AIR POLLUTION AND RESPIRATORY HEALTH: RE-ANALYSIS OF THE HAMILTON CHILDREN'S COHORT STUDY

AIR POLLUTION AND RESPIRATORY HEALTH:

RE-ANALYSIS OF THE HAMILTON CHILDREN'S COHORT STUDY

By

THEODORA POULIOU, B.Sc. (Hons), M.Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Theodora Pouliou, August, 2005

MASTER OF SCIENCE (2005) McMaster University

(School of Geography and Earth Sciences) Hamilton, Ontario

TITLE: Air Pollution and Respiratory Health: Re- Analysis of the Hamilton Children's Cohort Study

AUTHOR: Theodora Pouliou, B.Sc. (Hons) (University of Aegean), M.Sc. (National Technical University of Athens)

SUPERVISORS: Dr. Susan J. Elliott and Dr. Pavlos S. Kanaroglou

NUMBER OF PAGES: ix, 159

ABSTRACT

Respiratory diseases have been and still are one of the major challenges in health research. While many studies have demonstrated that respiratory health varies by socio-economic status and environmental exposure, consensus among the scientific community is still lacking regarding the sufficiency of evidence to infer a causal relationship therefore highlighting the need for improved assessment of both exposure and outcome. This thesis addresses these issues through a reanalysis of the Hamilton Children's Cohort (HCC) study, undertaken in the 1980's in Hamilton, Ontario, Canada. The HCC study was the first one that provided important evidence linking adverse respiratory health with smoking, hospitalization in infancy and air pollution. However, given the limited development of spatial analysis and GIS at the time of study, the spatial dimensions of respiratory diseases and air pollution were not fully explored. The objective of this thesis is to re-analyse the HCC data in order to investigate the spatial variation of air pollution and children's adverse respiratory health as well as the relative importance of other characteristics that may determine the burden of respiratory incidences. Children's exposure to air pollution is first re-estimated using kriging and land-use regression. Based on exposure assessment analysis, compared to kriging, the land-use regression model performs better in capturing local variation of particulate pollution (TSP). The results of the land-use regression model are then used in multivariate linear and logistic regression analysis for the assessment of children's respiratory health. Findings of the multivariate analysis verify and strengthen the results of the HCC study indicating that in addition to their findings, children's respiratory health is associated with chest illness in

siblings, use of gas for cooking, low income and household crowding. In conclusion, this thesis demonstrates the usefulness of spatial analysis and GIS in assessing children's exposure to air pollution and also strengthens the HCC study findings revealing statistically stronger associations between children's respiratory health and a number of covariates. In general, the results hold promise and in combination with space-time analysis may lead to the development of advanced exposure assessment models in order to improve our understanding of the potential determinants of respiratory health.

ACKNOWLEDGEMENTS

This thesis would not have been completed without the support of several individuals. First, I would like to thank my two supervisors Susan Elliott and Pavlos Kanaroglou for all their help and guidance. Through the last two years you have not only been my supervisors but also my colleagues and my mentors. I am extremely grateful for your patience, accessibility and most importantly friendship that have marked me as a researcher as well as a person.

I would also like to thank Dr. David Pengelly, the principal investigator of the Hamilton Children's Cohort Study, for his enthusiasm, support, and suggestions. You have taught me that you can find the work that you dream off - no matter what.

The departmental staff has been also tremendous; thank you Ann Wallace for keeping me organized, Darlene Watson, Kath Philips, Josie D'Allesandro and Charlene Muflinn for answering all of my questions and helping me with the TA's issues.

Thanks to my friends Hanna, Karen, Maria, Ruben and Stella; you have made the whole thing worthwhile for me. Thanks to my family, for supporting me through all my endeavours and not stopping me when I decided to come to Canada. You have my deepest gratitude and love, always. Finally, thanks to Dimitri for being with me in every little moment of my life. You mean the world to me.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER ONE

INTRODUCTION

1.1 Research Context

In December 1952, stagnant weather conditions in London (UK) caused a sharp increase in the concentrations of particulate-based smog and sulphur dioxide in the atmosphere resulting in more than 4000 deaths in just a few days. The observed excess in mortality was mostly due to bronchitis, pneumonia, and acute exacerbation of underlying cardiac and respiratory diseases (McCally, 2002). This event, dubbed as the "London 1952 Smog", called for the enforcement of a series of strict regulations to reduce ambient air pollution in the London region with the Clean Air Act (1956) being the most important (McCally, 2002). As a result of this incident, air quality standards-guidelines on both national (Health Canada, 2000) and international (World Health Organization, 2000) scales have been introduced to prevent similar severe air pollution episodes from reoccurring.

However, current levels of air pollution continue to pose a threat to public health. Specifically, scientific studies have shown that there have been alterations in the air pollution mixture, which is now characterized by high concentrations of nitrogen oxides and photochemical oxidants, with motor vehicles emerging as the most important air pollution source (Brunekreef and Holgate 2002; Katsouyanni, 2003).

The effects of exposure to air pollution are different among different groups of individuals. In fact, there is considerable evidence showing that children are particular vulnerable to air pollution, due to their developmental, physiological and behavioural characteristics (WHO, 2002a). Specifically, children have more lung surface area per weight than adults, which means that they breathe more air per weight. In addition, their immune system and their organs are immature and this increases their susceptibility to the harmful effects of air pollution. Furthermore, children are more active than adults, spending more time outdoors, which may result in longer exposure to air pollutants of higher concentrations.

While many studies have indicated the relation between air pollution exposures and health effects on children, there are still many areas of uncertainty (WHO, 2002b). Consensus among the scientific community and the policy analysts/makers is still lacking regarding the sufficiency of evidences to infer a causal relationship between exposures to ambient air pollution and children's health. Indeed, further research is needed to allow for more precise exposure assessment and to better understand the relationship between childhood illness and air pollution. Most of the studies conducted in this direction have -so far- adopted a classical epidemiological stance. In contrast, few studies have taken into consideration a geographic perspective, examining the variability of health outcomes and their potential determinants over space. Specifically, only a limited number of them have applied spatial analysis and

GIS methods that will provide more precise assessment of the environment and health[1] relationship (Brunekreef et al., 1997; Hoek et al 2002).

The present study attempts to address this deficiency, elucidating spatial interactions and relationships in an effort to understand the complex dynamics, affecting children's respiratory health. In particular, this work is based on a previous study, undertaken in the city of Hamilton ON, an area of poor air quality (Pengelly et al., 1984; Pengelly et al., 1987). The Hamilton Children Cohort (HCC) study investigated the effects of air pollution on the respiratory health of approximately 3,500 school children aged 7-10 years old. Data were collected using survey questionnaires and clinical tests over five time periods across 6 years (1979 – 1985). Also, for the purposes of this study, a pollution-sampling network was set to measure exposure to Total Suspended Particulates (TSP), Particle Size Distribution (PSD) and Sulphur Dioxide (SO_2). However, due to limited number of monitoring stations for the PSD and SO_2, we only used the TSP data to assign children's exposure to air pollution. In particular, regression models were applied to investigate the potential impacts of particulate air pollution on children's respiratory health, while controlling for a number of confounders such as smoking, use of gas stoves, hospitalization in infancy and chest illness in siblings. The HCC study has provided important evidence linking these variables with poor respiratory health. Specifically, the results of the study demonstrated that there are strong effects of maternal smoking and severe respiratory disease in infancy on the respiratory health of children and

[1] The environment here has been defined in its broader context including both its physical and social aspects.

relatively weak effects of ambient air pollution. As far as it concerns the pulmonary function of children, the associations with air pollution were positive suggesting that pulmonary functioning is improved at higher levels of TSP pollution.

However, an important limitation of the HCC study is that the spatial dimension of health-related data was not fully taken into account for the analysis which may either strengthen the study results or may even lead to new conclusions. In particular, during the study period, sophisticated methods of spatial analysis and GIS were not applied on health – related studies and especially in order to assess exposure to air pollution. The HCC study was the first to measure and present the effects of particle size on children's respiratory health. However, it only accounted for the general/global trend of the air pollution data resulting in a rough estimate of air pollution exposure.

Therefore one of the most important aspects of this study is the fact that the data were collected 20 years ago but now there is opportunity of re-analysing them using more sophisticated analytical techniques to assess air pollution exposure.

1.2 Research Objectives

The main objective of this study is to advance the current state of knowledge regarding the complex interactions between air pollution and children's health and also to examine the role of space in this relationship. Specifically there are two main research objectives that will be pursued with the re-analysis of the HCC study:

1. Investigate the spatial variation of adverse respiratory health of children in Hamilton
2. Investigate the relative importance of individual, social and environmental characteristics that may determine the burden of respiratory incidences by re-assessing exposure to air pollution using advanced methods of spatial analysis and geographical information systems (GIS)

Meeting these objectives will contribute to the development of a clearer conceptual understanding of the relationship between air pollution exposure and children's respiratory health. Such knowledge will hopefully help in the improvement of health policy decisions and interventions at both national and international levels.

1.3 Contributions

This thesis contributes to the existing body of evidence concerning the association between air pollution and children's respiratory. Particularly, with the application of advanced methods of spatial data analysis and GIS, this study re-assesses more precisely children's air pollution exposure. The use of such methods to estimate exposure to air pollution has been adopted by a number of health-related studies (Jerrett et al., 2003; Cakmak et al., 2003; English et al., 1999; Wilkinson et al., 1999). However, this study takes this one step further by focusing specifically on children's respiratory health.

Another contribution lies in the methods used to assess exposure to particulate air pollution. This thesis is the first to determine Total Suspended Particulates (TSP) exposure employing the land-use regression analysis in North America.

In addition, this research study contributes to the extension and/or refinement of the potential determinants of respiratory health that especially affect children's development and well-being. Guided by the population health perspective, this thesis is one of the few respiratory health studies that demonstrate the relative importance of factors -other than air pollution- such as parental smoking, use of gas for cooking, hospitalization during infancy on children's respiratory health.

Finally, this study could also serve to motivate individuals and related governmental and non-governmental institutions to establish prevention - oriented policies and actions focusing on children and having as the major goal to improve air quality and protect them from environmental exposure hazards.

1.4 Outline

This thesis consists of six chapters. **Chapter Two** provides a review of the literature on the determinants of health firstly by setting a broader research context; secondly by setting the study within the disciplines of health geography and spatial analysis; and finally through a critical appraisal of the related literature to date. **Chapter Three** consists of a detailed description of the study area such as its physical, social and demographic characteristics. This is followed by a description of the data as well as the methods used to analyze them. Specifically, visualization techniques, exploration methods and applied models are described in detail. **Chapter Four** presents the research findings regarding the assessment of exposure to particulate air pollution based on two methodological approaches: a) kriging analysis and b) land-use regression. This is followed by the production of the estimated air pollution surfaces using advanced GIS techniques. **Chapter Five** presents the prevalence of respiratory health outcomes. Then, the discussion focuses on the potential associations between children's respiratory health and a number of individual, environmental and socio-economic factors based on linear and logistic regression modeling. Finally, in **Chapter Six** summarizes the most important findings followed by a discussion of the major contributions and the recommendations for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The relationship between air pollution and human health has increasingly been the focus of concern for scientists, policy-makers and the general public. For example, the Ontario Medical Association estimated that for the year 2000, 1,925 premature deaths, 9,807 hospital admissions, 13,146 emergency room visits and approximately 46 million sick days for employees as a result of air pollution were experienced in Ontario. These health impacts involve about $10 billion in annual economic costs (OMA, 2000). Children are particularly susceptible to air pollution because of their developing anatomy, the amount of exposure time to ambient air pollution, and the absence of lifestyle confounders such as smoking. Consequently, a number of research studies have been conducted to date to investigate the health impacts of exposure to adverse air quality on children's respiratory health taking into account other potential risk factors such as social environment, family income, access to health care facilities and lifestyle behaviours.

This chapter reviews the literature regarding the potential determinants of children's respiratory health emphasizing the impacts of exposure to particulate air pollution. The research is informed by the population health perspective (Lalonde, 1974; Evans et al., 1994). A review of empirical studies regarding the association of air pollution and children's respiratory health is also included in this chapter as well as a brief overview of the application of spatial analysis and GIS in health - related research.

2.2 The Theoretical Context

There has been difficulty in arriving at a single definition of health but its heart is in understanding that health is *"a resource for every day living, and not the objective of living"* (WHO, 1998); *"a medium through which we are able to meet our needs and realize our aspirations"* (Hayes, 1996). In the past, when infectious diseases were the predominant cause of illness and death, health was defined according to the biomedical approach as *"the balance so as to be reasonably free from pain and disability"* (Engel, 1963).

In the late 1940s, new thinking proposed that health was more than just the absence of disease. Specifically, in 1948, the World Health Organization provided a pioneering definition of health *"as the state of complete physical, mental and social well-being, not merely the absence of disease or infirmity"*. However, this definition of health has been criticized as utopian and unfeasible in that it identifies all human activity as health-related (Frankish et al., 1999; Evans and Stoddart, 1990). Making health synonymous with human development and quality of life in general makes health indistinguishable from its determinants and therefore it becomes impossible to talk about the contribution of health to well-being and quality of life. As Gatrell (2002) pointed out, the WHO definition of health presents an ideal state and that most, if not all, people are unhealthy at all times. The World Health Organization responded to this quest for refinement in 1984 and expanded the definition of health as *"the extent to which an individual or a group is able to realize aspirations and satisfy needs, and to change or cope with the environment. Health is a resource for everyday life, not the objective of living; it is a positive concept, emphasizing social and personal resources as well as physical capabilities".* That is, health is intimately tied not only to biological aspects of human beings but also to personal circumstance that in turn are tied to social, cultural, economic and environmental influences (Frankish et al., 1999). This, definition, therefore, presupposes that health goes beyond the scope of the individual. This view provides the foundation for our understanding of a population health approach that recognizes the multiple factors and conditions that contribute to health.

2.2.1 Population Health Perspective

The population health perspective (PHP) is based on an understanding that the determinants of health are both broad and represent a complex range of interacting relationships whereby these determinants interact among themselves in their influences on health. It builds on a long tradition of public health and health promotion. Canada has been at the forefront in the development of the population health perspective through its internationally acclaimed work in the area of public health (Health Canada, 2001). Specifically, in 1974 the federal government's White Paper "A New Perspective on the Health of Canadians", known also as the Lalonde Report, provided a new conceptual framework to inspire a paradigm shift related to the determinants of health. Lalonde was strongly influenced by Thomas McKeown's radical finding that the decline in mortality rates in England during the eighteenth century was mainly due to changes in living standards and not only due to advancements in medicine (McKay, 2000). Specifically, Lalonde (1974) summarized McKeown's conclusions:

> *"that, in order of importance the major contributions to improvement in health in England and Wales were from limitation of family size (a behavioural change), increase in food supplies and healthier physical environment (environmental influence) and specific preventive and therapeutic measures".*

Lalonde's conceptual framework for health [1] gave rise to national and international initiatives in knowledge development, health advocacy, and public health advancement. Discussion now began to turn to the critical determinants of health and in particular the influence of the social, economic and environmental covariates as well as the importance of diminishing inequalities in health status among population groups. This shifted in 1986, when Canada hosted the First International Conference on Health Promotion. From this conference, the *"Ottawa Charter for Health Promotion"* (World Health Organization, 1986) and the *"Achieving Health for All: A Framework for Health Promotion"* (Epp, 1986) documents expanded the Lalonde Report by focusing on the broader determinants of health. Also, in 1989, the Canadian Institute for Advanced Research (CIAR) introduced the population health concept, altering

[1] Lalonde defines that framework as the Health Field Concept comprising four main elements: human biology, lifestyle, the organization of health care and the social and physical environments in which people live. A detailed discussion of these elements is taken up in the following sections of this chapter.

the research agenda beyond health care to the social, economic and cultural forces that serve as determinants of health, as well as the importance of values such as equity, solidarity and participation in a population context. More recently, members of CIAR conceptualized the basis of the determinants of health and their complex interactions (Evans and Stoddart, 1990) which was later repeated and amplified in a collaborative and critically acclaimed book *Why Are Some People Healthy and Others Not?* (Evans et al, 1994). Also in 1994, the Federal, Provincial, and Territorial Advisory Committee on Population Health in Canada, in a report entitled Strategies for *Population Health: Investing in the Health of Canadians* defined population health as:

> *"...the health of a population as measured by health status indicators and as influenced by social, economic, and physical environments, personal health practices, individual capacity and coping skills, human biology, early childhood development and health services. As an approach population health focuses on the interrelated conditions and factors that influence the health of populations over the life course, identifies systematic variations in their patterns of occurrence, and applies the resulting knowledge to develop and implement policies and actions to improve the health and well-being of those populations".*

The population[2] health perspective has been also defined by Frankish et al., (1996) as:

> *"...the epidemiological and social condition of a community (defined by geography or by common interests) that minimizes morbidity and mortality, ensures equitable opportunities, promotes and protects health and achieves optimal quality of life".*

---

[2]According to Health Canada (2001), the population health approach refers to the health of the entire population. However, this framework could also be used as a guide in a study where the population of interest is not the general population but a group with specific characteristics such as geographical setting, ethnicity, religion, socio-economic status, age and gender.

Figure 2.1 illustrates the population health framework. It emphasizes the social and physical environments, health care interventions as well as the individual genetic and behavioural characteristics that may determine an individual's health, well-being and general prosperity. According to Evans and Stoddart (1994) this figure is an attempt to extend the population health framework and its biological, behavioural and environmental determinants. Two major changes are introduced. First, health is defined both by the disease status which is recognized by the health care system but also by the individual's response and general behaviour. Second, according to figure 2.1, it is the individual's response that defines their health and well-being and general prosperity.

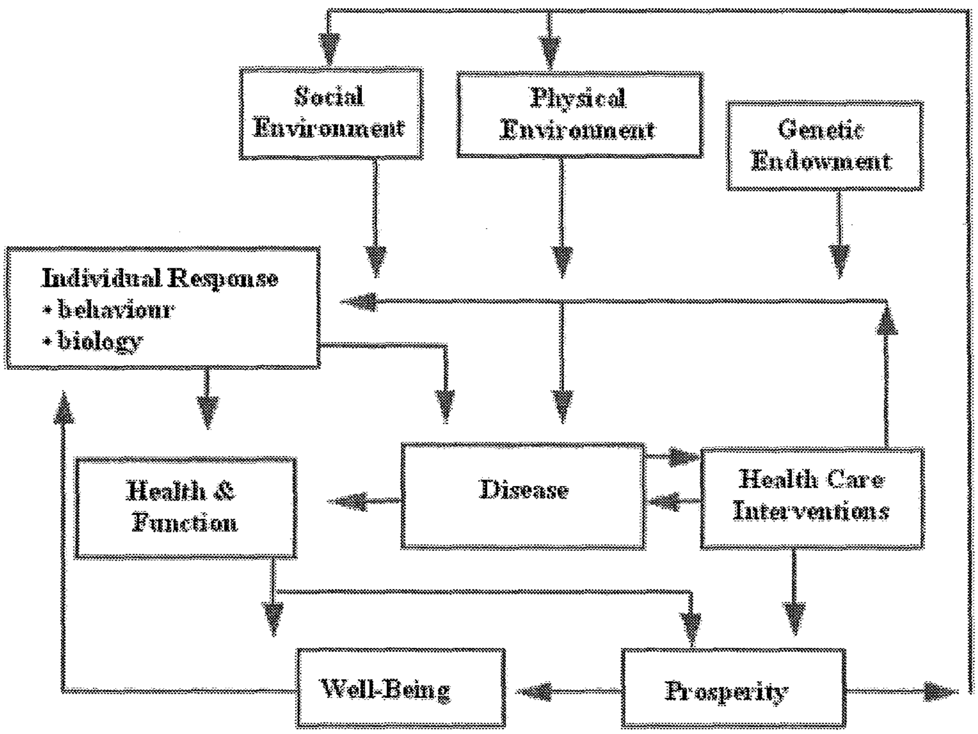

Figure 2.1: The population health framework (Health Canada)

The overarching goal of the Population Health Perspective is "to maintain and improve the health of the entire population and to reduce the inequalities in health between population groups" (Health Canada, 2001). It reflects the evolution related to the definition of health by recognizing that human biology, lifestyle, the social, economic and physical environment as well as the availability of health services may affect health.

The outcomes or benefits of a population health approach extend beyond improved health status by recognizing the complexity of the broader determinants of health and the interaction between them. According to this framework, a healthy population is more productive, increases national growth and development, requires less support in the form of health care and social benefits and is better able to support and sustain itself over the long term (Health Canada 2001). At the same time, the population health approach addresses health issues and focuses from prevention to health promotion, diagnosis, treatment care and protection, and integrates and balances action between them.

The population health perspective is not limited to Canada. It has also found its way as one of the components of the *Investment for health: a discussion of the role of the economic and social determinants* paper by the WHO European Office for Investment for Health and Development (2002). As well, the integration of a population health approach within the wider health care system is a key priority area for Australia and Great Britain's health policies (Health Canada, 2001).

Yet, the population health perspective has seen controversies in academic discourses and policy debates. For instance Zollner and Lessof (1998) pointed out that population health has been taken as justifying the diversion of resources from the health sector to support economic development or prioritizing those investments that impact health, such as socio-economic policies that address inequity. It has also been criticized for ignoring the broader social and political forces (for example, globalization, and class structure) that may warrant a comprehensive analysis of the determinants of health (Poland et al., 1998). Also, Hayes and Dunn (1998) indicated that findings from population health research addressing health inequalities have yet to be applied to health policy. Notwithstanding, the population health perspective has become a persuasive and pervasive framework for the Canadian and U.S. health and health promotion policies (Hamilton and Bhatti, 1996) as well as for research studies (Elliott et al., 1998; Eyles et al., 2001).

For this research, the population health framework acts as a guide for the investigation of the broad determinants of children's respiratory health with particular focus on the variability of respiratory health outcomes and their potential determinants over space. Specifically, this thesis is based on one of the major principles of the population health perspective, which recognizes that health development is influenced by the interactions between the genetic, physical, social, economic, and healthcare environments. Also, the role of space is incorporated by examining potential differences in air pollution exposure and health outcomes while controlling for a number of social, economic and lifestyle choice covariates for the city of Hamilton.

2.2.2 Determinants of Health

Population health is an approach that aims to improve the health of the entire population and to reduce health inequities among population groups. In order to reach these objectives, the broad range of factors and conditions that have a strong influence on health are taken into consideration (Health Canada, 2001). As noted before, it was the Lalonde report that introduced the Health Field Concept and consequently pointed out that in addition to health care, health status is linked to three other factors or "fields". These were identified as lifestyle, human biology and the environment. Since then, as knowledge is gained, the list with the determinants of health is undergoing constant clarification and integration. The following list of "influences" or determinants of health (Eyles et al., 2001) is derived from the population health model (Hamilton and Bhatti, 1996):

1) Income and social status
2) Social Support Networks
3) Education
4) Employment and Working Conditions
5) Physical environments
6) Biology and Genetic Endowments
7) Personal Health Practices and Coping Skills
8) Health Child Development
9) Health Services
10) Social Environments / Social cohesion (added by Evans et al., 1994)

The evidence as to the role of each of the determinants of health is rather complicated. However, brief observations regarding those factors that will concern this study later are offered below[3].

Income and Social Status

There is strong evidence regarding the relationship between socio-economic status and human health. For example, one Canadian study found that people in lower - income categories had a higher risk of death from non-accidental causes compared to those in the most favourable income categories (Finkelstein et al., 2003). Also, a British study found that for major[4] diseases, such as cancer and coronary heart disease, health increased with job rank (Marmot et al., 1987; in Health Canada, 1994). In fact, the social status and income factors seem to be among the most important influences for health. A wealth of evidence from a number of research studies across the world show that the size of the gap in social and economic status between groups within a population has a significant impact on the health status of a population (Wilkinson, 1996).

A very important aspect of this relationship is that it holds for different diseases, for men and women, and for people in different parts of the world. Therefore, this association needs to be interpreted with caution. Considerable

[3] The population of interest in this study are the children aged 8 - 12 years old. Therefore, the determinants employment and working conditions, personal health practices and coping skills as well as education level were based on their parent's status. Also, the determinants of social networks, health services and social environments/cohesion, were not taken into consideration due to data limitations and thus are not discussed in detail in this thesis.

[4] The term major here refers to the leading role of these diseases to worldwide mortality rates.

research had shown that the degree of control people have over life circumstances and their discretion to act are the key influences. Higher income and status generally result in more control and coping skills (Brown et al., 2005). There is still much to be learned but there is now considerable knowledge to improve population health by focusing on socio-economic status as well as the above mentioned risk factors.

Employment and working conditions

Unemployment is associated with poorer health. One Canadian study found the unemployed have significantly more psychological distress, anxiety, depressive symptoms, disability days, activity limitation, health problems, hospitalization and physician visits than the employed. Also, a major review from the World Health Organization found that high levels of unemployment and economic instability in a society cause significant mental health problems and adverse effects on the physical health of unemployed individuals, their families and their communities" (WHO, 2000a)

However, the effect of job rank on health shows the importance of status in the workplace, and that those with more control over their work circumstances are healthier. Health is also affected by stress related demands of the job such as the pace of work, the frequency of deadlines and reporting requirements. A recent study in Sweden' found that cardiovascular disease occurred most often among those with high job demands, low levels of control over their work, and low levels of social support at work.

Physical Environment

Population health is highly dependent on the physical environment in which people live. The air, water and soil conditions are among the most important elements of the physical environment.

Air pollution, including exposure to second hand tobacco smoke, is key aspect of the physical environment that has a significant association with health. Many studies have shown that both indoor (Williams et al., 2000, Monn, 2001, Kousa et al., 2001, Jones, 1999) and outdoor air pollution are significantly associated with health (Atkinson et al., 2001, Ayres, 2002, Boezen et al., 1999, Brunekreef and Holgate, 2002, Dockery et al., 1993, Filleul et al., 2003, Finkelstein et al., 2003, Harre et al., 1997, Hoek G. et al., 2002, Jalaludin et al., 2004, Katsouyanni, 2003, Koenig, 1999, Koop and Tole, in press, Krzyzanowski et al., 2002, Nafstad et al., 2003, Pope et al., 2002, Samet et al., 2000, Ward, 2004, Williams et al., 2000)

Pollutants of interest are mainly nitrogen dioxide, ozone, sulphur dioxide and particulate matter. However, much remains to be done to understand the complex associations between air pollution and human health. Cotton (1993) showed that the risks from small particles are greater compared to the risks from other pollutants such as ozone. Also, Schwartz and Neas (2000) demonstrated that fine particles ($<PM_{2.5}$) are more strongly associated with acute respiratory health effects in children compared to coarse particles ($PM_{2.5-10}$).

Biology and Genetic Endowment

The functioning of various body systems, the genetic endowment of an individual as well as the process of development and aging are among the most important determinants of health. Age and sex are considered to be the most important biologic attributes. Males and females at all ages and of all socio-economic strata have different life expectancies. The age of onset and the types of diseases are different among men and women. Vulnerability to significant health risks differs also with age. Children and the elderly are considered as the most vulnerable sub-populations to a number of environmental hazards, mainly because of their unique developmental and physiological characteristics (Filleul et al., 2003, Cohen Hubal et al., 2000). However, most of the foregoing differences are attributable only in part to biological sex and age. The main differences probably arise from differences in living conditions and roles in society. For example, Peters et al., (1999), studied chronic respiratory effects of four pollutants (ozone, particulate matter, acids and nitrogen dioxide) on schoolchildren. Rates of respiratory illness were higher for males and specifically those living in houses with pets, pests, mildew, and water damage, those whose parents had asthma, and those living in houses with smokers justifying in this way that health is affected by living and working environments as well as health care services and lifestyle behaviors.

Genetic endowment also provides an inherited predisposition to a wide range of individual responses that affect health status. Although socio-economic and environmental factors are important determinants of overall health, in some circumstances genetic endowment appears to predispose certain individuals to particular diseases or health problems.

Health Child Development

The effect of prenatal and early childhood experiences on health and well being in adulthood is very powerful (Health Canada, 1994). For example studies have shown that low birth weight is associated with problems throughout the lifespan (Dolk et al., 2000). In addition, many research studies have demonstrated that children exposed to environmental hazards such as tobacco smoke, air pollution from an industry in close proximity to the house and/or school are more likely to experience health problems during their lifetime (Braun-Fahrlaner et al., 1997; Hertzman et al., 1996; Hruba et al., 2001). Also, a limited number of studies have documented improved health measures following reduced exposure to air pollution (Heinrich et al, 2002; Kramer et al, 1999; Avol et al. 2001).

2.2.3 Environment and Health

As mentioned, the population health perspective advocates for the broad understanding of the determinants of health. It has provided a fresh way of looking at the multiplicity of factors affecting health. Among the many factors, environment has been recognised as one of the most important determinants of individual and population health in both framework building and empirical researches (Eyles, 1997).

Similar to health, the definition of environment has seen varying conceptualizations. Initially, it was defined as a synonym of nature, emphasizing its one-way relationship with society (Simons, 1993; Dickens, 1992: in Eyles, 1997). That is, human beings affect their physical environment, but their physical environment does not affect them. However, nowadays it is recognized that the relationship between environment and humans is multi-directional and that just as the environment (physical) affects humans, so too humans affect their environment. Furthermore, environment has broadened its context and is now seen as a synonym of ecosystem (Eyles, 1997), through which people interact with the environment, influenced by social and cultural forces (McMichael, 2001; page 17 – 22).

A number of papers have helped in shaping the relationship of environment and health within the population health perspective. Krieger (1994), in discussing "the web of causation" model proposed an eco-social framework as the alternative approach for the epidemiologic theory, according to which both social and biological determinants of health are taken into consideration. She also points out that an eco-social approach would challenge the definition of environment from "that which is exogenous to the organism[5]" to the social environment that is constructed by people and the natural environment that is the interplay of ecologies and global geologies and climatic forces that can be affected or even destroyed by people. Eyles (1999) recognizes also the importance of the environment in its broader context as a risk factor for health but also thinks that there is a necessity in integrating environment in

[5] According to the Dictionary of Epidemiology (1983)

the population health framework through an interdisciplinary approach. Specifically, to quote from Eyles (1999) "the investigative and analytic tools of toxicology, microbiology, epidemiology, environmental engineering and environmental psychology, among others, can help assess environmental hazards and the environmental burden of illness; those of ecology, economics, geography and environmental planning can help assess ecosystem conditions and human well-being and those of law, philosophy and the policy sciences can help assess environmental justice and the human condition". While Eyles and Krieger have highlighted the importance of the environment in determining health, Agius (2001) ventured to define environmental health as an integrated concept:

> *"those aspects of human health including quality of life that are determined by physical, biological, social and psychological factors in the environment. It also refers to the theory and practice of assessing correcting, controlling and preventing those factors in the environment that can potentially affect adversely the health of present and future generations".*

As mentioned in the first chapter, this thesis attempts to address the potential determinants of children's respiratory health and it is therefore guided by the population health perspective. The latter moves beyond the traditional definitions of health and opens the door to the consideration of broader determinants of health that include the social and physical environments as well as individual lifestyle behaviours. In this effort, it is important to set the stage for the role of geographers in environmental health research.

2.3 The Geographical Context

According to Elliott (1999) the geographer's role in health-related research is a result of a threefold evolution: the evolution of models and definitions of health; the evolution of medical to health geography; and the rise of the population health perspective.

At the outset, the idea that place and location may influence health is not exactly new. Meade et al., (1988) argued that the inclusion of geography in health research stems from ancient times, when Hippocrates stated that:

> *"Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year...then the qualities of waters...the mode in which the inhabitants* (of a city) *live, and what are their pursuits, whether they are fond of drinking and eating in excess...or are fond of exercise and labour... (Hippocrates, circa 400 B.C.)"*

Hippocrates was familiar with various aspects of the environment that may influence human health but it was not until the mid-fifties when the perspective and methodology of geography was officially applied to the study of health, disease and care (Meade et al., 1988; Johnston et al., 1994: page 374). Specifically, the systematic study of medical geography was concomitant with the disease ecology approach according to which environment was considered as the most important influence on health (Kearns and Moon, 2002). Furthermore in medical geography, "*subjects are viewed in holistic terms within a variety of cultural systems and a diverse biosphere*" (Meade et al, 1988).

The methods employed were mainly quantitative and based on the positivist stream. With regard to research, medical geography spanned the study of disease aetiology and diffusion (Fost, 1990; Dunn and Kingham, 1996; Beyea and Hatch, 1999; Dunn et al., 2001) and the study of health care organization and distribution across space (Peters, 1997; Walsh et al., 1997).

Later on, and as the evolution of more comprehensive frameworks and definitions for health emerged, medical geography "*required radical surgery*" (Mohan, 1989) and finally metamorphosed to "geography of health and health care" (Curtis and Taket 1996: page 13). The contemporary "geography of health and health care" reflects the new notions of health, as well as the importance and complexity of the various determinants of health. In this way, the population health perspective is also incorporated influencing as well the role of geographers in health research. With the development of alternative health frameworks that include socio-economic, environmental, lifestyle and biological influences on health (Evans et al., 1994), geographers are now called to address the significance of those influences and this is one of the major objectives of this study. In doing that, Geographical Information Systems (GIS) and spatial analysis can be used as they have recently emerged as an innovative and important component in many health studies (Scoggins et al., 2003; Jerrett et al., 2003; Cakmak et al., 2003(Cockings et al., 2004)

2.4 Spatial Analysis, GIS and Health

Bailey and Gatrell (1995) defined spatial analysis as "*the description and/or explanation of a process operating over space based on a sample of observations and its potential relationship to other spatial phenomena*". The application field and objectives of spatial analysis concern various issues such as transportation, environment and natural resources management, facilities and service planning (for example, education, police). The incorporation of spatial analysis in health-related issues started more than a century ago, when scientists started to explore the potential associations between location, environment and disease. A good example is Dr. John Snow's study about the cholera incidences in London in 1854. Dr. Snow drew maps showing the homes of people who had died of cholera and the locations of water pumps. He found that the density of incidences was increasing around a water pump on Broad Street. Dr. Snow's re search, which is also considered as one of the first studies in medical geography, illustrated that spatial association between disease and environment can illuminate environmental health processes (Figure 2.1).

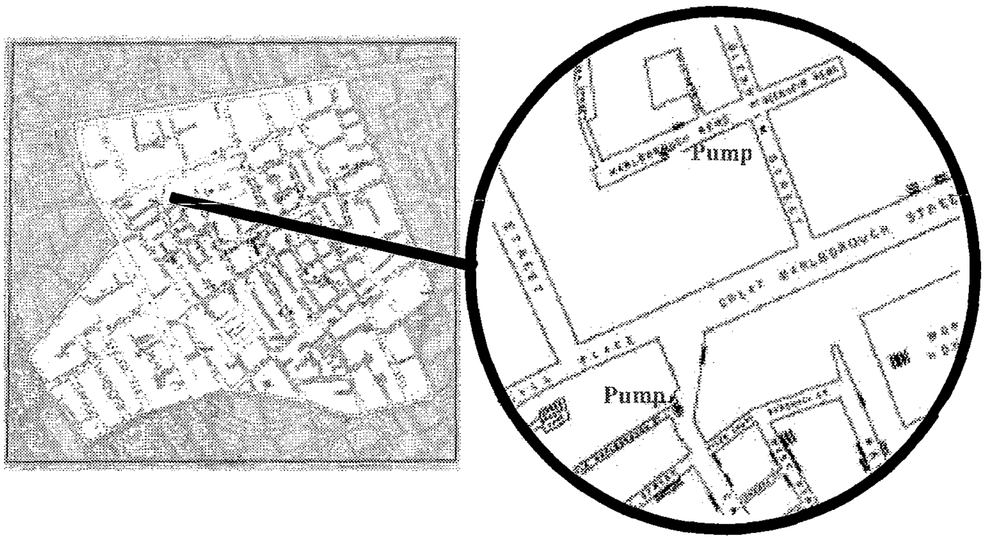

Figure 2.2: John Snow's cholera map, 1854 (Cliff and Haggett, 1988)

Nowadays, the combination of spatial analysis, GIS and epidemiology are powerful tools for improving our understanding of human health (Dunn et al., 2001). Specifically, epidemiology is concerned with the variation of a disease from time to time or from place to place (Birch et al., 2005). Spatial analysis facilitates the identification of high-risk disease areas and populations and the suggestion of potential influencing factors of a disease (Beggs et al., 1999; Jerrett et al., 2003) and the optimal allocation of health care facilities (Ali et all., 2001. Finally, the introduction of GIS expanded the implementation of spatial analysis in health research. GIS are seen as computer - based systems for collecting, editing, integrating, analyzing and displaying of spatially referenced data. They provide a sufficient framework for the analysis of health data by improving our understanding about the aetiology of current health problems as well as the potential relationships with other environmental, socio-economic and demographic covariates (Albert et al., 2000).

Many researchers have highlighted the importance and usefulness of spatial analysis and GIS within a health context (Gatrell, 2002; Elliott et al., 2000; Gatrell and Löytönen, 1998; Vine et al., 1997; Mayer, 1983). However, there is a gap between the potential use of those in health-related research and their actual use. For example, with regard to the application of spatial analysis and GIS to respiratory health, the literature was sparse until recently. Morris and Munasinghe (1994) studied the spatial variation of hospitalizations among the elderly in the United States. They applied spatial analysis to demonstrate that admission rates among the elderly showed marked geographic variation and are associated with regional indicators of socio-economic status (education level, income, household crowding), medical resources (hospital beds per capita) and smoking. Recently, the number of studies that focus on the potential association between air pollution and human health is increasing radically. For example, Jerrett et al., 2003 using spatial analytic techniques, studied the association of air pollution and mortality taking into account a number of environmental, socio-economic, demographic and lifestyle confounders that are representing potential risk factors for health. However, based on the results of the study, only sulphate and sulphur dioxide showed a significant association with mortality. While the analysis had made progress in improving our understanding of the influence of the above mentioned covariates as well as the usefulness of spatial analysis on analysing the air pollution – mortality relationship yet much methodological and conceptual work remains to be done in understanding the contextual spatial and temporal dimensions for the air pollution – health relationship.

In addition, a number of studies emphasize the application of spatial analysis and GIS in order to better estimate population's/individual's exposure to environmental contaminants (air pollutants). English et al., (1999) found that higher traffic flows may be related to an increase in repeated medical visits for asthmatic children in low income populations in San Diego County, California. Specifically, they applied GIS for the computation of traffic counts within a specific buffer area. Also, Kanaroglou et al., (2005) applied a location - allocation approach to establish an air pollution-monitoring network in order to assess intra-urban air pollution population exposure. Particularly, they locate the monitors based on land-use type, transportation infrastructure and the distribution of at-risk populations. In a following study, Jerrett et al., (2003) assessed exposure to air pollution by creating an air pollution surface for Toronto, ON. To create the surface, a land-use regression model based on the surrounding land-use, traffic volume and physical environment characteristics such as wind direction and topography was applied.

To assess children's exposure to particulate air pollution as well as investigate the potential risk factors determining their respiratory health, state-of-the-art spatial analysis methods were employed.

2.5 Exposure Assessment

According to Ott (1982) (taken from: Monn, 2001) exposure can be defined as "an event that occurs when a person comes in contact[6] with the pollutant". In this study, we are interested in the exposure of children to air pollution; therefore we will focus on the assessment of air pollution exposure.

As has been noted above, one of the main issues in health research is the accuracy of assessing air pollution exposure. This assessment is important for two reasons. First, several studies have uncovered significant positive health effects from exposure to environmental contaminants (Chen et al., 2004; Scoggins et al., 2003; Kryzanowski et al., 2002; Williams et al., 2000; Walker et al., 1999; Stern et al., 1994). Second, over the past ten years, advances in GIS and associated statistical techniques have expanded into the field of exposure analysis (Jerrett et al., 2004). Therefore, research on the health effects of

[6] This definition refers to a contact with a pollutant, which means that the person inhales, ingests or comes into skin contact with the pollutant for a specific period of time.

exposure to urban air pollution may improve our understanding for the outdoor and indoor air processes.

Human exposure to urban air pollution can be measured directly or indirectly. With respect to the former, exposure levels are determined on an individual basis, using personal monitors and/or a biological indicator (lead level in blood). On the other hand, in the indirect approach, exposure levels are either estimated by monitoring stations or determined by models (Monn, 2001). The majority of the studies associate air pollution and health using indirect techniques. The average value of the pollutant of interest is then assigned to the entire population, assuming that exposure is representative for the population of interest. A potential drawback of this approach, however, results from the variability of concentrations of pollutants over space and time. Few studies have addressed the question whether spatial and temporal variations in outdoor air pollution concentrations are reflected in similar variations in personal exposures (Brunekreef and Holgate, 2001). Potential spatial and temporal changes in air quality might pose particular difficulties in developing hypotheses about causal links between air pollution and an individual's health (Dunn and Kingham 1996). Therefore, the validity of exposure estimates based on ambient air monitoring should be thoroughly evaluated in each epidemiological study.

Another important issue in exposure assessment is related to the validity of outdoor air pollution exposure estimates on the basis that people spend more than eighty percent of their time indoors and most of that time at home (Brunekreef and Holgate, 2001). Additionally, indoor pollutant concentrations may differ from outdoor concentrations because of ventilation patterns and indoor sources.

Also, assessing exposure to air pollutants remains by definition extremely difficult due to the fact that pollutants are present as mixtures. Therefore, assessment of exposure has to rely on measurements of indicators of these mixtures such as nitrogen dioxide (NO_2), ozone (O_3), sulphur dioxide (SO_2) and particulate matter (Monn, 2001).

Finally, personal exposure measurements clearly reflect the actual individual exposure. However, one of the major disadvantages of using measurements of personal exposure is, that they are expensive, time consuming and difficult to apply to large populations. However, in the last few years smaller and less expensive devices have become available making personal monitoring more approachable (Gulliver and Briggs, 2004).

2.6 Environment and Respiratory Health

In most of the studies investigating the relationship between air pollution and human health, it is preferred to use health outcomes in order to evaluate such relationships, because of the difficulty in measuring health performance. For instance, in the case of respiratory health, most studies have measured hospitalization and mortality rates to explore the potential association between air pollution and respiratory health (Maheswaran et al., 2004; Finkelstein et al., 2003; Pope et al., 2002; Hoek et al., 2002; Erdas et al., 2000; Wilkinson et al., 1999; Schwartz, 1995). However, in this thesis, children's respiratory health is explored using both health performance and outcomes approaches, by studying the potential association of disease incidences as well as lung function capabilities with a number of risk factors. Yet, before, exploring the existing literature on the relationship between air pollution and respiratory health, it would be more appropriate to justify the selection of children as the particular population of interest.

2.6.1 Children as a vulnerable population

In the population health perspective the term population may refer to any group of individuals defined according to any number of general characteristics such as geographical location, ethnicity, religion, socio-economic status, age and gender. Children are among the most vulnerable sub-populations to environmental threats because of their unique developmental, physiological and behavioural characteristics. Particularly, children have more lung surface area compared to their body weight and inhale more air per body weight. Also, their immune system and their organs are immature, which increases their susceptibility to the harmful effects of a number of chemicals. Their breathing zone is lower than adults so they are more exposed to vehicle exhausts and heavier pollutants that concentrate at lower levels in the air (Canadian Association of Physicians for the Environment, 2000: see prints titled *Children Are At Greater Risk*). In addition, children are more active than adults and therefore breathe more rapidly. Finally, children often breathe through their mouth, bypassing the filtering effect of the nose and allowing more hazardous substances to be inhaled (Kleinman, 2000; Flynn et al., 2000).

2.6.2 Air Pollution and Children's Respiratory Health

There is a substantive body of evidence linking air pollution to adverse health effects in children. Based on this, assessing the relationship between air pollution and health depends largely on the specific disease or symptom in question. Health effects associated with both outdoor and indoor air pollution include increased hospital and emergency admissions due to acute and chronic respiratory illnesses (bronchitis and pneumonia), aggravation of asthma and chronic obstructive pulmonary diseases (Wilkinson et al., 1999; Pikhart et al., 2001; English et al., 1999; Hruba et al., 2001); decreased pulmonary function (Gauderman et al., 2004; Ward et al., 2002; Jalaludin et al., 2004; Moseler et al., 1994); and increased respiratory symptoms such as coughing and wheezing (Howel et al., 2001; Peters et al., 1999), as well as increased medical consultations (Jalaludin et al., 2004).

As far as it concerns the pulmonary function tests[7] applied to assess the general functioning and development of the lung, one of the most common is the Forced Vital Capacity (FVC). According to Cherniack (1977) and Miller (1982), FVC is the maximum volume of air, expired following a maximal inspiration and using a maximal effort to expel the air. FVC is largely a function

[7] These measures of pulmonary function are not the only ones that could be used. However, the ones that will concern this study are described in more detail later in this chapter.

of the number of alveoli, with differences in volume primarily attributable to differences in the number of alveoli, since their size remains relatively constant. In essence, deficits in the growth of FVC reflect a reduced number of alveoli in the lungs. This might be a plausible mechanism of the effect of air pollution on lung development. Another potential mechanism of the effect of air pollution on lung development is airway obstruction such as occurs in chronic obstructive pulmonary disease (COPD). One of the tests used to reflect both lung size and airways obstruction is the Forced Expiratory Volume in one second (FEV1), which is the volume of air, expired in one second and is used to reflect lung size and large airway obstruction. In addition, the maximum rate of expiration during FVC, otherwise known as Peak Expiratory Flow (PEF), has been used to reflect both lung size and airway obstruction. Finally, the Mean Forced Expiratory Flow between 25% and 75% ($MFEF_{25-75}$) of FVC and 75% and 85% ($MFEF_{75-85}$) of FVC as well as the Maximum Expiratory Flow after 50% (MEF_{50}) or 75% (MEF_{75}) of FVC has been applied to reflect only airway obstruction.

All these tests were used for a number of studies to demonstrate potential effects of air pollution on children's lung development. For instance, Gauderman et al., (2000, 2002, 2004) analyzed the FVC and FEV1 to measure the pulmonary function of children and found a significant negative association indicating that increased lung development is associated with lower levels of air pollution. Also, Stern et al., (1994) used the FVC, FEV1 and $MFEF_{25-75}$ to demonstrate that children living in Ontario (higher levels of air pollution than Saskatchewan) had significant mean decrements of 1.7% in FVC compared with children living in Saskatchewan. Finally, Hoek and Brunekreef found a weak negative association between concentrations of air pollution and pulmonary function tests of FVC, FEV1 and PEF. On the other hand, there are studies that did not find any significant association between air pollution exposure and limited pulmonary development. To be more specific, Moseler et al., (1994) were not able to provide any evidence on reduced lung functioning due to higher levels of air pollution in children without any asthmatic symptoms. The tests used were FVC, FEV1, MEF_{50} and MEF_{75}.

In general, the relationship between air pollution and children's respiratory health has been examined both in short-term studies, which relate day-to-day variations in air pollution and health (Howel et al., 2001; Ward et al., 2002; Jalaludin et al., 2004), as well as long-term studies (Peters et al., 1999; Hruba et al., 2001; Pikhart et al., 2001; Gauderman et al., 2004) that follow cohorts of exposed individuals over a period of time. The temporal dimension is of particular importance because of the nature of certain diseases. It is well known that certain diseases are chronic in nature and require a latency period before disease onset. For example, children exposed to tobacco smoke may develop cancer 10 to 60 years later (Hertzman and Wiens, 1996).

A plethora of research has demonstrated mostly positive, significant associations between short-term exposures to various pollutants and health outcomes. For example, Howel et al., 2001 studied the relationship between a number of respiratory health outcomes and symptoms (cough, wheeze, and asthma) of 2442 children aged 1 - 11 years and daily particulate levels in ten communities of Northern England. Five of them were in close proximity to operational opencast mining sites while the other five were located further away.

Based on the results of the study, there were positive associations between particulate air pollution and respiratory symptoms. However, it was not possible to make any inferences on the associations between PM_{10} and children's respiratory health because the strength of associations between all respiratory health symptoms that were taken into consideration and PM_{10} were similar between the near and further away from the coal mining sites.

Jalajudin et al., (2002) also studied the acute effects of air pollution (PM_{10}, ozone and nitrogen dioxide) on respiratory symptoms (wheeze, wet and dry cough), asthma attacks and doctor visits in 125 children in Sydney, Australia. The results of this study showed an association between PM_{10} and doctors visits for asthma and also between NO_2 and prevalence of wet cough. They were, however, unable to show that current levels of air pollution in Sydney have an adverse effect on children with a history of wheezing.

In addition, according to a systematic review of Ward and Ayres (2003), children with diagnosed asthma or persistent respiratory symptoms appeared less affected by PM_{10} levels than those without. Also, the majority of studies indicated that $PM_{2.5}$ are more strongly associated with respiratory health effects in children compared to PM_{10}.

The evidences of the long-term effects of air pollution on morbidity are also complicated (Maynard, 2000). Peters et al., (1999) in order to study chronic respiratory health effects of air pollutants initiated a 10-year prospective cohort of approximately 3,600 children in Southern California, of 9 to 16 years of age. The results of the study concerning outdoor air pollution suggested no consistent or large excesses of morbidity (asthma, bronchitis, wheeze) of children living in the most polluted communities and /or had the highest estimated exposures. On the other hand, Gauderman et al., (2004) studied the effect of air pollution on lung development of 1759 children cohort, 10 to 18 years of age, in twelve Southern California communities. Over an eight-year period, deficits in growth of lung function and specifically in Forced Expiratory Volume in one second (FEV1) were associated with exposure to PM_{10}, and elemental carbon. The results of this study were strengthened also from a previous effort of Gauderman et al., (2002), who investigated the association between air pollution and lung function growth in Southern California children cohort during a four-year period. They had also verified the association between particulate air pollution and growth of lung function.

Hruba et al., (2001) employed a study to investigate the long-term effects of air pollution (TSP) on respiratory symptoms (wheeze, cough) and hospitalizations (for asthma, bronchitis and pneumonia) of 667 children 7-11 years of age. Logistic regression modeling indicated a significant increase in hospital admissions for asthma, bronchitis and pneumonia associated with 15 μg/m3 increased in estimated TSP exposure after controlling for socio-economic differences, health care availability and other identified confounders.

Beside the selection of the health outcome(s), the selection and measurement of air pollutants is also critical in assessing the potential association between air pollution and respiratory health. In most cases, pollutants of interest include ozone, particulate matter, sulphur dioxide, nitrogen dioxide and carbon monoxide. Among those, the most significant effects of air pollution have been associated with particulate matter contaminants. Particulate air pollutants are usually classified by size. The major categories are those greater than 10 microns (Total Suspended Particulate - TSP), those smaller than 10 microns (PM_{10}) and finally those smaller than 2.5 microns ($PM_{2.5}$). Sources of particulate matter are mainly vehicles, stationary combustion sources (houses burning coal, cigarette smoking, biomass for cooking and heating, industrial plants, incinerators and fossil-fuel power plants), non-combustion sources (mining, cement plants) and finally natural sources (sea spray, forests, volcano emissions) (WHO, 2004).

2.7 Summary

This chapter sets the research and geographic contexts for an investigation of the potential determinants of children's respiratory health focusing more on the impacts of elevated particulate air pollution. A review of literature revealed the increasing recognition of the broader determinants of health that are codified in the population health perspective. Specifically, changing definitions of health were described, from the dominance of the biological aspects of human beings to a more holistic version that includes the social, cultural, economic and environmental influences providing at the end the foundation for a population health approach that recognizes the multiple factors and conditions that contribute to health. Following this, the population health framework, used to inform this research is described.

In the next part, the role of geographers is set in the context of the population health perspective. Specifically the metamorphosis of medical geography to the geography of health and health-care incorporates the population health framework because it reflects the new notions of health, as well as the importance and complexity of the various determinants of health. With the development of the geography of health and health care geographers are now called to address the significance of those influences through the incorporation of more inclusive methodologies and alternative approaches.

Having that in mind, the following two sections of this chapter addressed the application of spatial data analysis and GIS in health – related studies. Nowadays, the combination of spatial analysis, GIS and epidemiology are powerful tools for assessing exposure to environmental hazards and to improve our understanding on human health as well as the potential relationships with other environmental, socio-economic and demographic covariates. However, there is a gap between the potential use of those in health-related research and their actual use. For example, with regard to the application of spatial analysis and GIS to respiratory health, the literature was sparse until recently.

In the final section, research conducted on the potential determinants of respiratory health with particular emphasis on the associations between environment and children's respiratory health is reviewed. The review of the literature uncovered a mix of significant and non-significant associations especially regarding children's pulmonary development. The exact impact of air pollution on children's pulmonary function is still up for debate. That is, some studies point to an association between air pollution and children's respiratory health while others do not. In general, however most authors agree that air pollution influences children's respiratory health although the degree of significance and the long-term effects remain debatable.

Methodologically, the literature focuses on classical epidemiological approaches failing to embrace a geographical perspective. Only recently, spatial analysis and GIS are approaches have been employed in health-related research to account for the variability of children's respiratory health and their potential determinants over space. Still such analytic techniques have not been applied to assess children's health.

The remaining of this thesis describes an attempt to study the potential determinants of children's respiratory health using advanced spatial data analysis and GIS methods. The following chapter outlines the methods utilized to address the thesis objectives.

CHAPTER THREE

RESEARCH DESIGN AND METHODS

3.1 Introduction

This thesis builds on a previous research study, the Hamilton Children's Cohort (HCC) conducted in five time periods over 6 years (1979 – 1985) in Hamilton, Ontario. The HCC study provided important evidence linking air pollution with poor respiratory health of children. However, given the limited application of spatial analysis and GIS during the study period, the spatial dimension of health and air pollution data was not fully taken into account, which may either strengthen the study results or may even lead to new conclusions. This research, therefore, re-evaluates the study in an effort of developing a clearer conceptual understanding of the potential associations between air pollution exposure and children's respiratory health.

This chapter describes the methods utilized in this thesis in order to address the following research objectives:

1. Investigate the spatial variation of adverse respiratory health of children in Hamilton
2. Investigate the relative importance of individual, social and environmental characteristics that may determine the burden of respiratory incidences by re-assessing exposure to air pollution

using advanced methods of spatial analysis and geographical information systems (GIS).

The next section provides a detailed description of the study area emphasizing its industrial history and the point sources of air pollution. This is followed by a section discussing the data used paying particular attention to the clinical information and the questionnaire database. In the final section of this chapter, the statistical analysis is described in detail. Specifically, this last section is sub-divided into the exploratory data analysis and the regression modeling techniques.

3.2 The City of Hamilton

The focus area of this study is the City of Hamilton, Ontario. The City of Hamilton[1] is located on the west shore of Lake Ontario approximately 100 km southwest from the city of Toronto (Figure 3.1). Hamilton became a city in 1846 with a population of 6,832 (City of Hamilton, 2005). During the study period (1979 – 1985), the city of Hamilton was part of the Regional Municipality of Hamilton-Wentworth. The population of the municipality in 1976 was 312,003, while in 1981 it had decreased slightly to 306,434 (Statistics Canada, 1986).

The dominant geographic feature is the Niagara escarpment that separates the city into the lower coastal and the mountain area. The climate is

[1] The term City of Hamilton is used throughout this study to describe the city before amalgamation in 2001. The maps that follow display the boundaries of Hamilton's boundaries in 1980s.

generally continental, influenced by its proximity to the Great Lakes. Easterly winds off the open waters of Lake Ontario may add substantially to local snowfall however the prevailing winds are from the southwest during spring and summer but during autumn and winter the prevailing winds are from the northeast (Pengelly et al, 1987).

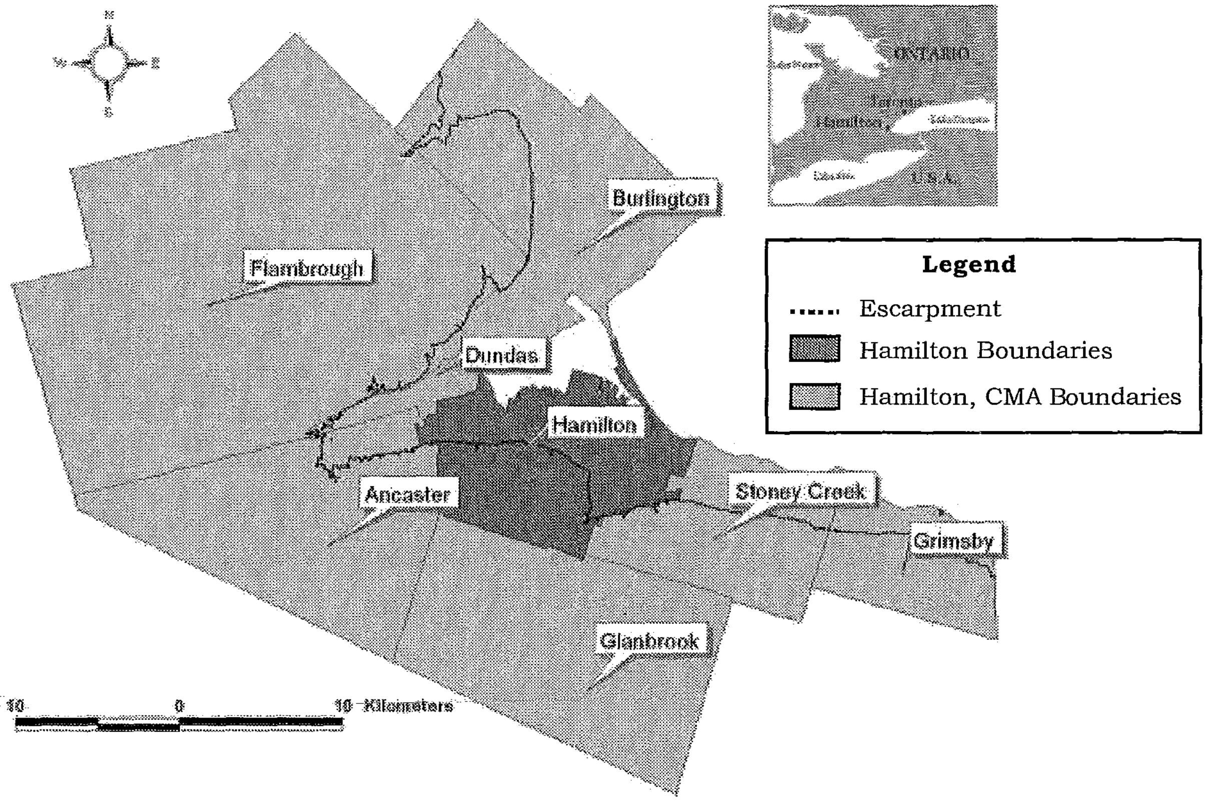

Figure 3.1: The City of Hamilton

Hamilton provides a useful case to study the potential associations between air pollution and children's respiratory health. Children are a particularly vulnerable part of the general population because of their unique physiological and behavioural characteristics. Therefore, studying the effects of particulate air pollution and other covariates on children's respiratory health is important in order to assess the potential health risks for the younger generations. To be more specific, Hamilton was considered to be one of the most polluted industrial cities worldwide during the study period (1979 – 1985). The presence of the largest two steel mills in Canada resulted into a substantial particulate pollution problem that was first documented in the mid-1950s (Pengelly et al., 1986). The industrial zone created higher levels of air pollution in the northeast and comparatively lower levels in the central to western parts of the city. Although the association between air pollution and respiratory health has been studied for Hamilton (Burnett et al., 1997), the spatial context of respiratory disease incidences and air pollution concentrations has not yet been considered as an important covariate.

In addition, the geography of Hamilton has played an important role in shaping its general development. During the 18th century, Hamilton was a booming city below the mountain; however as we drew closer to the 19th century, it began to expand towards the escarpment. This resulted in a special urban form with distinct spatial patterns of socio-economic status within the city, such that residents with higher socio-economic status were living in the south and west ends, while those with lower status tend to be located along the harbour and industrial areas in the north and northeast of the city.

Finally, the availability of the data from the HCC study along with the recent advancements in technology –such as GIS and spatial statistics- support the choice of the study area in the hope of addressing the potential role of geography in studying air pollution and health related issues.

3.2.1 Hamilton as the Steel Town

During the 1950s Hamilton was generally known as the "Steel City" due to its long history of industrial activity. Growth of the steel industry began with the establishment of two major steel companies in Canada as early as the 1900s. The heavy industrial core is located in the northeast part of the city.

Many factors contributed significantly to the development of heavy industry in the area, including the building of railways, the population explosion (the population of Hamilton increased by more than 53 percent from 1906 to 1911), low cost land, and the proximity to a source of steel located at the northeast part of the city (Dear et al., 1987; pages 119 – 124, 213). The Steel Company of Canada (Stelco) was formed in 1910 while in 1912, the Dominion Steel Castings Limited steel foundry was established, which later formed the Dominion Foundries and Steel Limited (Dofasco) (Figure 3.2).

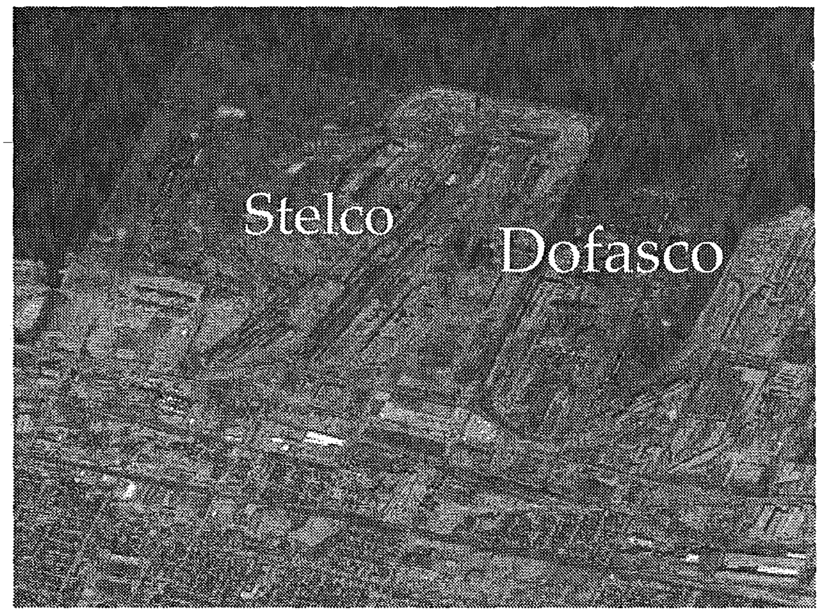

Figure 3.2: Hamilton Industrial Area (Source: Google Maps, 2005 / Imagery Digital Globe, Earth Sat)

During the First World War, all industries with war contracts declared high returns of profit; Stelco was confirmed as Hamilton's industrial giant. In 1914 the Dominion Steel Castings Limited steel foundry employed a total of 190 workers. By 1918, Dofasco had a work force of 2,283 (Dear et al., 1987). Following the war, industrial growth continued in the area. This was partially due to the availability of cheap electrical power in the city[2].

However, after the 1920's Hamilton's progress was brought to an abrupt halt by the onslaught of the Great Depression in 1929. During that time, many small industries failed and even Dofasco wavered on the edge of bankruptcy (Dear et al., 1987:130). Hamilton suffered severely during the Depression period due to the industry's dependence on national and international markets.

As the Second World War began, industrial production increased again. Following the war, the economy experienced only mild fluctuations. In 1941, 74 percent of all employed persons in the city were in the manufacturing sector and more than half of them in metalworking factories (Dear et al., 1987). Both steel companies expanded tremendously with new product lines, additional mills, and furnaces employing state-of-the-art technology.

The growth of Hamilton during the post-war era was the infilling of vacant-open space for residential and commercial land. For example, the first commercial mall, the Greater Hamilton Shopping Center was built in 1955.By 1960, much of the housing in the downtown area was razed and in general the downtown core was deteriorating (Evans, 1970).

---

[2] A group of businessmen in Hamilton were generating power for Hamilton 54km away from the city form the surplus water from Lake Erie (Dear et al., 1987 page 123)

In the following years, Hamilton remained the largest steel – producing centre in Canada. However, its heavy industry had entered into an economic downturn, common to most steel cities in the developed world. New industries had started to operate all over the world increasing the competition not only because of the escalating number but also because of their advanced technological infrastructure. If Hamilton's industries were to be competitive they had to take a larger slice of the international market share. However, the stagnant demand for steel products did not allow for an investment towards an improved infrastructure.

At the same time, health – related concerns started to emerge. Air pollution, mainly from the heavy industry and house heating[3] constituted one of the major problems in the city. As an emergency measure, the city decided to install special pollution filters in all heavy industrial facilities (Dear et al., 1987). But that was just the beginning. Up to the early 1970s, health issues related to air pollution were not taken into consideration, mainly because of the economic influence of steel industry for the city. Stelco and Dofasco were the major employers in the city, putting bread at the tables for many families. The HCC study was the first one that investigated the potential association between air pollution and children's respiratory health. However, this study was not able to provide sufficient evidence to infer a causal relationship. Since then, the majority of the studies (with few exceptions including Morris et al., 1994; Jerrett et al., 2003; Cakmak et al., 2003) have adopted a classical epidemiological stance, taking no account of the variability of health outcomes

[3] 95 percent of all homes were heated using coal or coke (Dear et al., 1987)

and their potential determinants over space. The studies that have applied these techniques have shown the importance of spatial variation in evaluating the association between air pollution and human health (Jerrett et al., 2003; references). Thus, through this thesis we try by re-analysing the HCC study to demonstrate the usefulness of spatial analysis and GIS in studying the potential associations between air pollution exposure and children's respiratory health in the ultimate hope to develop a clearer conceptual understanding of this relationship.

3.3 Sources of Data

This section describes the data available for the purposes of this thesis. The following paragraphs describe the spatial information collected from a number of sources as well as the respiratory health and air pollution data available through the HCC study (Pengelly et al., 1984, 1986, 1987; Kerigan et al., 1986). Based on data availability the base year of analysis was 1980. Specifically, respiratory health data from the questionnaires were based on the answers given during the second year of analysis, 1980. Also, the pollution data were available only from 1979 to 1981. Therefore, for the multivariate analysis data only from 1980 were taken into consideration.

3.3.1 Spatial Information

Spatial data were obtained mainly from three different sources. Digital geographic data such as boundary files and detailed road networks were extracted from the spatial databases of Desktop Mapping Technologies Incorporation (DMTI). Additionally, land-use information for 1980 was attained by the Hayman and Whetstone study (2000). However, land use was dassified under a different scheme compared to the one followed by Hayman and Whetstone (2000). Specifically, we aggregated into five land use types: Industrial, Commercial, Residential, Institutional, and Open Space. The projection of the data was transformed from Geographical Coordinate System (degrees) to the Universal Trans Mercator North American Datum (meters).

Children's home locations were geocoded based on the postal code information from the Canadian Census Analyser of the Computing in the Humanities and Social Sciences from the University of Toronto (Canadian Census Analyser, 1996).

3.3.2 Sample Size

As mentioned before, this thesis is based on the Hamilton Children's Cohort study (Pengelly et al., 1984, 1986, 1987; Kerigan et al., 1986). The HCC study recruited 3,505 children aged 7-10 by selecting randomly 30 out of 80 elementary schools operated by the Board of Education of the City of Hamilton, in September 1978 (Kerigan et al., 1986). At the end of the first year (1979), 3131 children participated in all the stages of this research study, which included a number of clinical tests and an interview of the parents/guardians of the child. However, during the second year of the study (1980), it was recognized that the industrial core area of the city, with the poorest air quality, was not well represented. Therefore, three schools were added from that area increasing the sample size from 3131 at the end of first year to 3439 children (Table 3.1). Also, during the second year, 474 children dropped out of the study either during the interview stage or during the testing stage.

Table 3.1: Cohort size during the first two years of study

Children	Number of Children at the start point of each year	Additions	Drops	Final
1979	3505	0	374	**3131**
1980	3131	782	474	**3439**

However, the data, available, included those children that participated in all five periods of study. In order to geocode the data, only the children that could also be identified by a unique postal code, based on the location of their home were taken into consideration. The children that fulfilled these criteria were 960. Yet, in order to have a good representation of the industrial core area, the children that were included during the second year of the study (782) but participated in the third year as well and could be identified by a unique postal code based on the location of their home (204 out of 782) were added into the sample. Based on that, the final number of children included in this study is n= 960 + 204 = 1164 children.

3.3.3 Respiratory Health – Questionnaire Database

The HCC data set contains individual level information regarding children's respiratory health and general medical background, as well as the family's medical/respiratory profile, quality of the dwelling (type of dwelling, number of rooms, and type of heating/cooking), socio-economic status and smoking habits (Kerigan et al., 1986). To collect this information regarding respiratory health of the eligible children, their parents were interviewed covering several aspects of their child and their household's profile. Within four weeks of the completion of the interview, pulmonary function testing was performed to those children whose parents have consented. The pulmonary function tests were performed at the children's school throughout the school year.

With regard to the pulmonary functioning tests, the focus of this study is to investigate the adverse effects of air pollution on pulmonary function (expressed by the forced expired manoeuvres) and development (expressed by the Spirometry tests) of lungs of the children. Therefore, Forced expired manoeuvres, such as Forced Expiratory Volume in one second (FEV1), Forced Vital Capacity (FVC), and Mid-Expiratory Flow (MEF) as well as the Spirometry test of Vital Capacity (VC) were taken into consideration.

3.3.4 Air Pollution Data

During the HCC study, a pollution-sampling network was established to effectively measure Total Suspended Particulate matter (TSP) at 23 sites in 1979 and at 28 sites in 1980 and 1981. The TSP sites were distributed throughout the city. The network was operated every 6 days. The five or six values from each site were used to estimate a monthly mean value for each site. In addition, there were 7-9 hi-vol samplers with Andersen 4-stage cascade impactors for the measurement of Particle Size Distribution (PSD) and 16 sites for sulphur dioxide (SO_2) monitoring. However due to limited number of monitoring stations for the PSD and SO_2, we only used the TSP data to assign children's exposure to air pollution.

According to the HCC study, Hamilton exhibited varied levels of air pollution. The highest levels were detected in the north part of the city. Specifically, the northwest part of Hamilton exhibited higher pollution levels compared to northeast Hamilton (Pengelly et al., 1984). This is probably associated with the location of the industrial core, the lake-breeze inversion, and the southwest prevailing winds.

3.4 Types of Variables

Before proceeding to the methods utilized in this thesis, the different types of variables used in the analysis, need to be defined. First 1, all variables can be grouped into categorical and continuous variables. Categorical variables are measures of differences in type rather than amount. Examples include race, gender, or color. These are also called qualitative variables because there is some quality that distinguishes these objects. A continuous variable is a variable that can take any value on the scale used to measure it. Any division on any unit on the scale produces a valid possible measure. Examples include things like height or weight. Statisticians often refer to the "levels of measurement" of a variable to distinguish between variables that have different properties. There are four basic levels: nominal, ordinal, interval, and ratio.

In this thesis, we used variables that were measured in all different levels of measurement. A summary of the variables is presented in table A1-1 (attached at the Appendix I).

3.5 Analytic Methods

This section describes the methods employed in this thesis. In general, the analytical techniques involved in this research, aim to the description of a process operating in space. Figure 3.3, demonstrates the steps followed and the methods employed in order to address the study's objectives. Based on Figure 3.3 *visualization* techniques were applied to reveal spatial patterns within the data set. Then, *exploratory* methods were utilized summarizing and investigating the spatial patterns and relationships while *modeling techniques* were used to first assess children's exposure to particulate air pollution and then investigate the potential determinants of children's respiratory health.

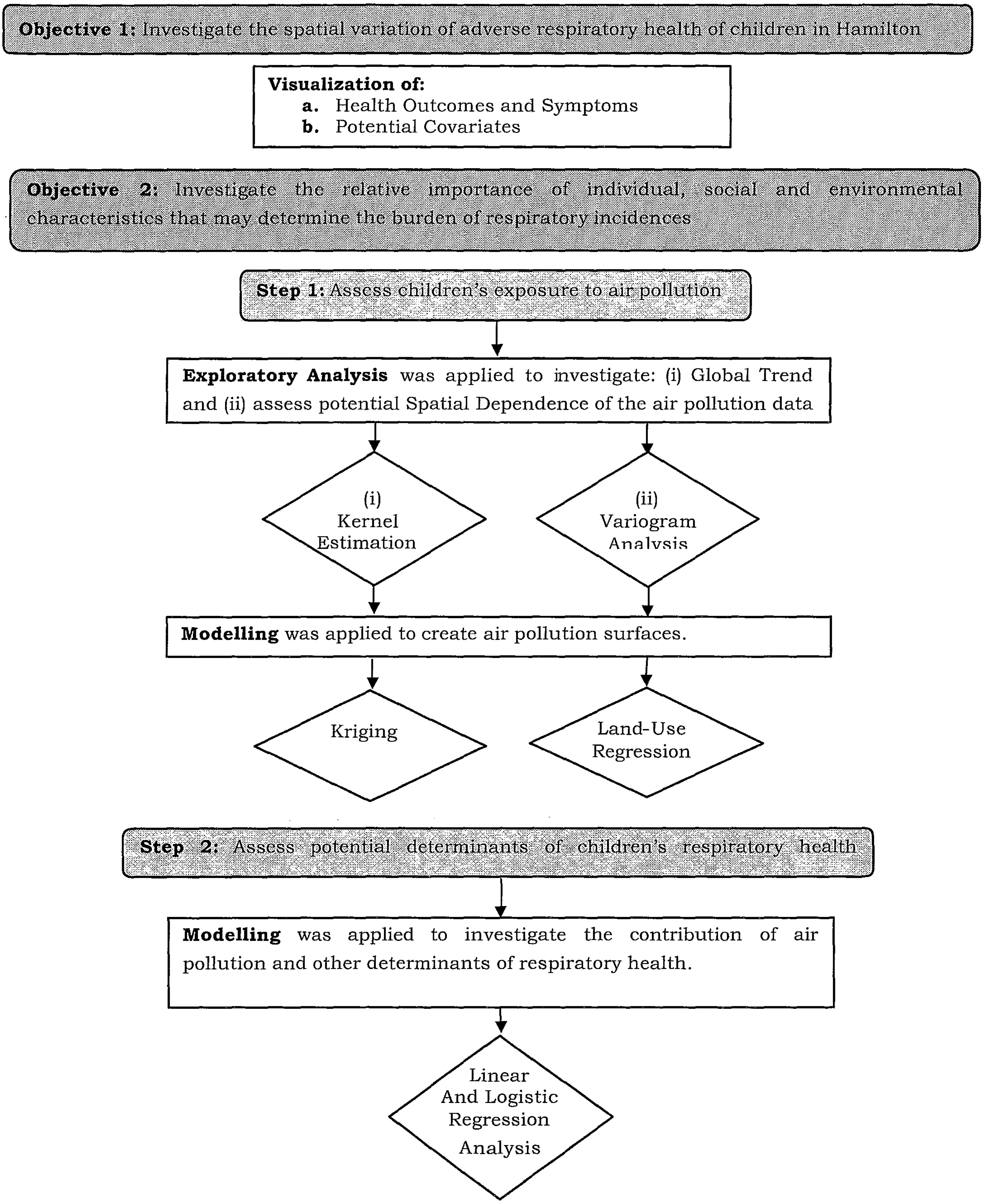

Figure 3.3: Methodological Framework

3.5.1 Visualization and Exploratory Data Analysis

The first objective of this thesis was to visualize and explore the spatial variation of children's respiratory health in the city of Hamilton. To address this, the next steps were followed.

3.5.1.1 Visualization

The first step in the analysis is to visualize the data. Maps and other graphical displays are the fundamental tools, concerned with evaluating patterns and generating hypotheses. For the purposes of this study, dot maps were created in order to visualize a number of health outcomes and symptoms as well as covariates such as family income, household smoking behaviour and use of gas for cooking that were used to investigate the potential determinants of children's respiratory health.

3.5.1.2 Exploratory Analysis

The next step in the analysis was to summarize or generally describe the data set. Exploratory methods involve identifying spatial patterns and generating hypothesis that might explain the pattern. Such techniques are characterized by making few *a priori* assumptions about the data, and many are designed specifically to be robust, that is, to be as resistant as possible to the effect of outliers. Some of the exploratory methods are more concerned with investigating the global trend (*first order effects*), others address the possibility of spatial dependence or local scale effects (*second order effects*). In general the exploratory methods may be in the forms of maps; or may involve graphical plots that are more conventional (Gatrell, 2002).

However, the dividing line between exploratory methods and visualization techniques is becoming rather blurred. Although, somewhat artificial, the distinction is generally a question of the degree of data manipulation, which the methods involve (Bailey, 1995).

3.5.1.3 Kernel Estimation

The Kernel estimation is a commonly used technique in order to identify the variation of the mean value of a set of observations over space. From an environmental perspective, kernel estimation techniques could help us to identify clusters of high or low air pollution levels. This technique estimates the average value, $\mu(\mathbf{s})$, of the observed values y_i, based on neighboring sample data points at a location $\mathbf{s}=(x,y)^T$. In such a case, the average value can be calculated as follows:

$$\hat{\mu}_\tau(s) = \frac{\sum_{i=1}^{n} k\left(\frac{(s-s_i)}{\tau}\right) y_i}{\sum_{i=1}^{n} k\left(\frac{(s-s_i)}{\tau}\right)} \tag{1}$$

Where,

$\hat{\mu}_\tau(s)$: Mean value of an attribute

k (): A suitably chosen bivariate probability density function (kernel function

t: The bandwidth which determines the amount of smoothing

n: Observed events

$\mathbf{s}$: Location in study area

$\mathbf{s}_i$: Typical event

y_i: Value of the attribute, sampled at location $\mathbf{s}_i$

At point $\mathbf{s}$ where the denominator is zero, $\hat{\mu}_\tau(\mathbf{s})$ is set also zero. In addition, $\sum_{i=1}^{n} w_i(\mathbf{s})$ is equal to one and $w_i(\mathbf{s})$ is analogous to h_i^{-a} where a, is parameter and h is the distance between two points. The effect of increasing the bandwidth, t, is to expand the region around **s**, within which the observed values y_i influence the estimate at **s**. For very large t, $\hat{\mu}(\mathbf{s})$ will appear flat and local features will be obscured; if t is small then $\hat{\mu}(\mathbf{s})$ tends to become a collection of spikes centered on the **s**.

For the purposes of the thesis, we employed the kernel estimate to explore the spatial variations of the mean of the pollutant of interest (TSP) and locate areas within the city with higher levels of air pollution.

3.5.1.4 Variogram Analysis

The visualization and exploratory analysis techniques that have been described in the previous sections were useful in order to address the first research objective that is to assess the spatial variation of adverse respiratory health of children in the city of Hamilton, ON. This part of exploratory analysis helps to address part of the second objective of this thesis because it helps investigate the spatial dependence between the air pollution data in order to assess air pollution exposure. Spatial dependence implies that observations that are close to one another have similar values compared to those that are far apart.

In order to explore and quantify the spatial dependence (also called spatial autocorrelation) between the data for any distance h we used the variogram function. The variogram is generally defined as the variance of the difference between the values of a variable at two locations. The estimator of a variogram can be described by the following equation:

$$2\gamma(h) = \frac{1}{n(h)} \sum_{s_i - s_j = h} (y_i - y_j)^2 \qquad (2)$$

Where,

2?(h): A function known as the variogram
h: The distance between all possible pair i, j
n(h): The summation of all pair of observed data
yi, yj: The attribute values of the variable of interest for all i, j pairs where the distance between them is h
si, sj: The location of the points i, j

The variogram model theoretically increases gradually and then flattens out. In general, it can be described by the nugget, the sill, and the range parameters (Figure 3.4).

Range: The distance at which the data are no longer auto-correlated
Sill: The value of the variogram function at a distance equal to range
Nugget: It represents the micro-scale variation or measurement error. It is estimated from the empirical variogram as the value ? (h) for h equals to zero.

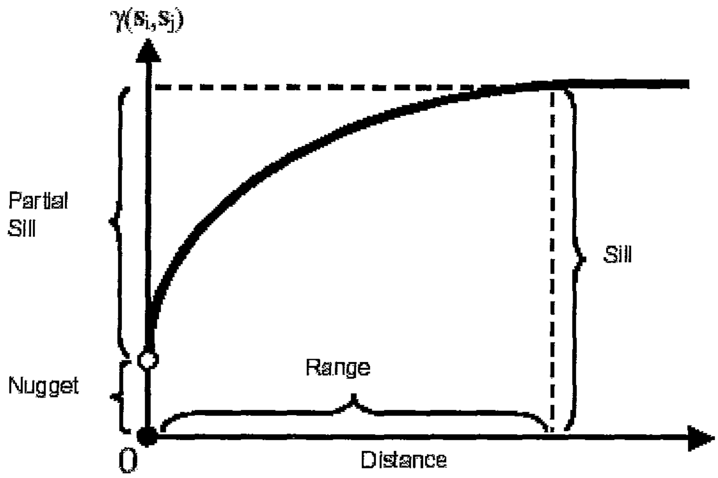

Figure 3.4: A typical representation of a variogram

In order to create a variogram we must define an appropriate lag increment for distance h, a tolerance value for the lag increment and the number of lags over which the variogram will be estimated. The lag increment defines the distances at which the variogram is calculated. The tolerance is added to and subtracted from a given distance, defining the upper and the lower boundary of a distance bin. The number of lags combined with the size of the lag increment will define the total distance over which the variogram is calculated.

The choice of the appropriate lag increment is subjective. It is up to the researcher to investigate the results using several different lags and selecting the one that best represents the pattern. A rule of thumb is to multiply the lag size times the number of lags, which should be about half of the largest distance among all points. However, if the lag increment is too large, the variogram will not capture potential local variation between neighboring sample points. Or, if the lag increment is too small, the variogram will rise sharply and then level off indicating little autocorrelation beyond the range. For this research, the maximum distance between two monitoring stations was 18 km. Thus, after a number of tests, the lag increment was set equal to 500m; the maximum distance was set equal to 10 km and the number of lags equal to 20.

According to Bailey and Gatrell (1995), the variogram analysis is a multi-step process. The first step is to explore the spatial dependence between the data by estimating the variogram cloud. However, in most cases, this is not insightful because it contains multiple values at discrete distances. More pertinent is the average values of these differences at different distances. Therefore, the following step is to estimate the omni-directional empirical variogram. In this case, the hypothesis is that the process under study is isotropic meaning that only the distance between two locations is taken into account. However, sometimes the values of the measured locations will contain a directional influence, which means that in certain directions closer things may be more alike than in other directions. This directional influence is known as anisotropy. In general, anisotropy can be statistically quantified and described by an ellipse. However, in order to control for it, a number of directional variograms need to be created. Specifically, in order to select the variogram that better represents the spatial pattern and thus use it to fit a theoretical variogram the direction and the anisotropy ratio for the directional variograms need to be defined based on the anisotropy ellipse. The direction can be based on the direction of the ellipse while the anisotropy ratio can be estimated based on the axis of the ellipse. The anisotropy ratio is described as follows (Isaaks and Srivastava, 1989):

$$Anisotropy\ Ratio = \frac{major\ axis\ of\ ellipse}{\min or\ axis\ of\ ellipse}$$

Based on that, we are then able to select the variogram that better represents the spatial pattern and thus use it to fit a theoretical variogram to model, in our case for example, particulate air pollution in the city of Hamilton.

3.5.2 Modeling

In some cases, a judicious choice of exploratory analyses, combined with appropriate visualization methods may suffice to answer the questions which a researcher wishes to ask of a particular set of data. In other cases, there may be a requirement to formally test certain hypotheses or to estimate with some precision the extent and form of relationships between the variables of interest. We are then forced to consider explicit statistical models for the data (Bailey and Gatrell, 1995).

A statistical model is concerned with phenomena or processes that are stochastic, and is used in order to specify a probability distribution for the random variable (or the variables) that represents the phenomenon. The specification of a model will involve a combination of both data and reasonable assumptions about the phenomenon under study. Such assumptions may arise for example from background theoretical knowledge about the behavior of a spatial phenomenon. How reasonable the assumptions are can be assessed by exploratory analyses of aspects of the observed data appropriate to those particular assumptions. Once the assumptions are specified, they provide a basic framework for the model. This general form is then fitted –that is values of the unknown parameters are estimated- by reference back to the observed data. The fitted model can then be evaluated which lead to modified assumptions and the fit of a different model.

This is an appropriate point to indicate where statistical hypothesis testing fits into our picture of statistical modeling. Testing a hypothesis is a question of comparing the fit to the data of two models one of which incorporates assumptions that reflect the hypothesis the other incorporating a less specific set of assumptions.

In the thesis, modeling includes procedures to explain the observed variation in the attribute values over the study area and also to predict the attribute values at un-sampled locations. This entails the integration of GIS with standard statistical methods.

3.5.2.1 Kriging – An interpolation technique

Modeling spatially continuous data engages the construction of specific models that predict the attribute values at un-sampled locations. Kriging is a widely used interpolation technique for spatial data. The input data are measurements/observations (**Y**) - of a spatially continuous process - at sample points distributed over the study area. The output is a surface or a highly dense network of point-locations of predicted values within the study area. Estimation of the continuous surface is based on the variogram and the spatial arrangement of measured values. Kriging estimators are called optimal, because they are statistically unbiased and they are minimizing the prediction-mean-square-error, a measure of uncertainty or variability in the predicted values. For a spatial stochastic process Y(**s**) within the area A ($s \in A$) the process of Kriging includes two distinct tasks:

- Quantify the structure, known as variography (Anselin, 2003), where you fit a theoretical variogram to the estimated empirical variogram. The most commonly used theoretical variograms are the spherical, the exponential, the Gaussian, the linear and the power model (Figure 3.5).
- Quantify the spatial structure of the data and produce a prediction. To make a prediction for an unknown value for a specific location, the fitted model from variography will be used.

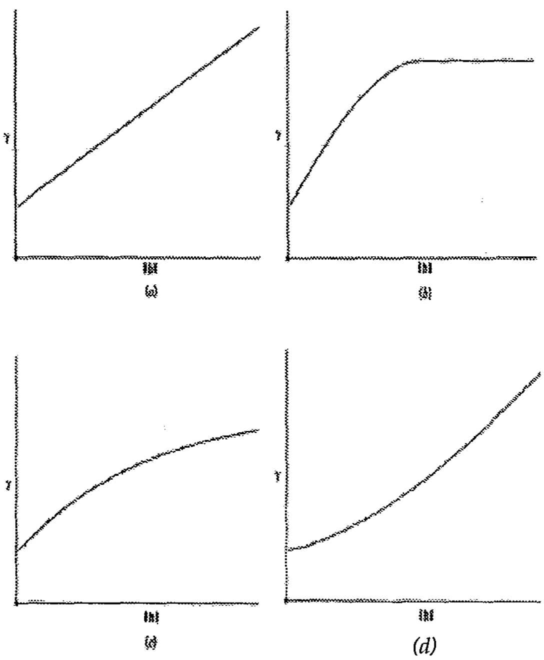

Figure 3.5: (a) Linear Semivariogram, (b) Spherical Semivariogram, (c) Exponential Semivariogram and (d) Power Semivariogram

The most commonly used types of Kriging are the Simple, Ordinary and the Universal kriging. The main difference between the three types is the assumptions about the mean value of the variable under study. The application of simple kriging requires a known mean (value), while ordinary kriging assumes a constant, but unknown, mean, where the deviations are spatially correlated. Universal kriging models local means as a sum of low order polynomial functions of the spatial coordinates and then estimates their coefficients in the model. The latter is more applicable in the presence of strong trends in the measurements (Krivoruchko and Gotway, 2004). The general form of the kriging model can be described by the equation:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}) \tag{3}$$

Where,

$Y(\mathbf{s})$: The variable of interest

$\mu(\mathbf{s})$: In simple Kriging $\mu(\mathbf{s})$ is a known constant mean value
In ordinary Kriging $\mu(\mathbf{s})$ is an unknown constant mean value and finally
In Universal Kriging $\mu(\mathbf{s})$ is some deterministic function of the location $(\mathbf{s})$

$e(\mathbf{s})$: A location-specific random error with zero mean, $E[\varepsilon(\mathbf{s})]=0$

One of the strengths of using this statistical approach is that it is possible to also calculate a statistical measure of uncertainty for the prediction by producing error or uncertainty surfaces.

For the purposes of this thesis Kriging was applied to the selected empirical model in order to create a surface of air pollution for the city of Hamilton in 1980. All calculations for the variogram analysis were conducted using S-Plus

2000 from Insightful Corporation as well as the Geostatistical Analyst extension from ESRI ArcGIS 8.3.

3.5.2.2 Regression Analysis

Regression analysis is a widely used statistical method that studies the relationships among variables. One purpose of regression may be to predict or estimate the value of one variable from known or assumed values of other variables related to it. In order to make prediction or estimates we must identify the effective predictors of the variable of interest. Thus, one of the most crucial tasks in a regression analysis study is to determine which variables are important indicators, which carry only a little information and which are redundant with other variables (Younger, 1985).

The simplest form of regression analysis is the prediction of one (dependent) variable Y from one other (independent) variable X. A regression using only one independent variable in order to predict the dependent variable is known as a *simple regression.* When there are two or more independent variables the analysis is called *multiple regression*. Multiple regression analysis is the most common tool in the modeling of non-spatial data. Based on the above the model can be described by:

$$Y = X\beta + U \qquad (4)$$

Where, Y is the dependent variable, X is the independent variable, ß is the regression coefficient and U is a zero mean vector of errors with variance – covariance matrix C according to which $E(U) = 0$ and $E(UU^T) = C$.

In our case, we applied regression analysis to create an air pollution surface for Hamilton (and compare it with the one we have created based on the Kriging technique) and also to address the final objective of this thesis, which was to investigate the potential factors impacting children's respiratory health for the city of Hamilton.

a) Land-Use Regression

When modeling the air pollution surface using a land-use regression (LUR) model the pollutant of interest - in our case TSP- is the dependent variable. Through this method, we seek to predict air pollution concentrations at a given location based on surrounding land use, traffic characteristics and physical environmental variables such as wind direction and topography.

To date, LUR models have been applied for health-related studies mostly in Europe (Briggs et al., 1997, Briggs et al., 2000, Brauer et al., 2003). Briggs et al., (1997, 2000) investigated the variation of traffic-related air pollution (NO2) in four urban areas of United Kingdom while Brauer et al., (2003) tried to compare particulate air pollution (PM2.5) in multiple European cities. In North America, only two studies, one American (Wentz et al., 2002) and one Canadian (Jerrett et al., 2003) have used this technique to model the variability of carbon dioxide and nitrogen dioxide respectively. In all cases, the models produced good predictions with R^2 values ranging from 0.54 - 0.87.

To our knowledge, this study is the first one that utilizes this method to model the variability of particulate air pollution (TSP) for a North American city. For the model, we created 93 variables using ArcView 3.2 and the SAS System for Windows 8.02 software (see Appendix II for a complete list of the variables). The variables we used may be grouped into the following categories: land use area, proximity to highways and major roads, geographic location (X, Y coordinates), and elevation. For the land use area variables we created buffers for different distances - ranging from 50m to 1750m - from the monitoring stations for the different types of land uses. For the proximity to highways and major roads variables we created binary type of variables to differentiate for the presence of not of a highway or a major road from a specific distance from a monitoring station. Finally, we used the results of a digital elevation model at a 30m resolution developed by DMTI to derive the elevation at the monitoring station site variable.

Due to data limitations we were not able to include any variables related to traffic and meteorology. At the same time, according to (McMichael, 2001) the main sources of particulate pollution (TSP) for an urban area are industry and waste facilities incinerators. Traffic-related pollution is not usually considered as a major source of TSP except for trucks. Therefore, we assumed that the exclusion of related variables might not influence the results of the model. Regarding the meteorological data, the direction of prevailing winds was known for the area and that was taken under consideration during the analysis and the interpretation of the results.

b) Health related regression analysis

When modeling the association between health and a number of other mediators such as socio-economic status of a household, exposure to indoor and outdoor air pollution and dwelling general characteristics, the choice of an outcome may vary from specific respiratory disease incidences, such as pneumonia, asthma, and bronchitis to health indicators such as measures of pulmonary function, symptom indicators and hospital admissions. In this case, both types of respiratory health outcomes were available. Therefore, we applied multivariate logistic regression in the case of a categorical dependent variable and multivariate linear regression in the case of a continuous dependent variable in order to model the relationship between a health outcome and a number of independent variables.

All variables that were significant at the level of 0.05 were entered into the model. All other variables were excluded in forward stepwise regression analysis starting with the variable with the highest significance levels.

3.6 Summary

This chapter described the methodological framework to meet the objectives of this study. The database was obtained mainly from the HCC study, which was then cleaned and geocoded based on the Canadian Census Analyser of the Computing in the Humanities and Social Sciences from the University of Toronto (Canadian Census Analyser, 1996). Spatial information was also extracted from Desktop Mapping Technologies Incorporation. Additionally, land-use information for 1980 was attained by the Hayman and Whetstone study (2000). The data were then examined using a variety of exploratory approaches in order to assess whether there were any trends in the data. Following this a number of analytical methods were applied to assess the potential factors that may impact children's respiratory health. The results of the analysis of these data are presented in the following chapter.

CHAPTER FOUR

AIR POLLUTION EXPOSURE ASSESSMENT

4.1 Introduction

This chapter presents the results of the analysis of the air pollution data in order to assess the exposure of children. In this chapter, exploratory analysis was used to address part of the second objective of this thesis, which is to assess exposure to air pollution spatial analysis methods.

The analysis is subdivided into three key areas: visualization, exploratory analysis and modeling. For visualization purposes, the TSP concentrations were mapped using the proportional symbols option provided by ESRI ArcView 3.2. This was followed by an exploratory analysis to investigate both the general trend (first order effects) and the variations on a local scale (second order effects). To examine the first order effects in the data, the kernel estimation technique was used while variogram analysis was applied for the second order analysis. Finally, in order to assess children's exposure to air pollution, two modeling methods, kriging and land-use regression[1], were employed.

4.2 Visualization of TSP Concentrations

In this section maps are presented to show the spatial distribution of the monitoring stations as well as the TSP variation over the City of Hamilton.

[1] For more details on the methods please refer to chapter three.

A dot map was created to present the distribution of monitoring stations (28 out of 30 were included in the analysis due to data availability). As we can see from figure 4.1, monitors are homogeneously distributed in Hamilton, since there are not many monitoring stations located in the southern part of the city. This is because the monitors were located based on children's homes and schools (Pengelly et al., 1987).

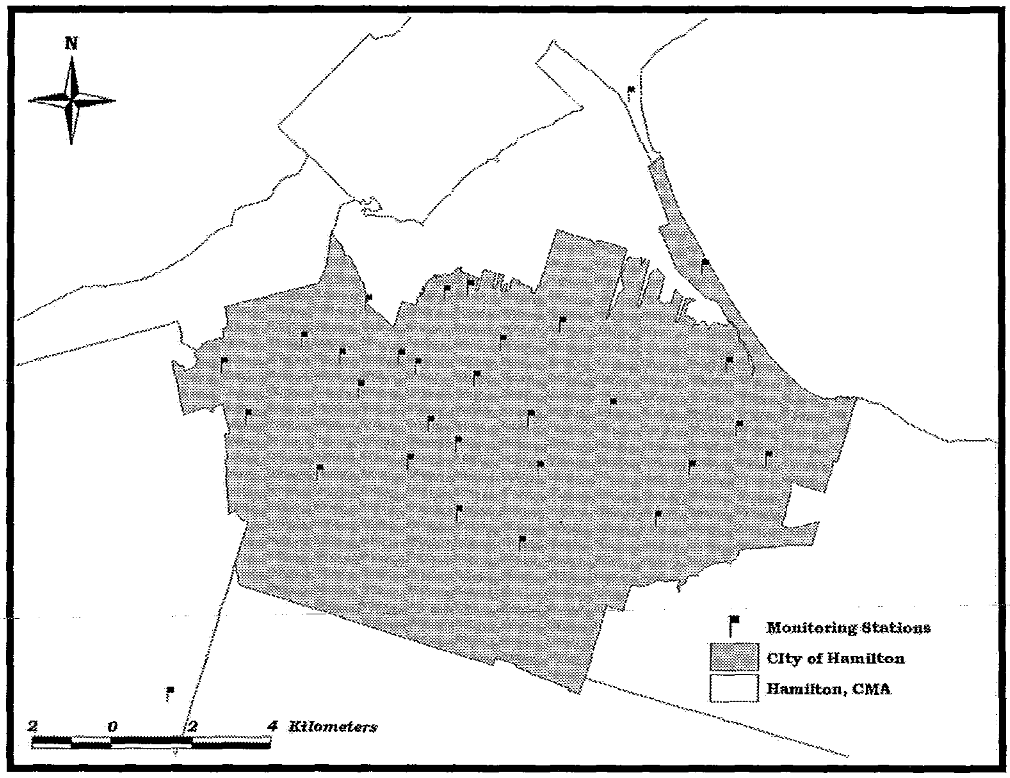

Figure 4.1: Location of monitoring stations in the City of Hamilton (n=28)

Additionally, we created a dot map using the proportional symbol option to visualize the observed TSP concentrations in the City of Hamilton (Figure 4.2). The concentrations of TSP vary between 35 -126 $\mu g/m^3$ with a mean value of

59.68 $\mu g/m^3$. Higher TSP concentrations occur close to the industrial core area at the centre north part of the city with a maximum value of over 120 $\mu g/m^3$.

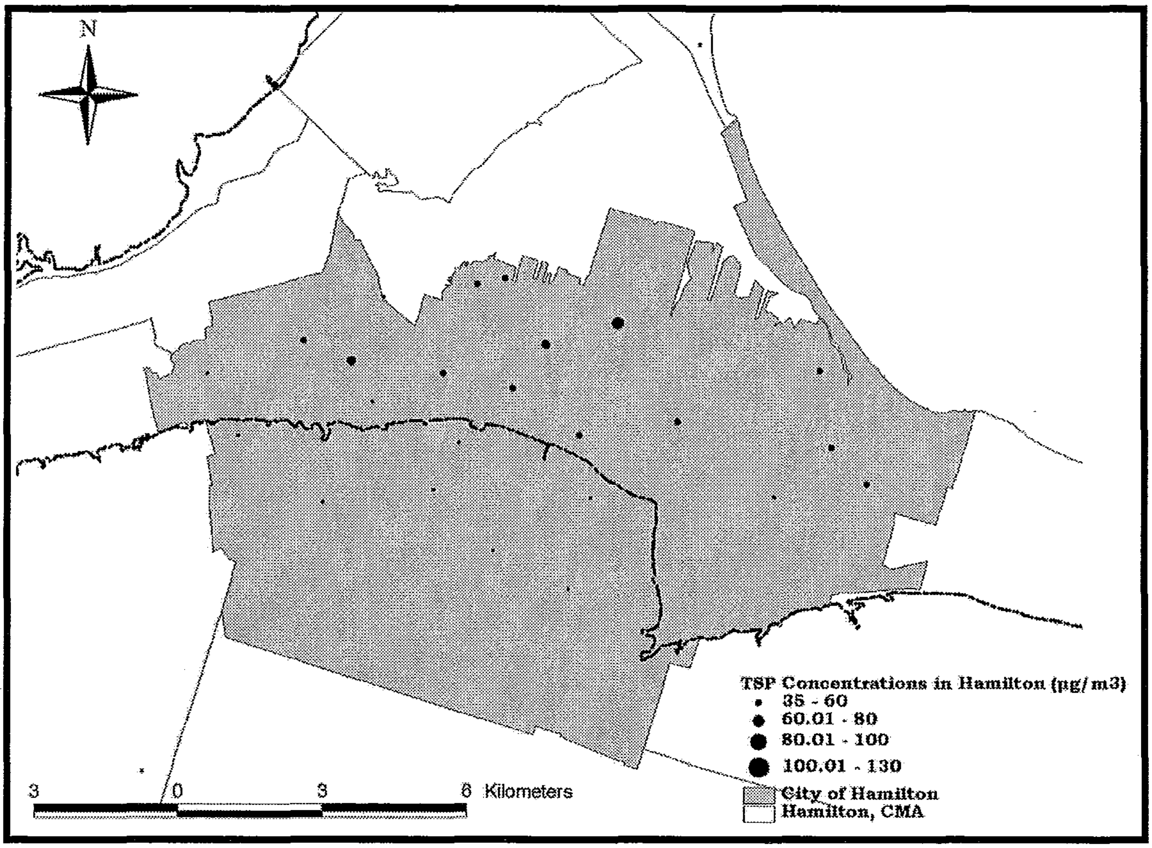

Figure 4.2: Proportional map of TSP concentrations in the City of Hamilton

4.3 Exploratory analysis

The visualization stage was useful in that we were able to observe the distribution of the data over space as well as the variation of TSP concentration values. However, it is difficult to come to any conclusions only on the basis of the visual analysis and thus we proceed to the next stage of the exploratory analysis.

The distribution of the observed values is shown in Figure 4.3. While the distribution of TSP concentrations is skewed to the right (Skewness = 1.28), the difference between the median (60.61) and the mean (59.98) is not significant. Therefore, there is no need to transform the data. Also, based on Figure 4.3 the last bar in the histogram is removed from the others by the width of one bar. The value that corresponds to this bar is the one from the monitor located close to the industrial core area indicating the maximum concentration of TSP observed.

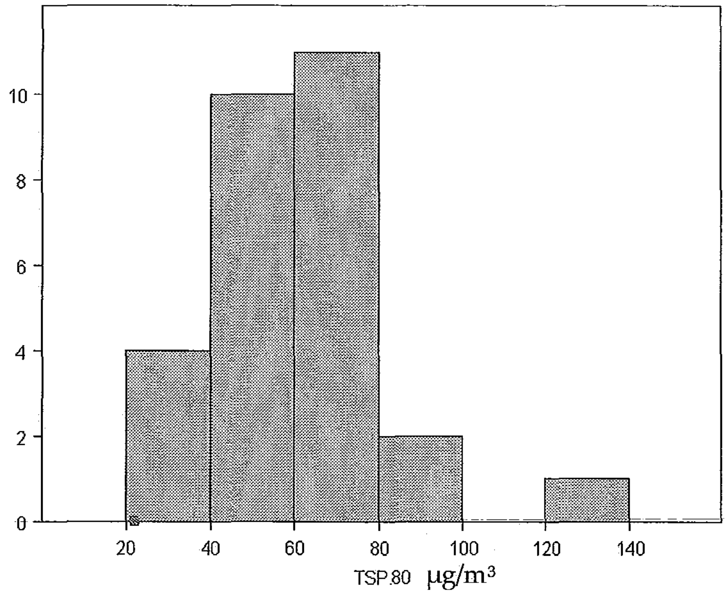

Figure 4.3: Histogram of TSP measurements in the City of Hamilton

Also, most of the TSP values are clustered in the left part of the distribution indicating a leptokurtic distribution with relatively thick tails. This is verified by the value of kurtosis (5.28) presented in Table 4.1.

Table 4.1: Summary Statistics for the TSP measurements μg/m³ in Hamilton

Min = 35.19	**1st Quartile** = 44.24
Max = 126	**3rd Quartile** = 69.93
Mean = 59.98	**Kurtosis** = 5.28
Median = 60.61	**Skewness** = 1.28
Number of Observ. = 28	**Std. Deviation** = 19.99

4.3.1 First-Order Effects

Figure 4.4, presents the kernel estimate for TSP concentrations in Hamilton. The output grid cell size for this estimation is 50m and the search radius 1 km. According to Figure 4.4, high levels of TSP can be traced predominantly at the northern part of the city. Specifically, the highest concentrations of TSP can be observed in close proximity to the industries (centre-north and north-west).

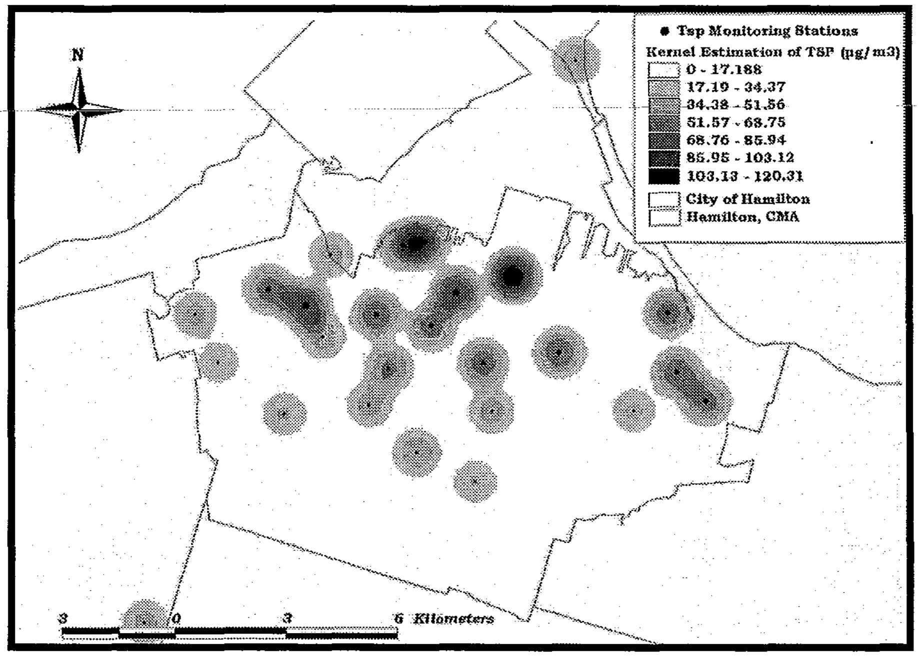

Figure 4.4: Kernel Estimation of TSP concentrations in the City of Hamilton

4.3.2 Second-Order Effects

One of the methods used to examine the spatial dependence between the data is the variogram analysis. The first step is to create the variogram cloud (Figure 4.5), which presents the distribution of the variance between the values at all pairs of monitoring stations for all possible distances. It is a diagnostic tool that can be used to look for potential outliers or trends, and to assess variability with increasing distance. In the case of TSP concentrations the variogram cloud can be seen in Figure 4.5. Based on this figure there are not any outliers or trends observed for the specific data set.

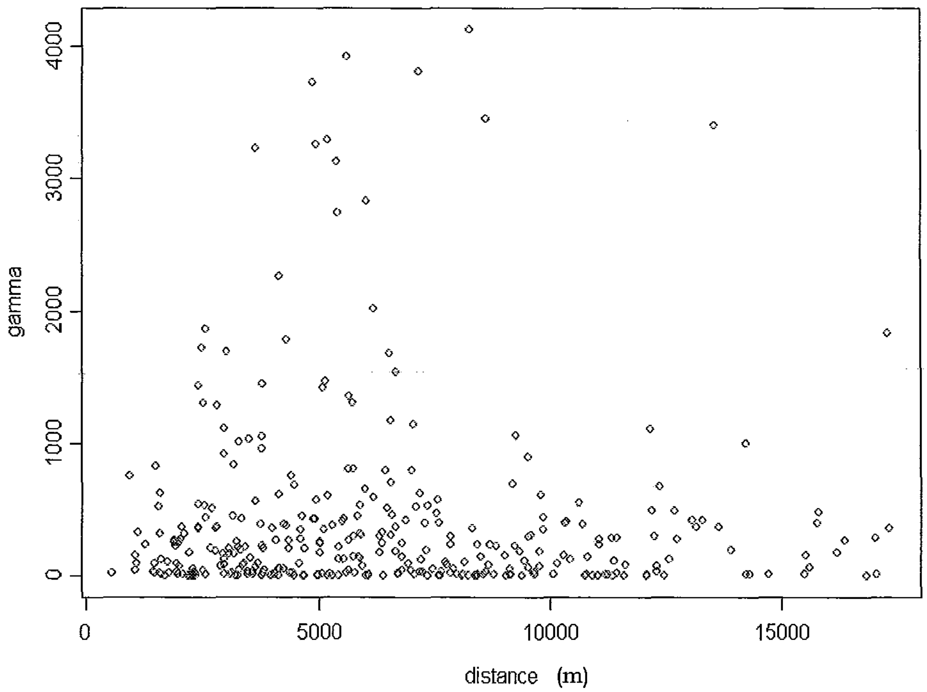

Figure 4.5: Variogram Cloud of TSP concentrations in Hamilton

The next step in the analysis of second - order effects is the estimation of the empirical variogram. To estimate this, we assumed that the process is isotropic, meaning that spatial dependence depends only on distance, not direction. This type of variogram is called omni-directional and for the TSP is presented in Figure 4.6. This figure actually presents the semi-variogram since it is a plot of gamma ?(h) and not 2?(h). After several tries, we set the lag increment equal to 1 km and the maximum distance to 7 km. As the distance between pairs of points increases the variogram increases as well implying that correlation between points decreases with distance. Specifically, the variance between the data keeps increasing up to 5km where it levels off.

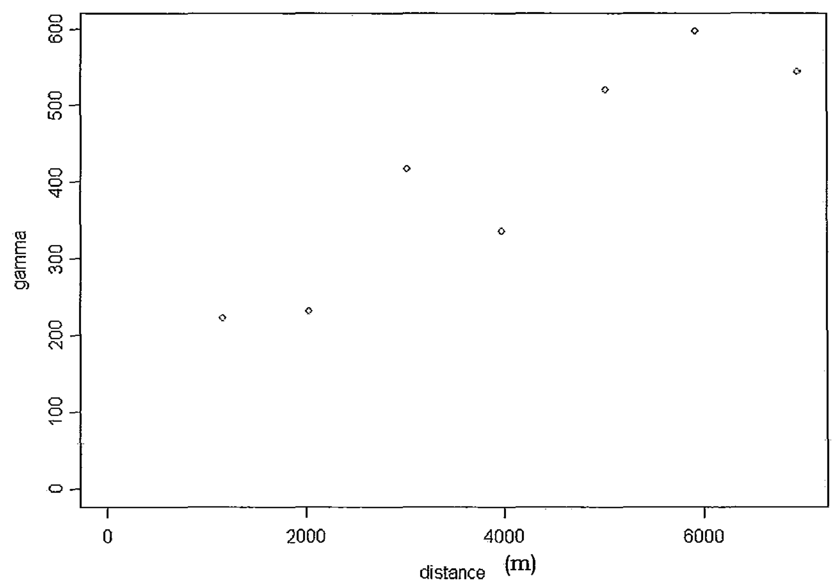

Figure 4.6: Omni-directional empirical variogram of TSP in Hamilton

As a following step in the analysis of the second - order effects, it would be interesting to explore the effects of direction on the spatial dependence between the data in order to assess if there is any particular trend that the data follow. Specifically, variograms for the directions of 0, 45, 90 and 135 degrees are presented in Figure 4.7.

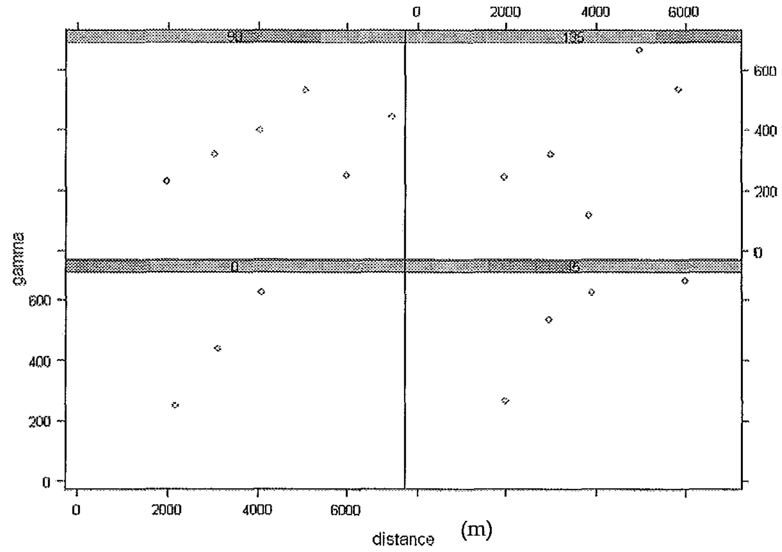

Figure 4.7: Directional Variograms for TSP in Hamilton

According to that figure, the spatial dependence changes with both distance and direction indicating the presence of anisotropy in the process under study. However, there are two types of anisotropy: the geometric and the zonal. Geometric anisotropy occurs when the range of the variogram changes in different directions, while the sill remains constant. Zonal anisotropy exists when the sill of the variogram changes with direction. Appropriately modeling and detrending the data may correct this type of anisotropy. We corrected for both types of anisotropy. Specifically, for geometric anisotropy, the variogram at 55 degrees with a ratio of 0.2 shows a clear variance structure (highlighted in Figure 4.8). However, prior to the selection of the final variogram, zonal anisotropy was also checked based on regression modelling techniques..

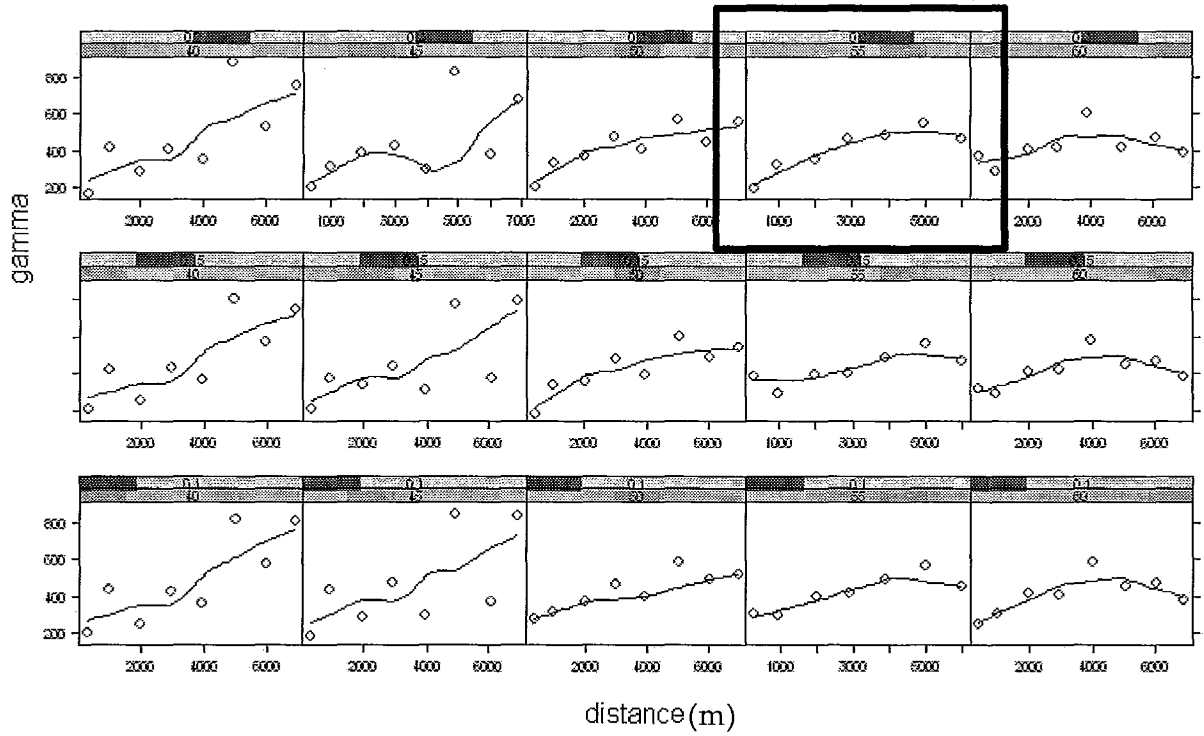

Figure 4.8: Variograms for geometric anisotropy

4.4 Modeling Air Pollution Data

4.4.1 Accounting for anisotropy

Variogram analysis is, in general, an iterative process. To account for zonal anisotropy and specifically investigate if there is any trend in the data, a linear regression model with the X and Y coordinates as the independent variables of the model is fitted.

Table 4.2: Output of the Linear Regression Model for TSP

Variables	Parameter	Std. Error	t-value	Pr(>|t|)
Intercept	-4495.4947	4542.0523	-0.9897	0.3318
Y Coordinate	0.0008	0.0010	0.8446	0.4063
X Coordinate	0.0009	0.0008	1.0624	0.2982
R^2 = 0.1317				

A second technique that is useful is to model trends in certain directions. To fit a model in every direction, we used the Local Regression (loess) function. LOESS combines much of the simplicity of linear least squares regression with the flexibility of nonlinear regression. It does this by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point.

The traditional weight function used for LOESS is the tri-cube weight function,

$$w(x) = \begin{cases} (1 - |x|^3)^3 & \text{for } |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases} \tag{1}$$

x is the predictor value associated with the response value to be smoothed. The weights have these characteristics:

- The data point to be smoothed has the largest weight and the most influence on the fit.
- Data points outside the span have zero weight and no influence on the fit.

The output from this model is presented on Table 4.3.

Table 4.3: Output of the Local (Loess) Regression Model for TSP

General Characteristics of the Model	**Results**
Number of Observations	28
Equivalent Number of Parameters	8.9
Residual Standard Error	15.27
R^2	0.67

According to Table 4.3 the R^2 value is 0.67 (compare to 0.13 from the linear regression model), and this could be considered as a good fit. On Figure 4.9, the surface created based on the Local Regression Model is presented.

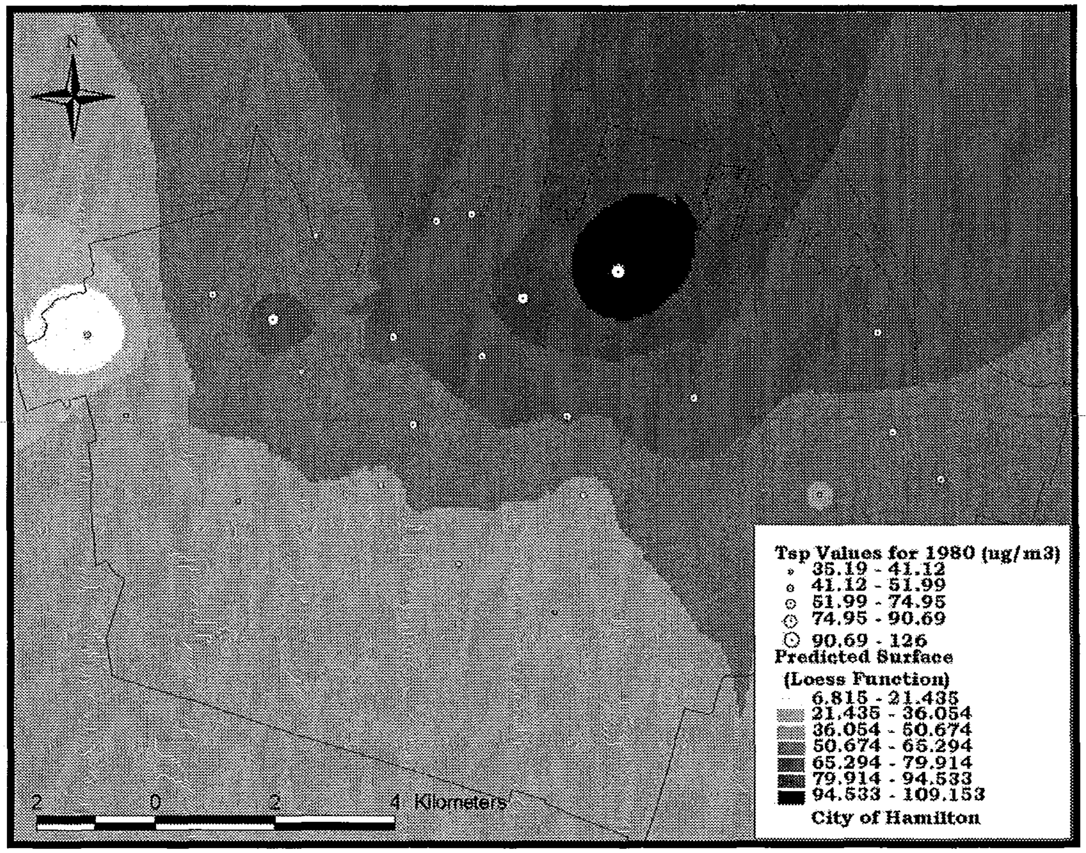

Figure 4.9: Surface created from the Local (Loess) Regression Model for TSP

Earlier, the spatial structure of the data using variogram analysis was examined. This was done with the variable itself (i.e., TSP). All of the previous variogram analysis can be replicated using residuals from the spatial loess model we generated. The reasoning behind the use of residuals is this; by applying a trend surface model to the data, the first order effect has been accounted for (i.e. the mean). The resulting residuals are used in a variogram since they contain whatever trend is remaining (second order effects). The directional variograms using the residuals from the loess regression model can be seen in Figure 4.10.

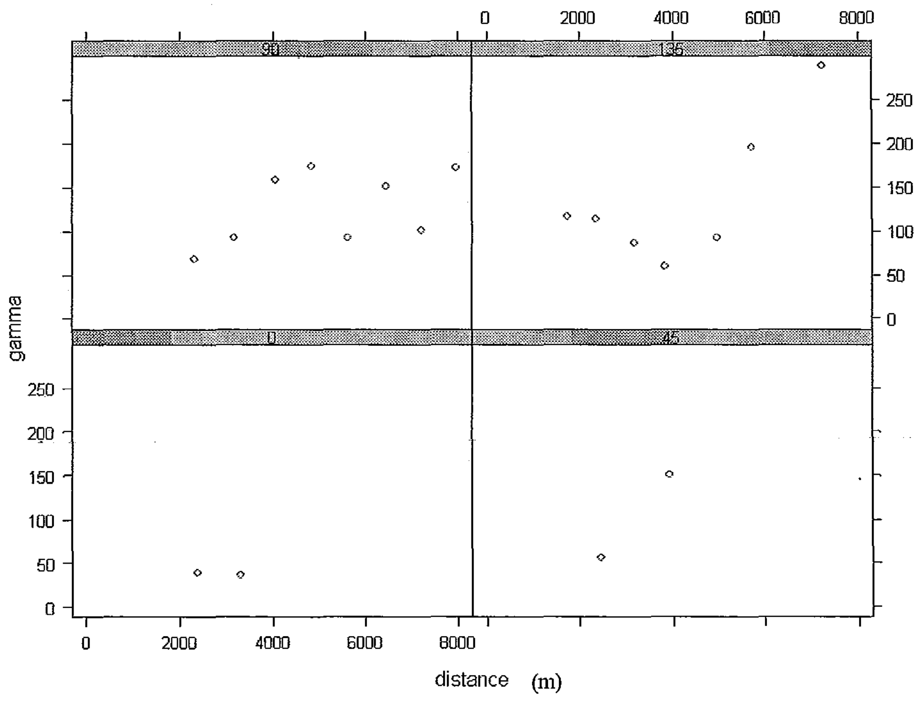

Figure 4.10: Directional Variograms based on the residuals from the Loess Model

After accounting for zonal anisotropy, we checked again for geometric anisotropy at the directional variograms created based on the residuals. Figure 4.11 presents the variograms corrected for geometric anisotropy for various angles and ratios. The variogram that we have selected for the analysis is the variogram at 90 degrees and 0.1 ratio. As we can see from Figure 4.11 after de-trending the data, the variogram that best describes the spatial autocorrelation between the data has a different direction (90 degrees versus 55 degrees).

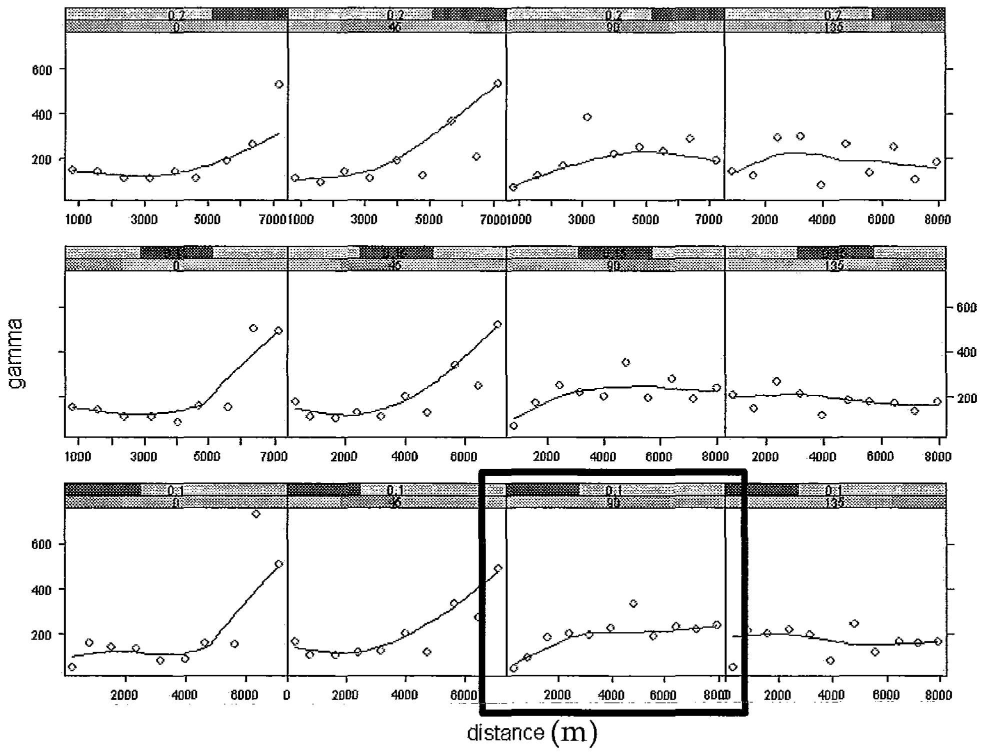

Figure 4.11: Variograms derived from residuals corrected for geometric anisotropy

The next step in our analysis is to fit a theoretical variogram to the observed data using the kriging technique.

4.4.2 Kriging

Kriging is a spatial interpolation method that allows predictions at unknown locations from observations at known locations. As mentioned on chapter three kriging includes two distinct tasks:

1. Model the Empirical Variogram
2. Quantify the spatial structure of the data and produce a prediction.

4.4.2.1 Modeling the Empirical Variogram

Fitting a theoretical variogram model requires the calculation of the empirical diagram. In this case, the variogram at an angle of 90 degrees with a ratio of 0.1 is used. We also applied the robust method of calculation along with a maximum distance of 8km.

Based on the shape of Figure 4.12 and comparing with the theoretical shapes in Figure 3.4 (see chapter three), we selected a spherical variogram.

Spherical Model Variogram	
Range	3238.001
Sill	231.581
Nugget	12.507
Objective	14510.019

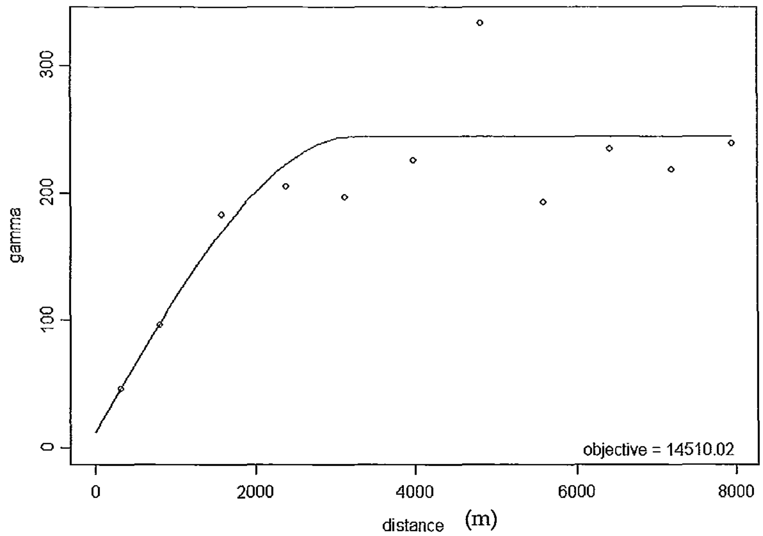

Figure 4.12: The spherical variogram model for TSP in Hamilton in 1980

The general form of a spherical model variogram is given by the following form (Bailey and Gatrell, 1995):

$$\gamma(h) = \begin{cases} \alpha + \sigma^2\left(\dfrac{3h}{2r} - \dfrac{h^3}{2r^3}\right) & \text{for} 0 < h \leq r \\ 0 \quad \text{for } h = 0 \\ \sigma^2 \quad \text{otherwise} \end{cases} \tag{2}$$

where, h is the distance between two point-observations, σ^2 is the sill, r is the range and α is the nugget effect. The estimated theoretical variogram of TSP concentrations has the following form:

$$\gamma(h) = \begin{cases} 12.507 + 231.581\left(\dfrac{3h}{2 * 3238.001} - \dfrac{h^3}{2 * (3238.001)^3}\right) & \text{for } 0 < h \leq r \\ 0 \quad \text{for } h = 0 \\ 231.581 \quad \text{otherwise} \end{cases}$$

4.4.2.2 Results of Kriging Analysis

The two types of kriging we applied for the purposes of this thesis are ordinary and universal kriging. Ordinary kriging uses a random function model of spatial correlation to calculate a weighted linear combination of observations at sampled points, for predictions of a nearby, un-sampled location. Universal kriging is an adaptation of ordinary kriging that accommodates spatial trends. In the case of the TSP data, the variogram model from the previous section informs the researcher that there is a trend in the data. If a trend is present, then the researcher has two options for kriging, the first is to remove the trend and work with the residuals using ordinary kriging. The second is to de-trend the data using universal kriging. For the purposes of this thesis we applied both ways.

Figures 4.13a and 4.13b show an ordinary kriging surface generated with a spherical variogram of the residuals from the Local (Loess) Regression Analysis for TSP (1980) in Hamilton as well as the standard errors. The number of monitoring stations included in the model is 28.

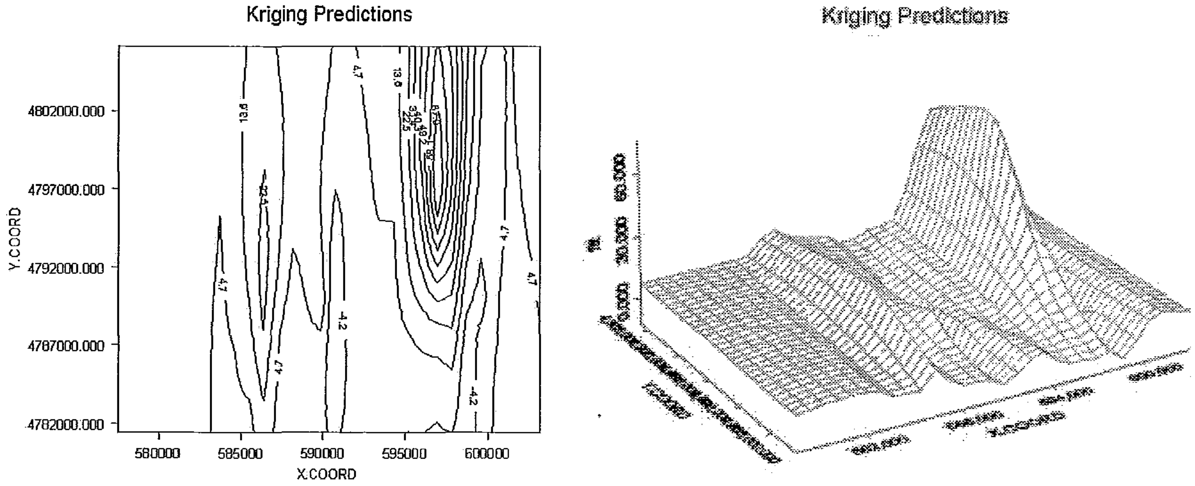

Figure 4.13a: Ordinary kriging Contour Plot and Predicting Surface for TSP

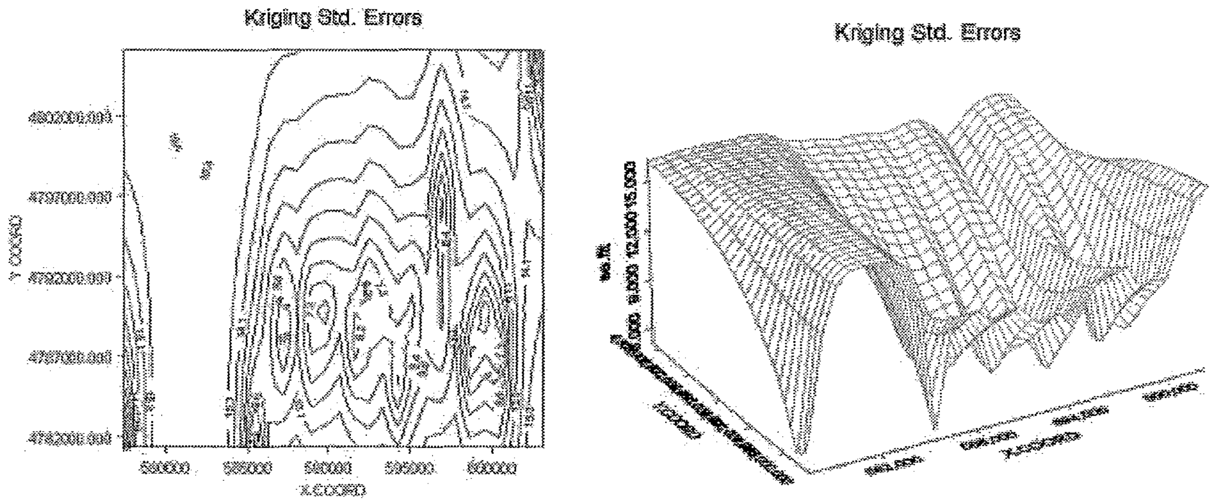

Figure 4.13b: Standard Errors Contour and Surface Plots

For visualization purposes, we created the surface in ArcView 3.2 based on the predicted values (Figure 4.14). Based on figures (4.13a and b and 4.14), the industrial core area (northeast part of the city) appears to have the highest levels of TSP.

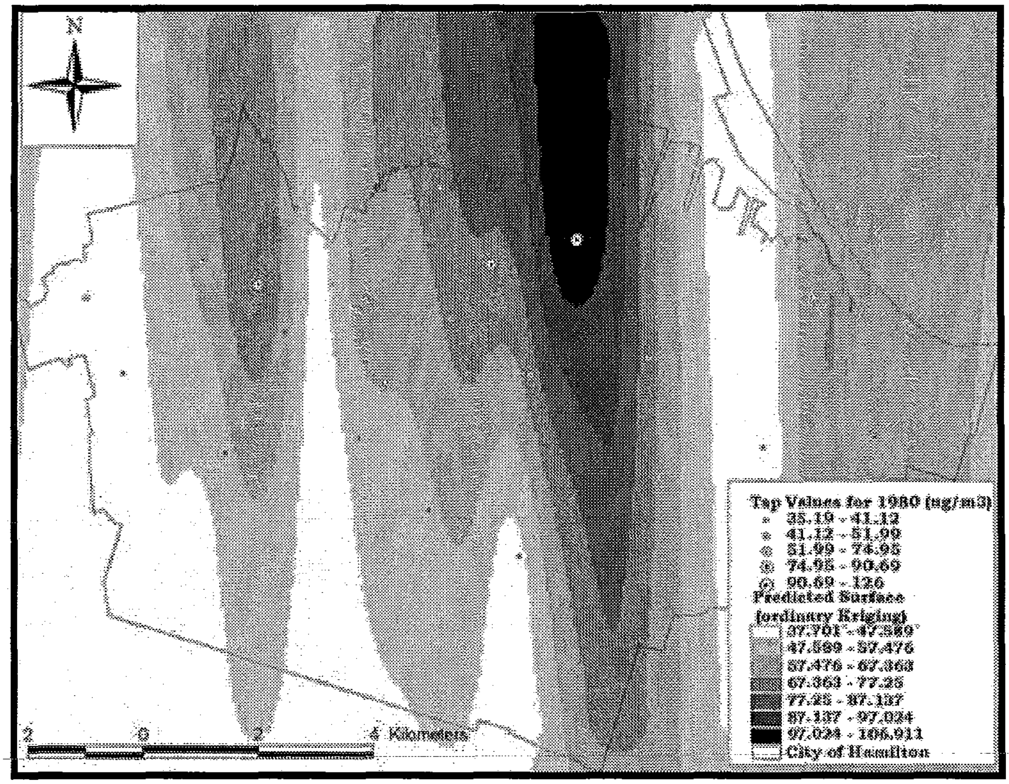

Figure 4.14: Ordinary Kriging Surface for TSP in Hamilton

Figure 4.15a and 4.15b show the results for universal kriging. The results produced are slightly different since this type of kriging incorporated a quadratic trend. In terms of the standard errors, the results are almost identical with larger standard errors occurring at the edges of the study area in both cases.

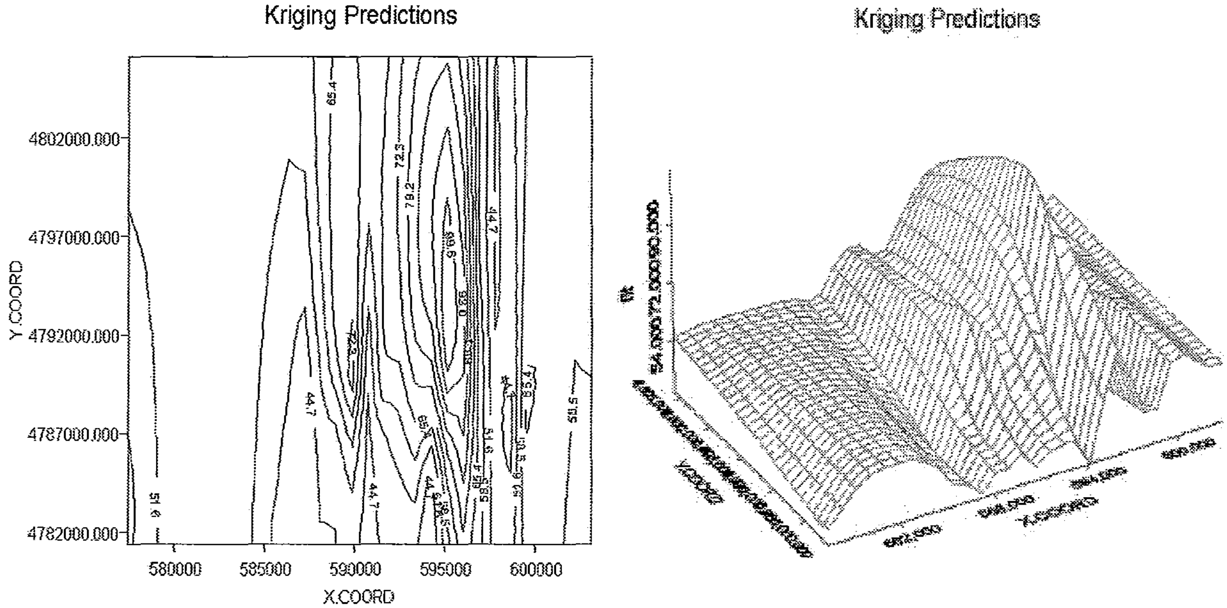

Figure 4.15a: Universal Kriging Predicting Contour Plot and Surface for TSP

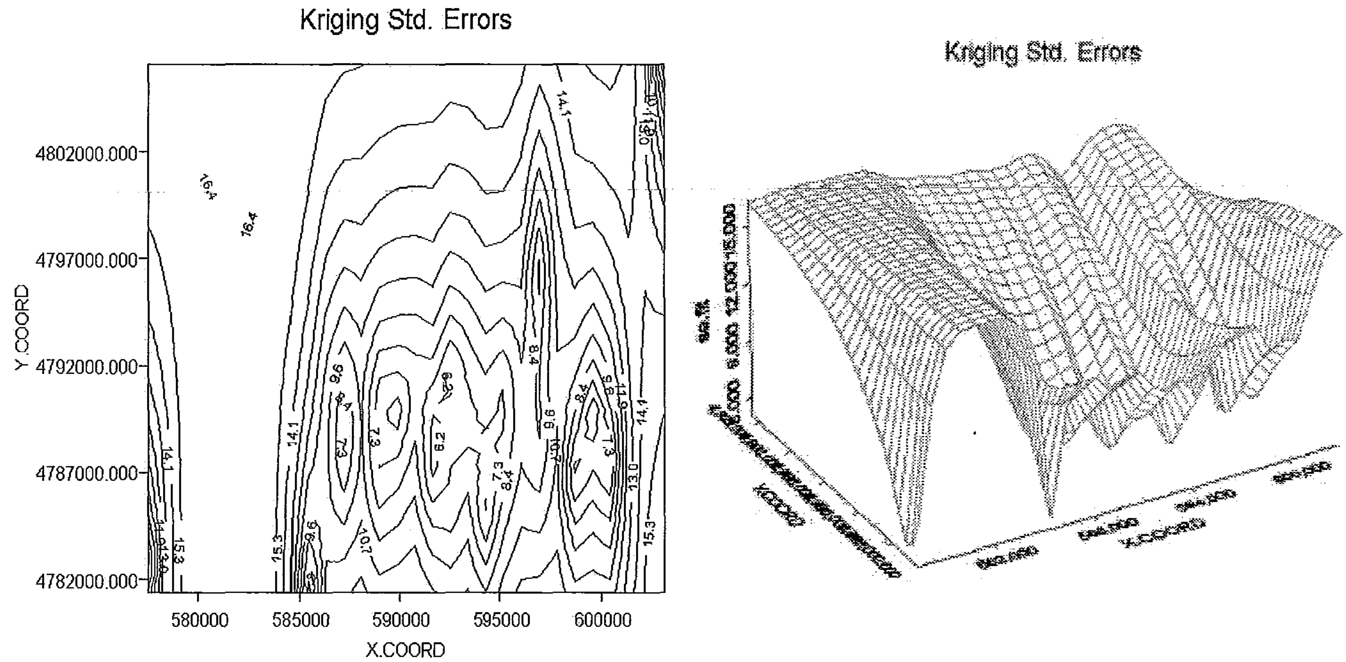

Figure 4.15b: Universal kriging Standard Errors Contour Plots and Surface for TSP

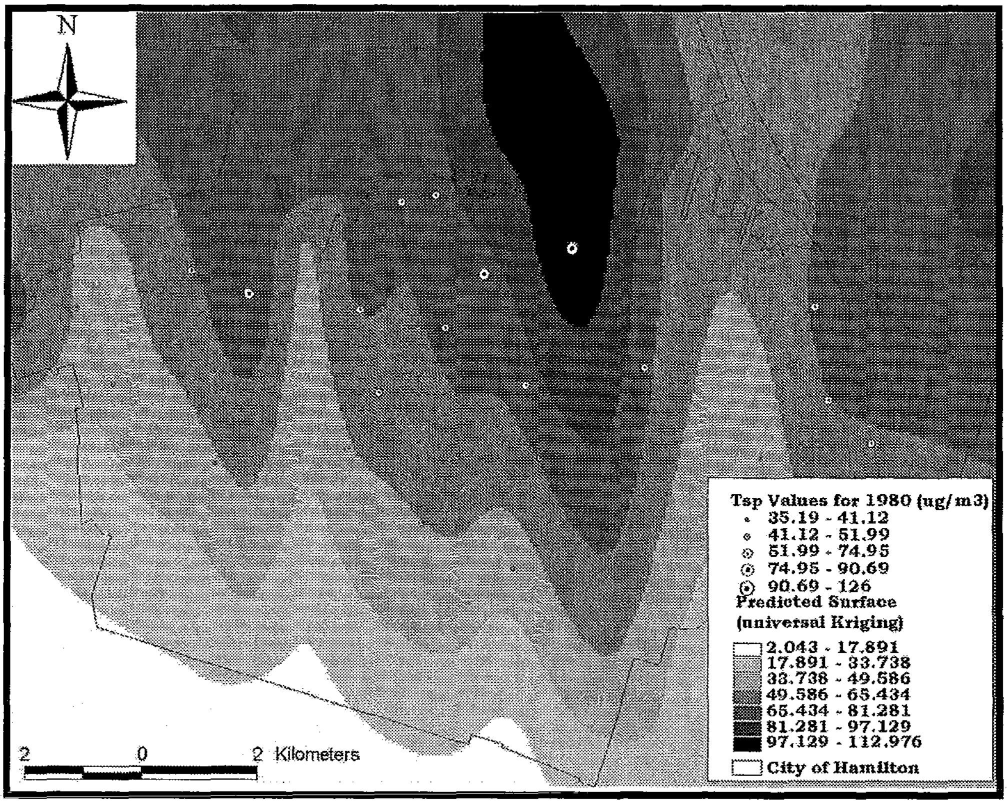

Figure 4.16: Universal Kriging Surface for TSP in Hamilton

Finally, we created the surface for Universal kriging in ArcView 3.2 (Figure 4.16). The results look similar, however the predicted surface produced by universal kriging appears to be closer to the concentration values of the monitoring stations in that the areas with high levels of TSP continue to the mountain area for ordinary kriging whereas for universal kriging high levels of TSP are mainly concentrated to the downtown and industrial core areas.

4.4.3 Land-Use Regression

Another method we applied in order to assess children's exposure to air pollution is the Land-Use Regression (LUR). In total, we created 161 variables or variations of different variables of land use using ArcView 3.2 and the SAS System for Windows 8.02. Each one of the independent variables was tested through a bivariate regression model using the SAS System for Windows 8.02. In this way we were able to identify those that were highly correlated with the TSP concentration values. In general, the variables with the highest association were the industrial type of land-use within 750m of a monitoring station (t-value = 5.15, R^2 = 0.51), the elevation at the monitoring site (t-value = -3.78, R^2 = 0.35) and the open space type of land use within 300m of a monitoring station (t-value =-2.01, R^2 = 0.13).

In the following step, we applied forward stepwise regression analysis starting with the variable with the highest significance levels (i.e. the industrial land use type within 750m of a monitoring station). All variables that were significant at the level of 0.05 produced the model shown in Table 4.4. According to Table 4.4, variables are not affected by multicolinearity according to the correlation matrix (attached at Appendix III). Also, all of the coefficients, except the variable "presence of a highway within 1km of a monitoring station" have the expected signs.

Table 4.4: Results of the LUR final model

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	74.37081	5.37366	13.84	<.0001
Commercial land use within 50m of a monitoring station	0.00256	0.00085953	2.98	0.0069
Industrial land use within 750m of a monitoring station	0.00005967	0.00000784	7.61	<.0001
Open space land use within 300m of a monitoring station	-0.00012114	0.00003317	-3.65	0.0014
Presence of a highway within 1km of a monitoring station	-10.50338	3.62635	-2.9	0.0084
Elevation at monitoring station	-0.1288	0.03182	-4.05	0.0005
R^2 = 0.8746 **Adj R^2 = 0.846**				

In order to understand the unexpected sign for the highway variable, we first re-run the model excluding the variable. The results are presented on Table 4.5. Based on this table, all the variables have a significant t-value (<0.05) and all of the coefficients have the expected signs. However, when tested for spatial autocorrelation within the residuals of the model, the Moran's I statistic was equal to 0.4233 with a p-value of 0.02.

Table 4.5: Results of the LUR where the HWY1000_80 variable has been excluded

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	68.70243	5.75264	11.94	<.0001
Commercial land use within 50m of a monitoring station	0.00314	0.00096083	3.27	0.0034
Industrial land use within 750m of a monitoring station	0.00005899	0.00000901	6.55	<.0001
Open space land use within 300m of a monitoring station	-0.00012147	0.00003812	-3.19	0.0041
Elevation at monitoring station	-0.1072	0.03555	-3.02	0.0062

R^2 = 0.8267
Adj R^2 = 0.7966

In addition, the appearance of distinct spatial patterns of the errors in Figure 4.17 suggests that they are not randomly distributed over the study area implying the presence of spatial autocorrelation. Specifically, negative values are concentrated at the northwestern part of the city. However, when the variable "presence of a highway within 1 km of a monitoring station" was included in the model the Moran's I statistic showed that spatial autocorrelation was not significant (Moran's I=0.06, pvalue<0.05). Therefore, the variable was included in the model mainly because this variable accounted for spatial autocorrelation within the residuals of the model. In addition, the negative sign of this variable could also indicate that at that period of time, the levels of TSP pollution at the areas close to the industry were much higher compared to those close to a highway, minimizing in this way the effect of highway pollution and resulting in the negative sign in the model.

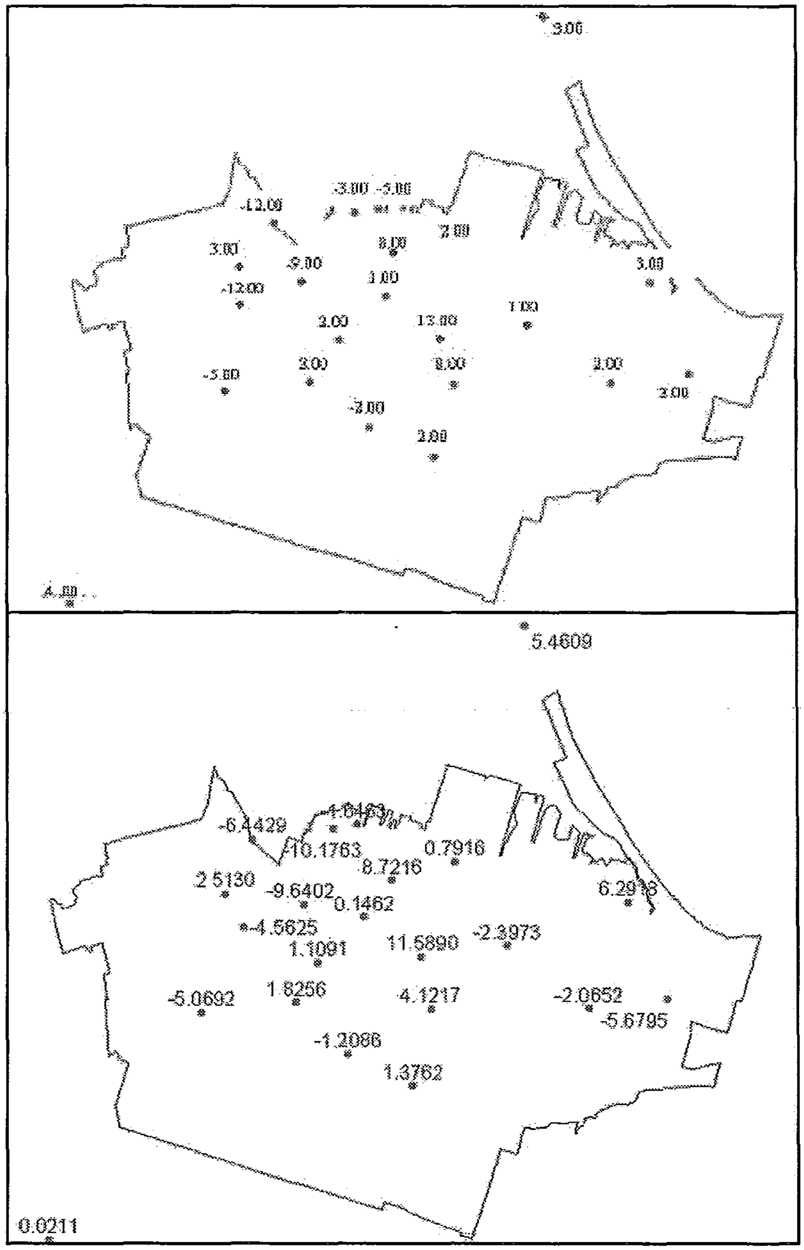

Figure 4.17: Spatial Distribution of Residuals from Linear Regression models:

a) The variable of "presence of a highway is not included in the model

b) The variable of "presence of a highway is included in the model .

After deriving the final LUR model (Table 4.3), the parameter estimates were used to map a predicted pollution surface using a 50 by 50 meter grid so that the land use and highway features could be represented accurately. For each one of the variables included in the final model, we estimated the appropriate buffers for each one of the variables (i.e. for the "Industrial land use within 750m of a monitoring station" we estimated a buffer of 750m, for the "commercial land use within 50m of a monitoring station" variable, a buffer of 50m for the "presence of a highway within 1km of a monitoring station" variable, a buffer of 1km and for the "parks-open space land-use within 300m of a monitoring station" variable a buffer of 300m). As a following step, the estimated buffers were intersected with the 50x50 grid. Finally, in order to generate the pollution surface the estimated parameters were incorporated using the ArcView Map Calculator. The final predicted surface was estimated based on the following equation (4.1) and mapped in ArcView 3.2 (Figure 4.18:

TSP = 74,3708.AsGrid + 0.00256.AsGrid * "Commercial land use within 50m of a monitoring station" + 0.00005967.AsGrid * "Industrial land use within 750m of a monitoring station" - 0.0001214.AsGrid * "Open space land use within 300m of a monitoring station" - 10.50338.AsGrid * "Presence of highway within 1km of a monitoring station" - 0.1288.AsGrid * "Elevation at monitoring station"

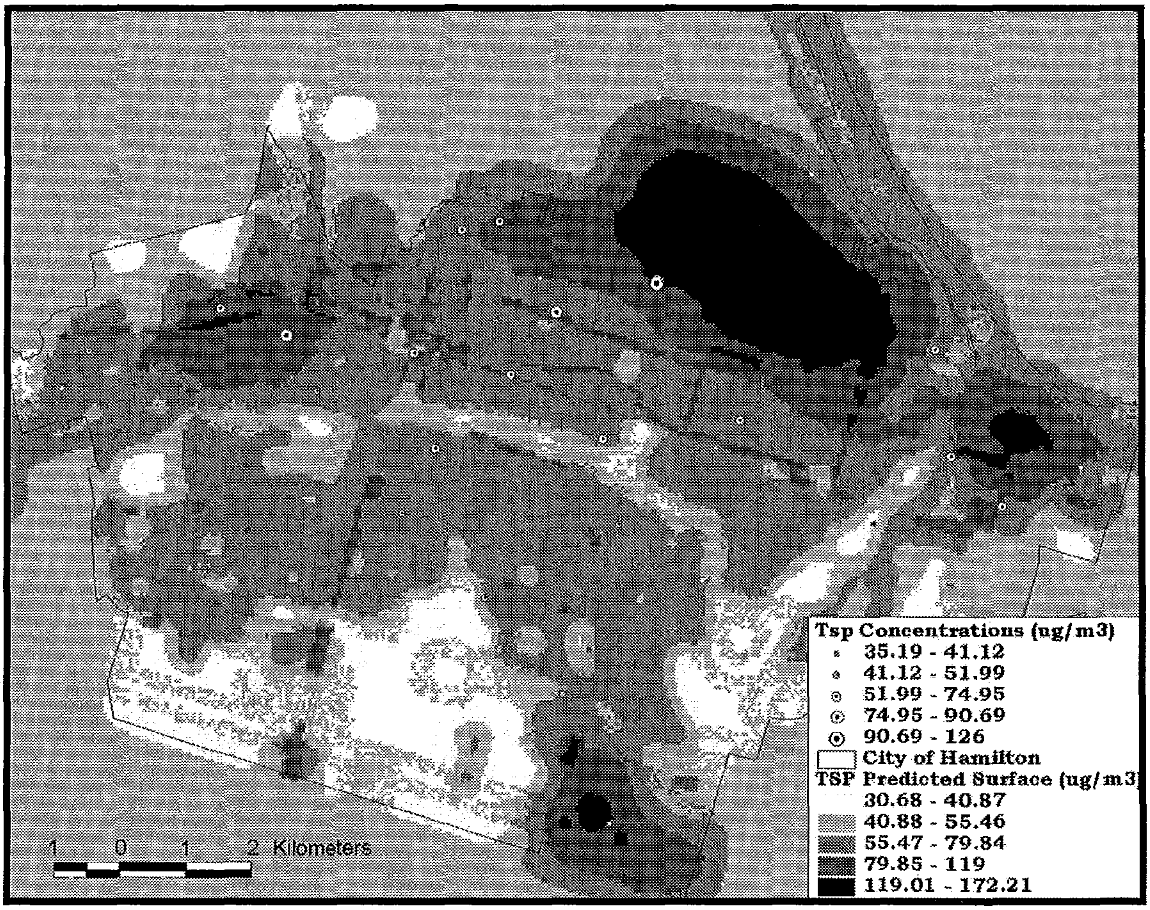

Figure 4.18: Predicted TSP Surface based on Land Use Regression for Hamilton

The land use regression model created a surface with high levels of air pollution close to the industrial core areas and the main commercial land use areas while the open space-parks areas exhibit lower levels of TSP. Compared to the surface created by kriging, this map shows more detailed spatial variation.

The final exposures from both surfaces (kriging and land-use regression) were assigned to each child based on their home location and using the "Get Grid Value" extension of ArcView 3.2.

4.5 Cross – Validation of Results

Cross-validation analysis was undertaken to confirm the validity of the results. Specifically, a random sample of five monitoring stations was selected from the original 28 monitoring stations and both kriging and land-use regression models were re-estimated to assess the performance of the models.

For Kriging, although several directional variograms were created at various ranges none of them could capture the spatial dependence between the data. Therefore direction was taken into consideration in this case and the variogram model that was produced using the 23 monitoring stations was based on an omni-directional variogram. Furthermore, a spherical model was used to fit the experimental omni-directional variogram. The resulting surface was then used to estimate the TSP concentrations at the five excluded sampling locations. The model generally under -predicted TSP especially at the areas with higher levels of air pollution. The difference in the average concentration comparing the predicted and the actual values is 8.06%. Yet, the results need to be interpreted with caution due to limited number of monitoring stations used to model the data. For land-use regression, using the 23 monitoring stations the model was re-estimated. The new model produced results that were remarkably similar to the model where all the monitoring stations were included. Then the coefficients of the new model were used in order to estimate the TSP concentrations at the five excluded sampling locations. The model over-predicted TSP for the locations with higher levels of air pollution and under -

predicted TSP at the areas with lower levels of air pollution. The difference in the average concentration comparing the predicted to the actual monitored values for the five stations was smaller than the average based on kriging (2.29% versus 8.06%) suggesting that LUR performs better compared to kriging in predicting the TSP concentration values over the city of Hamilton. The results from cross validation are presented in Table 4.6

Table 4.6: Comparison of the kriging and LUR new models to the actual TSP concentration values

Mon. Station	TSP 1980	Kriging	%Difference Kriging	LUR	%Difference LUR
29012	74.95	65.57	-12.5150	79.44	5.9907
29087	67.41	72.5	7.5508	65.72	2.5070
104	37.52	35.48	- 5.4371	39.1	4.2111
118	66.36	58.24	-12.2363	60.14	-9.3430
106	46.94	38.64	-17.6821	48.42	3.1530
Average			-8.0640		2.2894

4.6 Discussion

In this chapter, the results of the air pollution analysis were presented. In order to improve the validity of the data -in terms of the results of the analysis two different methods (kriging and land use regression) were employed to assess children's exposure. As described in section 4.2, TSP concentrations were available for 28 out of 30 monitoring stations for the kriging analysis and 25 out of 30 for the LUR analysis. It should be mentioned though that while a number of studies have used kriging to assess exposure to air pollution (Jerrett et al., 2003), according to Briggs et al., (1997), there is no clear consensus as to which approach provides a more accurate air pollution surface.

Before creating the final kriged surface, a number of steps were taken. Specifically, the estimation of the TSP concentration over the City of Hamilton requires the calibration of a model variogram (e.g., spherical, exponential, or gaussian). The first step in this procedure was to explore the data using the kernel estimate. The resulting surface from the kernel estimation showed that higher levels of TSP are present at the northern part of the city mainly due to the presence of industrial activity. As a following step, we estimated an omnidirectional and four directional variograms (for 0, 45, 90, 135 degrees) to produce the variogram that would best describe the spatial autocorrelation in the data.

The change of the sill and the range of the directional variograms presented in section 4.3.2, indicated the presence of both zonal and geometric anisotropy. Specifically, the industrial core area at the northeast part of the city produced a strong spatial trend at the 55 degrees in the measurements. In this case, the spatial coordinates serve as a proxy to this trend. Particularly, TSP concentrations change, not because of the coordinates themselves but mainly because of the presence of the industrial core area. Therefore, in order to de-trend the data, a local (loess) regression model was applied, which explained 67% of the TSP variation. After accounting for geometric anisotropy at the residuals from the loess regression and as the best variogram was finally selected at 90 degrees and 0.1 ratio. As a following step, a spherical model was selected to fit the final variogram. The choice of the theoretical variogram is generally based on experience and experimentation. The created TSP prediction surfaces are described in section 4.4.2.2. The predicted surface produced by universal kriging appears to describe better the spatial variation of TSP concentrations compared to ordinary kriging.

The LUR model estimated for this study incorporated three types of variables: a) land use information, b) elevation and c) road/highway network. Other variables that have been used in LUR and could potentially be incorporated in this model are traffic volume[2] (Briggs et al., 1997) and meteorological data (Jerrett et al., 2003).

The results of the LUR model suggest that although the model provides a relative logical estimation of TSP concentrations across the city of Hamilton, it also over- estimates the actual concentrations of TSP. To be more specific, the actual TSP values vary between 35 - 126 $\mu g/m^3$ whereas the predicted ones vary between 30 - 172 $\mu g/m^3$. This could however, be improved with calibration of the model at the local level, the addition of explanatory variables that have not been taken into account (for example meteorological data) and/or the incorporation of additional monitoring stations.

Following the estimation of the models, cross-validation was applied to assess the general performance of the models and their predictive abilities. In general the results of cross-validation indicate that both models perform relatively well. However, the LUR performed better than kriging, capturing the local variations of TSP concentrations values. Yet, the larger differences

[2] Due to the pollutant of interest (TSP) traffic volume data could emphasize more on trucks as - compared to cars- this is the major source of TSP.

between the actual and the predicted values may be attributed to the limited number of monitoring stations used for the cross-validation.

Finally, the sensitivity of the model to variable specification also requires our attention in the interpretation of the results. For example, for the "Industrial land use within 750m of a monitoring station" variable the model could produce a wide range of results, depending on the size of the buffer. However, based on the bivariate regression, the buffer of 750m from a monitoring station that included industrial land use type in it was the most significant and that is why it was selected as the first variable to enter the model in the multivariate forward stepwise regression analysis.

4.6 Summary

In general, the results presented in this chapter illustrate the complex nature of spatial variation of air pollution in urban areas. However, they also suggest that reasonable predicted surfaces can be derived for the City of Hamilton using both Kriging and LUR analysis. For LUR, it is worth noting that, to our knowledge, this is the first study in Canada that utilizes this method to model the variability of particulate air pollution (TSP). Also, compared with kriging, LUR seems to capture more effectively the local spatial variations of TSP for the City of Hamilton. These small area variations may prove to be crucial for the assessment of children's exposure to particulate air pollution and as a result may detect health effects that would have gone unnoticed if we only had employed kriging to estimate exposure. To further investigate exposure, as well as the potential factors that impact children's respiratory health, multivariate analysis was applied to the data. The results of this analysis are presented in the following chapter.

CHAPTER FIVE

DETERMINANTS OF RESPIRATORY HEALTH DATA

5.1 Introduction

The previous chapter described the assessment of air pollution. This chapter presents the results of the multivariate analysis performed in order to investigate the importance of children's exposure to air pollution along with a number of other covariates such as individual, socio-economic, and other environmental characteristics in determining their respiratory health. Specifically, the methods applied in this chapter have been used to address the second objective of this thesis: to investigate the determinants of children's respiratory health.

With respect to this objective, visualization techniques and exploratory/descriptive analysis have been employed to demonstrate the prevalence as well as the spatial distribution of the health outcomes and the covariates taken into consideration throughout the cohort and the study area. Also, in order to assess the potential determinants of children's respiratory health, we used bivariate and multivariate logistic and linear regression analysis.

5.2 Exploratory analysis

For this part of the analysis, maps were used to demonstrate the spatial distribution of a number of health outcomes and covariates over the study area. Specifically, once the geocoding was completed, several dot maps were created to assess any general trends and patterns across the study area. Figure 5.1 presents the home locations of all children selected for this study. From this figure, it is noticeable that the sample is uniformly distributed over the City of Hamilton.

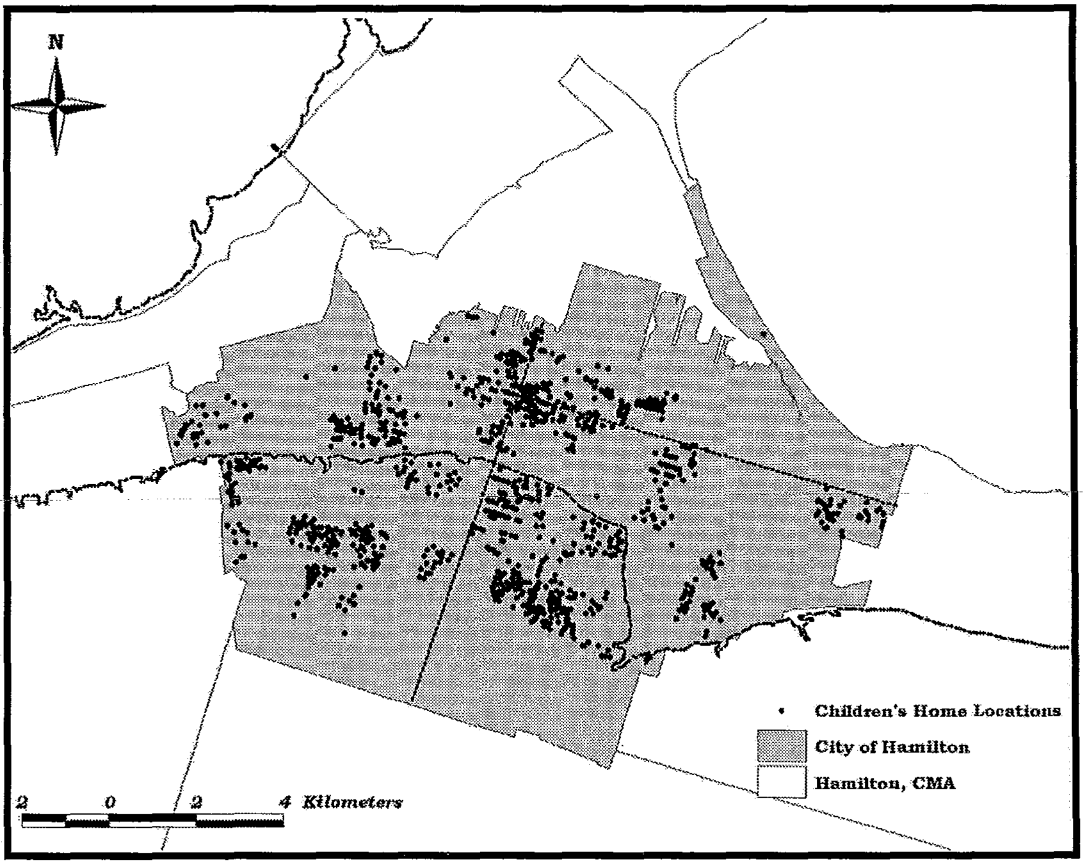

Figure 5.1: Location of children's homes in the City of Hamilton (n=1164)

Maps were also created to present the locations of schools. As we can see from Figure 5.2, schools are uniformly distributed all over the city of Hamilton based on the distribution of all schools located in the city.

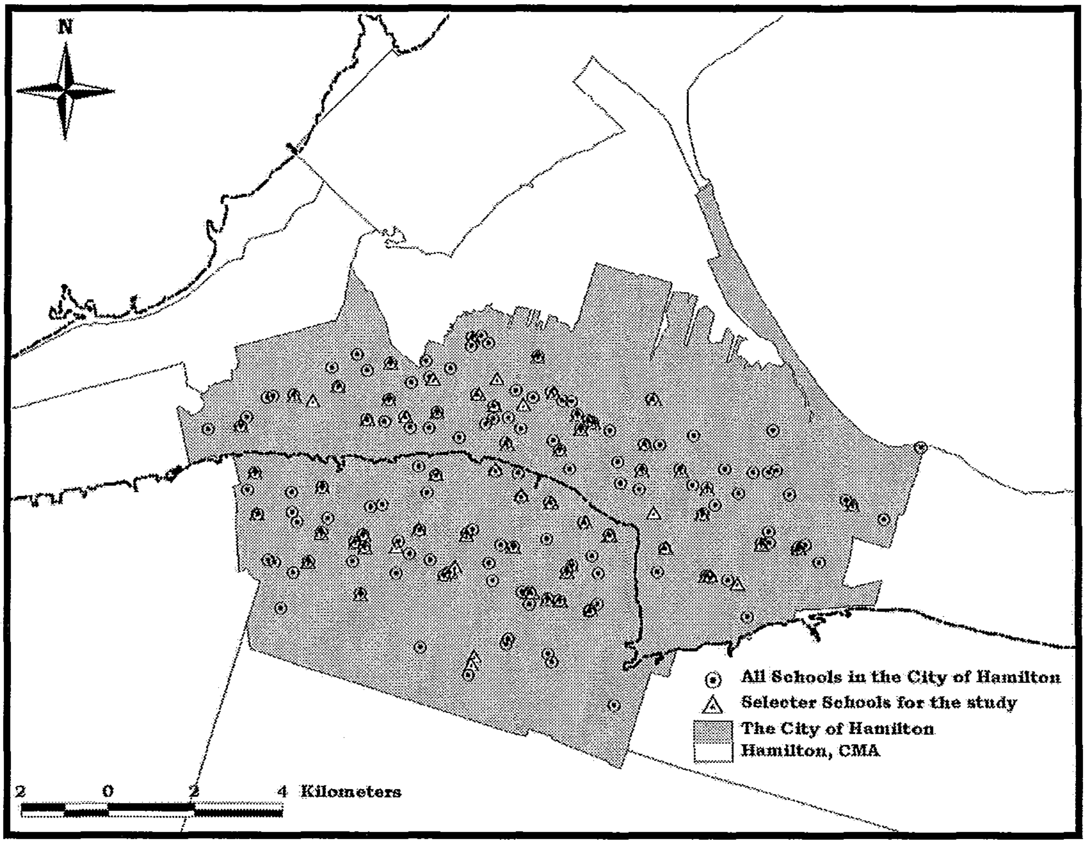

Figure 5.2: Location of schools selected in the City of Hamilton (n=69)

The prevalence rates for a number of health outcomes and covariates for 1980 are presented on Table 5.1.

Table 5.1: Prevalence rates for cohort health outcomes

Health Outcome	Yes	No	Prevalence rate
Wheezy	284	881	24.38
Chest Illness (pneumonia, bronchitis)	80	1085	6.87
Asthma	44	1121	3.78
Cold goes to chest	392	773	33.65

According to the above table, some interesting findings emerge. In particular, higher rates occur mostly for symptoms and not for actual health outcomes. For example, 33.65% of the children of the cohort -based on the answers of the questionnaire- when they have a cold, this usually goes to their chest. Also, 24.38% of the children's chests' sound wheezy. The highest prevalence rate for a respiratory disease incidence is that 6.87% of the children had a chest illness. The lowest rate appears for asthma attack incidences where only 3.78% of the children experience. Figures 5.3 and 5.4 present the spatial distribution of those children that when they have a cold; this usually goes to their chest and those children that had a chest illness such as pneumonia or bronchitis during the last year, respectively.

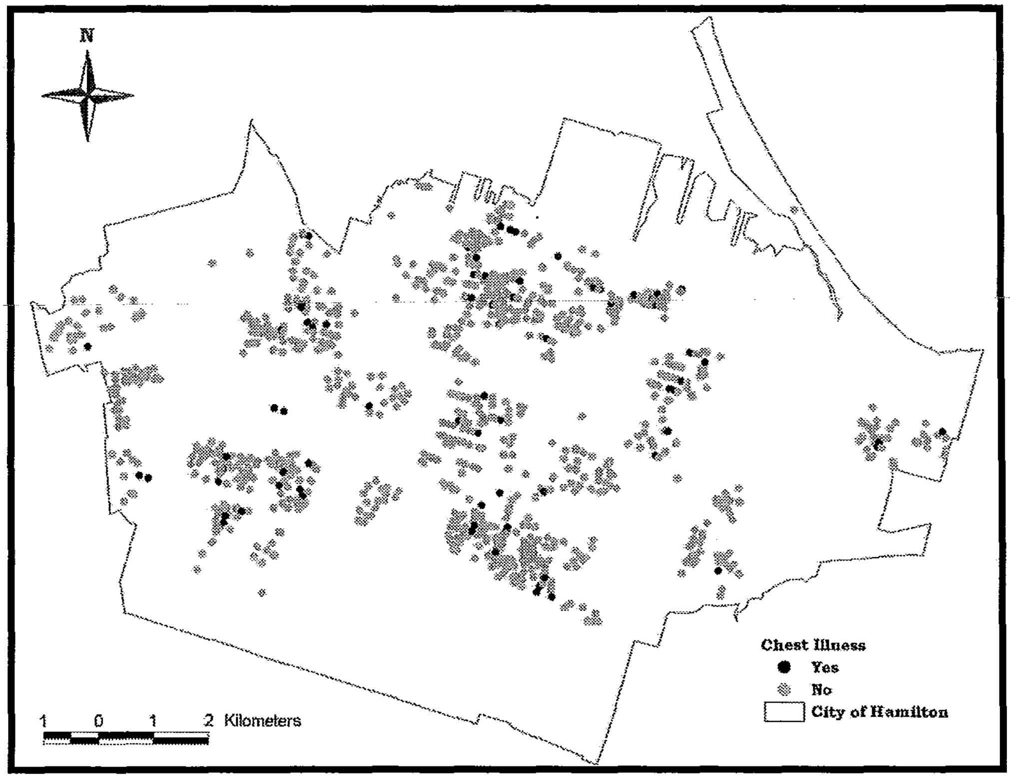

Figure 5.3: Spatial distribution of the children that had a chest illness in infancy

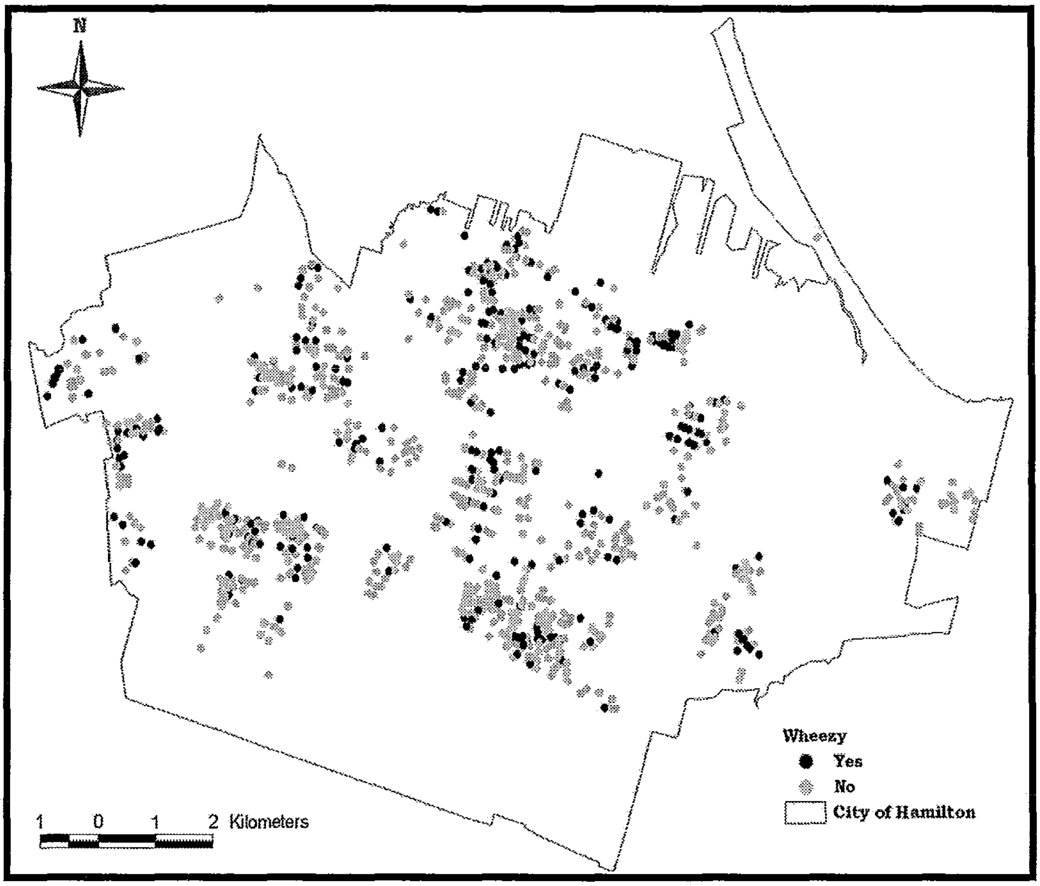

Figure 5.4: Spatial distribution of the children whose chest sounds wheezy

According to figure 5.3, the spatial distribution of the children who experienced a chest illness during the last year seems to be homogeneous over the City of Hamilton. However, based on Figure 5.4, this is not the case for the children whose chest sounds wheezy. Specifically, 128 of 284 (45%) of those children are located in close proximity to the industrial core area at the north-eastern part of the city of Hamilton.

In this sample, the industrial core area has the highest prevalence of gas stove cooking, more than two smokers in the house, and chest illness in siblings, hospitalization in infancy crowding and low income as well as the highest levels of particulate air pollution. Figures 5.5 and 5.6 indicatively demonstrate some of these characteristics.

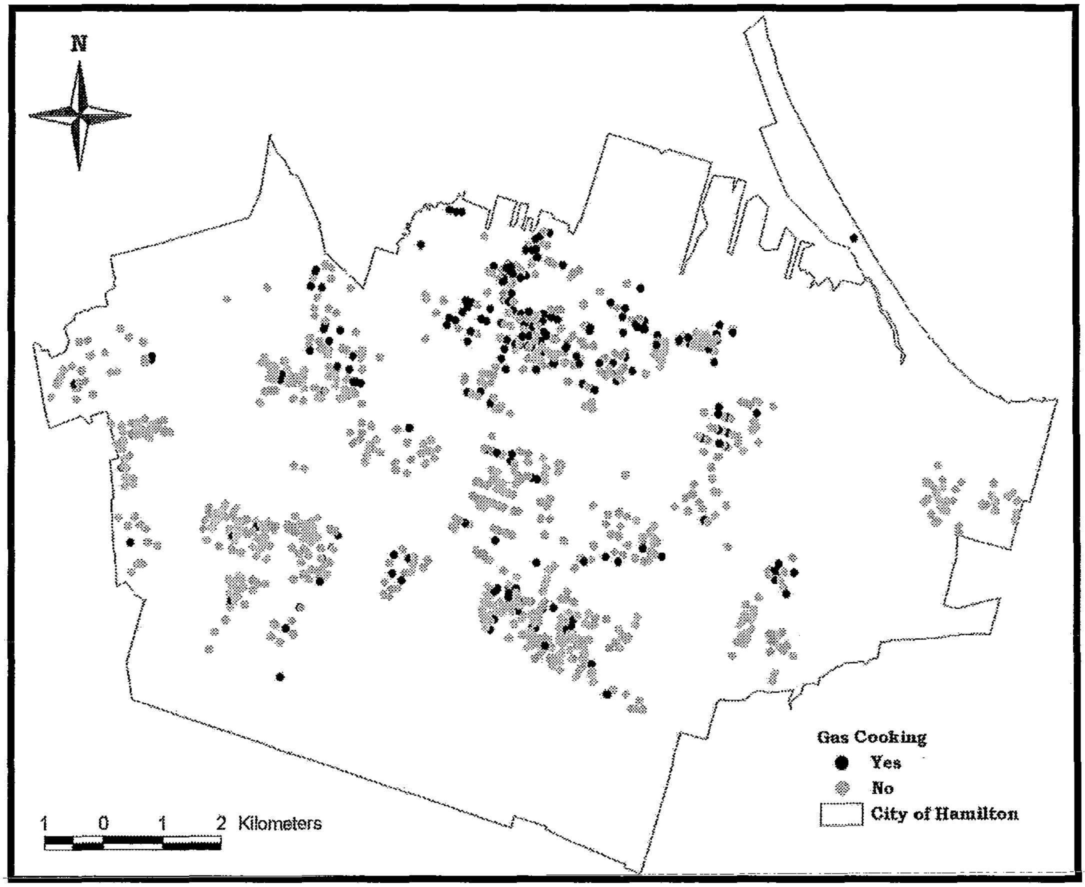

Figure 5.5: Prevalence of gas stove cooking across the city of Hamilton

Specifically, Figure 5.5 shows the prevalence of gas stove cooking across the city of Hamilton, while Figure 5.6 illustrates that if there are more than two smokers in the house, then most probably this household is located close to the industrial core area of Hamilton.

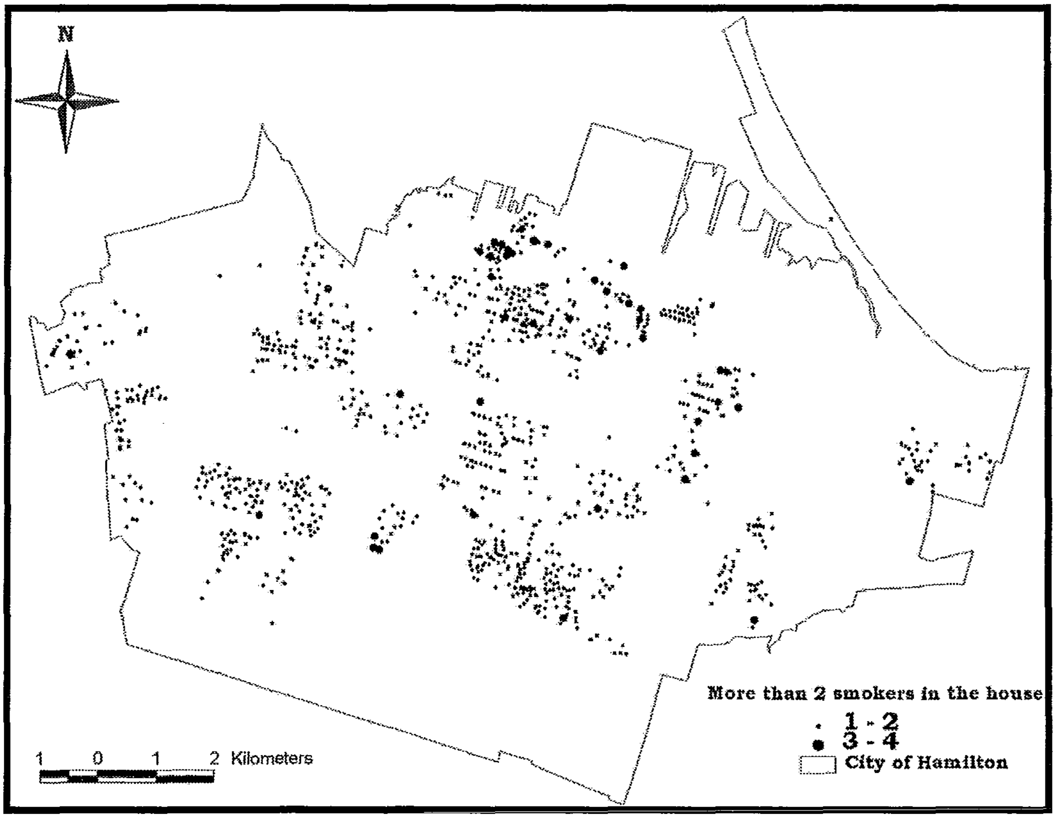

Figure 5.6: Prevalence of more than 2 smokers in the house across the city of Hamilton.

Finally, in order to examine the distribution of a number of covariates in the city of Hamilton, the city was divided in five neighbourhoods; the west upper (WU), the east upper (EU), the west lower (WL), the east lower (EL) and the industrial core (IC) areas. The division was based on the one applied for the HCC study (Kerigan et al., 1986). According to Table 5.2 there are common patterns in the distribution of the covariates. In particular, the industrial core are has the highest prevalence rates in this sample of hospitalization in infancy, more than two smokers in house, low income, use of gas stoves for cooking, crowding as well as the highest particulate pollution levels.

Table 5.2: Distribution of covariates (percentages - %) over the city of Hamilton

Variable	WU	EU	WL	EL	IC
Hospitalizations <2 years old	17	23	12	12	36
More than 2 smokers in House	12	18	11	21	38
Income (lowest income class)	16	11	21	16	36
Gas Cooking	10	17	19	12	42
Share Bedroom with 2 or more	21	22	15	17	26
More than 5 People live in the house	17	13	19	20	31
TSP (μg/m3)	43	43	61	60	98

5.3 Multivariate Analysis

Bivariate analysis was first applied and several variables had a significant relationship with children's respiratory health. Specifically, the variables with the highest association were the hospitalization before the age of 2 years, maternal smoking, chest illness in siblings and exposure to TSP based on LUR. The following step in the analysis was to combine the variables that were significant at the 0.05 level into a single model. Then, based on the type of the health variable either logistic or linear regression analysis was applied.

Multivariate Logistic Regression was employed for the categorical type of variables, which in this case were respiratory health symptoms and incidences. Also, for the variables based on the results of the pulmonary function tests (continuous variables) Multivariate Linear Regression analysis was employed. Models were run using a forward stepwise analysis starting with the most significant variable according to the bi-variate analysis.

5.3.1 Respiratory Symptoms

The results of the logistic regression of the respiratory symptoms "chest of child sounds wheezy" and "cold usually goes to chest" are presented in Table 5.3. The goodness of fit for logistic regression is represented by ?2, which is defined as one minus the ratio of maximized log likelihood values of the full model divided by the maximized log likelihood values of the constant-only-term model (Wrigley, 1985). ?2 usually ranges from zero to one; however, values between 0.2 and 0.4 represent a good fit of the model (McFadden, 1979).In this case, ?2 is 0.03 for the symptom "chest sounds wheezy", 0.227 for the "cold usually goes to chest" symptom and 0.29 for the symptom of "having a cough". In the first case, the value of ?2 is low, indicating a poor fit. However, for the other symptoms ?2 indicates a good fit of the model.

Table 5.3 presents the explanatory variables that were significant (p-value < 0.05). The Odds Ratio estimates and Confidence Limits associated with each variable are also reported in the table. The odds ratio is the factor that describes the likelihood of having the outcome variable (respiratory symptom or incidence) will change when the explanatory (independent) variables changes by one unit, or in the case of categorical variables, changes from one category to the other. The odds ratio is a measure of association in which a value of "1.0" means that there is no relationship between the variables. The value of an odds ratio can be less or greater than 1.0. The size of any relationship is measured by the difference (in either direction) from 1.0. An odds ratio less than 1.0 indicates an inverse or negative association. An odds ratio that is greater than 1.0 indicates a positive relation between the variables.

Table 5.3: Results of the Odds Ratio Estimates for the respiratory symptoms

Wheezy			
Effect	**Odds Ratio Estimate**	**95% Wald Confidence Limits**	
Chest Illness in siblings	1.638	1.059	2.534
Hospitalized < 2 years	2.597	1.498	4.501
Number of Smokers	1.01	1.003	1.016
TSP (LUR)	1.007	0.999	1.015

Log -Likelihood Constant only	**Log -Likelihood Constant and Covariates**	**?^{2}**
1248.467	1211.003	0.03

Cold usually goes to chest			
Effect	**Odds Ratio Estimate**	**95% Wald Confidence Limits**	
Hospitalized < 2 years	2.618	1.548	4.425
Mother is Smoking	1.307	1.01	1.693
TSP (LUR)	1.012	1.005	1.019

Log -Likelihood Constant only	**Log -Likelihood Constant and Covariates**	**?^{2}**
1485.644	1447.743	0.227

Cough			
Effect	**Odds Ratio Estimate**	**95% Wald Confidence Limits**	
Hospitalized < 2 years	3.427	3.256	3.692
Mother is Smoking	1.935	1.610	2.154
TSP (LUR)	1.564	1.345	1.798

Log -Likelihood Constant only	**Log -Likelihood Constant and Covariates**	**?^{2}**
1565.9567	1111.8293	0.29

Based on Table 5.3, children that were hospitalized in infancy had greater odds of experiencing the symptom of "chest sounds wheezy" compared to those that were not hospitalized. Also, a child was more likely to have the symptom if its siblings had a chest illness such as pneumonia or bronchitis. Finally, the association between children having the symptom and air pollution or smokers in the house -although significant (p-value<0.05) - is weak (1.01 and 1.007 respectively) and needs to be interpreted with caution.

For the symptom of "cold goes to chest" in children the odds were greater when the child was hospitalized before the age of 2 years (OR: 2.62) as well as when the mother was smoking (OR: 1.31). However, again the association between particulate air pollution and the occurrence of the symptom was weak (OR: 1.01). Specifically, the children were slightly more likely to experience the symptom if the TSP increases by $1\mu g/m^3$.

Finally, for the symptom of "having a cough" children that before the age of 2 years, their mother is smoking and they are exposed to higher levels of air pollution have a greater probability of experiencing this symptom. In this case the association with air pollution is the strongest among the symptoms tested (OR: 1.564).

5.3.2 Respiratory Diseases

The model of respiratory disease incidences (asthma or chest illness) is presented on Table 5.4. The ?^{2} is 0.13 for the asthma model and 0.08 for the chest illness model. In the both cases, the value of ?^{2} is low-especially for the second case- indicating that the fit of the model is not good and therefore the results need to be interpreted with caution.

According to table 5.4a children are **more likely** to have asthma attacks if they were hospitalized before the age of 2 years and their siblings have a chest illness. Also, children are **less likely** to have an asthma attack if their mother smokes, they use gas for cooking, and there are more people in the household than the number of bedrooms (crowding index). Finally, the association between air pollution and asthma was weak (OR: 0.98).

Table 5.4a: Results of the Odds Ratio Estimates

	Asthma (n=44)		
Effect	**Odds Ratio Estimate**	**Confidence Limits (95%)**	
Chest Illness in Siblings	5.271	2.461	11.291
Hospitalized < 2 years	7.723	3.082	19.348
Mother is Smoking	0.430	0.201	0.922
Gas Cooking	0.226	0.052	0.990
Crowding Index[1]	0.288	0.089	0.926
TSP (LUR)	0.976	0.953	0.999

Log -Likelihood (Constant)	**Log -Likelihood (Const. and Covar.)**	**?2**
371.814	323.737	0.129

[1] Crowding index is defined as the number of people in the household divided by the number of bedrooms

Table 5.4b: Logistic Regression Coefficient Estimates for asthma

Variable	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	-0.2692	0.9713	0.0768	0.7817
Chest Illness Siblings	1.6622	0.3887	18.2875	<.0001
Hospitalized < 2	2.0442	0.4686	19.0299	<.0001
Mother is Smoking	-0.8429	0.3887	4.7011	0.0301
Gas Cooking	-1.4868	0.7533	3.8951	0.0484
Crowding Index	-1.2460	0.5964	4.3653	0.0367
TSP (LUR)	-0.0242	0.0120	4.0426	0.0444

Table 5.4b presents the coefficient estimates from logistic regression. Based on this table not all the coefficients have the expected signs. For example, children with asthma are significantly and positively associated with having siblings with a chest illness as well as being hospitalized as an infant. However, the probability of a child having asthma was negatively associated with maternal smoking, use of gas stoves for cooking, TSP pollution, and the number of people in the household compared with the number of bedrooms. The opposite (positive sign was expected) could be due to the fact that most of the children who have asthma were located at the southern part of the city (mountain area) and not in close proximity with the industrial core area. A possible scenario for why this is happening could be that families with asthmatic children would most probably seek cleaner communities to live just because of the vulnerability of their children. Therefore, the mountain area (south part of the city) could be a candidate place to live due to improved air

quality compared to the industrial core and downtown areas (north part of the city). However, more data and research are needed in order to justify this scenario.

Table 5.5a presents the odds ratios for children who have a chest illness. Children are **more likely** to have a chest illness if they have siblings with a chest illness as well as if the household smoking rate and the total number of rooms in the household increase by one unit. For TSP, the association again is weak. In general, due to the small value of ?2 the results need to be interpreted with caution.

Table 5.5a: Odds Ratio Estimates for Children with Chest Illness

	Chest Illness (pneumonia, bronchitis)		
Effect	**Odds Ratio Estimate**	**Confidence Limits (95%)**	
Chest Illness in Siblings	5.254	3.032	9.106
Total Rooms	1.197	1.051	1.363
Household Smoking Rate[2]	4.826	1.110	20.971
TSP (LUR)	1.012	1.000	1.024

Log -Likelihood (Constant)	**Log -Likelihood (Const. and Cov.)**	**?2**
562.299	519.719	0.076

Table 5.5b presents the coefficient estimates from the logistic regression analysis. In this case, all the coefficients have the expected signs and magnitude.

[2] The Household smoking Rate is defined as the total number of smokers in the household divided by the number of adults in the household.

Table 5.5b: Logistic Regression Coefficient Estimates for chest Illness

Variable	Estimate	Std. Error	Wald-X^2	Pr > ChiSq
Intercept	-5.4126	0.708	58.4434	<.0001
Chest Illness in Siblings	1.6525	0.2807	34.6587	<.0001
Total Rooms	0.1805	0.0662	7.4301	0.0064
Household Smoking Rate[3]	1.5954	0.7504	4.5198	0.0335
TSP (LUR)	0.013	0.00624	4.3137	0.0378

5.3.3 Pulmonary Function

The results of the linear regression analysis for a number of pulmonary function tests are presented below. Specifically, based on Tables 5.6 to 5.11 several variables yielded a significant relationship with a number of pulmonary function tests. For example, the variable "having a chest illness before the age of two years" was negatively associated with FEV1 and FVC reflecting the impacts of early childhood experiences in the development of the children. Those children that had a chest illness in infancy had poorer pulmonary performance compared to those that did not experience that.

[3] The Household smoking Rate is defined as the total number of smokers in the household divided by the number of adults in the household.

Table 5.6: Linear Regression Results for FEV1

FEV1				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	2.0814	0.0405	51.34	<.0001
Chest Illness <2	-0.0817	0.0399	2.05	0.0409
Sex (girl = 1)	0.2099	0.0239	8.79	<.0001
TSP (LUR)	0.0024	0.00066039	3.61	0.0003
R-Square	0.0806	**Adj R-Square**	0.0782	

Table 5.7: Linear Regression Results for FVC

FVC				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	2.18301	0.06345	34.40	<.0001
Chest Illness <2	-0.08008	0.03995	2.00	0.0452
Sex (girl = 1)	0.20492	0.02397	8.55	<.0001
Mother is smoking	-0.01710	0.02515	-0.68	0.4966
People in the household	-0.02278	0.01102	-2.07	0.0390
TSP (LUR)	0.00262	0.00068026	3.86	0.0001
R-Square	0.0843	**Adj R-Square**	0.0803	

Hospitalization in infancy was negatively associated with the ratio of Forced Expiratory Volume in one second to Forced Vital Capacity (FEV1/FVC)[4] as well as the Peak expiratory Flow (PF).

To control for income, a binary variable was created categorizing the children to those whose family income was less than CN$15,000 and those whose family income was more than CN$15,000. The value of CN$15,000 was used as a proxy for the low-income cut off in 1980 for a family with 3 members living in a city of 300,000 people (Statistics Canada, 2001). This variable was negatively associated with FEV1/FVC and PF values. However, the sign of the variable for the PF test is opposite of the one expected. A potential explanation for that could be a potential correlation between the low income and the use of gas stoves variables. The maps created in section 5.2 support the argument of a potential association between low family income, crowding in the house, use of gas stoves and increased number of smokers.

Also, a crowding index variable was created (definition mentioned above) to control for potential health impacts of household volume. However, e included the number of people in a household variable because according to the results of bivariate regression analysis, the association between respiratory health of children and the number of people in a household was stronger, compared to the association of respiratory health and the crowding index. For the pulmonary function models, the number of people in the household was

[4] The ratio was used because it reflects better potential airway obstruction compared to the two tests separately.

negatively associated with FVC indicating a reduction in FVC values with an increase in the number of people in the household.

Table 5.8: Linear Regression Results for FEV1/FVC

FEV1/FVC				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	0.82424	0.00305	304.3769	73178.3
Hospitalized <2	-0.01937	0.00859	0.0212	5.09
Sex (1=girl)	0.02735	0.00382	0.2137	51.37
Gas for cooking	-0.01193	0.00496	0.0241	5.79
Low Income	-0.01381	0.00483	0.0339	8.16
R-Square	0.0576	**Adj R-Square**	0.0543	

Table 5.9: Linear Regression Results for PF

PF				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	3.65689	0.07535	1454.71639	2355.32
Hospitalized <2	-0.12112	0.04668	4.15863	6.73
Sex (girl = 1)	0.15227	0.06840	3.06094	4.96
Gas for cooking	-0.12724	0.04601	4.72259	7.65
Low Income	0.17139	0.04658	8.36380	13.54
R-Square	0.0282	**Adj R-Square**	0.0249	

Maternal smoking proved to be significantly associated with a decrease of the FVC, the Mean Forced Expiratory Flow between 25% and 75% points of FVC (FEF2575) as well as the Maximum Expiratory Flow after 50% of FVC has been exhaled (MEF75) values. Also, the use of gas stoves for cooking resulted in reducing the ratio of FEV1/FVC as well as the value of PF, reflecting potential airway obstruction.

Particulate air pollution (TSP) was significantly associated with FEV1, FVC, MEF75 and FEF2575. This variable was the second variable that took the opposite sign than the one expected. In particular, it was expected that increased levels of particulate air pollution (TSP) would be associated with decreased values of the pulmonary function tests associated. However, in this case positive sign indicates that pulmonary functioning is reduced at low levels as the TSP pollution. Yet, as Pengelly et al. (1986) suggested this could be interpreted in another way; particularly it could be said that FEV1, FVC and MEF75 are reduced with a reduction of particle size suggesting a "fine particle effect". In the case of the HCC study this type of association (positive) was found between FEV1 and the Mass Median Diameter[5].

Finally, another variable that showed a significant association with pulmonary functioning is sex of the child (FEV1, FVC, FEV1/FVC, PF, MEF75 and FEF2575). Specifically, in all cases, girls are positively associated with

[5] MMD (μm) was estimated in order to summarize the results of the measurements obtained by the Particle Size Distribution (PSD) Network. As we mentioned in chapter four we only used the TSP data to assign children's exposure to air pollution due to limited number of monitoring stations for the PSD

measures of health performance indicating better pulmonary functioning compared to boys of the same age.

Table 5.10: Linear Regression Results for MEF50

MEF75				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	0.94204	0.03388	90.21792	773.18
Mother is smoking	-0.06584	0.02142	1.10214	9.45
TSP (LUR)	0.00197	0.00059139	1.30124	11.15
Sex (girls = 1)	0.94204	0.03388	90.21792	773.18
R-Square	0.0234	**Adj R-Square**	0.0209	

Table 5.11: Linear Regression Results for FEF2575

FEF2575				
Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1.81896	0.07329	186.71094	615.98
Mother is smoking	0.12697	0.04924	2.01558	6.65
TSP (LUR)	-0.09948	0.03494	2.45683	8.11
Sex (girls = 1)	0.00307	0.00096726	3.04558	10.05
R-Square	0.0199	**Adj R-Square**	0.0173	

The overall performance of the models in terms of the R^2 values was poor. The highest Adjusted R^2 obtained was 0.0803 (FVC) indicating that 8% of the variation of the FVC values is explained by the variation of the independent variables included in the model. The lowest Adjusted R^2 value was 0.017(FEF2575) reflecting that only 1.7% of the variation of FEF2575 values is explained by the variation of the independent variables. Nevertheless, a potential explanation for the small values of R^2 could be the large sample size (n=1164); the bigger the sample size the larger the variation of the variables of interest.

5.4 Discussion

First, in terms of the application of a population health perspective, the results of the models clearly demonstrated that children's respiratory health is significantly associated with family income as well as the physical environment (both natural and built), biology and genetic endowments, and health child development.

Comparing the results of the original HCC study conducted in the early 80's and the re-analysis of the HCC conducted through this thesis, substantive issues arise (Table 5.12). According to the Table 5.12, the re-analysis of the HCC study resulted in strengthening the association of children's respiratory health with a number of individual, socio-economic, and environmental influences and behaviours.

Table 5.12: Comparing the results of the two efforts in analysing the HHC data

Explanatory Variables	HCC study in 1980	Re-analysis of HCC in 2005
Hospitalization <2 years	Cough	Cough, Wheezy, cold to chest
		Asthma
	MEF75	FEV1/FVC, PF
Chest Illness <2 years	Cough	Cough, cold goes to chest
		FEV1, FVC
Chest Illness in siblings		Asthma, Chest Illness
		Wheezy
		Cough, Wheezy, cold to chest
Maternal smoking		Asthma
	MEF75	FVC, MEF75, MEF2575
Household smoking rate		
		Chest Illness
Number of smokers in the house		Wheezy
TSP (MMD for HCC-1980 study)		
		Asthma, Chest Illness
	MEF75, PF	FEV1, FVC, MEF75, MEF2575
Sex		
		FEV1, FVC, FEV1/FVC, PF, MEF75 MEF2575
Gas for cooking		
		Asthma
		FEV1/FVC, PF
Low income		
		FEV1/FVC, PF
Number of people in the house		
		FVC
Number of rooms in the house		
		Chest Illness
Crowding Index		
		Asthma
Child shares the bedroom	PF	

Specifically, while the HCC study in 1980 found that respiratory health of children is associated with hospitalization in infancy, maternal smoking, and exposure to particulate air pollution, the re-analysis of the study verified these associations and also found links with chest illness in siblings, use of gas for cooking, low income and household crowding.

Also, the respiratory measures associated with the above mentioned risk factors for the HCC in 1980 study are less than those found to be significant in the re-analysis study.

Yet, one of the major issues in this study regarding the association between air pollution and children's respiratory health is that a direct exposure to air pollution - outcome relationship is not fully supported -especially in light of the results from the pulmonary function models- resulting in problems in the interpretation of the models. In particular, in both multivariate logistic and linear regression for symptoms, disease incidences and pulmonary function, the results from the models indicated a potential relationship between air pollution and respiratory health. However, the sign (positive or negative) of this relationship was not confirmed challenging at the end the nature of the association itself. To be more specific, based on the results of the logistic regression analysis particulate air pollution was (although weak) positively associated with a number of respiratory symptoms and diseases indicating that the presence of the symptoms and/or health outcomes was increasing with higher levels of air pollution. On the other hand, linear regression analysis also revealed a positive association between air pollution and pulmonary functioning. However, in this case the positive sign indicated that pulmonary functioning is reduced at low levels as the TSP pollution.

This inconsistency in the results also emerged at the HCC study raising important questions regarding the interpretation of findings. Specifically, the symptom "cold goes to chest" was associated with TSP concentration above the median (50μg/m3). Similarly positive associations were found between FEV1 and MEF75 and the Mass Median Diameter (MMD) suggesting that a reduction in pulmonary function is associated with a reduction in the diameter of particles. In addition, according to Pengelly et al., (1986) there is a positive association between TSP and MMD. Based on the above, the results of the pulmonary function were interpreted by Pengelly et al., (1986) in another way; particularly the results may indicate that FEV1, FVC and MEF75 are reduced with a reduction of particle size since TSP is positive associated with MMD.

Comparing the results with the respiratory health literature suggests that this relationship is rather complicated. For example, a number of studies (Gauderman et al., 2000, 2002, 2004; Stern et al., 1994) have demonstrated a potential association between air pollution and children's respiratory health. On the other hand, there are studies that did not find any significant association between air pollution exposure and children's respiratory health (Moseler et al., 1994). Therefore, in order to improve our understanding on disease pathogenesis more research is required in investigating the determinants of children's respiratory health.

5.5 Summary

This chapter has addressed the spatial distribution of various respiratory health measures and their potential covariates as well as the potential determinants of children's respiratory health emphasizing more on particulate air pollution. Based on the results, for respiratory symptoms and disease incidence in all cases the air pollution variable was significantly positively associated with the health outcome of interest. However, the association was weak. Also, the general fit of the models was poor supporting the need to interpret the results with caution. For pulmonary function tests, the association with air pollution was statistically significant only for FEV1, FVC, MEF75 and FEF2575. Nonetheless, the sign of the air pollution coefficient was opposite to the one expected, raising concerns regarding the real relationship between air pollution and children's health. An explanation given was that the positive association might be a result of a fine particle effect. The implications of this as well as a summary of the general findings of the thesis along with potential contributions and directions for further research are presented in the following chapter.

CHAPTER SIX

CONCLUSIONS

6.1 Introduction

This study re-analysed the HCC data in order to investigate the relative importance of air pollution exposure along with a number of individual, social and environmental characteristics in determining children's respiratory health. The following sections summarize the key findings with respect to each of the study objectives. Following the summary, the contributions of this research are discussed and directions for future research are suggested.

6.2 Summary of Results

This thesis presented the results of the re-analysis of the HCC study taking into consideration the spatial dimension in assessing children's exposure to particulate air pollution.

Two objectives have been addressed:

1. Investigation of the spatial variation of adverse respiratory health of children in Hamilton,
2. Investigation of the relative importance of individual, social and environmental characteristics that may determine the burden of respiratory incidences by re–assessing the exposure to air pollution using advanced methods of spatial analysis and geographic information systems (GIS).

The sample was defined based on the original cohort sample (the HCC study) and the following criteria:

1. Children must have participated in all years of the study, (1979 – 1985) and
2. The location of their home should be identified by a unique postal code
3. Also, if the children were added during the second year of the study (782) then selection was only carried out for those that:
    a. Participated also in the subsequent years, and
    b. Could be identified by a unique postal code based on the location of their home.

The final number of children included in this study is 1164. With respect to objective one, visualization and exploratory analysis was performed. First, the sample of children was geocoded based on the postal code information from the Canadian Census Analyser of the Computing in the Humanities and Social Sciences from the University of Toronto (Canadian Census Analyser, 1996). Mapping the sample as well as a number of respiratory health outcomes and covariates (section 5.2) revealed a city with diverse and unique population in terms of its socio-economic and environmental characteristics. In particular, the maps that were created verified the results of HCC study (Pengelly et al., 1984) and illustrated that the prevalence of a plethora of covariates, such as socio-economic status and air pollution, was not the same in different areas of the city of Hamilton. For example, the industrial core area appeared to have the highest levels of particulate air pollution as well as the highest prevalence of factors, such as gas use for cooking and heating, chest illness in siblings, crowding in the house and low family income. According to the literature many of those factors have been associated with adverse respiratory health (Hruba et al., 2001; Jerrett et al., 2003; Birch et al., 2000; Finkelstein et al., 2003). On the contrary, the southern part of the city was generally characterized by lower prevalence trends of these variables as well as by better air quality. Jerrett et al. (2001) also showed that Hamilton is a city with distinct patterns. Specifically, dwelling values were negatively associated with pollution exposure, indicating that those with higher socio-economic status are less likely to be exposed to high levels of particulate air pollution. In general, all of the above mentioned associations confirm the influence of the broader determinants of health as well as the complex interactions between them, recognizing the importance of the population health perspective. In addition, the use of GIS, in describing and presenting the data, provides an innovative approach and generally improves our understanding in linking health outcomes to their potential determinants.

Before moving on to investigate the potential determinants of children's respiratory health, the assessment of children's exposure to particulate air pollution was necessary (chapter four). The HCC study applied trend surface analysis in order to assess exposure to ambient particulate air pollution, controlling in this way only for the first order effects. In particular, the HCC study was one of the first that developed an air pollution network to effectively measure air pollution in order to assess adverse health effects, controlling at the same time for a number of covariates. However, as discussed, during the time of the study the application of advanced spatial analysis and GIS techniques that could account for local spatial variations of the pollutant (second-order effects) was limited if not absent in health-related studies. Therefore, the re-assessment of exposure to particulate air pollution was one of the major objectives in this thesis. Evidence from a growing body of literature has shown that exposure assessment studies have been quick to adopt advances in geographic information systems (GIS) and spatial analysis (Jerrett et al., 2004; Briggs et al., 1997). In this case, the application of advanced spatial analysis and GIS methods was employed in order to assess children's exposure to particulate air pollution. The comparison of two different techniques (kriging and land-use regression) offers new evidence with regard to the performance of those techniques.

For kriging, exposure assessment was based on a number of steps as described in section 4.3. Both ordinary kriging and universal kriging based on the TSP observations were applied. For the universal kriging, the data were de-trended based on a quadratic trend surface. Taking into consideration the errors resulting from both types of kriging (section 4.5), universal kriging was better at describing the spatial variation of TSP concentrations. The application of ordinary or universal kriging depends on the nature of the data and in this case due to a global trend that was present universal kriging performed better than ordinary kriging. Jerrett el al. (2001) also applied universal kriging to interpolate TSP from twenty–three monitoring stations in Hamilton for 1985 to 1994. The results of the Jerrett et al., (2001) study agree with the results of this thesis showing that higher concentrations of TSP are closer to the industrial core area.

A disadvantage of kriging is related to the availability of monitoring data. Geostatistical modeling in general, requires a dense network of sampling sites. The number of sites for an urban area typically ranges between 30 and 100 depending on the scale of analysis, variability of the pollutant, local emissions, topography of the study area and meteorological conditions (Jerrett et al., 2001). Although the HCC study was the first to establish a monitoring network to measure air pollution, limited variability of TSP, due to the limited number of monitoring stations, prohibited the estimation of an appropriate pollution surface with the application of kriging.

For LUR, the selection of the variables was based on the literature. Specifically, most of the studies have taken into consideration three types of variables: a) land use information, b) elevation and c) road/highway network (Briggs et al., 1997; Briggs et al., 2000; Jerrett et al., 2003). Due to data limitations, other variables such as traffic volumes[1] (Briggs et al., 1997) and meteorological data (Jerrett et al., 2003) were not incorporated into the model. However, we assumed that the exclusion of traffic-related pollution and meteorological variables do not influence the results of the LUR model, because TSP is not usually considered as a major by-product of traffic except for trucks. Regarding the meteorological data, the direction of prevailing winds was known for the area and was taken into consideration in the analysis and interpretation of the results. GIS was used to construct most of the variables incorporated in the final LUR model. The results of the model suggest that compared to kriging,

[1] Due to the pollutant of interest (TSP) traffic volume data could emphasize more on trucks - compared to cars-as this is the major source of TSP.

it generally provides a better estimation of TSP concentrations over the city of Hamilton. Also, it is worthwhile noting that, to our knowledge, this is the first study in Canada that utilizes this method to model the variability of particulate air pollution (TSP). To date, land-use regression has been applied to model the variability of nitrogen dioxide and in general gaseous pollutants; not particulate air pollution (Jerrett et al., 2003).

To address the second objective of this thesis and specifically the potential factors that impact children's respiratory health, multivariate logistic and linear regression analysis was applied to the data. The selection of the variables was based on the literature and specifically, the broader determinants of health according to the population health perspective (Table 6.1). The population of interest in this study are children 8-12 years of age. Therefore, the determinants of employment and working conditions and personal health practices and coping skills were based on the parents' status. Also, the determinants of social networks, health services and social environments/cohesion, were not taken into consideration due to data limitations.

Table 6.1: Variables included in the analysis and their relation to the broader determinants of health

Determinants of Health	**Variables included in the model**
Income and Social Status	Family Income, type of dwelling, number of rooms in the house, sharing of rooms
Social Support Networks	
Education	Children were at the school age
Employment and Working Conditions	Status of the parents
Physical Environments	TSP, parental smoking, gas cooking and heating
Biology and Genetic Endowments	Age, height, medical history of siblings
Personal Health Practices and Coping Skills	Parental smoking
Health Child Development	Diseases or hospitalization in infancy
Health Services	
Social Environments/Social Cohesion	

The results of the analysis show that hospitalization before the age of two years of age, low family income, household crowding, sex, parental smoking and particulate air pollution (TSP) are significantly associated with children's respiratory health. This thesis agrees with the results of Gauderman et al., (2000) study who also found that decreased pulmonary functioning is greater for the children spending more time outdoors, living in houses with gas-stoves and whose parents were smoking.

With regard to the potential association between air pollution and respiratory health, the results of the models for respiratory symptoms and pulmonary function are not consistent. Specifically, there is a positive but weak association between air pollution and prevalence of symptoms (i.e. chest sounds wheezy) or disease incidences (pneumonia and bronchitis). However, in the case of pulmonary functioning, the association is also positive indicating that children exposed to higher levels of air pollution have their lungs functioning better than those exposed to lower levels of air pollution. Yet, this association could be interpreted in a way that is consistent with the findings for the respiratory symptoms and disease incidences. In particular, using TSP to capture the potential impacts on human health is complicated because TSP would not be able to capture local area variations. That is because the main source of TSP is industrial activity.

This inconsistency in the results also emerged at the HCC study raising important questions regarding the interpretation of findings. Similarly positive associations were found between FEV1 and MEF75 and the Mass Median Diameter (MMD) suggesting that a reduction in pulmonary function is associated with a reduction in the diameter of particles. In addition, according to Pengelly et al., (1986) there was also positive association between TSP and MMD, which interpreted by the researchers as an indication of FEV1, FVC and MEF75 decreasing with a reduction of particle size.

Comparing the results of the analysis with the respiratory health literature suggests that this relationship is rather complicated. For example, a number of studies (Gauderman et al., 2000, 2002, 2004; Stern et al., 1994) have demonstrated a potential association between air pollution and children's respiratory health. On the other hand, there are studies that did not find any significant association (Moseler et al., 1994) therefore supporting the need for further research on this issue.

This study supports and strengthens the results of the HCC indicating an association between indoor and outdoor air pollution and children's respiratory health. In addition, this research provided better estimates of exposure as well as evidence of additional risk factors (Table 5.12). Yet, one of the future challenges for environmental health research is that of a better understanding of the factors that determine the distribution of environmental exposures, the reasons for increased susceptibility to environmental hazards among certain populations and the improvements to health that result from improvements to the environment.

6.3 Research Contributions

In addressing the objectives outlined in this study, this research has provided new information on children's respiratory health and the application of spatial methods in understanding this health issue. This thesis demonstrates the usefulness of spatial analysis and GIS in assessing children's exposure to air pollution, provides a comprehensive comparison between the two exposure models and also strengthens the HCC study findings revealing that in addition to their findings, children's respiratory health is associated with chest illness in siblings, use of gas for cooking, low income and household crowding.

The methodological contributions of this thesis relate to the use of spatial analysis and GIS methods. The utilization of these methodological approaches is now well established (Gatrell and Senior, 1999). Yet, few of the studies on children's respiratory health have employed these tools. In particular, the application of GIS and spatial analysis to estimate exposure to air pollution has been adopted by a number of health-related studies (Jerrett et al., 2003; Cakmak et al., 2003; English et al., 1999; Wilkinson et al., 1999). However, this study takes this one step further by exploring the relationship between air pollution and children's health embodying at the same time the spatial relationships and interactions between children and the environment as well as controlling for a number of socio-economic, demographic and environmental covariates. Another methodological contribution is that this study is the first one in Canada that utilizes LUR to model the variability of particulate air pollution (TSP) in order to assess exposure to air pollution.

Finally, this study by re-analyzing the HCC showed significant associations between air pollution and children's respiratory health but at the same time raised important questions/concerns regarding the sufficiency of evidences to demonstrate this cause - effect association supporting the need for further research on the determinants of respiratory health as well as the application of other approaches of spatial analysis and GIS along with the classical epidemiological study designs.

6.4 Future Research Directions

As outlined above, while this research has made several contributions, the research agenda in this area remains a challenging one. Three key areas for future research come out of the thesis findings. The first is related to exposure assessment. The emerging considerations outlined in the previous chapters highlight the substantial work being carried out in refining the existing air pollution exposure assessment techniques, which do not preclude the development of new and better ones. Specifically, in most of the studies conducted so far, human exposure to air pollution is considered to be a static process. Recently, the application of spatial analysis has seen a rapid development. However, in addition to the energy devoted in extending the scope of the exposure assessment to encompass important environmental, economic and land-use related covariates, a great deal of work is being given to a stronger and more dynamic framework. Exposure is now considered to be a dynamic process that occurs across space and over time. Therefore, at the heart of all these endeavours is the desire to apply spatio-temporal analytic techniques (Christakos et al., 1997), which may also shed light on the exposure assessment arena.

Second, another issue of concern that was also one of the obstacles for this study is the most appropriate number of monitoring stations. Kanaroglou et al., (2005) applied a location-allocation approach for sitting of monitoring stations; however, more research is required in order to determine the number of sites required for exposure estimation.

Finally, another area of further investigation is the temporal dimension of this research. Specifically, health-related studies that account for the long-term health effects of air pollution have rarely been the objective of research mainly because of the associated high financial costs and time requirements. However, the findings of recent studies (Schwartz, 1993; Abbey et al., 1995; Hruba et al., 2001; Brunekreef and Holgate, 2002) illustrate the importance of focusing more on the long-term health effects of exposure to ambient air pollution in order to gain important and policy-relevant knowledge on the link between air pollution and health. Therefore, one avenue for future research is a follow-up research study in order to investigate childhood chronic exposure to adverse air quality and health over the long-term. Elliott et al., (2002), already attempted to reconstruct the cohort of children using publicly available data search engines. According to the results of this effort, Elliott et al., (2002) have successfully located the parents of almost half the Hamilton's cohort of children.

In conclusion, while this research has contributed to a better understanding of the relationship between air pollution and children's respiratory health within the population health framework, it can only be useful to the society as a whole if it can also support policy making and analysis. This would not be an easy journey but will be a worthy endeavour and useful in improving the policy process.

Bibliography

Agius, R. (2001). *What is Environmental Health?,* Avaliable: http://agius.com/hew/resource/envhlth.htm *[Accessed on 18/04/2005].*

Ali, M., Emch, M., Tofail, F. and Baqui, A. H. (2001). Implications of Health care provision on acute lower respiratory infection mortality in Bangladeshi children *Social Science & Medicine,* **52,** 267 - 277.

Atkinson, R., Anderson, R., Sunyer, J., Ayres, J., Baccini, M., Vonk, J., Boumghar, A., Forastiere, F., Forsberg, B., Touloumi, G., Schwartz, J. and Katsouyanni, K. (2001). Acute Effects of Particulate Air Pollution on Respiratory Admissions: Results from APHEA 2 project *Am J Resp Crit Care Med,* **164,** 1860-1866.

Ayres, J. G. (2002). Chronic effects of air pollution *Occupational and Environmental Medicine,* **59,** 147 - 148.

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis,* Addison Wesley Longman Limited, Malaysia.

Beggs, P. (1999). *Spatial Analysis of Dwelling Crowding and Disease in Sydney, Australia, The 11th Annual Colloquium of the Spatial Information Research Centre,* University of Otago, Dunedin, New Zealand.

Beyea, J. and Hatch, M. (1999). Geographic Exposure Modeling: A Valuable Extension of Geographic Information Systems for Use in Environmental Epidemiology *Environmental Health Perspectives,* **107,** 181 - 190.

Birch, S., Jerrett, M., Wilson, K., Law, M., Elliott, S. and Eyles, J. (2005). Heterogeneities in the production of health: smoking, health status and place *HEALTH POLICY,* **72,** 301-310.

Boezen, H. M., Van der Zee, S., Postma, D. S., Vonk, J. M., Gerritsen, J., G., H., B., B., Rijcken, B. and Schouten, J. P. (1999). Effects of ambient air pollution on upper and lower respiratory symptoms and peak expiratory flow in children *Lancet,* **353,** 874 - 78.

Brauer, M., G., H., van Vliet, P. H. N., Meliefste, K., P., F. and Gehring, U. (2003). Estimating Long-Term Average Particulate Air Pollution Concentrations: Application of Traffic Indicators and Geographic Information Systems *Epidemiology,* **14,** 228 - 239.

Braun-Fahrlaner, C., Vuille, J. C., Sennhauser, H., Neu, U., Kunzle, T., Grize, L., Gassner, M., Minder, C., Schindler, C., Varonier, S., Wuthrich, B. and

Casrpol, T. (1997). Respiratory Health and Long-term Exposure to Air Pollutants in Swiss Schoolchildren *American Journal of Respiratory Critical Care Medicine,* **155,** 1042 - 1049.

Briggs, D., Collins, S., Elliott, P., P., F., Kingham, S., Lebret, E., Pryl, K., van Reeuwijk, H., Smallbone, K. and Van der Veen, A. (1997). Mapping urban air pollution using GIS: a regression - based approach *Int. J. Geographical Information Science,* **11,** 699 - 718.

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

Brown, P., Guy, M. and Broad, J. (2005). Individual socio-economic status, community socio-economic status and stroke in New Zealand: A case control study *Social Science & Medicine,* **61,** 1174 - 1188.

Brunekreef, B. and Holgate, S. T. (2002). Air pollution and Health *The Lancet,* **360,** 1233-1242.

Brunekreef, B., Janssen, N. and DeHartog, J. (1997). Air pollution from truck traffic and lung function in children living near motorways *Epidemiology,* **8,** 298 - 303.

Burnett, R., Brook, J., Yung, W., Dales, R. and Krewski, D. (1997). Association between ozone and hospitalization for respiratory diseases in 16 Canadian cities *Environmental Research,* **72,** 24 - 31.

Cakmak, S., Burnett, R., Jerrett, M., Goldberg, M. S., Pope, C. A. and Ma, R. (2003). Spatial Regression Models for large-cohort studies linking community air pollution and health *Journal of Toxicology and Environmental Health, Part A,* **66,** 1811 - 1823.

Chen, B. H., Hong, C. J. and Kan, H. D. (2004). Exposures and health outcomes from outdoor air pollutants in China *Toxicology,* **198,** 291-300.

Cherniack, R. M. (1977). *Pulmonary function testing,* Saunders, Philadelphia.

Clarke, K. C., McLafferty, S. L. and Tempalski, B. J. (1996). On epidemiology and Geographic Information Systems: A review and discussion of future directions *Emerging Infectious Diseases,* **2,** 85 - 92.

Cliff, A. and Haggett, P. (1988). *Atlas of Disease Distributions,* Blackwell, Oxford, UK.

Cockings, S., Dunn, C., Bhopal, R. S. and Walker, D. R. (2004). Users' perspectives on epidemiological, GIS and point pattern approaches to analysing environment and health data *Health & Place,* **10,** 169 - 182.

Cohen Hubal, E. A., Sheldon, L., Burke, J. M., McCurdy, T., Berry, M. R., Rigas, M. L., Zartarian, V. G. and Freeman, N. C. G. (2000). Children's Exposure Assessment: A Review of Factors Influencing Children's Exposure and the Data Available to Characterize and Assess that Exposure *Environmental Health Perspectives,* **108,** 475 - 486.

Cotton, P. (1993). Best data yet say air pollution below levels currently considered safe *Journal of American Medical Association,* **269,** 3087-3088.

Curtis, S. and Taket, A. (1996). *Health and Societies : Changing Perspectives,* Arnold, a division of Hodder Headline PLC, London, UK.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G. and Speizer, F. E. (1993). An association between Air Pollution and Mortality in Six U.S. Cities *The New England Journal Of Medicine,* **329,** 1753 - 1759.

Dolk, H., Pattenden, S., Vrijheid, M., Thakrar, B. and Armstrong, B. K. (2000). Perinatal and infant mortality and low birth weight among residents near cokeworks in Great Britain *Archives of Environmental Health,* **55,** 26 - 30.

Dunn, C. and Kingham, S. (1996). Establishing Links Between Air Quality and Health: Searching for the Impossible? *Social Science & Medicine,* **42,** 831 - 841.

Dunn, C., Kingham, S., Rowlingson, B., Bhopal, R. S., Cockings, S., Foy, C. J. W., Acquilla, S. D., Halpin, J., Diggle, P. and Walker, D. R. (2001). Analysing spatially referenced public health data: a comparison of three methodological approaches *Health & Place,* **7,** 1 - 12.

Dunn, J. R. and Hayes, M. V. (2000). Social inequality, population health and housing: a study of two Vancouver neighborhoods *Social Science & Medicine,* **51,** 563 - 587.

Elliott, P., Wakefield, J. C., Best, N. G. and Briggs, D. e. (2000). *Spatial Epidemiology: Methods and Applications.,* Oxford University Press, Oxford.

Elliott, S. J. (1999). And the Question Shall Determine the Method *The Professional Geographer,* **51,** 240 - 243.

Engel, G. L. (1963).A Unified Concept of Health and Disease In *Life and Disease*(Ed, Ingle, D. J.) Basic Books.

English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L. and Zhu, L. (1999). Examining Associations between Childhood Asthma and Traffic Flow Using a Geographic Information System *Environmental Health Perspectives,* **107,** 761 - 767.

Epp, J. (1986). *Achieving Health for All: A Framework for Health Promotion, First Internation Conference on Health Promotion,* Ottawa.

Erdas, B. and Hyndman, R. J. (2000). *The effect of air pollution and climate on hospital admissions for chronic obstructive airways disease: a non-parametric alternative.*

Evans, R. G., Barer, M. L. and Marmor, T. R. e. (1994). *Why are Some People Health and Others not? The Determinants of Health of Populations,* New York: Ardine de Gruyter.

Evans, R. G. and Stoddart, G. L. (1990). Producing Health, Consuming Health Care *Social Science & Medicine,* **31,** 1347 - 1363.

Eyles, J. (1997). Environmental Health research: setting an agenda by spinning our wheels or climbing the mountain? *Health & Place,* **3,** 1 - 13.

Eyles, J. (1999). Health, Environmental Assessments and Popualtion Health: Tools for a Complex Process *Canadian Journal of Public Health,* **90,** S31 - S34.

Eyles, J., Brimacombe, M., Chaulk, P., Stoddart, G. L., Pranger, T. and Moase, O. (2001). What determines health? To where should we shift resources? Attidutes towards the determinants of health among multiple stakeholder groups in Prince Edward Island, Canada *Social Science & Medicine,* **53,** 1611 - 1619.

Federal Provincial and Territorial Advisory Committee on Population Health (1994). *Strategies for population health: Investing in the health of Canadians: A report prepared for the Meeting of the Ministers of Health, Halifax, NS,* Health Canada Ottawa, ON.

Filleul, L., Dartigues, J. F. and Tessier, J. F. (2003). Risk factors among elderly for short term deaths related to high levels of air pollution *Occupational and Environmental Medicine,* **60,** 684 - 688.

Finkelstein, M. M., Jerrett, M., Deluca, P., Finkelstein, N., Verma, D. K., Chapman, K. and Sears, M. S. (2003). Relation between income, air pollution and mortality: a cohort study *CMAJ,* **169,** 397 - 402.

Fost, D. (1990). Using Maps to tackle AIDS *American Demographics,* **12,** 22.

Frankish, C. J., Green, L. W., Ratner, P. A., Chomik, T. A. and Larsen, C. (1996). *Health Impact Assessment as a tool for population health promotion and public policy: A report submitted to the Health Promotion Development Division of Health Canada,* University of British Columbia, Institute of Health Promotion Research Vancouver, BC.

Frankish, J., Veenstra, G. and Moulton, G. (1999). Population Health in Canada: Issues and Challenges for Policy, Practice and Research *Canadian Journal of Public Health,* **90,** S71 - S75.

Gatrell, A. C. (2002). *Geographies of Health,* Blackwell Publishers Ltd., Oxford.

Gatrell, A. C. and Löytönen, M. e. (1998). *GIS and Health,* Taylor and Francis, London.

Gauderman, J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., Margolis, H., Bates, D. and Peters, J. (2004). The Effect of Air Pollution on Lung Development from 10 to 18 Years of Age *The New England Journal Of Medicine,* **351,** 1057 - 1067.

Gauderman, J., Gilliland, F., Vora, H., Avol, E., Stram, D., McDonnell, R., Thomas, D., Lurmann, F., Margolis, H., Rappaport, E., Berhane, K. and Peters, J. (2002). Association between Air Pollution and Lung Function Growth in Southern California Children. Results from a Second Cohort *American Journal of Respiratory Critical Care Medicine,* **166,** 76 - 84.

Gauderman, J., McDonnell, R., Gilliland, F., London, S., Thomas, D., Avol, E., Vora, H., Berhane, K., Rappaport, E., Lurmann, F., Margolis, H. and Peters, J. (2000). Association between Air Pollution and Lung Function Growth in Southern California Children *American Journal of Respiratory Critical Care Medicine,* **162,** 1383 - 1390.

Gesler, W. (1986). The uses of spatial analysis in medical geography: a review *Social Science & Medicine,* **23,** 963-973.

Google Maps (2005). *Imagery Digital Globe, Earth Sat,* Avaliable: http:/www.google.ca/maps *[Accessed on 26/05/2005.*

Guidotti, T. (1997). Why are some people healthy and others not? A critique of the population health model *Annals , RCPSC,* **30,** 203 - 206.

Gulliver, J. and Briggs, D. (2004). Personal exposure to particulate matter in transport microenvironments *Atmospheric Environment,* **38,** 1 - 8.

Hamilton, N. and Bhatti, T. (1996). *Population Health Promotion: An Integrated model of population and health promotion,* Health Promotion Development Division, Heatlh Canada Ottawa.

Harre, E. S. M., Price, P. D., Ayrey, R. B., Toop, L. J., Martin, I. R. and Town, G. I. (1997). Respiratory effects of air pollution in chronic obstructive pulmonary disease: a three month prospective study *Thorax,* **52,** 1040 - 1044.

Hayes, M. V. (1996). Editorial *Health & Place,* **2,** iii - v.

Hayes, M. V. (1999). Population Health Promotion: Responsible Sharing of Future Directions *Canadian Journal of Public Health,* **90,** S15 - S17.

Hayes, M. V. and Dunn, J. R. (1998). *Population health in Canada: A systematic review,* Canadian Policy Research Network Study H/01 Ottawa, ON.

Health Canada (2000). *Regulations related to health and air quality: National Ambient Air Quality Objectives (NAAQOs),* Avaliable: http://www.hc-sc.gc.ca/hecs-sesc/air_quality/regulations.htm *[Accessed on 23/07/2005].*

Health Canada (2001). *The Population Health Template: Key Elements and Actions that Define a Population Heatlh Approach,* Strategic Directorate - Population and Public Health Brach.

Hertzman, C. and Wiens, M. (1996). Child Development and Long-term Outcomes: A Population Health Perspective and Summary of Successful Interventions *Social Science & Medicine,* **43,** 1083 - 1095.

Hoek, G. and Brunekreef, B. (1994). Effects of Low-Level Winter Air Pollution Concentrations on Respiratory Health of Dutch Children *Environmental Research,* **64,** 136 - 150.

Hoek, G., Brunekreef B., Goldbohm S., Fischer P. and P., B. (2002a). Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study *The Lancet,* **360,** 1203-1209.

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

Hruba, F., Fabianova, E., Koppova, K. and Vandenberg, J. J. (2001). Childhood respiratory symptoms, hospital admissions, and long-term exposure to airborne particulate matter *Journal of Exposure Analysis and Environmental Epidemiology,* **11,** 33 - 40.

Isaaks, E. H. and Srivastava, R. M. (1989).An introduction to Applied Geostatistics Oxford University Press, Oxford, UK.

Jalaludin, B. B., O'Toole, B. I. and Leeder, S. R. (2004). Acute Effects of urban ambient air pollution on respiratory symptoms, asthma medication use, and doctor vistis for asthma in a cohort of Australian children *Environmental Research,* **95,** 32 - 42.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglou, T., Morrison, J. and Giovis, C. (2004). A review and evaluation of intra-urban air pollution exposure models *Journal of Exposure Analysis and Environmental Epidemiology,* E-print published online August 4, 2004.

Jerrett, M., Burnett, R., Goldberg, M. S., Sears, M. S., Krewski, D., Catalan, R., Kanaroglou, P., Giovis, C. and Finkelstein, N. (2003). Spatial Analysis for environmental health research: concepts, methods and examples *Journal of Toxicology and Environmental Health, Part A,* **66,** 1783 - 1810.

Johnston, R. J., Gregory, D., Pratt, G. and Watts , M. e. (1994). *The Dictionary of Human Geography,* Blackwell Publishers, Oxford, UK.

Jones, A. P. (1999). Indoor air quality and health *Atmospheric Environment,* **33,** 4535 - 4564.

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

Katsouyanni, K. (2003). Ambient air pollution and health *British Medical Bulletin,* **68,** 143 - 156.

Kearns, R. and Moon, G. (2002). From medical to health geography: novelty, place and theory after a decade of change *Progress in Human Geography,* **26,** 605 - 625.

Kerigan, A. T., Goldsmith, C. H. and Pengelly, L. D. (1986). A three-Year Cohort Study of the Role of Environmental Factors in the Respiratory Health of Children in Hamilton, Ontario *AM REV RESPIR DIS,* **133,** 987 - 993.

Koenig, J. Q. (1999). Air pollution and asthma *J Allergy Clin Immunol,* **104,** 717 - 722.

Koop, G. and Tole, L. (in press). Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *Journal of Environmental Economics and Management*

Kousa, A., Kukkonen, J., Karppinen, A. and Aarnio, P. (2002). A model for evaluating the population exposure to ambient air pollution in an urban area *Atmospheric Environment,* **36,** 2109 - 2119.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine,* **39,** 887 - 903.

Krzyzanowski, M., Cohen, A., Anderson, R. and WHO Working Group (2002). Quantification of health effects of exposure to air pollution *Occupational and Environmental Medicine,* **59,** 791 - 793.

Maheswaran, R., Chan, D., Fryers, P. T., McManus, C. and McCabe, H. (2004). Socio-economic deprivation and excess winter mortality and emergency hospital admissions in the South Yorkshire Coalfields Health Action Zone, UK *Public Health,* **118,** 167-176.

Mayer, J. D. (1983). The role of spatial analysis and geographic data in the detection of disease causation *Social Science & Medicine,* **17,** 1213 - 1221.

McCally M. (ed.) (2000). *Life Support: The Environment and Human Health,* MIT Press, Cambridge, MA.

McGrail, K. M., Ostry, A., Thomas, V. S. and Sanmartin, C. (1998). *Determinants of population health: A synthesis of the literature,* National Health Research and Development Program, Health Canada Ottawa.

McKay, L. (2000). *Making the Lalonde Report: Towards a New Perspective on Health Project, Background Paper,* Health Network, CPRN.

McMichael, T. (2001). *Human frontiers, environments and disease: Past patterns, uncertain futures,* Cambridge University Press, Cambridge.

Meade, M., Florin, J. and Gesler, W. (1988). *Medical Geography,* The Guilford Press, New York.

Miller, A. e. (1987.). *Pulmonary function tests : a guide for the student and house officer,* Grune & Stratton, Orlando.

Mohan (1989). Medical Geography: Competing diagnoses and prescriptions *Antipode,* **21,** 176.

Monn, C. (2001). Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmospheric Environment,* **35,** 1-32.

Morris, R. D. and Munasinghe, R. L. (1994). Geographic Variability in Hospital Admission Rates for Respiratory Disease Among the Elderly in the United States *Chest,* **106,** 1172 - 1181.

Moseler, M., Hendel - Kramer, A., Karmaus, W., Forster, J., Weiss, K., Urbanek, R. and Kuehr, J. (1994). Effect of NO2 Air Pollution on the Lung Funtion of Children with Asthmatic Symptoms *Environmental Research,* **67,** 109 - 124.

Nafstad, P., Haheim, L. L., Oftedal, B., Gram, F., Holme, I., Hjermann, I. and Leren, P. (2003). Lung cancer and air pollution: a 27 year follow up of 16,209 Norwegian men *Thorax,* **58,** 1071 - 1076.

Ontario Medical Association (2000). *The Illness Costs of Air Pollution in Ontario: A Summary of Finding,* Avaliable: http://www.oma.org/Health/smog/icap.asp *[Accessed on 18/03/2005].*

Pengelly, D. L., Kerigan, A. T., Goldsmith, C. H. and Inman, E. M. (1984a). The Hamilton Study: Distribution of Factors Confounding the Relationship between Air Quality and Respiratory Health *JAPA,* **34,** 1039 - 1043.

Pengelly, L. D., Goldsmith, C. H., Kerigan, A. T., Furlong, W. and Toplack, S. A. (1986).The Hamilton Study: Effect of Particle Size on Respiratory Health in Children In *Aerosols,* Vol. 56 (Eds, Lee, S. D., Schneider, T., Grant, L. D. and Verkerk, P. J.) Lewis Publishers, Michigan USA.

Pengelly, L. D., Kerigan, A. T., Goldsmith, C. H. and Inman, E. M. (1984b). The Hamilton Study: Distribution of Factors Confoundign the Relationship between Air Quality and Respiratory Health *JAPCA,* **34,** 1039 - 1043.

Peters, M. J. (1997). *A GIS application: Spatial and temporal patterns of ambulance performance, Integrating Spatial Information Technologies for Tomorrow, GIS ' 97,* Fort Collins.

Peters, M. J., Avol, E., Navidi, W., London, S. J., Gauderman, W. J., Lurmann, F., Linn, W. S., Margolis, H., Rappaport, E., Gong, H. and Thomas, D. C. (1999). A Study of Twelve Southern California Communities with Differing Levels and Types of Air Pollution I. Prevalence of Respiratoty Morbidity *Am J Resp Crit Care Med,* **159,** 760 - 767.

Poland, B., Coburn, D. and Eakin, J. (1998a). Wealth, equity and health care: A critique of a "population health" perspective on the determinants of health *Social Science & Medicine,* **46,** 785-798.

Poland, B., Coburn, D. and Eakin, J. (1998b). Wealth, equity and health care: A critique of a population health perspective on the determinants of health *Social Science & Medicine,* **46,** 785 - 798.

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Kazuhiko, I. and Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality and Long-Term Exposure to Fine Particulate Air Pollution *Journal of American Medical Association,* **287,** 1132-1141.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I. and Zeger, S. (2000). Fine Particulate Air Pollution and Mortality in 20 U.S. Cities 1987-1994 *The New England Journal Of Medicine,* **343,** 1742-1749.

Schwartz, J. (1995). Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease *Thorax,* **50,** 531 - 538.

Schwartz, J. and Neas, L. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren *Epidemiology,* **11,** 6 - 10.

Scoggins, A., Kjellstrom, T., Fischer, G., Connor, J. and Gimson, N. (2003). Spatial analysis of annual air pollution exposure and mortality *The Science of the Total Environment,* **(in press).**

Stern, B. R., Raizenne, M. E., Burnett, R., Jones, L., Kearney, J. and Franklin, C. A. (1994). Air Pollution and Childhood Respiratoty Health: Exposure to Sulfate and Ozone in 10 Canadian Rural Communities *Environmental Research,* **66,** 125 - 142.

Vine, M. F., Degnan, D. and Hanchette, C. (1997). Geographic Information Systems: Their Use in Environmental Epidemiologic Research *Environmental Health Perspectives,* **105,** 598 - 605.

Walker, S. E., Slordal, L. H., Guerreiro, C., Gram, F. and Gronskei, K. E. (1999). Air pollution exposure monitoring and estimation Part II: Model evaluation and population exposure *J Environ Monit,* **1,** 321 - 326.

Walsh, S. J., Page, P. H. and Gesler, W. M. (1997). Normative Models and healthcare planning: Network - based simulations withing geographic information system environment *Health Services Research,* **32,** 243 - 260.

Ward, D. J. (2004). Particulate air pollution and panel studies on children: a systematic revirew *Occupational and Environmental Medicine,* **61,** electronic paper.

Wentz, E., Gober, P., Balling, J. R. C. and Day, T. A. (2002). Spatial Patterns and Determinants of Winter Atmospheric Carbon Dioxide Concentrations in an Urban Environment *Annals Assoc Amer Geo,* **92,** 15 - 28.

Wilkinson, P., Elliott, P., Grundy, C., Shaddick, G., Thakrar, B., Walls, P. and Falconer, S. (1999). Case-control study of hospital admission with asthma in children aged 5 - 14 years: relation with road traffic in north west London *Thorax,* **54,** 1070 - 1074.

Wilkinson, R. G. (1996). *Unhealthy Societies: The Afflictions of Inequalities,* Routledge: London.

Williams, R., Creason, J., Zweidinger, R., Watts, R., Sheldon, L. and Shy, C. (2000). Indoor, outdoor, and personal exposure monitoring of particulate air pollution: the Baltimore elderly epidemiology-exposure pilor study *Atmospheric Environment,* **34,** 4193 - 4204.

World Health Organization (1948). *Preamble of the Consitution of the World Health Organization, International Health Conference,* New York, (19 - 22 June 1946). Signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

World Health Organization (1986). *Ottawa Charter on Health Promotion, First Internation Conference on Health Promotion,* Ottawa.

World Health Organization (2000). *Air Quality Guidelines for Europe, 2nd Edition,* WHO Regional Office for Europe Copenhagen.

World Health Organization (2002a). *A discussion of the role of economic and social determinants. Studies on social and economic determinants of population health, No. 1,* WHO Regional Office for Europe Copenhagen.

World Health Organization (2002b). *The World Health Report, 2002,* World Health Organization Geneva.

Zollner, H. and Lessof, S. (1998). *Population Health - Putting Concepts into Action. FInal Report,* World Health Organization, Regional Office for Europe

APPENDIX I

LIST OF VARIABLES USED FOR THE HEALTH – RELATED REGRESSION ANALYSIS

Table A1-1: Variables used for the re-analysis of the HCC study

#	Variable Coding	Type	EXPLANATION
1	POSTAL	Char	Postal Code
2	X_COORD1	Num	X coordinate (UTM-NAD 1983 ZONE 17N)
3	Y_COORD1	Num	Y coordinate (UTM-NAD 1983 ZONE 17N)
4	ID1	Num	Unique Identifier
5	TSP80KR	Num	TSP exposure assessment based on the kriging technique for 1980
6	TSP80LUR	Num	TSP exposure assessment based on the LUR technique for 1980
7	WHEEZY	Binary	The chest of the child sounds wheezy = 1 otherwise 0
8	DURATION_W	Binary	If the child is experiencing the symptom of "chest sounds wheezy" for most of the days/nights of the yea r= 1 otherwise 0
9	ASTHMA	Binary	The child is suffering from asthmatic attacks = 1 otherwise 0
10	GOCHEST	Binary	When the child has cold, it usually goes to the chest = 1 otherwise 0
11	COLD_LASTM	Binary	The child had a cold in the last month = 1 otherwise 0
12	CHEST_YEAR	Binary	The child had a chest illness such as bronchitis or pneumonia that has kept him/her at home for one week or more = 1 otherwise 0
13	HEATING	Binary	The home of the child is heated by air -dust forced air = 1 otherwise 0
14	WITHMOTHER	Binary	The child is living with the mother = 1 otherwise 0
15	SHARE_RM	Binary	The child is sharing the bedroom = 1 otherwise 0
16	OTHERCHILD	Binary	Not counting the child that participates in the survey other children of the household usually cough during day or night =1 otherwise 0
17	OTHERCHILD1	Binary	Not counting the child that participates in the survey other children of the household had a chest illness during the last 12 months =1 otherwise 0
18	HOSP2	Binary	The child has been hospitalized before the age of 2 years = 1 otherwise 0
19	CHEST2	Binary	The child had a chest illness before the age of 2 years = 1 otherwise 0
20	BSEX	Binary	Girl = 1 Boy = 0

#	Variable Coding	Type	EXPLANATION
21	BMSMOKE	Binary	If the mother is smoking =1 otherwise 0
22	BFSMOKE	Binary	If the father is smoking =1 otherwise 0
23	BMAPAK	Binary	If the mother is smoking > 1 packs of cigarettes per day= 1 otherwise 0
24	BPAPAK	Binary	If the father is smoking > 1 packs of cigarettes per day= 1 otherwise 0
25	BGAS	Binary	If the household is using gas for cooking = 1 otherwise 0
26	BANYSMKS	Binary	If there is at least one smoker in the house = 1 otherwise 0
27	TOTALROOMS	Num	The total number of rooms in the house (not counting the bathroom)
28	BEDROOMS	Num	The total number of bedrooms in the house
29	PPL_HLD	Num	The total number of people living in the house including the child
30	CHILD_HLD	Num	The total number of children under the age of 15 living in the house
31	CROWD_INDEX	Num	The total number of people in the house divided by the total nubmer of bedrooms
32	LW_ INCOME	Binary	The income of the family is less than CAN$15,000 = 1 otherwise 0
33	BNSMOKE2	Num	The total number of smokers in the house
34	BHT	Num	The height of the child
35	BFVC	Num	Forced Vital Capacity
36	BFEV1	Num	Forced Expiratory Volume in one second
37	BFEF2575	Num	Forced Expiratory Flow between 25% and 75% (FEF_{25-75}) and 75% and 85% (FEF_{75-85}) of FVC
38	BPF	Num	Peak Flow
39	BMEF50	Num	Maximum Expiratory Flow after 50%
40	BMEF75	Num	Maximum Expiratory Flow after 75%

#	Variable Coding	Type	EXPLANATION
41	BCIGS	Num	The total number of cigarettes smoked in the house
42	BCIGSPAK	Num	The total number of packs of cigarettes smoked in the house
43	HSHLD_SMOK_RATE	Num	The total number of smokers in the house divided by the total number of people in the house
44	BCVVC	Num	Ratio of Closing Volume to Vital Capacity
45	BFEV1FVC	Num	Ratio of the Forced Expiratory Volume in one second to Forced Vital Capacity

APPENDIX II

LIST OF VARIABLES USED FOR THE LAND – USE REGRESSION ANALYSIS

Table A2-1: Variables used for the land use regression analysis of the HCC study

#	Variable	Type	EXPLANATION
1	TSP_ID	Num	Unique Identifier
2	TSP_80	Num	Actual concentration values of TSP for 1980
3	X_COORD	Num	X coorkdinate (UTM-NAD 1983 ZONE 17N)
4	Y_COORD	Num	Y coordinate (UTM-NAD 1983 ZONE 17N)
5	ELEV80	Num	Elevation at the monitoring station
6	HDIST300	Binary	Presence of a highway within 300 m buffer = 1, 0 otherwise
7	HDIST500	Binary	Presence of a highway within 500 m buffer = 1, 0 otherwise
8	HDIST750	Binary	Presence of a highway within 750 m buffer = 1, 0 otherwise
9	HDIST1000	Binary	Presence of a highway within 1000 m buffer = 1, 0 otherwise
10	HDIST1500	Binary	Presence of a highway within 1500 m buffer = 1, 0 otherwise
11	RDIST100	Binary	Presence of a major road within 100 m buffer = 1, 0 otherwise
12	RDIST300	Binary	Presence of a major road within 300 m buffer = 1, 0 otherwise
13	RDIST200	Binary	Presence of a major road within 200 m buffer = 1, 0 otherwise
14	RDIST500	Binary	Presence of a major road within 500 m buffer = 1, 0 otherwise
15	IND50	Num	Industrial land use area within 50m buffer from a monitoring station
16	CM50	Num	Commercial land use area within 50m buffer from a monitoring station
17	INSTIT50	Num	Institutional land use area within 50m buffer from a monitoring station
18	PARKS50	Num	Parks land use area within 50m buffer from a monitoring station
19	RES50	Num	Residential land use area within 50m buffer from a monitoring station
20	CEMETE100	Num	Cemetery land use area within 100m buffer from a monitoring station
21	CM100	Num	Commercial land use area within 100m buffer from a monitoring station
22	IND100	Num	Industrial land use area within 100m buffer from a monitoring station
23	INSTIT100	Num	Institutional land use area within 100m buffer from a monitoring station

#	Variable	Type	EXPLANATION
24	PARKS100	Num	Parks land use area within 100m buffer from a monitoring station
25	SPACE100	Num	Open-land land use area within 100m buffer from a monitoring station
26	RES100	Num	Residential land use area within 100m buffer from a monitoring station
27	WATER200	Num	Water land use area within 200m buffer from a monitoring station
28	RS200	Num	Residential land use area within 200m buffer from a monitoring station
29	CM200	Num	Commercial land use area within 200m buffer from a monitoring station
30	INDUST200	Num	Industrial land use area within 200m buffer from a monitoring station
31	INSTITU200	Num	Institutional land use area within 200m buffer from a monitoring station
32	PARKS200	Num	Parks land use area within 200m buffer from a monitoring station
33	CEMETE200	Num	Cemetery land use area within 200m buffer from a monitoring station
34	SPACE200	Num	Open-land land use area within 200m buffer from a monitoring station
35	CM250	Num	Commercial land use area within 250m buffer from a monitoring station
36	RS250	Num	Residential land use area within 250m buffer from a monitoring station
37	IND250	Num	Industrial land use area within 250m buffer from a monitoring station
38	WATER250	Num	Water land use area within 250m buffer from a monitoring station
39	CEMETE250	Num	Cemetery land use area within 250m buffer from a monitoring station
40	INSTIT250	Num	Institutional land use area within 250m buffer from a monitoring station
41	SPACE250	Num	Open-land land use area within 250m buffer from a monitoring station
42	PARKS250	Num	Parks land use area within 250m buffer from a monitoring station
43	WATER300	Num	Water land use area within 300m buffer from a monitoring station
44	RS300	Num	Residential land use area within 300m buffer from a monitoring station
45	CM300	Num	Commercial land use area within 300m buffer from a monitoring station
46	INDUST300	Num	Industrial land use area within 300m buffer from a monitoring station
47	INSTITU300	Num	Institutional land use area within 300m buffer from a monitoring station
48	PARKS300	Num	Parks land use area within 300m buffer from a monitoring station
49	CEMETE300	Num	Cemetery land use area within 300m buffer from a monitoring station

#	Variable	Type	EXPLANATION
50	SPACE300	Num	Open-land land use area within 300m buffer from a monitoring station
51	WATER500	Num	Water land use area within 500m buffer from a monitoring station
52	RES500	Num	Residential land use area within 500m buffer from a monitoring station
53	CM500	Num	Commercial land use area within 500m buffer from a monitoring station
54	INDUST500	Num	Industrial land use area within 500m buffer from a monitoring station
55	INSTITU500	Num	Institutional land use area within 500m buffer from a monitoring station
56	PARKS500	Num	Parks land use area within 500m buffer from a monitoring station
57	CEMETE500	Num	Cemetery land use area within 500m buffer from a monitoring station
58	SPACE500	Num	Open-land land use area within 500m buffer from a monitoring station
59	WATER750	Num	Water land use area within 750m buffer from a monitoring station
60	RS750	Num	Residential land use area within 750m buffer from a monitoring station
61	CM750	Num	Commercial land use area within 750m buffer from a monitoring station
62	INDUST750	Num	Industrial land use area within 750m buffer from a monitoring station
63	INSTITU750	Num	Institutional land use area within 750m buffer from a monitoring station
64	PARKS750	Num	Parks land use area within 750m buffer from a monitoring station
65	CEMETE750	Num	Cemetery land use area within 750m buffer from a monitoring station
66	GOLF750	Num	Golf land use area within 750m buffer from a monitoring station
67	SPACE750	Num	open-land land use area within 750m buffer from a monitoring station
68	WATER1000	Num	Water land use area within 1000m buffer from a monitoring station
69	RS1000	Num	Residential land use area within 1000m buffer from a monitoring station
70	CM1000	Num	Commercial land use area within 1000m buffer from a monitoring station
71	INDUST1000	Num	Industrial land use area within 1000m buffer from a monitoring station
72	INSTIT1000	Num	Institutional land use area within 1000m buffer from a monitoring station
73	PARK1000	Num	Parks land use area within 1000m buffer from a monitoring station
74	CEMETE1000	Num	Cemetery land use area within 1000m buffer from a monitoring station
75	GOLF1000	Num	Golf land use area within 1000m buffer from a monitoring station

#	Variable	Type	EXPLANATION
76	SPACE1000	Num	Open-land land use area within 1000m buffer from a monitoring station
77	RLENGTH50	Num	Length of major road within 50 m buffer from the monitoring station
78	RLENGTH100	Num	Length of major road within 100 m buffer from the moni toring station
79	RLENGTH200	Num	Length of major road within 200 m buffer from the monitoring station
80	HLENGTH200	Num	Length of major highway within 200 m buffer from the monitoring station
81	HLENGTH300	Num	Length of major highway within 300 m buffer from the monitoring station
82	RLENGTH300	Num	Length of major road within 300 m buffer from the monitoring station
83	HLENGTH400	Num	Length of major highway within 400 m buffer from the monitoring station
84	RLENGTH400	Num	Length of major road within 400 m buffer from the monitoring station
85	HLENGTH500	Num	Length of major highway within 500 m buffer from the monitoring station
86	RLENGTH500	Num	Length of major road within 500 m buffer from the monitoring station
87	HLENGTH600	Num	Length of major highway within 600 m buffer from the monitoring station
88	RLENGTH600	Num	Length of major road within 600 m buffer from the monitoring station
89	HLENGTH750	Num	Length of major highway within 750 m buffer from the monitoring station
90	RLENGTH750	Num	Length of major road within 750 m buffer from the monitoring station
91	OPENSPACE100	Num	Openspace land use area within 100m buffer from a monitoring station
92	OPENSPACE200	Num	Openspace land use area within 200m buffer from a monitoring station
93	OPENSPACE250	Num	Openspace land use area within 250m buffer from a monitoring station
94	OPENSPACE300	Num	Openspace land use area within 300m buffer from a monitoring station
95	OPENSPACE500	Num	Openspace land use area within 500m buffer from a monitoring station
96	OPENSPACE750	Num	Openspace land use area within 750m buffer from a monitoring station
97	OPENSPACE1000	Num	Openspace land use area within 10 00m buffer from a monitoring station

APPENDIX III

CORRELATION MATRIX FOR THE LAND-USE REGRESSION VARIABLES