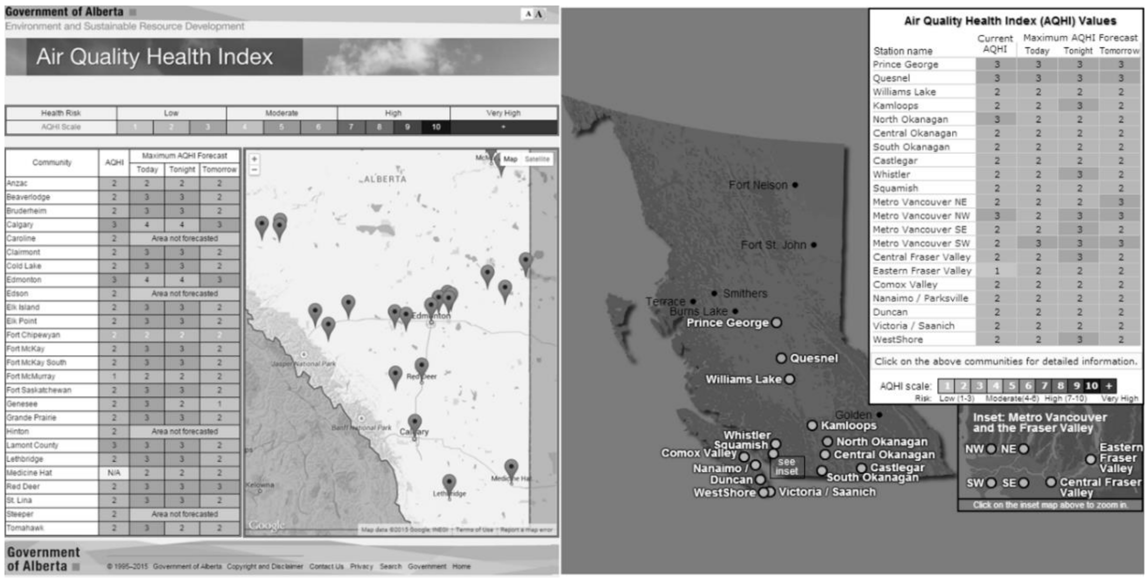

Figure 4.2 Air Quality Health Index Maps for Alberta, Canada (Left) and British Columbia, Canada (Right)

Mobile air pollution monitoring technologies allow for a greater spatial coverage of a city compared to traditional fixed stations. Their limitation is that the monitoring data do not provide a continuous time-series of observations. These discontinuous time-series data can become useful with the use of appropriate modelling techniques.

In this paper, we propose the use of artificial neural network models (neural networks) in a land use regression framework to predict the spatial variation in AQHI for Hamilton that could be updated in real-time. Land use regression is a spatial interpolation modelling technique that makes use of land-use characteristics around air pollution observations. These characteristics may include types and amounts of land use within a buffer of the monitoring locations, information about roads and road use, and meteorological information (Johnson et al., 2010). Typically the dependent variable in such models consists of observations that represent long-term average concentrations at several locals within the study area. However, this approach would limit our observations in Hamilton to the four stationary monitors. We have extended this approach with neural networks to utilize the mobile monitoring data as the dependent variable and predict these values with values from co-occurring stationary monitoring for air pollution and meteorology, land use characteristics, congestion, road types and amounts. We can use this model to predict air pollution concentrations for any location in real-time, with data connections to the stationary monitors in the area. Our modelling process is presented in Figure 4.3.

Neural networks are mathematical models that mimic biological neural networks, and can be used for modelling complex relationships. In our research, we have applied neural networks as an approach to generalize the linear regression model in an attempt to capture non-linear relationships. We found that the use of linear regression models with our data demonstrated poor performance. For a general discussion on artificial neural networks and their design, please refer to Venables & Ripley (2002).

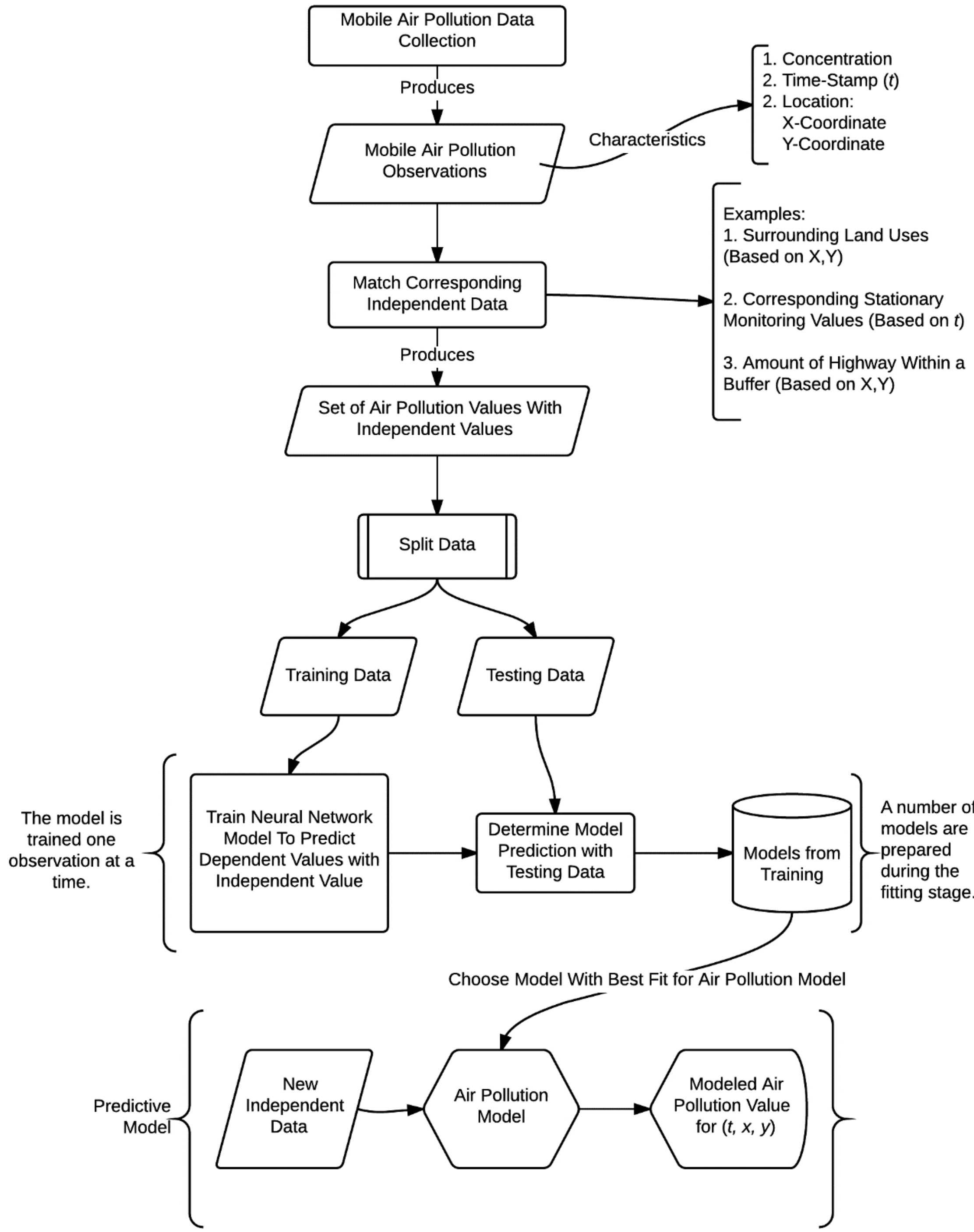

Figure 4.3 Air Pollution Modelling Structure

Neural networks have been demonstrated as an effective method to forecast air pollution concentrations (Barrón-Adame et al., 2012; Solaiman et al., 2009). We specified two neural network models, one each for $PM_{2.5}$ and NO_2. A neural network model is often described as a black box method, which requires an appropriate technique to ensure the model is not over-fit to the data set. Over-fit models are known to exhibit inferior predictability when used with data that deviate from the data used to train the models. The allowable complexity in the modeled relationships determines the degree of over-fitting. The most common approach to ensure the model generalizes well to data that was not used during fitting (trained on), is to reserve a portion of the data set for validation purposes; we will refer to these as testing data.

A feedforward neural network was applied, which does not cycle data in the model like recurrent neural networks, so there is no memory of the previous state of the network. Recurrent neural networks are beneficial for forecasting. We selected the feedforward approach because we lacked a continuous series of mobile air pollution data at every location over time. The network chosen is a multi-layer perceptron network, which consists of multiple layers of computational units (neurons). Each of the neurons in a layer has a directed connection to each of the neurons of the adjacent layers. Our network is composed of three layers. First is the input layer; each independent variable in the model is represented in this layer as a neuron. Second is the hidden layer, which maps the input values to the output layer using weights and logistic functions. The output layer is the dependent variable. The structure of the neural network model is presented in Figure 4.4.

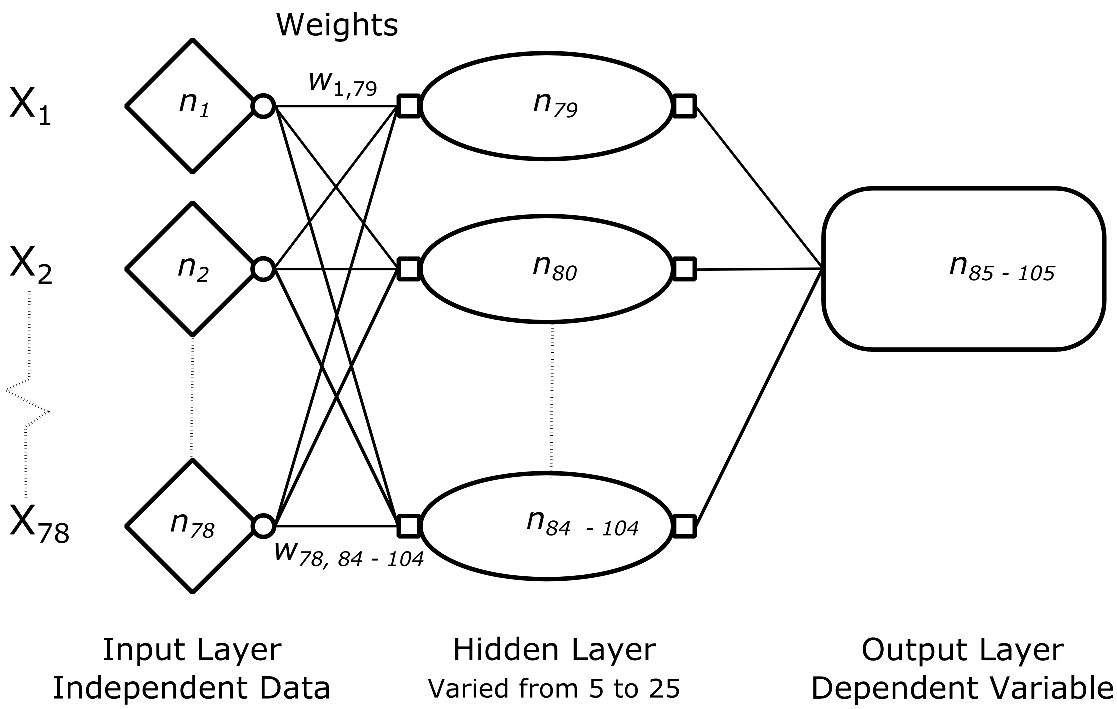

Figure 4.4 Multi-layer perceptron neural network framework. The input layer can be considered the independent data, the output layer is the dependent variable, with the hidden layer functioning to model the relationship between the two.

Our model consisted of 78 input neurons, one for each independent variable; one output layer neuron (dependent value); and we varied the number of hidden neurons during training from 5 up to 25. The activation function in the hidden layer was a logistic function.

All statistical modelling was conducted in R: A language and environment for statistical computing (R Core Team, 2014). The neural network models were fit with "Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS", which is a port of the Stuttgart Neural Network Simulator to R (Bergmeir and Benitez, 2012). We fit the model with back propagation, which is a supervised learning technique that fits the network by iterating through the data and propagating the errors back to the inputs, adjusting the network with the objective of minimizing these errors. During fitting we varied the number of hidden layer neurons from 5 to 25, this process utilized the 'caret' package for R (Kuhn, 2013) and produced a number of potential models for both $PM_{2.5}$ and NO_2. To prevent overtraining of the model the maximum training iterations were limited to 2,000. The optimal models were selected based on the minimum RMSE.

The data consisted of 7,047 and 7,060 observations for $PM_{2.5}$ and NO_2, respectively. Twenty percent of the data were retained as testing data to be applied following all fitting processes. These data were only used after the model with the lowest RMSE was selected, ensuring the model's prediction ability was based on data it had not seen during the supervised training. The coefficient of determination between the testing values predicted by the model and their actual values are used to determine the model's predictive ability.

4.4 Dependent Variable

Mobile air pollution sampling campaigns were conducted between 2005 and 2013. During the campaigns, air pollution concentrations were recorded to data loggers simultaneously with GPS coordinates. Air pollution concentrations were reported by the instrumentation as 2-minute rolling averages, derived from one-second observations. We purged this data to only retain the first record of every 60-second period to reduce redundant observations. With this mobile collection, we recorded data for both particulate matter 2.5 microns or smaller in aerodynamic diameter ($PM_{2.5}$) and nitrogen dioxide (NO_2). $PM_{2.5}$ concentrations were obtained with a Grimm laser optical scanner model 1.1.07 monitor with an operation range of 0-6,500 μg/m^3 and a precision of 1 μg/m^3. NO_2 concentrations were obtained with a Thermo Scientific Model 42i chemiluminescence monitor, with an operating range of 0 – 1,000 ppb and a precision of 0.4 ppb. The monitors were housed within a modified van with an air intake extended up through the roof of the vehicle, protected with a rain shield. The air intake is 3 m above ground and pointed in the direction of travel. All instruments were properly calibrated prior to mobile monitoring campaigns; a full description of mobile monitoring techniques can be found in Wallace et al. (2009). This data will be further referred to as the dependent variable. We present the location of mobile monitoring campaigns in Figure 4.5.

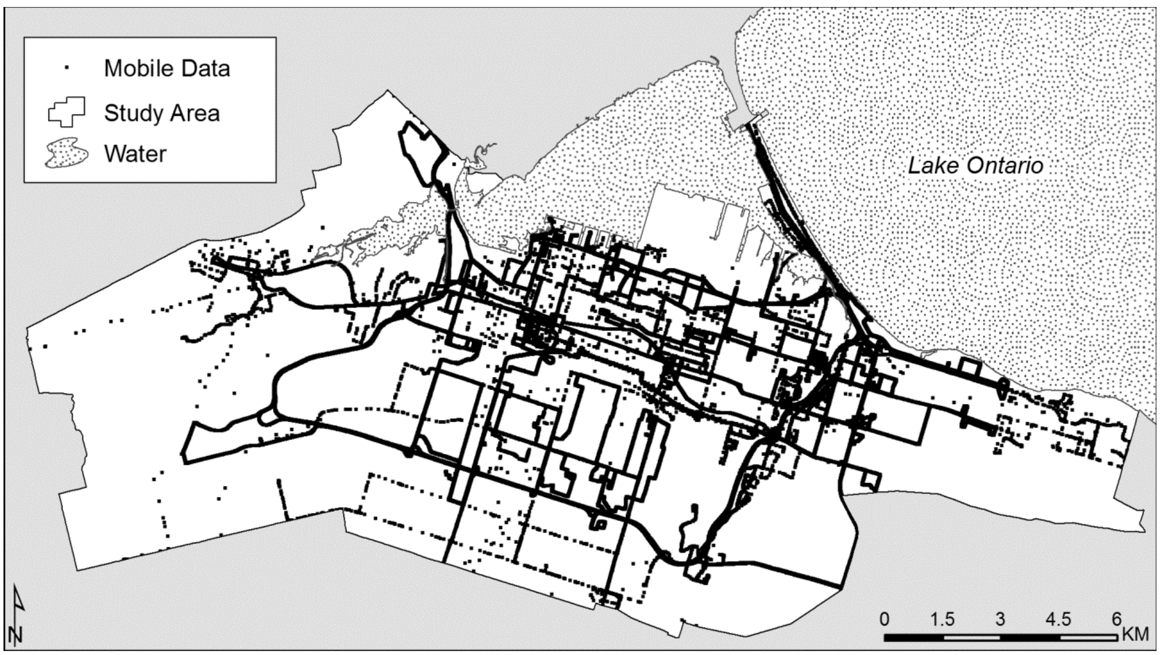

Figure 4.5 Mobile monitoring locations within the study area.

4.5 Independent Variables

The independent variables were selected for this study based on the predictor variables in past land use regression models (Hoek et al., 2008; Ryan and LeMasters, 2007). Unlike linear regression models, leaving an independent variable that provides no predictive ability in a neural network will not reduce the effectiveness of the model. This is because independent variables with no effect on the outcome are given negligible weights in the model.

To identify the ideal set of candidate variables we required both time-series and spatially varying data. Our set of predictor variables were diverse and included land use information, transportation-related characteristics, air pollution and meteorology; the independent variables are presented in Table 4.1. For predictors focused on the spatial component, we borrowed from indicators commonly used in land use regression air pollution studies. For the time-series component, we obtained data at stationary monitors for meteorology and air pollution concentrations.

Five of the predictor attributes were obtained by a separate regional wind model. Wind direction and speed data were obtained for the study area and surrounding region from Environment Canada's National Climate Data and Information Archive (Environment Canada, 2012), and the Hamilton Air Monitoring network. Wind measurements were recorded with North being 0°. We obtained hourly wind data from 29 monitoring stations; on average 18 monitors had available data during monitoring campaigns. Wind data were separated into U_{met} (U) and V_{met} (V) components and then each component was interpolated with kriging. The U component was calculated as $U = -S * \sin(D * (\pi / 180))$, and the V component with $V = -S * \cos(D * (\pi / 180))$, where S was speed and D was the direction between 1 and 360 degrees. After interpolation wind direction was determined as: $D = \text{atan2}(-U,-V) * 180 / \pi$. Wind data were coded to four directions: 0 to 89 degrees (NE), 90 to 179 degrees (SE), 180 to 269 degrees (NW), and 270 to 359 degrees (SW). The wind model is programmed in R: A language and environment for statistical computing (R Core Team, 2014). The interpolation applied the automated kriging approach implemented in the "automap" package for R (Hiemstra et al. 2008).

Vehicle emissions generate significant volumes of particulate matter and other pollutants. Land-use regression models typically retain a traffic congestion variable as a predictor (Hoek et al. 2008). Some common measures include (1) traffic volume, (2) traffic intensity, and (3) traffic within buffer distances. We utilized data from INRIX, a Seattle-based vendor, which provides observed speeds recorded by GPS-enabled vehicles and stationary traffic sensors. The data is represented as annual averages every 15 minutes for all road links in the study area for 2011. We aggregate the data on each road link by the hour and compute congestion severity as a variable in our model similar to Texas Transportation Institutes Travel Time Index and INRIX's Travel Time Tax. Congestion is calculated by dividing average hourly travel time by free-flow travel time. This index reflects the extra time it takes to traverse a road segment compared to free-flow for each hour of the day. Using GIS software, we apply a 300 m buffer to the road network and intersect the maximum value of congestion severity to a 100 m grid. The 300-meter buffer is chosen because air pollution levels tend to disperse from the roadways to ambient background concentrations over a distance of about 300 meters (Pratt et al., 2013).

The local and regional background air pollution monitors are operated by the Ontario Ministry of the Environment, and we obtained their hourly averaged data for the study period from their historic database. The regional background monitor was chosen for this study area on their advice. The elevation was obtained from a 10 m DEM, and the land use and roadway data were from DMTI Spatial Inc, a commercial GIS data provider.

4.6 Results

4.6.1 Mobile and Stationary Monitoring Comparison

The data obtained by mobile collection for $PM_{2.5}$ and NO_2 both had distributions shifted to the right when compared to the stationary monitoring data, which is presented in Figure 4.6. The mobile air pollution monitoring data were collected in a range of areas within the city, including on-road measurements. The stationary monitoring units are located to reduce the direct emissions from vehicles, which is achieved by their location in parks or open areas. Both the Hamilton Mountain and Hamilton Downtown monitors are located in parks. This variation in the collection sites suggests that this approach of combining datasets can provide additional information within the model from the mobile data than simply relying on the stationary monitoring units.

Table 4.1 Description of model predictor variables.

Predictor Variable	Description
X- Coordinate	X coordinate from the universal transverse Mercator coordinate system
Y- Coordinate	Y coordinate from the universal transverse Mercator coordinate system
Congestion	Congestion index within 300 meters.
Elevation	Elevation at monitored location (mobile)
Wind Speed	Wind Speed from regional wind model.
Industrial Land Use Commercial/Government Land Use Parks or Open Area Land Use Residential Land Use	Binary indicators: Value of 1, if the mobile observation was located on the land use.
Minor Roads (Streets) Major Roads (Arterials) Highway (Freeway and Expressways)	The amount of each road type within circular buffers of 25, 100, 400 and 1600 meters.
Residential Commercial Parks and Recreation Resource and Industrial Government and Institutional Open Area	The amount of each land use within circular buffers of 50, 100, 200 and 400 meters.
North East Winds North West Winds South West Winds South East Winds	Binary indicator: Value of 1 if the wind direction was from that direction during collection.
Background Pollution Monitoring	A monitor southwest of Hamilton, Ontario measuring $PM_{2.5}$, NO_2, NO, NO_X, & O_3. Suitable for the measurement of air pollution from the USA due to westerlies.
Local Air Pollution Monitors	Four air pollution monitors located in Hamilton and Burlington, Ontario monitoring $PM_{2.5}$ (4), NO_2 (2), NO (1), NO_X (1), CO (1), SO_2 (1) & O_3 (4). The number monitoring each pollutant are in brackets.

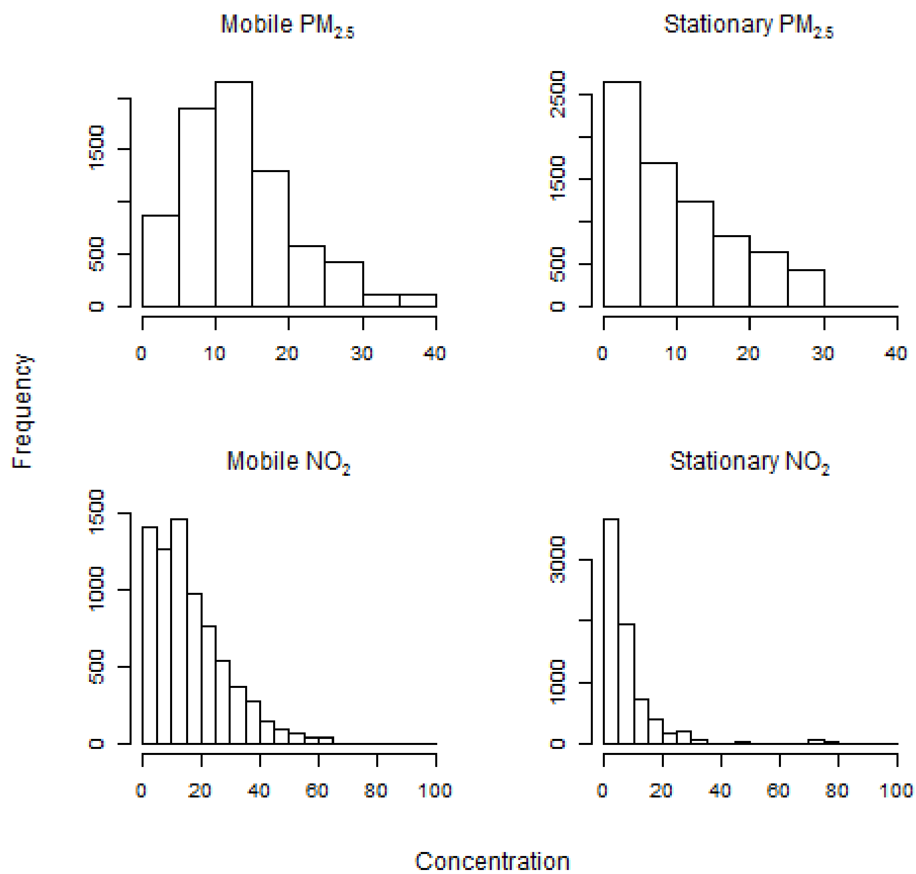

Figure 4.6 $PM_{2.5}$ and NO_2 histograms for mobile and stationary monitoring data. Stationary data from the downtown Hamilton monitor.

The $PM_{2.5}$ mobile data had a Pearson's *r* coefficient of 0.36 to the $PM_{2.5}$ stationary data collected in downtown Hamilton. The NO_2 mobile data was similar in the strength of the correlation to the stationary downtown Hamilton data with a Pearson's *r* coefficient of 0.33. Comparing both the $PM_{2.5}$ and NO_2 mobile data with the stationary monitoring data representing background concentrations from the United States (The Chatham Ontario Monitors), the correlation was much stronger for $PM_{2.5}$ than NO_2 at 0.40 and 0.12, respectively.

4.6.2 Model Fit

The models were both fit using a boot-strapping approach that varied the number of hidden layer neurons, which produced a number of potential models. The optimal number of hidden neurons was chosen based on the model that minimized the RMSE for the prediction of data that were set aside for each training event. We selected a model with eleven hidden neurons for the $PM_{2.5}$ model and a model with seven hidden neurons for NO_2.

Linear regression models allow the modeller to understand the relationship between each independent variable and the dependent variable, particularly for the variable's effect and strength by analyzing the standardized coefficients. The weights in a neural network model, which map the data to the hidden layer neurons, are partially analogous to the coefficients in a regression model. To identify the relative importance of each independent variable in the model we have applied a method described by Garson (1991) and then later in detail by Goh (1995) to determine the relative importance of each independent variable in the models. The technique generalizes all weights for each variable to each hidden neuron, which totals 858 and 546 weighted-connections for the $PM_{2.5}$ and NO_2 models respectively, into a value that describes the relationship to the response variable. We utilized the implementation of Garson's method that is in the "NeuralNetTools" package for R (Beck, 2015). The relative weight results are presented in Figure 4.7.

Pearson's *r* correlation coefficient was calculated between the two sets of relative weights values from both models for the independent variables, which was $r = 0.31$. The natural logarithms of the values were used to approximate normal distributions. A high correlation coefficient would indicate the same variables have high relative importance in both models. A lower value indicates that the variables in one model are not important variables in the other model.

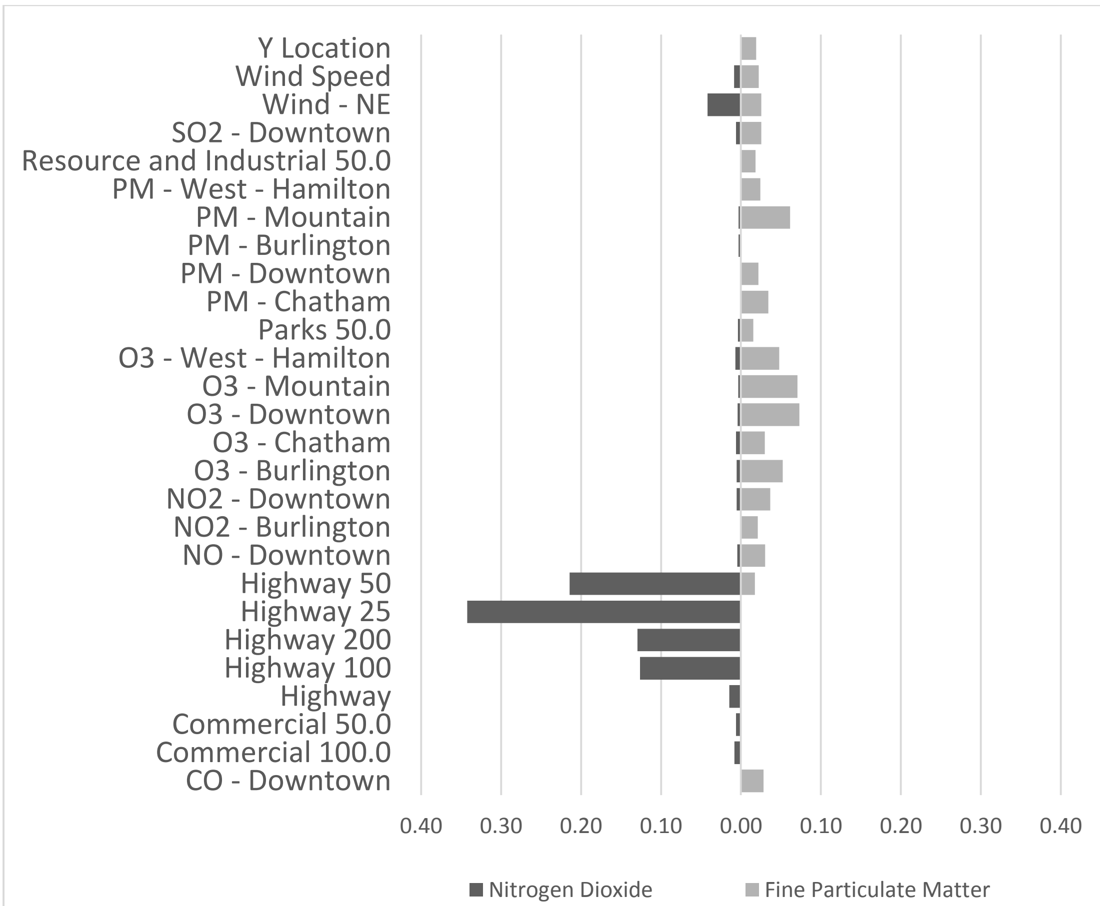

Figure 4.7 Model variables' relative weights. Relative weights of the highest twenty variables for each of the two models. The relative weights for Nitrogen Dioxide are graphed to the left of centre, and the relative weights of the $PM_{2.5}$ model are graphed to the right of centre. All weight values are non-negative. If no bar appears on a side, this indicates the variable was not one of the highest twenty variables in the model.

The particulate matter model performed better than the nitrogen dioxide model for predicting the testing data that had been withheld during model training, which is analogous to predicting future events. The $PM_{2.5}$ model performed well for predicting unseen conditions with a Pearson's r of 0.88, a coefficient of determination of 0.78, and an RMSE of 3.5 μg/m³. The NO_2 model performed well, but was not as strong in its predictive ability to unseen conditions as the $PM_{2.5}$ model with a Pearson's *r* of 0.59, a coefficient of determination of 0.34, and a RMSE of 10 ppb. The model validation fits are presented visually in Figure 4.8, which includes plots of the predicted values and the actual values of the testing data that was unseen during the model fitting procedure.

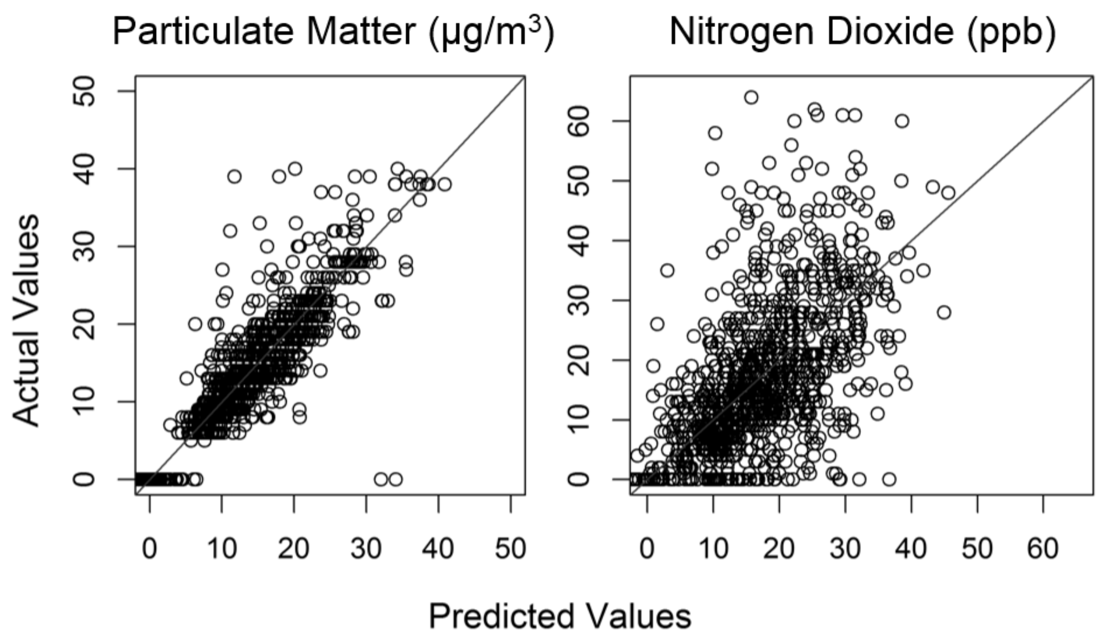

Figure 4.8 Model validation data plots. Plots of the actual values and those estimated from the model for the testing datasets for both $PM_{2.5}$ and NO_2.

4.7 Discussion

Air quality (health) indices are an important tool, providing citizens with the knowledge of air pollution risk in their region, which is particularly important for vulnerable populations (Chen et al., 2013). However, the current methods of disseminating real-time air pollution information are limited, which typically includes point values on a map. These maps are valuable, but can be significantly improved upon by presenting the spatial variation of air pollution in the region.

In our research, we produced two neural network models. Both pollutants that were modelled are included in Canada's air quality health index (Steib et al., 2008). Applying a neural network based land use regression approach, we were able to demonstrate effective fits to the mobile air pollution data for the purposes of prediction of additional events. A direct comparison to other researchers' models is challenging because our study employed a unique collection method compared to past land use regression studies (Hoek et al., 2008; Ryan and LeMasters, 2007). As well, we are unique in that our model is focused on modelling the individual events as opposed to long-term averages, which leads to a larger information content in the data set and a larger variation that must be explained by the model. Models of $PM_{2.5}$ for long-term averages across similarly sized study areas typically have similar or poorer model fits with land use regression mapping approaches for long-term averages compared to what we obtained in this research (Hoek et al., 2008). When modelling long-term average concentrations, it should be expected that the R^2 should be higher because it has averaged out a lot of the unique information in the dataset. We argue that because of the high information content, which is due to our data collection methods that the NO_2 model is still of sufficient fit to provide valuable air pollution estimates.

$PM_{2.5}$ is a pollutant known to be controlled by local conditions and long-range transport of particulate matter (Wang et al., 2015). NO_2 is highly linked to vehicular sources, which is often identified in land use regression studies (Briggs et al., 2000; D. J. Briggs et al., 1997; Gilbert et al., 2005; Stedman et al., 1997). This difference in process scales leads to the improved fit in the $PM_{2.5}$ model compared to the NO_2 model because background conditions are easier to incorporate into a model. Ross et al. (2007) modeled $PM_{2.5}$ using a land use regression approach and identified that county emissions played a role in their model. Regional air pollution patterns for $PM_{2.5}$ were drivers in our model, which was identified by the Chatham stationary variable ranked high in importance. The NO_2 was driven by vehicle emissions, which is noted with the high importance of the highway variables.

Seventy-eight variables were generated for inclusion in the neural network models and of the top twenty most important variables for each of the models fourteen variables were the same. This indicates that similar critical land use characteristics are affecting both pollutants throughout Hamilton, Ontario. The sum of the relative weights of the five most important variables for the NO_2 model was 0.853. The particulate matter model has a much reduced relative importance of the top five variables with their relative importance summing to 0.306. These two values can be multiplied by their model's respective R^2 to demonstrate that the top five variables of the models accounted for the explanation of 29% and 23% of the modeled variation for NO_2 and $PM_{2.5}$ respectively.

Examining Figure 4.8, we find that the NO_2 model has a reduced ability to predict the extreme events in the data set, particularly the values of 40 ppb and above. Our sampling approach of including on-road measurements in locations across the city resulted in elevated concentrations due to traffic congestion. These highly localized events are challenging to model because they are unique and not very generalizable, which is similar to the events that are monitored in vehicle chase air pollution monitoring studies (Westerdahl et al., 2005).

Our model is capable of estimating real-time air pollution across Hamilton; however, its utility is not limited to modelling future pollution concentrations. Many land use regression models are developed to assign air pollution exposure in epidemiological studies. It has been noted that exposure modelling should incorporate human time-activity patterns, which cannot be incorporated into many current models (Baxter et al., 2013b). Because our models predict on an hourly basis, with time-activity patterns, we could assign the correct hourly exposure to individuals in the study population.

We recognize that the structures of cities are always evolving. For example, land uses changes occur because of development, transportation networks grow and shrink, and the emissions from vehicles on these networks vary over time. The lifespan of the model has not been evaluated in this work, but we do identify that as we move forward in time without additional updates to this model it will eventually be of little utility. To maximize the models' lifespans we suggest for prediction locations to use the most recent land use data, the mobile monitoring campaigns should continue, and the model should be continually recalibrated with additional information. We applied data from 2005 through 2013, which produced a model fit with eight years of monitoring. We have yet to explore if this period is optimal for the model, or if the period could be lengthened with additional data or should be reduced to account for any non-stationary processes that are occurring.

4.8 Conclusion

Air pollution is a concern in Ontario particularly during elevated episodes that tend to occur within the summer months. This research provides the foundation for the development of a spatially resolved AQHI for Hamilton, Ontario, Canada. The models are well fit with the provided predictor data and can model air pollution concentrations in real-time when real-time connections are established to the stationary monitoring units in the region.

This technique of applying neural network models in a land use regression framework is an extension of land use regression modelling but allows for complex non-linear relationships that may be present in the data. The neural network approach has allowed for the complex space-time relationships of air pollution to be captured sufficiently, and our findings may suggest that neural networks will serve as an improved modelling approach for modelling with land use regression. Neural network models are often considered deficient because of the potential "black-box" issues, but with the utilization of the techniques described by Garson (1991) and Goh (1995) for identifying the relative weights of the input variables, this issue is partially overcome.

4.9 References

Adams, M. D., De Luca, P. F., Corr, D., & Kanaroglou, P. S. (2012). Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005–2010. *Social Indicators Research*, 108(2), 351–364. doi:10.1007/s11205-012-0061-5

Ainslie, B., Reuten, C., Steyn, D. G., Le, N. D., & Zidek, J. V. (2009). Application of an entropy-based Bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants. *Journal of Environmental Management*, 90(8), 2715–29. doi:10.1016/j.jenvman.2009.02.016

Barrón-Adame, J. M., Cortina-Januchs, M. G., Vega-Corona, a., & Andina, D. (2012). Unsupervised system to classify SO2 pollutant concentrations in Salamanca, Mexico. *Expert Systems with Applications*, 39(1), 107–116. doi:10.1016/j.eswa.2011.05.083

Baxter, L. K., Dionisio, K. L., Burke, J., Ebelt Sarnat, S., Sarnat, J. A., Hodas, N., et al. (2013). Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *Journal of Exposure Science and Environmental Epidemiology*, 23(6), 654–9. doi:10.1038/jes.2013.62

Beck, M. W. (2015). NeuralNetTools: Visualization and Analysis Tools for Neural Networks. http://cran.r-project.org/package=NeuralNetTools

Bergmeir, C., & Benitez, J. M. (2012). Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *Journal of Statistical Software*, 46, 1–26.

Briggs, D. (2006). The role of GIS: coping with space (and time) in air pollution exposure assessment. *Journal of Toxicology and Environmental Health, Part A.*, 68(13-14), 1243–61. doi:10.1080/15287390590936094

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., et al. (1997). Mapping urban air pollution using GIS: a regression-based approach. International *Journal of Geographical Information Science*, 11(7), 699–718. doi:10.1080/136588197242158

Briggs, D. J., De Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., & Smallbone, K. (2000). A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments. *Science of the Total Environment*, 253(1-3), 151–167. doi:10.1016/S0048-9697(00)00429-0

Chen, R., Wang, X., Meng, X., Hua, J., Zhou, Z., Chen, B., & Kan, H. (2013). Communicating air pollution-related health risks to the public: an application of the Air Quality Health Index in Shanghai, China. *Environment International*, 51, 168–73. doi:10.1016/j.envint.2012.11.008

Garson, G. D. (1991). Interpreting neural network connection weights. Artificial *Intelligence Expert*, 6(4), 46–51.

Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., & Jerrett, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in Montréal, Canada, with a land-use regression model. *Journal of the Air & Waste Management Association*, 55(8), 1059–1063. doi:10.1080/10473289.2005.10464708

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*. doi:10.1016/0954-1810(94)00011-S

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33), 7561–7578. doi:10.1016/j.atmosenv.2008.05.057

Jerrett, M., Burnett, R. T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., & Brook, J. R. (2001). A GIS - environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, 33(6), 955–973. doi:10.1068/a33137

Johnson, M., Isakov, V., Touma, J. S., Mukerjee, S., & Özkaynak, H. (2010). Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmospheric Environment*, 44(30), 3660–3668. doi:10.1016/j.atmosenv.2010.06.041

Kanaroglou, P. S., Adams, M. D., De Luca, P. F., Corr, D., & Sohel, N. (2013). Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model. *Atmospheric Environment*, 79, 421–427. doi:10.1016/j.atmosenv.2013.07.014

Ozkaynak, H., Baxter, L. K., Dionisio, K. L., & Burke, J. (2013). Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science and Environmental Epidemiology*, 23(6), 566–72. doi:10.1038/jes.2013.15

Plaia, A., & Ruggieri, M. (2010). Air quality indices: a review. Reviews in Environmental Science and Bio/Technology, 10(2), 165–179. doi:10.1007/s11157-010-9227-2

Pratt, G. C., Parson, K., Shinoda, N., Lindgren, P., Dunlap, S., Yawn, B., et al. (2013). Quantifying traffic exposure. *Journal of Exposure Science and Environmental Epidemiology*, 1–7. doi:10.1038/jes.2013.51

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria. http://www.r-project.org/.

Ross, Z., Jerrett, M., Ito, K., Tempalski, B., & Thurston, G. (2007). A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41(11), 2255–2269. doi:10.1016/j.atmosenv.2006.11.012

Ryan, P. H., & LeMasters, G. K. (2007). A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation toxicology*, 19 Suppl 1, 127–33. doi:10.1080/08958370701495998

Solaiman, T. a., Coulibaly, P., & Kanaroglou, P. (2009). Ground-level ozone forecasting using data-driven methods. *Air Quality, Atmosphere & Health*, 1(4), 179–193. doi:10.1007/s11869-008-0023-x

Stedman, J. R., Vincent, K. J., Campbell, G. W., Goodwin, J. W. L., & Downing, C. E. H. (1997). New high resolution maps of estimated background ambient NO(x) and NO2 concentrations in the U.K. *Atmospheric Environment*, 31(21), 3591–3602. doi:10.1016/S1352-2310(97)00159-3

Steib, D. M., Burnett, R. T., Smith-Doiron, M., Brion, O., Hyun Shin, H., & Economou, V. (2008). A New Multipollutant, No-Threshold Air Quality Health Index Based on Short-Term Associations Observed in Daily Time-Series Analyses. *Journal of the Air & Waste Management Association*, 58(3), 435–450. doi:10.3155/1047-3289.58.3.435

Tang, R., Blangiardo, M., & Gulliver, J. (2013). Using building heights and street configuration to enhance intra-urban PM10, NOX and NO2 land use regression models. *Environmental Science & Technology*. doi:10.1021/es402156g

Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., & McCarry, B. (2009). Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring*, 11(5), 998–1003. doi:10.1039/b818477a

Wallace, J., & Kanaroglou, P. (2008). Modeling NOx and NO2 emissions from mobile sources: A case study for Hamilton, Ontario, Canada. *Transportation Research Part D: Transport and Environment*, 13(5), 323–333. doi:10.1016/j.trd.2008.04.001

Wang, L., Liu, Z., Sun, Y., Ji, D., & Wang, Y. (2015). Long-range transport and regional sources of PM 2.5 in Beijing based on long-term observations from 2005 to 2010. *Atmospheric Research*, 157, 37–48. doi:10.1016/j.atmosres.2014.12.003

Westerdahl, D., Fruin, S., Sax, T., Fine, P. M., & Sioutas, C. (2005). Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment*, 39(20), 3597–3610. doi:10.1016/j.atmosenv.2005.02.034

Chapter 5: Are children living in lower socioeconomic status neighbourhoods exposed to higher doses of particulate matter air pollution during their trip to school, at home or at school?

5.1 Introduction

Air pollution is associated with many negative health effects in children that include reduced lung function (Brunekreef et al., 1997; Gauderman, 2000; Wallner et al., 2012), lung development (Gauderman et al., 2004), cognitive performance (Hutter et al., 2013; Suglia et al., 2008) and function (Freire et al., 2010). See Wong et al. (2003) for a meta-analysis of children's' health benefits that can be attributed to reduced air pollution exposure. A large body of research also indicates a relationship between air pollution and asthma induced hospital admissions (Wong et al., 2003).

Most research examining the effects of air pollution on health are observational studies that use ambient air pollution concentrations as the exposure metric (Heck et al., 2013; Jerrett et al., 2007; Moridi et al., 2013; Urman et al., 2013). Air pollution exposure during travel is variable and can be affected by many factors, for example, ambient air pollution exposure can change because of transportation mode choice (Panis et al., 2010; Zuurbier et al., 2010), and inside of a vehicle air pollution concentrations may be greater (Gulliver and Briggs, 2004) or less than (Briggs et al., 2008) the surrounding ambient conditions. Recent research suggests that exposure analysis should include activity patterns when estimating exposure to improve the estimation of the effect (Baxter et al., 2013a; Buonanno et al., 2013).

Ambient conditions are a suitable exposure metric if the population has the same activity levels; however, changing the exposure metric to dose instead of ambient conditions allows for a more refined exposure metric. The dose exposure is largely based on the volume of air consumed and the ambient conditions, which can account for changes in respiration rate due to changes in activity level. As a person increases their intake of air, they will have increased potential exposure to pollution. This is particularly important when comparing air pollution exposure between active transportation modes, because when a person engages in strenuous activity their respiration rate increases to obtain more oxygen, which results in higher doses of air pollution as they inhale a higher amount of total air.

In the Greater Toronto and Hamilton Area (GTHA) more than half of the car trips taking students to school are single purpose trips (Metrolinx, 2010), which could be avoided with students using active transportation. Many programs have been developed to encourage students to use active transportation to travel to school in both the GTHA and most of North America. These programs have been developed to increase children's activity to reduce obesity. These programs as timely in the GTHA, as one-quarter of parents, who drove their children to school and were surveyed agreed that the distance was suitable for walking (Metrolinx, 2010). While the push to return students active modes of transportation has potential benefits; it is common for parents to have concerns for their children to walk to school due to potential traffic incidents and stranger danger.

Lower income families have fewer options for housing compared to higher incomes households because of financial burdens. It has been demonstrated that low-cost housing may be located in the less desirable regions, for example this housing may be located in the area of higher than average air pollution concentrations (Harrison and Rubinfeld, 1978; Nelson, 1978; Smith and Huang, 1993; Zheng et al., 2014). The social justice literature has looked at this issue to determine if an effect is occurring where people of low-income households are exposed to higher levels of air pollution. This particular issue in Hamilton, Ontario, Canada was first identified as having an effect where low-income households were exposed to greater air pollution concentrations than higher income residents with exposure data from 1984 – 94 (Jerrett et al., 2001). A subsequent analysis in the city identified that higher early mortality was occurring in areas Hamilton with lower socioeconomic conditions (Jerrett et al., 2004). The relationship between socioeconomic status and environmental pollutants is a complex relationship when examined across different regions, in a Spanish study examining the exposure to environmental pollutants during childbirth there was no clear evidence of lower socioeconomic groups demonstrating a higher exposure (Vrijheid et al., 2012).

In our paper, we determine if students living in lower-income or higher density neighbourhoods of Hamilton, Ontario, Canada are faced with higher air pollution exposures, calculated as the dose during their trip to and from school if they utilize active transportation. This non-discretionary trip occurs twice daily for students between September and June (inclusive). We compare the results of two modes of active transportation, walking and cycling to identify if either mode results in a lower dose of air pollution. Household and school ambient concentrations are compared; as well if students from lower-income neighbourhoods have higher ambient school or home concentrations.

5.2 Methods

5.2.1 Study Area

Hamilton, Ontario, Canada, is an area with a long history of research on air pollution (Adams et al., 2012; Arain et al., 2007, 2009; Buzzelli et al., 2003; Jerrett et al., 2004, 2001; Kanaroglou et al., 2013; Sahsuvaroglu et al., 2006; Wallace and Kanaroglou, 2009, 2008; Wallace et al., 2009). With a population of about 520,000 (Statistics Canada, 2012), the city is diverse in land use with a major industrial complex along the city's northern edge, a downtown core, subdivision surrounding the core and rural regions at the extremities. Two inter-city freeways pass through the city, one along the north end of the city and one along the western side of the city. Two intra-city freeways complete the freeway network, with one freeway in the city's southern region and another in the west. The city is separated by an escarpment, which is a significant factor in accessibility as active transportation is limited to long and steep staircases. The escarpment is also responsible for variable air pollution conditions; for example elevated concentrations often occur in the lower city due to temperature inversions that are avoided in the upper city (Wallace et al., 2010). Hamilton has significant spatial variation in socio-economic factors, with detrimental health outcomes (Patrick F. DeLuca et al., 2012). The study area is presented in with the 2011 population density included in Figure 5.1.

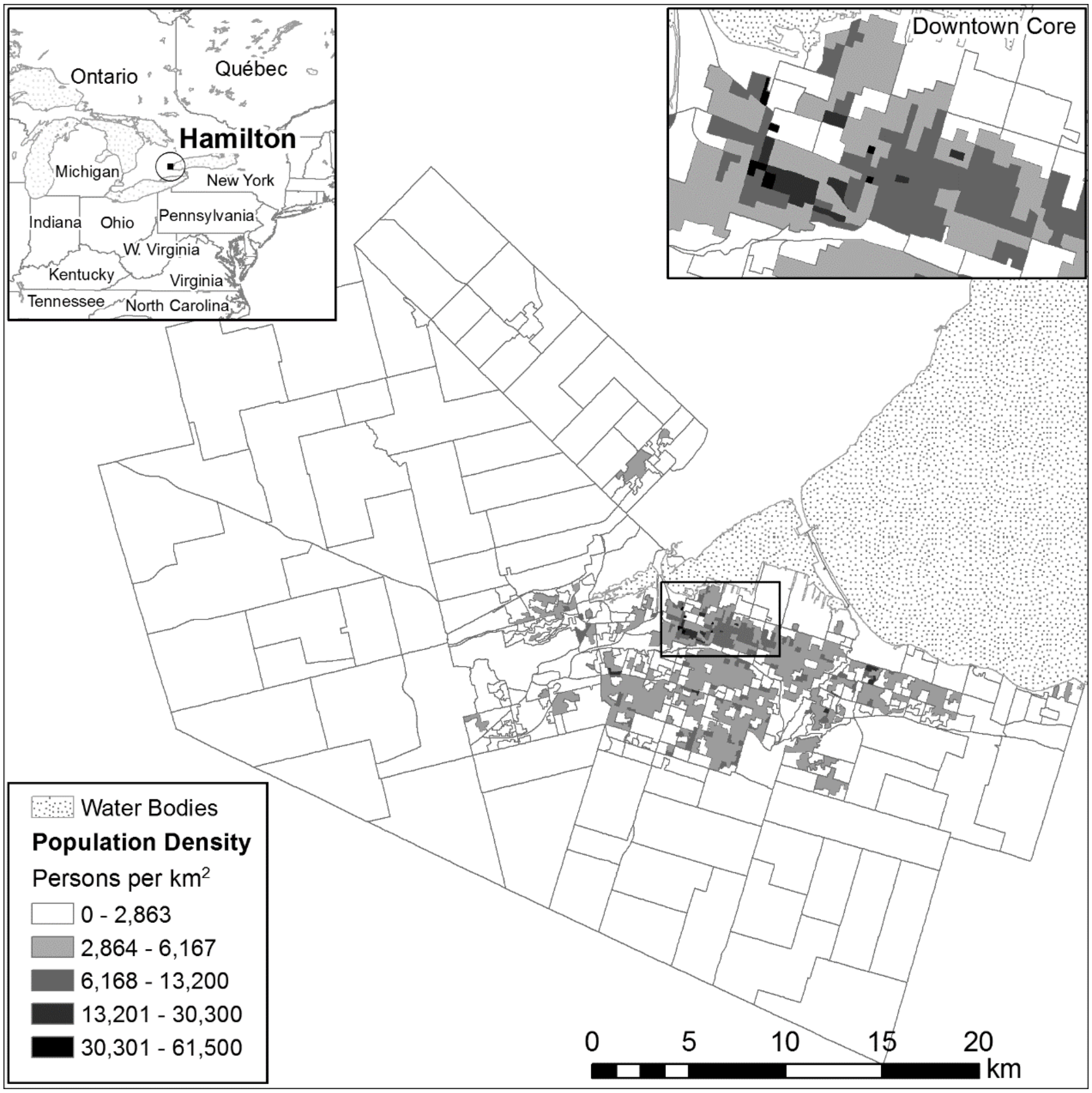

Figure 5.1 Hamilton, Ontario with 2011 population density

5.2.2 Home and School Shortest Path Analysis

The routes to school were provided by Bennet and Yiannakoulias (2015). We provide an overview of their shortest path analysis in this paper. It has been suggested that the routes children take between their home and school is determined by the shortest path

(Cooper et al., 2010; Hill, 1984); however, there are some contradictory findings that indicate that children may not take the shortest path to school and that the built environment can affect the route choice (Buliung et al., 2013). For our analysis, we utilized routes to schools that are based on the shortest path. We accept any potential conflicts until a better understanding of Hamilton's students' routes is defined. The seventy-three primary schools (destinations) involved in the study were part of the Hamilton Wentworth District School Board, where 80% of children reside within the catchment area of their primary school (Hamilton Wentworth District School Board, 2011). During route generation all potential routes ensured students were taking routes to the school within the catchment area. Some students will travel to a school outside of their catchment area because either a parent or the school board chooses to send that student to a different school. In the region, active transportation is only likely for students living within 1.6 km of their school, beyond this distance students are provided transportation by the school board. School catchment boundaries were obtained from the Hamilton Wentworth District School Board (Hamilton Wentworth District School Board, 2012).

Origins were defined using a parcel land database that identified the lot of each house in the city (City of Hamilton, 2010). Parcels were randomly selected within each dissemination area based on the number of children (Statistics Canada, 2008). For example, if 150 school aged children were identified as living within a dissemination area, 150 houses were selected as origins.

The routes between the origins (land parcels) and the destinations (primary schools in the Hamilton Wentworth District School Board) were derived along a pedestrian network, which included additional route options from the standard vehicular network; particularly, green-pathways that include parks, schoolyards, walking trails and other corridors. The base network included all the road links in the study area (City of Hamilton, 2010), which was supplemented with green-pathway data (City of Hamilton, 2005) and manual edits using on-screen digitizing of short-cuts and unmarked pedestrian infrastructure with satellite imagery (Google Inc., 2013). Road links that prohibit walking were removed, which included expressways and major highways.

5.2.3 Air Pollution Model

Particulate matter 2.5 microns ($PM_{2.5}$) concentrations were measured with mobile air pollution monitoring campaigns using a roving scan approach, see Wallace et al., (2009) for a detailed description of the technique. In general, the roving scan approach uses a van with air pollution monitoring units mounted inside, and an air intake mounted on top of the vehicle pointing in the direction of travel. The vehicle moves throughout the city to collect air pollution concentrations in a number of different locations. Data collection occurred between 2005 and 2013 covering the city with a high-spatial resolution, scans were conducted along expressways, major and minor roads, in the downtown core and the surround suburbs, and the city's rural regions. Previous analysis of these data for SO_2 concentrations that were co-collected with the $PM_{2.5}$ data demonstrated spatial variation in Hamilton (Kanaroglou et al., 2013). In the SO_2 study, data were used in a land use regression model to predict the air pollution concentrations across Hamilton. Land use regression, models air pollution concentrations using information on surrounding land use as independent variables, which may include land use types, population densities, amount of nearby road, and surrounding traffic density with a regression model. These models may be used to estimate exposure for long or short term analyses (Dons et al., 2014). We are modelling short-term exposures with the use of artificial neural network models instead of linear regression models for additional modelling flexibility. A multilayer perception feed-forward neural network model, which is a fully connected model, was trained with the back propagation learning function. Model processing was done in R (R Core Team, 2014), and the neural network models were fit using the Stuttgart Neural Network Simulator ported to R (Bergmeir and Benitez, 2012). A detailed description of an early version of the model is available in Adams et al. (2013).

Our predictor data included: (1) Land use data, calculated as the area of each land use within circular buffers of 50, 100, 200, and 400 meters centered on the location of air pollution monitoring. The land use classes included residential land, commercial, parks and recreation, resource and industrial, open area, and government and institutional. (2) Using a set of local wind monitors situated in Hamilton we coded a set of dummy variables for the four quadrants of direction. (3) A set of indicator variables for each of the land use types were included, a value of one if the monitoring occurred in that land use type, zero otherwise. (4) The x and y coordinate values from the Universal Transverse Mercator coordinate system. (5) The total length of each freeways and highways, major roads, and minor roads within 25, 50, 100, 200, 400, 800, and 1600 meter buffers. (6) Air pollution concentrations at the time of mobile monitoring observed at five different pollution monitors operated by the Ontario Ministry of Environment, presented in Table 5.1.

Table 5.1 Pollutants measured at stationary monitors.

Monitor (ID)	$PM_{2.5}$	O_3	NO	NO_2	NO_X	SO_2	CO
Burlington	Y	Y		Y			
Chatham	Y	Y	Y	Y	Y		
Hamilton Mountain	Y	Y					
Hamilton West	Y	Y					
Hamilton Downtown	Y	Y	Y	Y	Y	Y	Y

The validation of the model was determined by withholding 15% of the data, which was the most recently collected 15% of the data. This approach should reduce the correlation in the training and validation datasets and help confirm a stationary process. These data were predicted with the model, and the R^2 was calculated to estimate prediction ability, $R^2 = 0.78$. The model was applied to predict the air pollution concentrations at locations across the children's routes during every day of the school year in 2013.

5.2.4 Dose Exposure

In this work, we calculate the exposure for two modes of active transportation, which includes walking and cycling. The exposure metric is the total dose of air pollution that the child is exposed to during their trips to and from school. Air pollution dose is largely dependent on the ambient concentration and the inhalation rate of the person, which is often measured by the number of litres of air per minute and known as the minute ventilation rate (MVR).

We derived minute ventilation rates from the literature based on metabolic equivalent of task (MET) for both cycling and walking, which is a measure of the energy cost of physical activity and is strongly associated with ventilation rates. The EPA's Child-Specific Exposure Factors Handbook (United States Environmental Protection Agency,

2008) provides minute ventilation rates based on MET. To obtain these values we applied the following steps:

1. Determine the MET for both activates, cycling and walking.
    a. Arvidsson et al. (2007) included a review of children's activities and the associated MET values, along with their own the determination of MET values for many activities. We use the rate of 12 km/hr for the cycling speed and 4 km/hr for walking speeds, which have corresponding MET rates of 5.9 and 3.3 respectively in children (Arvidsson et al., 2007).
2. Translate the MET values to MVR from the EPA's Child-Specific Exposure Factors Handbook, which provides percentile data of MVR for different ranges of METs values. Both of the actives fell within the same range of MET values, which was 3 to 6. To obtain these values we:
    a. Determined the percentile of each of our MET values between 3 and 6.
    b. Utilized the percentile of the MET values along with the MVR data to extract the MVR value at the matching percentile provided by the EPA. For example, walking with walking have a percentile value of 0.1, we determine what the MVR value was at the 0.1 percentile to use as the MVR rate.

The MVR rates and associated velocities applied are included in Table 5.2.

Table 5.2 MET and MVR rates for both modes with velocity.

Activity	METs	MVR (m^3 / minute)	Velocity (m/min)
Cycling 12 km/hr	5.9	0.035	200
Walk 4 km/hr	3.3	0.017	66

5.2.5 Exposure Assignment

The ambient air pollution concentrations along the routes to school vary as the student travels, which can be caused by many factors including their distance from major roads and changing land use conditions. The route data to and from schools consisted of a set nodes connected with links (lines). Nodes were placed during route assignment when direction changed and the lines indicated the distance travelled between nodes, an example is presented in Figure 5.2 panel A. Our air pollution model is designed to predict concentrations at individual locations, e.g. an x-y location and not along a line.

To assign air pollution concentrations along the routes, we split the route into sections. When the distance between the start and end nodes of a section was less than 10 meters, it was kept intact and assigned a section number based on the id of its start node. If the distance was greater than 10 meters from the start to end node of a potential section, we added additional nodes every 10 meters until there were less than 10 meters to the end node. This process is exemplified in Figure 5.2 panel B. Sections were assigned the air pollution concentration calculated from the model at section's start node. The dose calculation for routes is presented in equation 3.

$$D_{rm} = \sum_{i=1}^{n} \frac{S_i}{V_m} * MVR_m * C_i \quad (3)$$

Where, D_{rm} is the dose for route r of mode m in micrograms, m is either walking or cycling, S_i is the length of the section of path counted from i =1 to the number of sections in meters, V_m is the velocity of mode m in meters per second, MVR is the minute ventilation rate of mode m in meters cubed per minute, and C_i is the air pollution concentration assigned to S_i in micrograms per meter cubed.

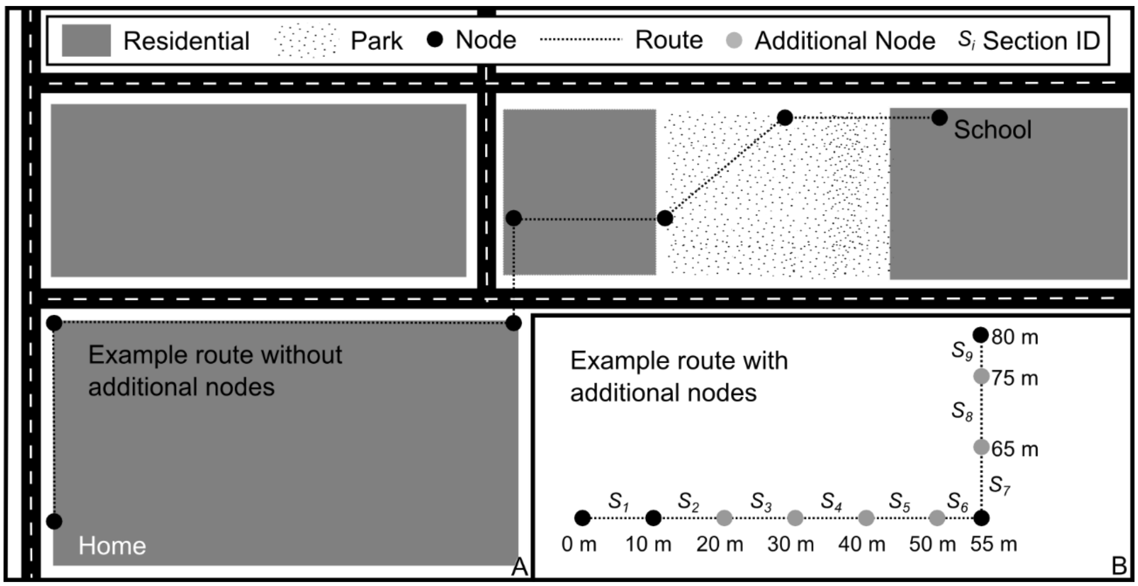

Figure 5.2 An example of a route to school and additional nodes. (A) Route to school example. (B) An example of additional nodes placed at ten-meter intervals along a route for air pollution concentration estimates.

5.2.6 Statistical Analysis

In our analysis, descriptive statistics were used to describe pollution concentrations and doses, and socioeconomic status variables. The dose exposure data were not transformed because they approximated normal distributions. In our analysis, we randomly selected 250 routes from the set of generated routes to assign air pollution exposure. These were one-way routes from the home to the school for the morning estimates and then from school to home for the afternoon journey. For the starting location of each route (the child's home), we extracted the neighbourhoods' median and mean income from the 2011 National Household Survey conducted by Statistics Canada. Data were available at the Dissemination Area level of geography, which are regions consisting of 400 – 700 people (Statistics Canada, 2015). Univariate and multivariate linear regression models were applied to determine if lower SES conditions were related to higher doses during the trips to or from school or the ambient conditions at home or school. Following an analysis of the total population, we stratified the population into two groups based on the mean income of their neighbourhoods. The two groups were composed of: 1) students with mean incomes less than the 20^{th} percentile, and 2) students with mean incomes higher than the 80^{th} percentile. These two groups were compared based on their distances travelled, household and school exposures, and dose exposures with t-tests. All statistical analysis was conducted in R (R Core Team, 2014)

5.3 Results

The morning average ambient air pollution concentrations at the students' households ranged between 10.4 and 27.1 μg/m^3 with a mean of 15.7 μg/m^3. The afternoon demonstrated a larger variation with a range of 9.3 – 31.2 μg/m^3 with a similar mean of 15.6 μg/m^3. We applied a paired student's *t*-test to determine if there was a significant difference between the households' ambient conditions in the morning or the afternoon, the test indicated no significant difference, $t = 1.1$, df = 249, $p = 0.27$. These household ambient conditions are presented spatially across Hamilton in Figure 5.3 panel A for the morning and panel B for the afternoon.

Ambient conditions for students' schools during the morning ranged from 13.3 – 27.6 and from 11.8 – 32.3 μg/m^3 during the afternoon with a morning mean of 18.3 μg/m^3 and an afternoon mean of 19.3 μg/m^3. A paired student's *t*-test indicated a significant difference between the two sets of the morning and afternoon school ambient air pollution concentrations (t = -8.8, df = 249, $p < 0.001$). These values are presented spatially across Hamilton in Figure 5.3 panel C for the morning and panel D for the afternoon.

The mean household income varied between $15,220 (Canadian Dollars) and $80,920 with a household average income of $37,340. These values are presented spatially across Hamilton in Figure 5.3 panel E. The median household income varied between $6,699 and $60,630 with an average of $30,830. These values are presented spatially across Hamilton in Figure 5.3 panel F.

We compared the ambient concentrations of students' households and their respective schools' with a paired *t*-test. Both the morning and afternoon school concentrations were significant higher ($p < 0.001$) than the household concentrations.

The population density of student household locations varied between 62 and 34,450 people per square kilometer, with a mean of 4,068 people per square kilometer in the City of Hamilton. The population density is presented in Figure 5.1.

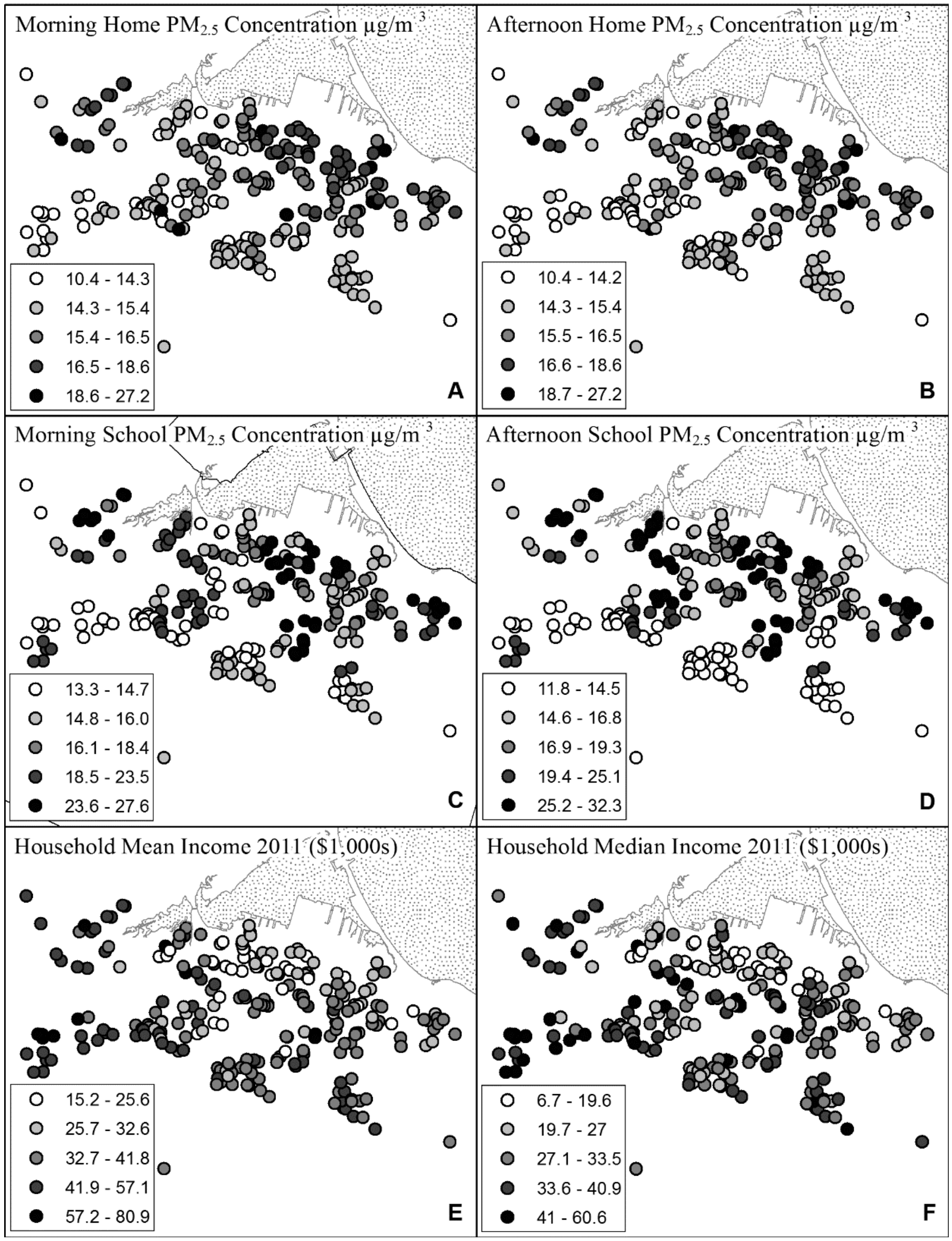

Figure 5.3 Students' morning home air pollution ambient conditions (A), afternoon home air pollution ambient conditions (B), morning school air pollution ambient conditions (C), afternoon school air pollution ambient conditions (D), household mean income (E), and median household income (F).

Air pollution doses for cycling to school varied in the mornings between 0.085 and 5.67 µg with a mean dose of 2.17 µg. The afternoon doses' distribution was significantly different than the morning doses' distribution, identified with a paired *t*-test (t = -2.3667, df = 247, p-value = 0.018), the mean of the afternoon doses was higher at 2.19 µg with a range of 0.097 to 5.61 µg.

For walking trips between home and school, doses ranged in the mornings between 0.126 and 8.327 µg with a mean dose of 3.19 µg. The afternoon doses' distribution was significantly different than the morning doses' distribution identified with a paired *t*-test (t = -2.3667, df = 247, p-value = 0.018), the mean of the afternoon doses was 3.23 µg with a range of 0.14 to 8.26 µg.

Linear regression models were applied to determine if a relationship occurred between the morning and afternoon dose exposure of $PM_{2.5}$ on the trips to and from school and the socioeconomic conditions of the students' neighbourhoods. First, we fit univariate regressions of cycling dose exposure, walking dose exposure, school exposure and household exposure for both the morning and afternoon value with each socioeconomic indicator as independent variables. The results of the morning are presented in Table 5.3 and the afternoon are presented in Table 5.4.

Multivariate linear regression models were fit with the independent variables that demonstrated a significant relationship to either cycling exposure or walking exposure, which were the exposure outcomes with multiple significant univariate models. The four multivariate models that were combinations of population density and route distance to predict morning cycling exposure, morning walking exposure, afternoon cycling exposure, and afternoon walking exposure did not perform any better than the univariate model fit only with route distance. The R^2 value was the same as the univariate model for route distance, and the population density predictor was insignificant in all models ($p > 0.05$).

Table 5.3 Univariate linear regression results for morning trips.

Dependent	Independent	Adj. R^2	*p*	Coefficient
Cycling Exposure	Median Income ($10,000s)	0.003	0.16	-0.1043
Cycling Exposure	Mean Income ($10,000s)	0	0.87	-0.1061
Cycling Exposure	**Population Density**	**0.4**	**0.001**	**-0.26026**
Cycling Exposure	**Route Distance (Natural Logarithm)**	**0.95**	**<0.001**	**2.746e-03**
Walking Exposure	Median Income ($10,000s)	0.003	0.16	-0.1534
Walking Exposure	Mean Income ($10,000s)	0	0.87	-0.1561
Walking Exposure	**Population Density**	**0.4**	**0.001**	**-0.3831**
Walking Exposure	**Route Distance (Natural Logarithm)**	**0.95**	**<0.001**	**4.042e-03**
Household Pollution	Median Income ($10,000s)	0.01	0.10	-0.2078
Household Pollution	**Mean Income ($10,000s)**	**0.02**	**0.03**	**-0.2430**
Household Pollution	Population Density	0.002	0.23	-0.1698
School Pollution	Median Income ($10,000s)	0	0.39	-0.2488
School Pollution	Mean Income ($10,000s)	0	0.87	-0.0402
School Pollution	Population Density	0	0.56	0.1832

Models significant at $p < 0.05$ are bolded.

Table 5.4 Univariate linear regression results for afternoon trips.

Dependent	Independent	Adj. R^2	*p*	Coefficient
Cycling Exposure	Median Income ($10,000s)	0.01	0.08	-0.1320
Cycling Exposure	Mean Income ($10,000s)	0	0.53	-0.0405
Cycling Exposure	**Population Density**	**0.03**	**0.01**	**-0.2257**
Cycling Exposure	**Route Distance (Natural Logarithm)**	**0.91**	**< 0.01**	**2.691e-03**
Walking Exposure	Median Income ($10,000s)	0.001	0.07	-0.1942
Walking Exposure	Mean Income ($10,000s)	0	0.53	-0.0597
Walking Exposure	**Population Density**	**0.03**	**0.01**	**-0.3321**
Walking Exposure	**Route Distance (Natural Logarithm)**	**0.91**	**< 0.01**	**3.961e-03**
Household Pollution	Median Income ($10,000s)	0.007	0.09	-0.2849
Household Pollution	**Mean Income ($10,000s)**	**0.02**	**0.01**	**-0.3821**
Household Pollution	Population Density	0	0.75	-0.06065
School Pollution	Median Income ($10,000s)	0	0.25	-0.4542
School Pollution	Mean Income ($10,000s)	0	0.59	-0.1833
School Pollution	Population Density	0	0.42	0.3491

Models significant at $p < 0.05$ are bolded.

We stratified the population into two groups, the first group includes those students with mean incomes in the 20th and lower percentile and the second group with mean incomes of the 80th and higher percentile. The mean values for the distances travelled to school; and the morning and afternoon school, household, walking dose and cycling doses are presented in Table 5.5.

Table 5.5 Comparison of students' trips between the highest and lowest 20% of students based on their neighbourhood incomes.

		Group Mean Values		*t*-test	
	Time	20^{th}	80^{th}	*t* statistic	*p* value
Distance	-	739.2	795.4	-0.717	0.48
School Pollution	Morning	18.7	19.1	-0.407	0.68
	Afternoon	20.6	20.4	0.121	0.90
Home Pollution	**Morning**	**16.2**	**15.4**	**2.043**	**0.04**
	Afternoon	**16.8**	**15.4**	**2.747**	**0.007**
Walking Dose	Morning	3.27	3.26	0.022	0.98
	Afternoon	3.43	3.27	0.464	0.64
Cycling Dose	Morning	2.22	2.21	0.022	0.98
	Afternoon	2.34	2.22	0.464	0.64

t-tests significant at $p < 0.05$ are bolded.

5.4 Discussion

In Hamilton, it was identified that lower income residents were exposed to greater air pollution concentrations than higher income residents with exposure data from 1984 – 94 (Jerrett et al., 2001). This conclusion is very logical, as it is known that lower value housing, which people having lower incomes can afford, occur in the less desirable portions of a city that can include areas with higher air pollution conditions (Harrison and Rubinfeld, 1978). Our analysis was only concerned with the hours during which students travel to and from school; however, when we regressed both school air pollution concentrations against both the median or mean household income no relationship existed. Regressing the same independent variables against household air pollution during both the morning and afternoon periods, we find that the 2% of the variation in the data set can be accounted for by the variation in income; however, this is not an indication of a causal factor between low-income and an increase in ambient air pollution concentrations. This indicates that

Hamilton's lowest income students/families are not being exposed to any greater concentrations during the two periods we studied at their schools, but they are slightly more exposed to air pollution concentrations at their household. This effect we identified is likely attenuated because we only have mean and median household incomes at the dissemination area level of geography from Statistics Canada. These areas are generally small but comprise of 400 – 700 people. Individual level income data may elicit a greater effect.

We were interested in examining the air pollution dose that is inhaled during the trip to school and during the trip back to home when students used active transportation. In the Toronto and Hamilton region, these active trips account for 38% and 41% percent of students travelling to and from school respectively. During the trip to school, cycling accounts for only 2.6% of the active mode trips and for the trip home from school cycling only accounts for 2.5% of the active mode trips (Metrolinx, 2010). Our results indicate that no significant difference in $PM_{2.5}$ dose occurs between students who are travelling either to or from school when their homes are in lower income neighbours of Hamilton when compared with the remaining population. A significant relationship does exist between air pollution doses inhaled during the trip to school and the density of the housing where the student resides. This effect of density on the dose can only explain four percent in the variation of morning cycling and walking trips and three percent of the variation in air pollution dose for afternoon walking and cycling trips. The primary agent for the resulting air pollution dose during a trip to or from school in Hamilton is simply the length of the trip, which during the morning accounts for 95% of the variation in both cycling and walking morning trips and 91% of the variation in afternoon cycling and walking trips.

Interestingly, we find in Hamilton that when we compare the students' household morning and afternoon ambient air pollution concentrations, no difference in the air pollution concentrations occurs. However, the schools that the students are attending have a slightly significantly higher concentrations during the afternoon than in the morning. This increase is only one microgram per metered cubed of air, which would account for small differences in health outcomes. We do make note that the test applied, the paired *t*-test is sensitive to small variations due to a high statistical power. Along a similar line, the dose for both walking and cycling to school was significantly different when compared to the dose during the afternoon. The difference between the means was 0.02 μg. We view this difference as having no true effect on health outcomes and that the dose of air pollution inhaled does not vary for students in Hamilton between their morning and afternoon trips.

When we stratified the population into the highest and lowest household income groups we found that no significant difference occurred between their air pollution dose exposures for either cycling or walking; however, the lower income groups' routes are only 93% of the distance to school of the higher income groups' routes. Though this distance is not statistically significantly longer it does indicate, since the higher incomes groups' dose exposure is lower for walking and cycling in both the morning and afternoon, that the lower-income students are walking shorter distances with higher ambient air pollution conditions.

Traditional analysis of air pollution exposure relies solely on monitoring units that are permanently fixed, often few in number (Ryan and LeMasters, 2007), and may have not been located for exposure analysis or to capture the variation in air pollution across the spatial extent of the city (Adams et al., 2012). Our research presents a modern approach to exposure analysis that includes both time-activity patterns and models capable of estimating finely resolved air pollution concentrations, which is being called for by the epidemiological literature (Baxter et al., 2013a; Buonanno et al., 2013).

5.5 Conclusion

Air pollution exposure during the morning and afternoon trips between home and school are not environmentally unjust when considering income as a socioeconomic indicator in Hamilton, Ontario during our study period. As well, only a minor effect is demonstrated with lower income households, having higher air pollution ambient concentrations during the hours when students are travelling between home and school. A comparative analysis of historic data would lend insight into the potential drivers of these results, which is not possible due to significant monitoring changes over time. However, we postulate that the reason students from lower-income households are not affected by higher air pollution concentrations in Hamilton is the gentrification of many neighbourhoods has been occurring in Hamilton, similar to the land use changes in the nearby City of Toronto (Skaburskis, 2012), a reduced industrial sector, and land use change.

We conclude that the doses of air pollution in Hamilton during active transportation to school is primarily controlled by the route length. When students arrive at their school, they are exposed to higher ambient air pollution concentrations than they were at home. Our results indicate that students who ride their bikes to school when they engage in active transportation are having reduced doses of air pollution during their journey.

The primary policy implication from this work is that programs that encourage active transportation should focus on encouraging cycling as the active mode to school.

5.6 References

Adams, M. D., De Luca, P. F., Corr, D., & Kanaroglou, P. S. (2012). Mobile Air Monitoring: Measuring Change in Air Quality in the City of Hamilton, 2005–2010. *Social Indicators Research*, 108(2), 351–364. doi:10.1007/s11205-012-0061-5

Adams, M. D., Luca, P. F. D. E., Corr, D., & Kanaroglou, P. S. (2013). AIR QUALITY HEALTH INDEX MAPPING : A DATA DRIVEN MODELLING APPROACH. In *Proceedings of the 13th International Conference on Environmental Science and Technology* (p. 8pp). Athens, Greece.

Arain, M. A., Blair, R., Finkelstein, N., Brook, J., & Jerrett, M. (2009). Meteorological influences on the spatial and temporal variability of NO 2 in Toronto and Hamilton. *Canadian Geographer*, 53(2), 165–190. doi:10.1111/j.1541-0064.2009.00252.x

Arain, M. a., Blair, R., Finkelstein, N., Brook, J. R., Sahsuvaroglu, T., Beckerman, B., et al. (2007). The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment*, 41(16), 3453–3464. doi:10.1016/j.atmosenv.2006.11.063

Arvidsson, D., Slinde, F., Larsson, S., & Hulthén, L. (2007). Energy cost of physical activities in children: validation of SenseWear Armband. *Medicine and science in sports and exercise*, 39(11), 2076–84. doi:10.1249/mss.0b013e31814fb439

Baxter, L. K., Burke, J., Lunden, M., Turpin, B. J., Rich, D. Q., Thevenet-Morrison, K., et al. (2013). Influence of human activity patterns, particle composition, and residential air exchange rates on modeled distributions of PM2.5 exposure compared with central-site monitoring *data. Journal of Exposure Science and Environmental Epidemiology*, 23(3), 241–7. doi:10.1038/jes.2012.118

Bennet, S. A., & Yiannakoulias, N. (2015). Motor-vehicle collisions involving child pedestrians at intersection and mid-block locations. *Accident Analysis & Prevention*, 78, 94–103. doi:10.1016/j.aap.2015.03.001

Bergmeir, C., & Benitez, J. M. (2012). Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *Journal of Statistical Software*, 46, 1–26.

Briggs, D. J., de Hoogh, K., Morris, C., & Gulliver, J. (2008). Effects of travel mode on exposures to particulate air pollution. *Environment International*, 34(1), 12–22. doi:10.1016/j.envint.2007.06.011

Brunekreef, B., Janssen, N., Hartog, J. de, Harssema, H., Knape, M., & van Vliet, P. (1997). Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology*, 8(3), 298–303. http://www.jstor.org/stable/10.2307/3702257. Accessed 3 February 2014

Buliung, R. N., Larsen, K., Faulkner, G. E. J., & Stone, M. R. (2013). The “path” not taken: Exploring structural differences in mapped-versus shortest-network-path school travel routes. *American Journal of Public Health*, 103(9), 1589–1596. doi:10.2105/AJPH.2012.301172

Buonanno, G., Stabile, L., & Morawska, L. (2013). Personal exposure to ultrafine particles: The influence of time-activity patterns. *The Science of the Total environment*, 468-469C, 903–907. doi:10.1016/j.scitotenv.2013.09.016

Buzzelli, M., Jerrett, M., Burnett, R., & Finklestein, N. (2003). Spatiotemporal perspectives on air pollution and environmental justice in Hamilton, Canada, 1985–1996. *Annals of the Association of American Geographers*, 93(3), 557–573. doi:10.1111/1467-8306.9303003

Cooper, A. R., Page, A. S., Wheeler, B. W., Griew, P., Davis, L., Hillsdon, M., & Jago, R. (2010). Mapping the Walk to School Using Accelerometry Combined with a Global Positioning System. *American Journal of Preventive Medicine*, 38(2), 178–183. doi:10.1016/j.amepre.2009.10.036

DeLuca, P. F., Buist, S., & Johnston, N. (2012). The Code Red Project: Engaging Communities in Health System Change in Hamilton, Canada. *Social Indicators Research*, 108(2), 317–327. doi:10.1007/s11205-012-0068-y

Dons, E., Van Poppel, M., Panis, L. I., De Prins, S., Berghmans, P., Koppen, G., & Matheeussen, C. (2014). Land use regression models as a tool for short, medium and long term exposure to traffic related air pollution. *The Science of the Total Environment*, 476-477C, 378–386. doi:10.1016/j.scitotenv.2014.01.025

Freire, C., Ramos, R., Puertas, R., Lopez-Espinosa, M.-J., Julvez, J., Aguilera, I., et al. (2010). Association of traffic-related air pollution with cognitive development in children. *The Journal of Epidemiology and Community Health*, 64(3), 223–8. doi:10.1136/jech.2008.084574

Gauderman, W., Avol, E., Gilliland, F. D., Vora, H., Thomas, D., Berhane, K., et al. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *The New England Journal of Medicine*, 351(11), 1057–1067. http://www.nejm.org/doi/full/10.1056/nejmoa040610. Accessed 3 February 2014

Gauderman, W. J. (2000). Association between air pollution and lung function growth in southern California children. *American Journal of Respiratory and Critical Care Medicine*, 162(4 Pt 1), 1383–90. doi:10.1164/ajrccm.162.4.9909096

Gulliver, J., & Briggs, D. J. (2004). Personal exposure to particulate air pollution in transport microenvironments. *Atmospheric Environment*, 38(1), 1–8. doi:10.1016/j.atmosenv.2003.09.036

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1), 81–102. doi:10.1016/0095-0696(78)90006-2

Heck, J. E., Park, A. S., Qiu, J., Cockburn, M., & Ritz, B. (2013). An exploratory study of ambient air toxics exposure in pregnancy and the risk of neuroblastoma in offspring. *Environmental Research*, 1–6. doi:10.1016/j.envres.2013.09.002

Hill, M. R. (1984). Walking straight home from school: Pedestrian route choice by young children. Transportation Research Record (No. HS-038 232).

Hutter, H.-P., Haluza, D., Piegler, K., Hohenblum, P., Fröhlich, M., Scharf, S., et al. (2013). Semivolatile compounds in schools and their influence on cognitive performance of children. *International Journal of Occupational Medicine and Environmental Health*, 26(4). doi:10.2478/s13382-013-0125-z

Jerrett, M., Arain, M. A., Kanaroglou, P., Beckerman, B., Crouse, D., Gilbert, N. L., et al. (2007). Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology and Environmental Health, Part A.*, 70(3-4), 200–12. doi:10.1080/15287390600883018

Jerrett, M., Burnett, R. T., Brook, J., Kanaroglou, P., Giovis, C., Finkelstein, N., & Hutchison, B. (2004). Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, Canada. *The Journal of Epidemiology and Community Health*, 58(1), 31–40.

Jerrett, M., Burnett, R. T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., & Brook, J. R. (2001). A GIS - environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, 33(6), 955–973. doi:10.1068/a33137

Kanaroglou, P. S., Adams, M. D., De Luca, P. F., Corr, D., & Sohel, N. (2013). Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model. *Atmospheric Environment*, 79, 421–427. doi:10.1016/j.atmosenv.2013.07.014

Metrolinx. (2010). Metrolinx Greater Toronto and Hamilton Area School Travel Household Attitudinal Study Report. Canada.

Moridi, M., Ziaei, S., & Kazemnejad, A. (2013). Exposure to ambient air pollutants and spontaneous abortion. *Journal of Obstetrics and Gynaecology Research*. doi:10.1111/jog.12231

Nelson, J. P. (1978). Residential choice, hedonic prices, and the demand for urban air quality. *Journal of Urban Economics*, 5(3), 357–369. doi:10.1016/0094-1190(78)90016-5

Panis, L. I., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., et al. (2010). Exposure to particulate matter in traffic: A comparison of cyclists and car passengers. *Atmospheric Environment*, 44(19), 2263–2270. doi:10.1016/j.atmosenv.2010.04.028

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria. http://www.r-project.org/.

Ryan, P. H., & LeMasters, G. K. (2007). A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation toxicology*, 19 Suppl , 127–33. doi:10.1080/08958370701495998

Sahsuvaroglu, T., Arain, A., Kanaroglou, P., Finkelstein, N., Newbold, B., Jerrett, M., et al. (2006). A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario, Canada. *Journal of the Air & Waste Management Association*, 56(8), 1059–69.

Skaburskis, A. (2012). Gentrification and Toronto’s Changing Household Characteristics and Income Distribution. *Journal of Planning Education and Research*, 32(2), 191-203.

Smith, V. K., & Huang, J. C. (1993). Hedonic models and air pollution: Twenty-five years and counting. *Environmental and Resource Economics*, 3(4), 381–394. doi:10.1007/BF00418818

Suglia, S. F., Gryparis, A., Wright, R. O., Schwartz, J., & Wright, R. J. (2008). Association of black carbon with cognition among children in a prospective birth cohort study. *American Journal of Epidemiology*, 167(3), 280–6. doi:10.1093/aje/kwm308

United States Environmental Protection Agency. (2008). Child-Specific Exposure Factors Handbook. www.epa.gov/ncea. Accessed 11 February 2014

Urman, R., McConnell, R., Islam, T., Avol, E. L., Lurmann, F. W., Vora, H., et al. (2013). Associations of children's lung function with ambient air pollution: joint effects of regional and near-roadway pollutants. *Thorax*, 1–8. doi:10.1136/thoraxjnl-2012-203159

Vrijheid, M., Martinez, D., Aguilera, I., Ballester, F., Basterrechea, M., Esplugues, A., et al. (2012). Socioeconomic status and exposure to multiple environmental pollutants during pregnancy: evidence for environmental inequity? *The Journal of Epidemiology and Community Health*, 66(2), 106–113. doi:10.1136/jech.2010.117408

Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., & McCarry, B. (2009). Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring*, 11(5), 998–1003. doi:10.1039/b818477a

Wallace, J., Corr, D., & Kanaroglou, P. (2010). Topographic and spatial impacts of temperature inversions on air quality using mobile air pollution surveys. *The Science of the Total Environment*, 408(21), 5086–96. doi:10.1016/j.scitotenv.2010.06.020

Wallace, J., & Kanaroglou, P. (2008). Modeling NOx and NO2 emissions from mobile sources: A case study for Hamilton, Ontario, Canada. *Transportation Research Part D: Transport and Environment*, 13(5), 323–333. doi:10.1016/j.trd.2008.04.001

Wallace, J., & Kanaroglou, P. (2009). The sensitivity of OMI-derived nitrogen dioxide to boundary layer temperature inversions. *Atmospheric Environment*, 43(22-23), 3596–3604. doi:10.1016/j.atmosenv.2009.03.049

Wallner, P., Kundi, M., Moshammer, H., Piegler, K., Hohenblum, P., Scharf, S., et al. (2012). Indoor air in schools and lung function of Austrian school children. *Journal of Environmental Monitoring*, 14(7), 1976–82. doi:10.1039/c2em30059a

Wong, E. Y., Gohlke, J., Griffith, W. C., Farrow, S., & Faustman, E. M. (2003). Assessing the Health Benefits of Air Pollution Reduction for Children. *Environmental Health Perspectives*, 112(2), 226–232. doi:10.1289/ehp.6299

Zheng, S., Cao, J., Kahn, M. E., & Sun, C. (2014). Real Estate Valuation and Cross-Boundary Air Pollution Externalities: Evidence from Chinese Cities. *Journal of Real Estate Finance and Economics*, 48(3), 398–414. doi:10.1007/s11146-013-9405-4

Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., van den Hazel, P., & Brunekreef, B. (2010). Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, 118(6), 783–9. doi:10.1289/ehp.0901622

Chapter 6: Conclusions

6.1 Introduction

Many nations have worked to limit releases of air contaminants because of the implications for human health (Kampa and Castanas, 2008). However, these reductions or limitations have yet to eliminate this global health issue (Evans et al., 2013; van Donkelaar et al., 2010). Recently a briefing in the BMJ (British Medical Journal) named air pollution the public health problem that won't go away (Hawkes, 2015). Currently, many countries such as China, are facing, what can only be termed, extreme air pollution events (Tie and Cao, 2009). With all of the current issues, air pollution is still considered a global human health hazard that warrants continued research.

The intention of this thesis was to explore and provide value to the field of air pollution modelling, specifically to incorporate mobile air pollution monitoring data into air pollution models. As well, improve upon the techniques that are applied to assign air pollution exposure. The need to refine exposure assignment was identified in the epidemiological literature, which calls for new modelling approaches to reduce exposure misclassification. The ability to identify health effects is attenuated with exposure misclassification. Epidemiologists note that new models should incorporate human activity patterns for the assignment of air pollution exposure, utilize spatially variable input datasets, and continually refine the spatial precision (Ozkaynak et al., 2013). The assignment of air pollution exposure is a fundamental step in air pollution epidemiological research. It is, therefore, necessary to determine the relationship between contaminant concentrations and negative health outcomes.

We explored the utility of mobile air pollution monitoring units in this work, which are becoming an accepted research tool in the literature (Bukowiecki et al., 2002; Lightowlers et al., 2008; Qi and Shimamoto, 2012; Wallace et al., 2009; Xu et al., 2007). In this work, mobile monitoring has been utilized in a similar way to traditional stationary monitoring units, with the main difference of the units being able to be relocated within the study area. It is prudent to mention that mobile monitoring units have other purposes, which includes measuring in-situ emissions during car chase and mobile emissions studies (Westerdahl et al., 2005).

We have presented research that can lead to the implementation of improved risk awareness systems of air pollution containments. The provision of such a system will not necessarily result in a change in people's behavior, but how they perceive the risk will be the driver for behavioral change. In Hamilton, individuals perceived risk of air pollution varies considerably, which is partially attributable to their location in the city but also their socio-demographics (Simone et al. 2012). It is suggested that as an individual's risk exposure is increased, their perceived risk lowers over time. In many cities, the air pollution risk that is presented to individuals may not relate to their personal risk exposure because of the location of monitoring units relative to their own. This dissconnect may lead to reduced perceived risk over time. By localizing risk information, it may be possible to reduce the attenuation of perceived risk that is due to incorrect risk warnings.

6.2 Contribution to the Air Pollution Literature

The primary contribution to the literature of this thesis is the demonstration of a model that is able to combine both mobile and stationary air pollution data to model the spatial variation in a city that can be used to assign air pollution exposure to specific human activities, such as the trip to work or school (as was demonstrated in this thesis).

6.2.1 Supplementing Stationary Monitoring Units with Mobile Units

- Mobile data collection allows for air pollutant concentrations to be obtained with a larger spatial coverage and higher density than is possible with stationary monitors because of the ability to relocate a single unit.
- Mobile monitoring coupled with stationary monitoring can provide a more detailed spatial distribution of air quality, which can be used to identify changes in the air shed that may have been missed solely with stationary monitoring units. This was demonstrated for Hamilton, Ontario, which has a monitoring network clustered around the historic industrial region.
- Seasonal fluctuations that occur in the mobile air pollution datasets, because of their discontinuous time-series, can be reduced with the use of a centrally located stationary monitoring unit. This allows for various regions in the city to be monitored with a single mobile unit.

6.2.2 Calculating long-term exposure with discontinuous mobile time-series data

Researchers can design the collection of mobile air pollution monitoring data to minimize error in calculating the long-term exposure.

- To reduce the error in a long-term estimate, mobile monitoring campaigns should balance the number of visits to a site with the length of monitoring during each visit.

Maximizing either the number of visits to a site or the length of a visit results in greater error than a balance of the two.

- If the monitoring campaign collects less than twenty-five percent of all possible observations at a single location, adjusting the data based on a central monitor can improve the long-term mean estimate.
- The classical error in estimating the long-term mean concentration can be estimated *a prior* if a few stationary monitoring units are present. This estimate can be used to help identify the need for further funding or to determine how many sites can be visited.
- Fixed location monitoring units that are the basis for adjustment of mobile time-series data should be situated in the same land use as the mobile units to ensure similar temporal variations.

Classical error will occur when estimating long-term concentrations with discontinuous time-series data; however, our findings provide valuable methods to both estimate *a prior* and minimize this error.

6.2.3 Modelling locally and temporally variable air pollution concentrations with mobile monitoring units.

In leveraging a number of new technologies and techniques a model capable of estimating the spatial distribution of $PM_{2.5}$ and NO_2 was developed. This model has a lot of flexibility in that it uses the information content from a number of different datasets:

a. Discontinuous time-series mobile air pollution monitoring data with a high spatial coverage.

b. Continuous time series stationary air pollution monitoring data. We leveraged monitors located in the city to obtain local temporal variations and background monitors to capture the temporal variations from long-distance transport.
c. Land use information, many studies have been able to equate land use information as predictors of ambient air pollution variation (Dons et al., 2014; Jerrett et al., 2005; Ryan and LeMasters, 2007). This is because many activities on specific land uses are air pollution generators, for example, vehicles on transportation networks and the operations in the industrial sector.
d. Meteorological data to control for variations in the wind patterns that affect the dispersion strength and direction of air contaminants.
e. Congestion data to identify areas susceptible to vehicular emissions during peak traffic hours.

This flexibility in data types was harnessed with neural network models predicting spatial fields of air pollution using a land use regression framework. We found that using neural network models as opposed to linear regression models improved modelling accuracy, likely due to the additional flexibility in the relationships between the predictor and outcome variables.

6.2.4 Activity Based Exposure Analysis

The $PM_{2.5}$ model developed in chapter four was applied to determine exposure in an activity-based analysis of active transportation trips for students between home and school. The air pollution model avoided the concerns with assigning air pollution concentrations from a central location or a sparse network (Ozkaynak et al., 2013). This assessment identified the following results:

- Air pollution dose exposure is not related to household income during the morning and afternoon trips between home and school in Hamilton, Ontario, during our study period.
- A minor effect is demonstrated with lower income households exposed to higher ambient air pollution concentrations during the hours when students are travelling between home and school. Their shorter route length offsets the total dose compared to higher income households.
- The dose of air pollution exposure is controlled primarily by the length of the route to school in Hamilton.
- Students arriving at their schools are exposed to higher ambient air pollution concentrations than they were at home.
- Students who ride their bikes to school when they engage in active transportation are exposed to reduced doses of air pollution during their journey.

6.3 Recommendations for Future Research

Two primary recommendations for future research are:

1. This research is solely concerned with the modelling of outdoor air pollutants; however, it is necessary to continue to refine the methods for assigning indoor exposure (Allen et al., 2012). A complete exposure model should be able to assign exposure incorporating variations in outdoor concentrations, indoor infiltration of contaminants, and indoor concentrations of air pollutants in buildings, vehicles and other structures.
2. The research contained within this thesis is limited to a selection of the criteria air contaminants. Many other air contaminants are present at elevated levels that can affect human health. It will be necessary to identify, if the techniques presented are transferable to pollutants with different source profiles, such as volatile organic compounds and toxic metals.

6.4 References

Allen, R. W., Adar, S. D., Avol, E., Cohen, M., Curl, C. L., Larson, T., et al. (2012). Modeling the residential infiltration of outdoor $PM_{2.5}$ in the multi-ethnic study of atherosclerosis and air pollution (MESA Air). *Environmental Health Perspectives*, 120(6), 824–830. doi:10.1289/ehp.1104447

Bukowiecki, N., Dommen, J., Prevot, A. S. ., Richter, R., Weingartner, E., & Baltensperger, U. (2002). A mobile pollutant measurement laboratory — measuring gas phase and aerosol ambient concentrations with high spatial and temporal resolution. *Atmospheric Environment*, 36, 5569–5579. doi:10.1016/S1352-2310(02)00694-5

Dons, E., Van Poppel, M., Panis, L. I., De Prins, S., Berghmans, P., Koppen, G., & Matheeussen, C. (2014). Land use regression models as a tool for short, medium and long term exposure to traffic related air pollution. *The Science of the Total Environment*, 476-477, 378–386. doi:10.1016/j.scitotenv.2014.01.025

Evans, J., van Donkelaar, A., Martin, R. V., Burnett, R., Rainham, D. G., Birkett, N. J., & Krewski, D. (2013). Estimates of global mortality attributable to particulate air pollution using satellite imagery. *Environmental Research*, 120, 33–42. doi:10.1016/j.envres.2012.08.005

Hawkes, N. (2015). Air pollution in UK: the public health problem that won't go away. *BMJ*, 350, h2757–h2757. doi:10.1136/bmj.h2757

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., et al. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Science and Environmental Epidemiology*, 15(2), 185–204. doi:10.1038/sj.jea.7500388

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, 151(2), 362–7. doi:10.1016/j.envpol.2007.06.012

Lightowlers, C., Nelson, T., Setton, E., & Peter Keller, C. (2008). Determining the spatial scale for analysing mobile measurements of air pollution. *Atmospheric Environment*, 42(23), 5933–5937. doi:10.1016/j.atmosenv.2008.03.033

Ozkaynak, H., Baxter, L. K., Dionisio, K. L., & Burke, J. (2013). Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science and Environmental Epidemiology*, 23(6), 566–72. doi:10.1038/jes.2013.15

Qi, J., & Shimamoto, S. (2012). A cognitive mobile sensor network for environment observation. *Telematics and Informatics*, 29(1), 26–32. doi:10.1016/j.tele.2011.03.001

Ryan, P. H., & LeMasters, G. K. (2007). A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation toxicology*, 19 (Suppl 1), 127–33. doi:10.1080/08958370701495998

Simone, D., Eyles, J., Newbold, K. B., Kitchen, P., & Williams, A. (2012). Air Quality in Hamilton: Who is Concerned? Perceptions from Three Neighbourhoods. *Social indicators research*, 108(2), 239-255.

Tie, X., & Cao, J. (2009). Aerosol pollution in China: Present and future impact on environment. *Particuology*, 7(6), 426–431. doi:10.1016/j.partic.2009.09.003

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, 118(6), 847–855. doi:10.1289/ehp.0901623

Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., & McCarry, B. (2009). Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring*, 11(5), 998–1003. doi:10.1039/b818477a

Westerdahl, D., Fruin, S., Sax, T., Fine, P. M., & Sioutas, C. (2005). Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment*, 39(20), 3597–3610. doi:10.1016/j.atmosenv.2005.02.034

Xu, X., Brook, J. R., & Guo, Y. (2007, November). A statistical assessment of saturation and mobile sampling strategies to estimate long-term average concentrations across urban areas. *Journal of the Air & Waste Management Association*. http://www.ncbi.nlm.nih.gov/pubmed/18069463