4-30-2009

Air Toxics and Equity: A Geographic Analysis of Environmental Health Risks in Florida

Angela Gilbert
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Air Toxics and Equity:
A Geographic Analysis of Environmental Health Risks in Florida

by

Angela Gilbert

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Arts
Department of Geography
College of Arts and Sciences
University of South Florida

Major Professor: Jayajit Chakraborty, Ph.D.
Pratyusha Basu, Ph.D.
Steven Reader, Ph.D.

Date of Defense:
April 30, 2009

Keywords: environmental justice, air pollution, NATA, geographically weighted regression, public health

© Copyright 2009, Angela Gilbert
ACKNOWLEDGEMENTS
This thesis would not be possible without the guidance, clarity, and confidence of Jay Chakraborty. Thank you.

Much gratitude goes to my family for their kindness, love, and generosity.

Finally, I’d like to acknowledge Wilco, whose music provided sanity and perspective during the long, late hours of reading, writing, and analyzing.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter 1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2. Background and Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Measuring Exposure to Air Pollution</td>
<td>7</td>
</tr>
<tr>
<td>2.1.1 From Individual Source Analysis to a Cumulative Approach</td>
<td>7</td>
</tr>
<tr>
<td>2.1.2 Measuring the Extent of Exposure to Air Pollution</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Statistical Methods</td>
<td>15</td>
</tr>
<tr>
<td>2.3 Summary</td>
<td>18</td>
</tr>
<tr>
<td>Chapter 3. Study Area and Methodology</td>
<td>20</td>
</tr>
<tr>
<td>3.1 Study Area</td>
<td>20</td>
</tr>
<tr>
<td>3.2 Data Sources and Variables</td>
<td>21</td>
</tr>
<tr>
<td>3.2.1 Dependent Variables</td>
<td>23</td>
</tr>
<tr>
<td>3.2.2 Explanatory Variables</td>
<td>25</td>
</tr>
<tr>
<td>3.3 Methods</td>
<td>28</td>
</tr>
<tr>
<td>Chapter 4. Conventional Environmental Justice Analysis of Estimated Cancer Risks</td>
<td>32</td>
</tr>
<tr>
<td>4.1 Descriptive Choropleth Mapping and Statistics</td>
<td>32</td>
</tr>
<tr>
<td>4.2 Bivariate Correlation Analysis</td>
<td>40</td>
</tr>
<tr>
<td>4.3 Traditional Multiple Regression Analysis</td>
<td>42</td>
</tr>
<tr>
<td>Chapter 5. Geographically Weighted Regression Analysis of Estimated Cancer Risks</td>
<td>46</td>
</tr>
<tr>
<td>5.1 Cumulative Cancer Risk from All Sources</td>
<td>48</td>
</tr>
<tr>
<td>5.2 Lifetime Cancer Risk from Major Point Sources</td>
<td>57</td>
</tr>
<tr>
<td>5.3 Lifetime Cancer Risk from Other Point Sources</td>
<td>62</td>
</tr>
<tr>
<td>5.4 Lifetime Cancer Risk from On-Road Mobile Sources</td>
<td>69</td>
</tr>
<tr>
<td>5.5 Lifetime Cancer Risk from Non-Road Mobile Sources</td>
<td>72</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 3.1 Definitions of Explanatory Variables and Expected Statistical Association with Cancer Risk ............................................................................ 28
Table 4.1 Descriptive Statistics for Variables Analyzed ........................................ 40
Table 4.2 Pearson’s Correlation Coefficients of Lifetime Cancer Risk by Source and Explanatory Variables ................................................................................ 41
Table 4.3 Multivariate Regression Results: Cumulative Cancer Risk from All Sources........................................................................................................................................... 43
Table 4.4 Multivariate Regression Results: Cancer Risk from Four Known Sources........................................................................................................................................... 44
Table 5.1 GWR Results for Cumulative Cancer Risk from All Sources............... 49
Table 5.2 GWR Results for Cancer Risk from Major Point Sources ................. 58
Table 5.3 GWR Results for Cancer Risk from Minor Point Sources ............... 63
Table 5.4 GWR Results for Cancer Risk from On-Road Mobile Sources ......... 70
Table 5.5 GWR Results for Cancer Risk from Non-Mobile Sources............... 73
LIST OF FIGURES

Figure 4.1 Estimated Lifetime Cancer Risk From All Sources by Census Tract, 1999

Figure 4.2 Estimated Lifetime Cancer Risk From Major Point Sources by Census Tract, 1999

Figure 4.3 Estimated Lifetime Cancer Risks From Minor Point Sources, 1999

Figure 4.4 Estimated Lifetime Cancer Risk From On-Road Mobile Sources, 1999

Figure 4.5 Estimated Lifetime Cancer Risk From Non-Road Mobile Sources, 1999

Figure 4.6 Proportion Non-Hispanic Black by Census Tract, 2000

Figure 4.7 Proportion Hispanic by Census Tract, 2000

Figure 4.8 Proportion Asian by Census Tract, 2000

Figure 4.9 Homeowner Occupied Housing Units by Census Tracts, 2000

Figure 4.10 Proportion of Population Below the Federal Poverty Rate, 2000

Figure 4.11 Population Density by Census Tract, 2000

Figure 5.1 Local R-Squared from GWR Model of Cancer Risk from All Sources

Figure 5.2 Distribution of t-values for Proportion Black and Cancer Risk from All Sources

Figure 5.3 Distribution of t-values for Proportion Hispanic and Cancer Risk from All Sources

Figure 5.4 Distribution of t-values for Proportion Asian and Cancer Risk from All Sources

Figure 5.5 Distribution of t-values for Home Ownership Rate and Cancer Risk from All Sources

Figure 5.6 Distribution of t-values for Poverty Rate and Cancer Risk from All Sources

Figure 5.7 Distribution of t-values for Population Density and Cancer Risk from All Sources

Figure 5.8 Local R-Squared of Cancer Risk from Major Point Sources

Figure 5.9 Distribution of t-values for Proportion Black and Cancer Risk from Major Point Sources
Figure 5.10 Distribution of t-values for Proportion Hispanic and Cancer Risk from Major Point Sources ................................................................. 60
Figure 5.11 Distribution of t-values for Proportion Asian and Cancer Risk from Major Point Sources ................................................................. 60
Figure 5.12 Distribution of t-values for Home Ownership Rate and Cancer Risk from Major Point Sources ............................................................... 61
Figure 5.13 Distribution of t-values for Poverty Rate and Cancer Risk from Major Point Sources ..................................................................................... 61
Figure 5.14 Local R-Squared of Cancer Risk from Other Point Sources .... 65
Figure 5.15 Distribution of t-values for Proportion Black and Cancer Risk from Other Point Sources ........................................................................... 65
Figure 5.16 Distribution of t-values for Proportion Hispanic and Cancer Risk from Other Point Sources ........................................................................... 66
Figure 5.17 Distribution of t-values for Proportion Asian and Cancer Risk from Other Point Sources ........................................................................... 66
Figure 5.18 Distribution of t-values for Home Ownership Rate and Cancer Risk from Other Point Sources ........................................................................... 67
Figure 5.19 Distribution of t-values for Poverty Rate and Cancer Risk from Other Point Sources ........................................................................... 67
Figure 5.20 Distribution of t-values for Population Density and Cancer Risk from Other Point Sources ........................................................................... 68
Figure 5.21 Local R-Squared of Cancer Risk from On-Road Mobile Sources .... 71
Figure 5.22 Distribution of t-values for Poverty Rate and Cancer Risk from On-Road Mobile Sources ........................................................................... 71
Figure 5.23 Local R-Squared of Cancer Risk from Non-Road Mobile Sources .... 74
Figure 5.24 Distribution of t-values for Proportion Black and Cancer Risk from Non-Road Mobile Sources ........................................................................... 74
Figure 5.25 Distribution of t-values for Proportion Hispanic and Cancer Risk from Non-Road Mobile Sources ........................................................................... 75
Figure 5.26 Distribution of t-values for Proportion Asian and Cancer Risk from Non-Road Mobile Sources ........................................................................... 75
Figure 5.27 Distribution of t-values for Home Ownership Rate and Cancer Risk from Non-Road Mobile Sources ........................................................................... 76
Figure 5.28 Distribution of t-values for Poverty Rate and Cancer Risk from Non-Road Mobile Sources ........................................................................... 76
Figure 5.29 Distribution of t-values for Population Density and Cancer Risk from Non-Road Mobile Sources ........................................................................... 77
Figure A-1 Frequency Histogram of Cumulative Cancer Risk per Million People .......................... 96
Figure A-2 Frequency Histogram of Major Point Sources of Cancer Risk per Million People ........................................................................................................ 96
Figure A-3 Frequency Histogram of Other Point Sources of Cancer Risk per Million People ........................................................................................................ 97
Figure A-4 Frequency Histogram of On-Road Mobile Sources of Cancer Risk per Million People ........................................................................................................ 97
Figure A-5 Frequency Histogram of Non-Road Mobile Sources of Cancer Risk per Million People ........................................................................................................ 98
Figure A-6 Frequency Histogram of Proportion Black by Census Tract, 2000 .......... 98
Figure A-7 Frequency Histogram of Proportion Hispanic by Census Tract, 2000 .... 99
Figure A-8 Frequency Histogram of Proportion Asian by Census Tract, 2000 ........... 99
Figure A-9 Frequency Histogram of Poverty Rate by Census Tract, 2000 ............... 100
Figure A-10 Frequency Histogram of Home Ownership Rate by Census Tract, 2000 ...................................................................................................................................... 100
Figure A-11 Frequency Histogram of Population Density by Census Tract, 2000 ....... 101
Figure A-12 Frequency Histogram of the Natural Log of Population Density by Census Tract, 2000 ...................................................................................................... 101
AIR TOXICS AND EQUITY:
A GEOGRAPHIC ANALYSIS OF ENVIRONMENTAL HEALTH RISKS
IN FLORIDA

Angela Gilbert

ABSTRACT

A large number of quantitative studies have examined social inequities in the geographic distribution of air pollution. Although previous research has made strides towards understanding the nature and extent of inequities, they have been limited methodologically in three ways. First, the presence of pollutants have been rarely linked to their adverse health effects, with many studies using proximity to sources as a proxy for risk. Second, there has been a tendency to study a single pollution source instead of assessing multiple types of sources. Finally, conventional statistical methods such as multivariate regression have been limited by their inability to discern spatial variations in the relationships between dependent and explanatory variables.

This thesis addresses these gaps in environmental justice analysis of air pollution by using data from U.S. Environmental Protection Agency’s 1999 National-Scale Air Toxics Assessment in combination with 2000 U.S. Census data to evaluate inequities in the geography of cancer risks from hazardous air pollutants in Florida. The objective is to determine if there are racial/ethnic inequities in the distribution of estimated cancer risks from outdoor exposure to point and mobile sources of air pollutants, after controlling for well-documented contextual variables. The first phase of the study utilizes traditional correlation and regression techniques to reveal that cancer risk from most air pollution
sources are distributed inequitably with respect to race, ethnicity, and socioeconomic state. In the second phase, geographically weighted regression is used along with choropleth mapping to explore the spatial nonstationarity of regression model parameters and geographic variations in the statistical association between cancer risks and various explanatory variables. Results indicate that while Black and Hispanic proportions remain consistent indicators of cancer risk from most pollution sources, these relationships vary across space within Florida. This thesis contributes to environmental justice analysis by demonstrating that conventional multivariate regression can hide important local variations in the relationships between environmental risk and explanatory variables such as race, ethnicity, and socioeconomic status. Since this spatial nonstationarity can be significant within an entire region or a single urban area, understanding its nature and extent is imperative to advancing environmental justice goals.
CHAPTER 1: INTRODUCTION

Increasingly in the last few decades, geographers have been concerned with the unequal spatial distribution of both the benefits and negative by-products of modern society. Of particular interest is the distribution of the adverse effects of various technological and natural hazards (Walker and Bulkeley 2006). By documenting and analyzing the characteristics of communities that are exposed to the externalities of industrial and commercial growth such as air pollution and its adverse health outcomes, geographers hope to uncover distributional inequities of environmental risks and advance the aims of social justice.

Environmental justice is concerned with spatial and social inequities in the distribution of environmental pollution and adverse health consequences of industrial activities and environmental policies. Environmental justice began as a movement stemming from an incident in Warren County, North Carolina. In 1982, the state of North Carolina selected a predominately African-American and low-income area to locate a toxic waste landfill for the disposal of polychlorinated biphenyls (PCBs) dumped illegally in other areas of the state. The site chosen was declared by experts to be unsafe and unsuitable. The community, believing they had been targeted due to their poverty and racial composition, engaged in legal action and a campaign of civil disobedience to keep the landfill out of their neighborhood. Despite their protests and more than 500 arrests, the landfill was placed in the planned location in 1983 (McGurty 2000). Although the residents of Warren County failed to keep the hazardous waste out of their community,
they succeeded in attracting national attention to the issue of environmental justice. As a result, the U.S. General Accounting Office (1983) launched an investigation into the distribution of hazardous facilities in the South and found that Black residents comprised the majority of the population in three of the four communities that contained landfills. This report was followed by a more comprehensive national study conducted by the United Church of Christ (UCC) Commission for Racial Justice (1987). This study indicated that race was the most significant factor in determining the location of commercial hazardous waste facilities and that three out of every five Black and Hispanic individuals in the U.S. lived in communities containing uncontrolled waste sites. The 1987 UCC report set the wheels in motion for various quantitative and qualitative studies that attempted to seek empirical evidence for the claims made by environmental justice activists.

Quantitative environmental justice research has sought to provide statistical evidence of environmental inequity, or the disproportionate distribution of environmental burdens on people and places. These studies have led to the implementation of policies at the national, state, and local level, including President Clinton’s 1994 Executive Order that requires federal agencies to include environmental justice considerations in their plans, programs, and all aspects of operation. The U.S. Environmental Protection Agency (EPA) has recently expanded the definition of environmental justice to the “the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies.” (EPA 2008a). Even with these policies in place, an update of the UCC study on the state of environmental justice shows that race
remains the most statistically significant indicator of the presence of hazardous waste sites in the U.S. (UCC 2007).

While various types of hazardous facilities and undesirable land uses have been investigated, the distributional impacts of air pollution remain a persistent public health and social concern. Although previous empirical studies have made important strides towards understanding the causes and consequences of the inequities in the geography of air pollution, they have been limited methodologically in three critical ways.

First, there has been a lack of connection between the presence of emissions and the adverse health risks they engender. Many studies have utilized proximity to pollution sources as a proxy for risk (e.g., Pollock and Vittas 1995; Cutter et al. 1996; Perlin et al. 2001; Pastor et al. 2004), but this approach has several drawbacks. An exclusive focus on proximity ignores the quantity, toxicity, and environmental fate of released chemicals. Local meteorological conditions and other factors that heavily influence the direction and distance traveled by pollutants are also overlooked. While specific studies have focused on modeling exposure to toxic pollution (e.g., Chakraborty and Armstrong 1997; Bevc et al. 2007), few have attempted to examine whether unequal exposure patterns lead to disproportionate health risks among minority and low-income communities.

A second pitfall of environmental justice research on air pollution is the tendency to study a single pollution source, particularly industrial facilities, instead of cumulatively assessing exposure to multiple chemicals and emission sources. It is equally important to consider mobile emission sources, smaller emitters, and preexisting background concentrations of naturally occurring and persistent air pollution. While other overlooked polluters such as dry-cleaning facilities, auto-body shops, and off-road mobile sources
may release fewer quantities of air pollutants than their large counterparts, they potentially and cumulatively contribute to adverse health risks in residential areas (Fitos and Chakraborty 2003).

Finally, statistical methods such as multivariate regression have been used to examine the inequity hypothesis by evaluating the association between magnitude of pollution and well-documented socioeconomic and demographic variables. Conventional regression techniques, however, are limited by their inability to discern local variations in the relationships between the dependent and independent variables. When assessing environmental injustice across a large area such as a nation, region, or state, the lack of geographic specificity can obfuscate underlying patterns of inequity.

This thesis seeks to address these three methodological gaps in quantitative environmental justice analysis of air pollution through a case study that examines adverse health risks from multiple types of pollution sources in the state of Florida. The research utilizes an EPA database of ambient air emission information and a regression methodology that allows the evaluation of geographic variation in analytical results. The specific objective is to determine if there are racial/ethnic and socioeconomic inequities in the distribution of estimated cancer risks from outdoor exposure to both stationary and mobile sources of air pollutants, after controlling for well-documented contextual variables. The specific research questions investigated in Florida are as follows:

(a) Is cancer risk from outdoor exposure to hazardous air pollutants from various known sources distributed inequitably with respect to race/ethnicity and socioeconomic status?
How does the strength and significance of the statistical relationships between cancer risks from various known sources and race/ethnicity or socioeconomic status vary across the state?

The key data source is the EPA’s 1999 National-Scale Air Toxics Assessment (NATA), which integrates information from the local, state and federal levels in order to produce health risk estimates at the census tract level from four sources of air pollution: major stationary sources, other stationary sources, on-road mobile sources, and non-road mobile sources. A pertinent set of variables describing population and housing characteristics at the census tract level from U.S. Census 2000 is used to analyze inequities in modeled health risks. The first phase of the analysis utilizes traditional techniques such as bivariate linear correlation and multivariate regression analysis to examine the relationship between estimated cancer risk from the four source categories and relevant racial/ethnic, demographic, and socioeconomic characteristics. The second phase uses a local spatial statistical technique known as geographically weighted regression (GWR) in conjunction with choropleth mapping to address the problems inherent in conventional regression and investigate the spatial nonstationarity of model parameters and model performance. GWR can be used to examine geographic variations in estimated regression parameters and provides localized coefficients for analytical units in a given study area (Fotheringham et al. 2002).
CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

The research literature on the inequitable distribution of the adverse effects of technological hazards, specifically air pollution, can be generally classified into two distinct categories: (a) studies that document the historical processes leading to environmental inequities and grassroots movements to achieve environmental justice, and (b) studies that attempt to determine if current patterns of environmental risk resulting from past decisions are inequitable with respect to race, ethnicity, or income. Research concerned with the historical production of spatial and social inequities has relied on qualitative methods such as interviews, participant observation, and archival document analysis (e.g., Pulido 2000, Boone 2002, Grineski et al. 2005). Quantitative methods are typically used to describe the nature and extent of the statistical association between a measure of the presence or magnitude of environmental pollution and the racial/ethnic or socioeconomic characteristics of the potentially exposed population.

Undesirable land uses and hazards that have been studied in environmental justice literature include landfills (e.g., Been 1994, Liu 1997), industrial manufacturing facilities releasing toxic chemicals (e.g., Mitchell et al. 1999, Pastor et al. 2004), hazardous waste treatment, storage, and disposal facilities (e.g., Anderton et al. 1994, Pastor et al. 2001), facilities storing chemicals classified as extremely hazardous substances (e.g., Chakraborty 2001), nuclear power plants (e.g., Knezevic and Chakraborty 2004), and noise pollution (e.g., Most et al. 2004). Despite the abundance of hazards that could potentially impact human health and welfare, air pollution was one of the first public
concerns to be directly addressed by legislation in the U.S. with the passage of the Clean Air Act in 1970. Inequities in the spatial distribution of air pollution have been studied since the 1970s (Freeman 1972, Asch and Seneca 1978, and Gianessi et al. 1979), even before the emergence of the environmental justice movement and research agenda in the 1980s. In addition to often being colorless and odorless, air pollution can be difficult to ascertain due to its quick dispersion, leaving those impacted by it potentially unaware of its presence.

Analyzing the environmental justice implications of exposure to air pollution using quantitative methods requires: (a) the selection of the measurement used to detect the presence of emissions, and (b) methods for estimating the magnitude of exposure to the pollution. Both decisions come with potential problems that will be explored in this literature review. The first section will review research concerning the selection of the source of exposure, then the magnitude of the exposure to air pollution. The second section will provide an overview of common statistical methods used in environmental justice research to measure relationships between variables.

2.1 Measuring Exposure to Air Pollution

2.1.1 From Individual Source Analysis to a Cumulative Approach

Air pollution is the amount of contaminants in the air or the presence of one or more contaminants at concentrations high enough to cause adverse health effects (Godish 2004). Air pollution is produced by many different types of sources, and while large factories that release ominous plumes may attract more policy and media attention, other sources of air pollution that have a negative impact on public health are just as important to assess. In addition to point sources such as factories and power plants, there are less
conspicuous point sources like auto body and paint shops and dry cleaning facilities. There are also mobile sources, including cars, trucks, and motorcycles, and off road mobile sources such as airplanes, all-terrain vehicles, and snowmobiles. Additionally, naturally occurring and lingering background concentrations of chemicals in the air can contribute to pollution levels. Environmental justice research, however, has analyzed major point sources more than any other type of source due to the ease of availability of data and their conspicuous presence.

Facilities listed in the EPA’s annual Toxic Release Inventory (TRI) represent one specific point source that has been studied extensively in the environmental justice research literature. This database includes industrial manufacturing facilities that employ more than ten people and either manufacture or process more than 25,000 lbs of chemicals. TRI sites and emissions data are particularly well suited for environmental justice analysis. Facilities that meet certain threshold requirements must report annually to the EPA the quantities of more than 600 toxic chemicals that they release directly to air, water, or land, and/or transport to offsite facilities. This means that a central agency can collect and manage all data nationwide. TRI data is also publicly accessible and fairly straightforward to download and use. Since the 1990s, the TRI has emerged as the most comprehensive data source on industrial toxic emissions in the U.S. (Chakraborty 2004).

The environmental justice impacts of TRI facilities and releases have been studied in Florida (Pollock and Vittas 1995), South Carolina (Cutter et al. 1996, Mitchell et al. 1999), California (Pastor et al. 2004), New Jersey (Mennis and Jordan 2005), Texas (Tiefenbacher and Hagelman 1999), Des Moines, Iowa (Chakraborty and Armstrong 1997), Minneapolis, Minnesota (Sheppard et al. 1999), U.S. counties (Daniels and
Friedman 1999), U.S. states (Chakraborty 2004), and U.S. metropolitan areas (Ash and Fetter 2004). The scope and resolution of the studies vary, but TRI facilities remain an important part of environmental justice assessment.

An exclusive focus on TRI facilities or other large industrial sources, however, is likely to bias the assessment of environmental risk burdens within a community, because emissions from automobiles, smaller industries, and various other sources that contribute to air pollution are ignored. Smaller point source emitters have received less attention from environmental justice researchers. Fitos and Chakraborty (2003) examined inequities in the spatial distribution of dry cleaning facilities in Hillsborough County, Florida. They found that drycleaners are less likely to locate inside minority and impoverished neighborhoods but more likely to be found in densely populated areas near these neighborhoods. The authors pointed out that while drycleaner emissions may be less apparent and less regulated than those from more conspicuous industrial facilities, small point source polluters can be dangerous due to their tendency to be located near residential neighborhoods, increasing the risk of exposure. Also, small point sources do not fall under as strict EPA guidelines and monitoring may be nonexistent. Other studies that analyze non-TRI point sources include a study in Orlando, Florida that utilized the EPA’s aerometric information retrieval system (AIRS) in order to assess the environmental justice implications of children’s exposure to air pollution (Chakraborty and Zandbergen 2007). AIRS contains data for automobile mechanics, dry cleaning facilities, hospitals, and manufacturing facilities too small to be regulated as a TRI facility. By including small and large point sources, the researchers provided a more comprehensive profile of the present risks.
Point sources, large or small, are easier to analyze because they are stationary. The presence of air pollution from mobile sources, however, is difficult to assess, therefore distance to roadways is commonly used as an indication of risk. In a case study of New York City, Jacobson et al. (2005) used proximity to highways as an indicator of exposure, recognizing that proximity to a highway can be both a benefit and a health hazard. Using advanced simulated pollution models, Kingham et al. (2007) analyzed the environmental justice implications of traffic related air pollution in Christchurch, New Zealand. They found evidence of distinct socioeconomic inequity in the distribution of vehicle emissions, with the areas that had the highest vehicle ownership demonstrating lower air pollution concentrations. Wu and Batterman (2006) assessed the environmental justice implications of the locations of schools near high traffic roadways in the Detroit area. The study found that Black and Hispanic children were more likely to attend schools near these high traffic roads, which can increase the chances of various health problems such as asthma. As with point sources, the mobile sources analyzed in the reviewed literature required simplistic assumptions about the nature and environmental fate of emitted air pollutants.

From this overview of source-by-source studies, it is evident that environmental justice analysis could benefit from a more cumulative approach that incorporates multiple pollution sources. Recent studies have acknowledged the need to include multiple types of sources and emissions that can be found in a given study area. Morello-Frosch et al. (2001) modeled estimated lifetime cancer risks derived from TRI sites, small point sources, and automobile emission data released from the EPA’s Cumulative Exposure Project (CEP). According to these authors, small point and mobile sources are
problematic to regulate due to their dispersion and diversity, but the lack of guidelines makes their health impacts all the more important to consider. Concern over the absence of detailed information on multiple sources led several studies to use a new data set released by the EPA that incorporates both mobile and point sources (Pastor et al. 2005, Apelberg et al. 2005, Linder et al. 2008). However, these studies compared 1996 emission data with Census 2000 socio-demographic data - a temporal mismatch that can be avoided by using more recent information on air pollution. In light of these points, this thesis uses a data set from 1999 that incorporates multiple sources of outdoor air pollution in order to create a profile of resulting adverse health risks within the state of Florida.

2.1.2 Measuring the Extent of Exposure to Air Pollution

Environmental justice research on air pollution distribution began by measuring the adverse risk from air pollution as a function of distance to the pollution source. While many studies disclaimed their findings with the caveat that proximity to pollution did not necessarily reflect the actual health risks imposed on the population, this proxy for risk was widely used. The earliest studies used proximity to hazardous facilities as an indicator for the risks minority populations encountered, regardless of the ways these pollutants and facilities may adversely impact public health (e.g., Asch and Seneca 1978, Gianessi et al. 1979). These studies asserted that the mere presence of these facilities had a detrimental effect on the community economically and socially by implying that the neighborhood is not a desirable place in which to live.

The study of air pollution with regards to its impact on the well being of the surrounding community is different because immediate dispersion of releases makes
emissions harder to contain and keep track of. Ambient pollution is not intended to remain at the location in which they are released. Nevertheless, Pollock and Vittas (1995) used proximity to facilities as an indicator of risk in their study of toxic release inventory (TRI) sites in Florida. The analysis explored the environmental justice implications of potential pollution exposure to TRI emissions, measuring exposure as a non-linear function of distance to the facility. While the authors did acknowledge that risk of exposure is not fully explained by distance from the source, no possible avenues for improvement were provided.

Several researchers have used GIS-based circular buffers around point sources of pollution to obtain a more valid or consistent spatial definition of the affected area, instead of relying on administrative units such as census tract or zip code boundaries. Glickman (1994) explored how a GIS can be utilized to incorporate actual risk into EJ analysis, not just proximity to the source. Part of the analysis uses a proximity-based measurement in the form of buffers placed around each source to create a circle of potential risk and also includes risk-based measurements derived from information from extremely hazardous substances facilities. Perlin et al.’s 1999 study also used circular buffers around TRI facilities in block groups, studying the Kanawha Valley in West Virginia, the Baton Rouge to New Orleans corridor in Louisiana, and the Baltimore, Maryland metropolitan area. Buffers of radii 0.5, 1.0, 1.5, and 2.0 miles were placed around the sources in order to assess the impacts these arbitrary distances had on the results. Although racial/ethnic and economic inequities were found in all these areas, these authors acknowledged that proximity to TRI facilities does not necessarily mean exposure to pollution is the result.
Keeping with the trend of moving towards the buffer analysis method in order to accurately predict the location of impacted areas and populations, Sheppard et al. (1999) experimented with buffer sizes in order to assess the inequity associated with TRI locations in Minneapolis and to ascertain the statistical outcomes of different buffer distances on the results. Buffers in all these studies are used to determine the exposed area around a facility, moving away from municipal boundaries such as census tracts that may not accurately reflect the range of the air pollution plume.

Chakraborty and Armstrong (1997) focus on the environmental fate and transport of pollution after its release and translating that information to spatially define areas potentially exposed to adverse health risks. These authors asserted the circular buffers around air pollution emitters used in other studies may be inadequate because chemical dispersion plumes are not released in a perfect circle, but are affected by local wind patterns. This is important because while some people may be in closer proximity to the source, it does not mean they are in more danger when circumstances of release such as wind speed and direction is taken into account. Also, buffer sizes are based on arbitrary distances that may not reflect the nature and quantity of substances emitted (Chakraborty and Armstrong 2004).

Research concerning the most effective way to approximate areas exposed to air emissions contributed to exposure assessment by moving beyond mere proximity based analysis towards more sophisticated methods that incorporate the behavior of air pollutants and their known health effects. Environmental justice researchers have begun to incorporate more accurate estimates of health risks that accompany exposure. In 2001, Morello-Frosch et al. confronted this challenge of making the connection between the
presence of a polluting facility and the resulting health risks in Southern California. Air toxics concentration estimates were gathered from the U.S. EPA’s Cumulative Exposure Project, which estimates the annual average outdoor concentrations of a class of pollutants. The researchers then analyzed lifetime cancer risks derived from data obtained from the U.S. EPA and the California Air Pollution Control Officers Association. This article makes an important step towards understanding why the presence of pollution alone is not enough to assess risk and health risk estimates is a better measure of potential inequities.

Apelberg et al. (2005) analyzed health risks derived from EPA estimates at the census tract level in Maryland. The risk studied was lifetime cancer risk for 29 hazardous air pollutants. This database incorporated fate and transport of the chemicals, the resulting ambient exposure, and the associated health risks for each census tract. Pastor et al. (2005) utilized the same database for a study on California and Linder et al. (2008) worked with an updated and expanded version of the data to assess cancer risks in the Houston area. These three studies have utilized the most complete profile of the actual health risks associated with air toxics for environmental justice analysis. While these studies incorporated health risk estimates, they did not examine the environmental justice implications of health risks from different emissions sources.

This thesis project uses the most recent version of the database used by Apelberg et al. (2005) and Pastor et al. (2005) and the one used by Linder et al. (2008). As described further in Chapter 3, this advanced database of pollution emissions endeavors to improve upon past methodologies and provide a more precise portrait of adverse health risk from air pollution by incorporating local meteorology, chemical dispersion patterns,
and circumstances of release. This information will be used to derive human exposure concentrations and estimated cancer risks for four different sources of air toxics.

2.2 Statistical Methods

Once the source of environmental pollution and the method for approximating exposure to the risk is selected in a quantitative environmental justice study, a statistical method can be used to analyze the collected data and to draw conclusions about the inequities in the demographic and socioeconomic distribution of exposure. These methods can range from straightforward descriptive and correlation statistics, to multiple regression models, and newly emerging statistical methods that account for spatial processes and effects.

Correlation analysis can be used to explore the linear relationship between individual variables, such as the association between the percentage of Black residents and the quantity of toxic air releases. Correlation can provide evidence of positive or negative associations between two variables, which can be useful in order to seek evidence to support an inequity hypothesis. Bivariate correlation tests have been used by Bowen et al. (1995) to analyze TRI locations and emissions in Ohio, by Cutter et al. (1996) to evaluate the location of TRI facilities, toxic storage and disposal sites, and inactive hazardous waste sites in South Carolina, and by Tiefenbacher and Hagelman (1999) to assess acute and chronic toxic air releases in urban counties of Texas. However, in situations where variables may overlap with each other, bivariate correlation statistics do not account for how the effect of an independent variable (e.g., minority percentage and poverty rate) may change in the presence of another. These correlation tests can be
used for initial exploration of data, but is primarily used today in variable identification in order to improve subsequent statistical analysis.

Multivariate regression methods express how the relationship between the dependent variable and an explanatory variable changes in the presence of several other explanatory variables in the same model. Environmental justice research encompasses social, economic, and demographic explanatory variables that are often related to each other and difficult to isolate. Prior studies have used the multiple regression approach to analyze the equity implications of hazardous waste sites (e.g., Boer et al. 1997), TRI sites (e.g., Bowen et al. 1995, Daniels and Friedman 1999, Mennis 2002, Pastor et al. 2004) Superfund sites (e.g., Bevc et al. 2005) and treatment, storage, and disposal facilities (e.g., Mennis 2002, Zahran et al. 2008). Multivariate regression analysis also allows for the creation of multiple combinations of explanatory variables that can provide additional depth to the exploration of the relationship between the nature or extent of pollution and relevant explanatory factors.

A problem arises, however, when conventional linear regression is applied to spatial data. Standard regression produces a global model where a single equation is provided to explain all variation in the relationships between the dependent and explanatory variables. The processes being examined are thus assumed to be constant or stationary over space. In the context of environmental justice analysis, there are two reasons to consider spatial variability in the relationships between variables (Mennis and Jordan 2005). First, differences may occur due to misspecification of the model because of missing variables such as metropolitan designation. However, the attempt to account for such variation can lead to oversimplification of regional differences and the
assumption that it is possible to categorize complex local characteristics. Second, there
may be nothing wrong with the model; the variation can be attributed to the unique
characteristics of different locations.

There has been a very recent trend towards the application of spatially oriented
statistical methods. Geographically weighted regression (GWR) is a local spatial
statistical technique used to analyze spatial nonstationarity, or when the measurement of
relationships among variables differs from location to location (Fotheringham et al.
2002). The analytical utility of GWR has been demonstrated in geographic studies of
disaster declarations (Schmidlein et al. 2008), poverty in urban areas (Longley and Tobon
2004), patterns of violent crime (Cahill and Mulligan 2007), and local politics and voting
(Calvo and Escolar 2003). These studies utilized GWR because of the inherent spatial
nature of the data and the relationships being analyzed.

It is important to consider that environmental justice is an explicitly spatial
problem, concerned with the geographic distribution of exposure to pollution and its
relationship to explanatory variables that are rarely distributed uniformly across any
study area. However, only one published study has used GWR to examine the
environmental inequity hypothesis (Mennis and Jordan 2005). These authors applied
GWR in order to assess the relationship between the location of TRI facilities and
race/ethnicity and other socioeconomic variables in the state of New Jersey. GWR
allowed Mennis and Jordan to map model parameters that represented the statistical
relationship between the dependent and relevant independent variables. The resulting
maps showed significant racial/ethnic inequity in the distribution of TRI facilities in
urban areas of the state, while simultaneously finding a lack of inequity for Blacks in
Newark and evidence of disparity for Hispanics in rural Vineland. The results provided different explanations for the presence of these facilities in each part of the state. This study was limited by its use of a single pollution source (TRI sites) and proximity as a surrogate for exposure and adverse health risk-- limitations that will be addressed in this proposed thesis.

2.3 Summary

This literature review has traced the methodological history of quantitative environmental justice research in order to elucidate three problems encountered in the search for answers to questions of environmental inequity. First, there has been a lack of cumulative source assessment, with many studies focusing on human exposure to only one pollution source, despite the potential oversight of equally dangerous sources. Past empirical studies have focused primarily on inequities associated with major stationary sources of pollution, thus ignoring mobile emission sources and smaller emitters that also pollute the local environment. The failure to consider emissions from automobiles, smaller industries, and other less conspicuous sources that contribute to air pollution is likely to distort the assessment of inequities in the distribution of environmental risk burdens. Several recent environmental justice studies have emphasized the need for going beyond locational inequities and a focus on stationary emission sources towards a more cumulative exposure approach that considers the health risks that a community may face from various types of pollutants and emissions sources. A risk modeling approach that considers multiple sources of pollution is also consistent with the emerging policy focus on cumulative exposure assessment. Second, estimating exposure to air pollution has been problematic due to the difficulty in obtaining the abundance of data necessary to
model emissions and the resulting health risks. A large number of studies have used proximity to hazardous facilities or pollution sources as a proxy for potential human exposure, instead of assessing the nature and quantity of pollutants emitted, local meteorological conditions, and other factors that influence exposure. While specific studies focused on modeling exposure to toxic pollution, few have attempted to examine whether unequal exposure patterns lead to disproportionate health risks among minority and low-income communities. Finally, commonly used statistical methods produce a single model to explain all variation in the data thus ignoring the fact that statistical associations between environmental health risks and explanatory factors can vary from place to place within a study area. Geographically weighted regression, a technique that examines local variation in model parameter estimates, has rarely been utilized in environmental justice studies to investigate the spatial non-stationarity of model parameters and understand local and regional differences.

These three gaps in environmental justice methodology present an opportunity for further research and improvement. The following chapter outlines the data sources and methodology used in a case study that focuses on evaluating the geographic distribution of estimated health risks from air pollution in the state of Florida.
CHAPTER 3: STUDY AREA AND METHODOLOGY

This section describes the study area, data sources used, and the methodology followed in order to assess the equitable distribution of various known sources of cancer risk in Florida. First, the study area is introduced and the source of the data and the process used to derive the key variables are outlined. Next, the variables used in the case study are defined and described, along with their data sources. Finally, the methods that were chosen to address the limitations of previous studies are explained.

3.1 Study Area

The geographic scope of this analysis includes the entire state of Florida. Florida provides a suitable setting for a study concerning the environmental justice implications of air pollution, in part, because of its rapidly growing population. It is the fourth largest state in the U.S. in terms of total population and one of the fastest growing states in the nation. Since 1970, nine million people have moved to Florida and an average of 800 individuals relocate to the state daily (Chapin 2006). This state experienced a 23.5 percent increase in population from 1990 to 2000 accompanied by swift economic growth in the past few decades that has led to unchecked commercial, residential, industrial, and infrastructure development in some cases. As Florida continues its growth, it is important to ensure that growth for some residents of Florida does not mean adverse and disproportionate health risks for others. Florida is also a diverse state in terms of race/ethnicity, with the three largest minority groups, Hispanic, non-Hispanic Black, and Asian collectively comprising over a third of the entire population. As the populations of
these minority groups continue to grow, the urgency to assess the environmental justice implications of the health risks from air pollution becomes even more pronounced.

Systematic quantitative research on the environmental justice implications of exposure to air pollution has been limited at the state level. Past environmental justice studies in the state of Florida include Pollock and Vittas (1995), who focused on analyzing distance to TRI facilities in the state, based on demographic and socioeconomic characteristics of census block groups. Racial and ethnic minority sub-populations, particularly African American households, were found to reside closer to polluting sources. Stretesky and Hogan (1998) examined the spatial relationship between Superfund sites and socio-demographic characteristics of census tracts surrounding these sites in a longitudinal study of Florida. They found that Blacks and Hispanics were more likely to live near these sites and this association increased with time. While these previous studies have investigated inequities associated with major point sources, more research is necessary to analyze exposure to other types of pollution sources in Florida. By utilizing the 1999 NATA, this thesis will attempt to improve the information available by including estimated health risks and incorporating several different sources of emissions.

3.2 Data Sources and Variables

The Clean Air Act of 1990 separated air pollutants into two distinct categories, criteria air pollutants and air toxics. Criteria air pollutants are a narrow classification of common pollutants including ozone, sulfur dioxide, carbon monoxide, nitrogen dioxide, particulate matter, and lead. The EPA has set criteria, or threshold levels, at which these substances become potentially harmful to public health. Air toxics, however, do not have
set levels at which the EPA deem them dangerous; instead, any exposure to air toxics is considered unsafe. Air toxics are a group of 188 air pollutants that are known to cause or are suspected of causing cancer and other serious health problems such as respiratory, reproductive, and neurological damage. Sensitive populations such as the elderly and children are particularly vulnerable to health risks caused by air toxics. Air toxics include metals such as cadmium, mercury, and chromium, as well as chemicals such as asbestos, benzene, and dioxin.

While the environmental justice implications of criteria air pollutants have been extensively studied (Wernette and Nieves 1992; Jerrett et al. 2001; Kingham et al. 2007), air toxics have received less attention. Beginning in 2002, the EPA developed a nationwide database describing the release of and health risks associated with air toxics. The first release of the National-Scale Air Toxics Assessment (NATA) profiled emissions from 1996. This assessment included 33 air pollutants identified as the most dangerous to public health in the largest urban areas (EPA 2008b). Because the EPA updates its emission databases every three years, the next assessment featured 1999 data. The 1999 NATA increased the number of air toxics included in the database to 177, with health risk estimates for 133 of these pollutants. The NATA estimates health risks at the census tract level providing a geographically detailed resolution that can be matched with population and housing data from Census 2000.

The 1999 NATA utilizes a four-step process in order to obtain and calculate health risk estimates at the census tract level. First, the EPA gathers emission data and estimations directly from the sources of pollution as well as state and tribal agencies responsible for tracking outdoor emissions. For some sources, the EPA estimates
emissions using models and measurements. This information is compiled into the National Emissions Inventory (NEI). The NEI draws its data from four sources: major point sources, such as industrial facilities and waste incinerators, minor point sources, including dry cleaning facilities and auto-body shops, mobile on-road sources like cars and trucks, and mobile off-road sources, including boats and all-terrain vehicles. Next, ambient air concentrations are estimated using a Gaussian dispersion equation known as the Assessment System for Population Exposure Nationwide (ASPEN) model. Using the 1999 NEI as the input data, the ASPEN computer model estimates the concentrations in every census tract in the U.S. by factoring in location of release, height of release, local meteorological conditions, and the behavior of the particular chemical once it is released, including the conversion of one chemical into another and the settling out of chemicals in the atmosphere. Background concentrations, or levels of pollutants that preexist in the environment either due to natural formation or lingering presence from previous release, are also included. Third, inhalation exposure is estimated utilizing the Hazardous Air Pollutant Exposure Model, Version 5 (HAPEM5). This model incorporates ASPEN ambient air pollution estimates, average indoor/outdoor activity patterns, climate data, and census data to derive estimated inhalation exposure. Finally, the EPA estimates specific types of health risks associated with inhalation of air toxics according to current data on health effects, EPA risk assessment characterization, and exposure estimates (EPA 2008c).

3.2.1 Dependent Variables

The result of the EPA’s four-step process is a database containing tract-level estimates for three types of public health risks: cancer risk, respiratory risk, and
neurological risk. Each risk is further divided by emission source category: major point sources, minor point sources, on-road mobile sources, off-road mobile sources, and background concentrations. The NATA also provides cumulative exposure estimates for each census tract that collectively includes all sources of toxics. This analysis will be focusing on estimated lifetime cancer risks.

Cancer risks are expressed as individual lifetime excess risk from a lifetime of exposure to one unit of a pollutant. The NATA estimates cancer risk on the basis of the inhalation unit risk (IUR) factor, a measure of the cancer-causing potential of each air toxic. Lifetime cancer risk for each pollutant in each census tract is calculated by:

\[ R_{ij} = C_{ij} \times IUR_j \]

Where

\[ R_{ij} = \text{estimate of individual lifetime cancer risk from air pollutant } j \text{ in census tract } i \]

\[ C_{ij} = \text{Concentration in micrograms of pollutant per cubic meter of air of air pollutant } j \text{ in census tract } i \]

\[ IUR = \text{inhalation unit risk estimate for pollutant } j. \]

Cancer risks are assumed to be additive and lifetime cancer risk from all air toxics present in a tract are summed to obtain the total estimated lifetime cancer risk for the tract. The estimated lifetime cancer risk variable is expressed as \( N \) in one million people, or the number of people expected out of one million to contract cancer if exposed to the 1999 concentrations estimated for that census tract for 24 hours a day over seventy years, which is the EPA’s estimate of an average life span. These estimates are in addition to cancer cases unrelated to air toxic exposure, not including them (EPA 2008d).
In order to explore spatial inequities in cancer risk from outdoor air pollution in detail, this thesis analyzes cumulative lifetime cancer risk from all sources in Florida, as well as risk from the four following individual sources of air toxic emissions:

- **Major point sources:** These are defined by the Clean Air Act as facilities that have the potential to emit 10 tons of one air toxic or 25 tons of a combination of air toxics.

- **Other point sources:** These are also known as area/other point sources and include facilities below the ‘major’ threshold levels such as auto body shops and dry-cleaning facilities and also include sources such as wildfires.

- **On-road mobile sources** include motorized vehicles that normally operate on public roadways and comprises passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses.

- **Non-road mobile sources** are mobile sources not found on roads and highways, such as trains, airplanes, lawn mowers, construction vehicles, and farm machinery (EPA 2008e).

### 3.2.2 Explanatory Variables

All explanatory variables for this study include well-documented variables from previous environmental justice studies extracted from Census 2000 data at the census tract level. These are defined in Table 3.1, along with their expected statistical relationship with the dependent variables. The explanatory variables can be divided into three categories: racial/ethnic variables, socio-demographic variables, and density. In light of the primary concern of race and ethnicity in environmental justice research, the three largest racial/ethnic minority groups in the state of Florida will be included. Persons self-identifying as non-Hispanic Black, Hispanic/Latino of any race, and Asian together
make up about 37 percent of the population in Florida (U.S. Census 2006) and their inclusion in this study is crucial to exploring equity. Racial and ethnic minorities are generally expected to have increased exposure to health risks when compared to their white counterparts due to discrimination and institutional racism in the housing market (Pulido 2000, Pastor et al. 2001).

Urban areas have a legacy of residential segregation in the U.S. and Florida that invariably contribute to the current location pattern of the Black population at the metropolitan scale. Lingering racial discrimination in the housing market and employment opportunities, along with various educational and economic disparities, can lead to limited housing choices and restrict their locations to areas disproportionately exposed to adverse health risks from technological hazards (Stretesky and Hogan 1998, Pastor et al. 2001). While Hispanic populations do not have as deeply ingrained legacies of urban segregation, these communities can experience similar outcomes and exposure to toxic pollution (Pulido 2001). Particularly in Florida, migrant Hispanic workers may be at increased risk of housing discrimination, possibly contributing to increased health risks (Pastor et al. 2005). While Asians have been included in the minority category in previous environmental justice research, few studies have separately analyzed the relationship between Asian populations and environmental health risks (Pastor et al. 2001). This study assumes similar racial discriminatory practices are at work and examines whether Asian individual are disproportionately exposed to environmental health risks.

Several socioeconomic variables will be included to provide other avenues through which inequity can occur and to ensure that any unjust distribution found among
minorities is not due to differences in economic status across the various racial/ethnic groups. The proportion of residents below the federal poverty level indicates financial resources, which can become a limiting factor not only with regards to the areas available in which to live, but the ability to move out of tracts that may become more risky over time. The proportion of homes owned by their occupants is another economic indicator that gives an impression of immediate financial resources within a tract and residents’ wealth and power. Some researchers use home ownership as an indicator of political participation (Pastor et al. 2005), believing that home ownership leads to increased investment in local decision-making processes. The environmental justice literature indicates that tracts with relatively lower levels of home ownership and higher poverty rates will be expected to have a higher risk of exposure to the adverse health effects caused by air toxics (Boer et al. 1997, Morello-Frosch et al. 2001, Pastor et al. 2005).

Population density is a commonly used control variable in environmental justice studies because densely populated areas are more likely to contain polluting facilities and activities. While population density is commonly measured as the number of people per square mile, the natural logarithm of this value will be taken in order to account for the diminishing effect of higher numbers, as suggested by Mennis (2002) and Pastor et al. (2005).

The explanatory variables (proportion non-Hispanic Black, proportion Black, proportion Asian, proportion below poverty, proportion owner occupied, and population density) have been obtained from the 2000 Census at the tract level, matching up with the resolution and time frame of the 1999 NATA dependent variables.
Table 3.1 Definitions of Explanatory Variables and Expected Statistical Association with Cancer Risk.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Expected Statistical Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Non-Hispanic Black</td>
<td>Ethnically non-Hispanic individuals identifying themselves as Black expressed as a proportion of the census tract’s total population.</td>
<td>Positive</td>
</tr>
<tr>
<td>Proportion Hispanic/Latino</td>
<td>Individuals identifying themselves as Hispanic/Latino (of any race) expressed as a proportion of the census tract’s total population.</td>
<td>Positive</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>Ethnically non-Hispanic individuals identifying themselves as Asian expressed as a proportion of the census tract’s total population.</td>
<td>Positive</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>Proportion of the population within a census tract with an annual family income below the federal poverty level.</td>
<td>Positive</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>Proportion of occupied housing units that are owner-occupied within a census tract.</td>
<td>Negative</td>
</tr>
<tr>
<td>Population Density</td>
<td>Number of people per square mile in a census tract.</td>
<td>Positive</td>
</tr>
</tbody>
</table>

3.3 Methods

In order to assess the distribution of estimated cancer risks from air toxics in Florida, this thesis project implements three stages of statistical analysis. First, the linear association between the dependent and each explanatory variable is measured using Pearson’s correlation coefficient at the census tract level. Bivariate correlations provide an initial indication of the relationships that exist between the health risks and the various racial/ethnic, and socioeconomic variables utilized in this study. Conventional multiple regression is used to develop five sets of models, one for each source category of estimated cancer risk and one for cumulative cancer risk from all sources, based on the ordinary least squares method. These models express the relationship all independent
variables possess with each source of cancer risk simultaneously, at the census tract level.

The multivariate regression models can be summarized by the following equation:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + \epsilon \]

Where \( y \) is the dependent variable, \( x_1, x_2, \ldots, x_k \) are the independent variables, and \( \beta_0, \beta_1, \beta_2, \ldots, \beta_k \) are the model parameters.

The model parameters indicate the nature and strength of the association between the particular independent variable and the dependent variable, negative or positive, when all other variables are also taken into account. Multivariate regression allows the influence each independent variable exerts over the dependent variable to be examined, after controlling for the effects of the other independent variable.

While multivariate regression has been used extensively in environmental justice studies, this methodology does not account for geographic variations in signs and coefficients of model parameters with a given study area, as described previously. In order to detect trends within the state of Florida, the second phase of analysis uses geographically weighted regression (GWR), a local spatial statistical technique used to analyze spatial nonstationarity (Fotheringham et al. 2002). Instead of calibrating a single or global regression equation, GWR generates a separate regression equation for each observation. Each equation is calibrated using a different weighting of the observations contained in the data set. Each GWR equation may be expressed as

\[ y_i = \beta_0(a_i, r_i) + \beta_1(a_i, r_i)X_{i1} + \beta_2(a_i, r_i)X_{i2} + \ldots \beta_k(a_i, r_i)X_{ik} + \epsilon_i \]

Where \((a_i, r_i)\) is the location of \( i \). The assumption is that nearby observations have greater influences on one another’s parameter estimates than observations farther apart. The weight assigned to each observation is based on a distance decay function that is centered
on observation \(i\). In the case of areal data (e.g., census tracts), the distance between observations is calculated as the distance between polygon centroids. The distance decay function, which may take multiple forms, is modified by a bandwidth setting at which distance the weight approaches zero.

GWR software version 3.0.18 provides the user with three options to assist with calibration of the model to achieve optimal results. Generally, all observations are weighed by a function that decreases with distance from point \(i\). This weighted window is referred to as a kernel.

\[
W_{ij} = \exp\left[-\frac{d_{ij}}{b}\right]^2
\]

Where \(W_{ij}\) is the weight of data point \(j\) at regression point \(i\), \(d_{ij}\) is the Euclidean distance between \(i\) and \(j\), and \(b\) is the bandwidth. The farther away data point \(j\) is from regression point \(i\), the lower its influence will be. Choosing the appropriate bandwidths is essential to producing a dataset that will not oversimplify regional differences by being too large, or be too detailed to provide a clear pattern by being too small. The bandwidth can be manually set by the user after experimenting with previous models, or it can be automatically selected by cross-validation or minimizing the Akaike Information Criterion (AIC), which is the method used in this study (Fotheringham et al. 2002).

\[
AIC = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{\frac{\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)}}{n}\right\}
\]

Where \(n=\) the sample size, \(\hat{\sigma}\) = the estimated standard deviation of the error term, and \(\text{tr}(S)\) = the trace of the hat matrix which is a function of the bandwidth. The lower the AIC, the better the fit of the model. The AIC method has the advantage of considering the
fact that the degrees of freedom may vary among regression models centered on different observations.

In the application of GWR, the user may choose a fixed bandwidth that is used for every observation (e.g., Mennis and Jordan 2005) or a variable bandwidth that expands in areas of sparse observations and shrinks in areas of dense observations (e.g., Calvo and Escolar 2003). This analysis uses a variable kernel bandwidth that adapts for the density of data in each location due to the variability in the sizes of census tracts.

Because the regression equation is calibrated independently for each analytical unit, a separate parameter estimate, t-value, and goodness-of-fit is calculated for each census tract in Florida. These values are mapped using GIS software, thus providing a way to visually interpret the geographic distribution of the nature and strength of the relationships between explanatory and dependent variables. Indeed, in many cases the GWR output would be inscrutable without the ability to map the results. Fotheringham et al. (2002) describe several different methods for displaying the data, including point symbols, area symbols, contour plots, and pseudo-3D displays. This study utilizes choropleth mapping techniques to clearly display significance of local regression coefficients and is also suitable for comparing descriptive choropleth maps of the dependent variables. These maps provide the basis for local and regional analysis of the relationship between cancer risks from various sources of air toxics and various explanatory variables, as well as comparisons between the analyzed results of conventional regression and GWR.
CHAPTER 4: CONVENTIONAL ENVIRONMENTAL JUSTICE ANALYSIS OF ESTIMATED CANCER RISKS

This chapter focuses on the use of traditional statistical techniques to explore the environmental justice implications of estimated cancer risk from inhalation exposure to air toxics in Florida. First, descriptive choropleth mapping and summary statistics are used to explore the spatial distribution of the dependent and explanatory variables at the census tract level. Second, bivariate correlations are derived to examine the statistical association between each explanatory variable and the five dependent variables: lifetime estimated cancer risk from all sources and the four known sources of ambient air toxics: major point sources, other point sources, on-road mobile sources, and non-road mobile sources. Finally, conventional multivariate regression analysis is employed to estimate each dependent variable as a function of the racial/ethnic and socioeconomic variables in a single model. By analyzing cumulative cancer risk and risks from four different sources of air toxics in Florida, it is possible to assess how statistical associations between relevant explanatory variables and magnitude of cancer risk varies by emission source.

4.1 Descriptive Choropleth Mapping and Statistics

The five dependent variables are displayed as classified choropleth maps in Figures 4.1 to 4.5. Census tracts in Florida are grouped into four quartiles based on estimated values of lifetime cancer risk. Lifetime cancer risk from all sources of air toxics is concentrated primarily in the largest metropolitan areas of Florida (Figure 4.1) such as Jacksonville, Orlando, Tampa, and Miami. Because most pollution-generating activities
take place in urban areas, this pattern is not surprising. However, when cancer risks are separated by release source, different patterns emerge. Cancer risk from major sources alone (Figure 4.2) is concentrated in both urban centers such as Jacksonville and in rural areas. Lifetime cancer risk from minor point sources (Figure 4.3) is comparably lower in the central southern portion of the state. Lifetime estimated cancer risk from both on-road mobile sources (Figure 4.4) and from non-road mobile sources (Figure 4.5) are highly concentrated in the metropolitan areas of Florida, reflecting areas of dense vehicular traffic and other modes of transportation such as airports and railroads.

**Figure 4.1** Estimated Lifetime Cancer Risk From All Sources by Census Tract, 1999
**Figure 4.2** Estimated Lifetime Cancer Risk From Major Point Sources by Census Tract, 1999

**Figure 4.3** Estimated Lifetime Cancer Risks From Minor Point Sources, 1999
Figure 4.4 Estimated Lifetime Cancer Risk From On-Road Mobile Sources, 1999

Figure 4.5 Estimated Lifetime Cancer Risk From Non-Road Mobile Sources, 1999
The geographic distributions of the six explanatory variables are summarized in Figures 4.6 to 4.11. These choropleth maps classify each variable into four categories based on the quartile method, and suggest substantial spatial variation for each variable. The Black population (Figure 4.6) is concentrated mostly in Florida’s panhandle region and major metropolitan areas. Additionally, some rural tracts near Lake Okeechobee show comparatively higher proportions of this variable. The Hispanic population (Figure 4.7) is concentrated in the southern portion of the state, both in urban and rural tracts. Higher proportions of Asian populations exist in and around major cities like Tampa, Orlando, and Jacksonville (Figure 4.8).

**Figure 4.6** Proportion Non-Hispanic Black by Census Tract, 2000
Figure 4.7 Proportion Hispanic by Census Tract, 2000

Figure 4.8 Proportion Asian by Census Tract, 2000
Figure 4.9 Homeowner Occupied Housing Units by Census Tracts, 2000

Figure 4.10 Proportion of Population Below the Federal Poverty Rate, 2000
Places with comparatively higher rates of home ownership are mostly in rural and suburban areas of Florida (Figure 4.9). Higher proportions of persons living below the poverty line can be observed in rural tracts as well as major urban centers (Figure 4.10). As expected, areas of high population density coincide with the largest cities in Florida (Figure 4.11).

Summary statistics for all variables used in this study are provided in Table 4.1, with frequency histograms included in Appendix A. The mean values of the dependent variables suggest that major point sources of air toxics pose the lowest cancer risk, while on-road mobile sources are responsible for the highest. Most of the explanatory variables suggest substantial variability in their values across census tracts in Florida. The average proportion of Blacks and Hispanics approximately equal 0.15, but these values range from 0 to 0.99 and 0.95, respectively. The Asian proportion has a much lower mean of only 0.02 and a considerably smaller range. The mean proportion of owner occupied
housing units (home ownership rate) is 0.63, while proportion below poverty has a mean of 0.13 among census tracts in the state of Florida. Population density appears to have the most skewed distribution, with a mean value of 3,223 people per square mile that is lower than its standard deviation and almost ten times smaller than its maximum value.

**Table 4.1** Descriptive Statistics for Variables Analyzed

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(estimated lifetime cancer risk in persons per million)</em>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Sources</td>
<td>3,154</td>
<td>37.295</td>
<td>0.000</td>
<td>108.133</td>
<td>11.998</td>
</tr>
<tr>
<td>Major Point Sources</td>
<td>3,154</td>
<td>0.463</td>
<td>0.000</td>
<td>32.397</td>
<td>1.335</td>
</tr>
<tr>
<td>Other Point Sources</td>
<td>3,154</td>
<td>5.818</td>
<td>0.000</td>
<td>54.77</td>
<td>3.728</td>
</tr>
<tr>
<td>On-Road Mobile Sources</td>
<td>3,154</td>
<td>9.295</td>
<td>0.000</td>
<td>81.995</td>
<td>5.628</td>
</tr>
<tr>
<td>Non-Road Mobile Sources</td>
<td>3,154</td>
<td>2.083</td>
<td>0.000</td>
<td>52.503</td>
<td>1.862</td>
</tr>
<tr>
<td><strong>Independent Variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion non-Hispanic Black</td>
<td>3,151</td>
<td>0.154</td>
<td>0.000</td>
<td>0.990</td>
<td>0.227</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>3,151</td>
<td>0.146</td>
<td>0.000</td>
<td>0.954</td>
<td>0.196</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>3,151</td>
<td>0.016</td>
<td>0.000</td>
<td>0.202</td>
<td>0.016</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>3,151</td>
<td>0.628</td>
<td>0.000</td>
<td>1.000</td>
<td>0.236</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>3,151</td>
<td>0.130</td>
<td>0.000</td>
<td>0.768</td>
<td>0.105</td>
</tr>
<tr>
<td>Population Density</td>
<td>3,151</td>
<td>3,223</td>
<td>0.000</td>
<td>34,289</td>
<td>3,500</td>
</tr>
</tbody>
</table>

### 4.2 Bivariate Correlation Analysis

Bivariate parametric correlations are used to assess the linear relationship between each explanatory variable and the five dependent variables, at the census tract level. Pearson’s correlation coefficients (*r*-values), presented in Table 4.2, indicate the strength and significance of each individual exploratory variable’s statistical association with estimated lifetime cancer risk, and do not account for the presence of other variables. The
The strength and direction of correlations differ when individual sources of estimated cancer risk are analyzed. Cancer risk from major point sources is positively and significantly correlated with the proportion of Black and Hispanic residents, as well as with poverty rate, which corresponds with the expected behavior of this variable in light of the environmental justice framework. However, the Asian proportion and population density are not considered significantly correlated. Home ownership rate behaves as expected and is negatively and significantly correlated with estimated cancer risks from major point sources. Statistical associations between lifetime cancer risk from other point sources and the proportion of Blacks, Hispanics, Asians, poverty rate, and population

| Proportion non-Hispanic Black | 0.253** | 0.054** | 0.280** | 0.232** | 0.089** |
| Proportion Hispanic | 0.467** | 0.084** | 0.240** | 0.401** | 0.342** |
| Proportion Asian | 0.195** | -0.028 | 0.087** | 0.148** | 0.072** |
| Proportion Owner Occupied | -0.251** | -0.042* | -0.186** | -0.257** | -0.231** |
| Proportion Below Poverty | 0.242** | 0.096** | 0.278** | 0.245** | 0.143** |
| Population Density (natural log) | 0.618** | -0.012 | 0.276** | 0.547** | 0.370** |

*p<0.05; **p<0.01
density remain positive matching the findings for cancer risk from all point sources. Additionally, home ownership rate remains negatively and significantly correlated with estimated cancer risk from other point sources. For both on-road and non-road mobile sources, home ownership rate demonstrates a significant and negative relationship, while all other variables indicate a significant and positive correlation with estimated cancer risk.

In summary, the explanatory variables display the expected statistical correlations not only with cumulative cancer risk from all sources of air toxics, but also with those from both types of mobile sources and minor point sources. Major point sources differ in that the proportion of Asians and population density do not exhibit a significant relationship with cancer risk. In the next section, multivariate regression is used to explore the relationship between each of the five dependent variables and the combined presence of all six explanatory variables.

4.3 Traditional Multiple Regression Analysis

In order to assess the simultaneous effects of all explanatory variables on each of the five dependent variables, conventional multivariate regression based on the ordinary least squares (OLS) method is utilized. These regression results are summarized in Tables 4.3 and 4.4. To check for multicollinearity, the condition index was calculated for each multiple regression model. None of the five models indicated a condition index greater than 30, suggesting the absence of serious collinearity problems.

Table 4.3 provides the regression results for cumulative cancer risk from all sources of air toxics. The ANOVA F-test indicates statistical significance for the overall model ($p<.001$) and the value of the adjusted multiple $R$-squared exceeds 50 percent,
suggested a reasonably high goodness-of-fit. After controlling for the effects of the other independent variables, population density and all racial and ethnic variables are significantly and positively associated with cancer risk, while home ownership and poverty rates are not significantly related to cancer risk from all sources of air toxics. These results match the bivariate correlations reported previously, with the exception of the non-significance exhibited by the socioeconomic variables.

Table 4.3 Multivariate Regression Results: Cumulative Cancer Risk From All Sources

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion non-Hispanic Black</td>
<td>12.51</td>
<td>13.13**</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>21.26</td>
<td>25.02**</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>82.86</td>
<td>8.61**</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>-1.26</td>
<td>-1.30</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>-3.15</td>
<td>-1.29</td>
</tr>
<tr>
<td>Population Density (natural log)</td>
<td>3.41</td>
<td>31.75**</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>F-statistic</td>
<td></td>
<td>566.87**</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>3145</td>
</tr>
</tbody>
</table>

*p<0.05; **p<0.01

The multiple regression models associated with the different sources of air toxics indicate substantial variation in model performance and significance of model parameters (Table 4.4). Although the ANOVA F-test indicates overall significance for all these OLS models, the adjusted R-squared values are generally smaller compared to the multiple regression for cumulative cancer risk. In presence of the other explanatory variables, only the Hispanic proportion and poverty rate show an expected and significantly positive effect on cancer risk from major point sources. The only other significant variable is the
natural log of population density, which exhibits a negative association with cancer risk.

All other explanatory variables are not significant in the regression model for major point sources. Additionally, this model yields the lowest R-squared value, indicating unsatisfactory goodness-of-fit.

**Table 4.4 Multivariate Regression Results: Cancer Risk from Four Known Sources**

<table>
<thead>
<tr>
<th></th>
<th>Major Point Sources</th>
<th>Minor Point Sources</th>
<th>On-Road Mobile Sources</th>
<th>Non-Road Mobile Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Black</td>
<td>0.11</td>
<td>0.70</td>
<td>3.64</td>
<td>9.37**</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>0.57</td>
<td>4.18**</td>
<td>3.73</td>
<td>10.75**</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>-0.62</td>
<td>-0.40</td>
<td>21.94</td>
<td>5.59**</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>0.10</td>
<td>0.66</td>
<td>0.44</td>
<td>1.10</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>1.01</td>
<td>2.59**</td>
<td>3.85</td>
<td>3.85**</td>
</tr>
<tr>
<td>Population Density (ln)</td>
<td>-0.04</td>
<td>-2.04*</td>
<td>0.36</td>
<td>8.25**</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.01</td>
<td>0.18</td>
<td>0.40</td>
<td>0.21</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.90**</td>
<td>118.48**</td>
<td>350.20**</td>
<td>141.71**</td>
</tr>
<tr>
<td>N</td>
<td>3145</td>
<td>3145</td>
<td>3145</td>
<td>3145</td>
</tr>
</tbody>
</table>

*p<0.05; **p<0.01

The regression model for estimated lifetime cancer risk from minor point sources indicates positive and significant effects for all three racial/ethnic variables, poverty rate, and population density, as indicated by the bivariate correlations. The R-squared value of 0.18 is higher than model for major point sources, but is considerably smaller than the model for cumulative cancer risk from all sources. The multiple regression model for on-road mobile sources shows a significant and positive relationship between cancer risk and proportion of Black, Hispanic, and Asian, residents, as well as population density. Home ownership rate behaves as expected, demonstrating a significant and negative relationship.
with cancer risk from on-road mobile sources of air toxics. This model yields a relatively high R-squared value of 0.40, indicating a more satisfactory fit than those based on point sources. The multivariate model for non-road mobile sources of cancer risk provides similar results to the one derived from on-road sources, with two exceptions. The proportion of Asians is not significant and proportion below poverty is negatively significant, which is incongruous with the expectations of environmental inequity.

In summary, the presence of Black and Hispanic residents and higher population density represent the most consistent and significant indicators of lifetime cancer risk from all emission source categories, after accounting for the independent effects of other explanatory variables. The socioeconomic variables suggest mixed results based on the source of air toxics. Cancer risk from point source emissions is significantly greater in areas of high poverty rate, while cancer risk from mobile sources is greater in areas of low home ownership. It is important to consider, however, that these conventional regression models assume spatial stationarity in the relationships between estimated cancer risk and each explanatory variable. Given the spatial variability observed in the choropleth maps for the dependent and independent variables, it is necessary to examine if the statistical relationships indicated by traditional regression (i.e., the five global models) remain consistent across the state of Florida. The next chapter thus focuses on the application of geographically weighted regression to analyze spatial variation in the OLS regression model performance and model parameters.
CHAPTER 5: GEOGRAPHICALLY WEIGHTED REGRESSION ANALYSIS OF ESTIMATED CANCER RISKS

This chapter focuses on exploring the statistical relationship between the dependent variables representing cancer risk and explanatory variables on the basis of geographically weighted regression (GWR), a method that allows a unique local regression model to be produced for every individual census tract in Florida. Again, cumulative cancer risk from all sources is evaluated, followed by cancer risks associated with the four known sources of ambient air toxic emissions. The local estimates of regression parameters and regression model performance provided by GWR are compared to the results of traditional multiple regression analysis that were reported in the previous chapter.

In order to determine the appropriate bandwidth for GWR analysis, an adaptive kernel method is used for this study, instead of a fixed bandwidth setting. Previous studies indicate that choosing a fixed bandwidth is difficult when census enumeration units are used, because these vary in size and shape according to the population density of the area (Mennis and Jordan 2005; Mennis 2006; Kavanagh et al. 2006). A spatially adaptive kernel selects an optimal number of neighboring units for the analysis, relying upon contiguity rather than distance to seek a specific number of nearest neighbors to ensure a constant size of local samples (Zuhuang 2006). Where census tract centroids are farther apart (e.g., rural areas), the bandwidth becomes larger, and when tracts are closer together (e.g., urban areas), the bandwidth becomes smaller. The optimal bandwidth for
this study is determined through an iterative process to minimize the Akaike information criterion (AIC), following previous applications of GWR (e.g., Longley 2004; Cahill and Mulligan 2007; Ali et al. 2007). In this method, the software calculates the AIC for several different local sample sizes, finally selecting the minimum value. The AIC is considered to be the most appropriate method for implementing the adaptive kernel method due to its ability to take goodness of fit and degrees of freedom into account (Fotheringham et al. 2002).

In order to maintain a consistent spatial definition of nearest neighbors for census tracts near the boundary of the study area, tracts from the neighboring states of Alabama and Georgia that are sufficiently close to Florida are also included in the GWR analyses. The data set used to estimate the GWR models for this study thus encompasses all tracts in Florida (n=3,151) and additional tracts from neighboring states (n=359) whose centroid falls within a 50-mile radius of the centroid of a Florida tract located at the state border. While the GWR analyses utilize 3,510 total observations, the results for only the census tracts within Florida are presented here to allow comparison with the respective global regression models.

The output for GWR includes model parameter estimates and t-values for each explanatory variable, along with local R-squared values to indicate model performance. While this is the same output provided by a global regression model, GWR produces this set of diagnostics for each data point, or in this case, a census tract. This makes it possible to map the regression results and visualize how relationships vary across space, rather than relying upon one set of parameters to express the relationship for an entire study area. In this chapter, the GWR results associated with each dependent variable (cancer
risk by source) are organized and summarized in the form of a table, a map depicting the spatial variability in model performance, and a series of maps that show the distribution of relevant regression model parameters across Florida.

5.1 Cumulative Cancer Risk from All Sources

The numerical results associated with the GWR model for cancer risk from all sources are summarized in Table 5.1. This table allows the global regression parameters presented in the previous chapter to be compared with the range of results produced by GWR. For example, traditional multiple regression indicated that cancer risk from all sources is positively and significantly associated with proportion Black among census tracts in Florida. While the global regression model estimated the parameter at 12.51 for the entire state (Table 4.3), the GWR model shows that it ranges from 0.09 to 32.97, with a median of 11.34 (Table 5.1). This variability in the model coefficient suggests that the relationship between cancer risk from all sources of air toxics and the proportion of Black residents is not static across the state.
### Table 5.1: GWR Results for Cumulative Cancer Risk from All Sources

<table>
<thead>
<tr>
<th>Activity</th>
<th>Global Regression Coefficient</th>
<th>GWR Coefficients</th>
<th>Statistical Significance of t-values Across Census Tracts in Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Median</td>
<td>Max</td>
</tr>
<tr>
<td>Proportion Black</td>
<td>12.51**</td>
<td>0.09</td>
<td>11.34</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>21.26**</td>
<td>-28.23</td>
<td>16.46</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>82.86**</td>
<td>-88.29</td>
<td>104.61</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>-1.26</td>
<td>-23.99</td>
<td>-4.82</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>-3.15</td>
<td>-63.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Population Density (Ln)</td>
<td>3.41**</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Adjusted R squared</td>
<td>0.52</td>
<td>0.14</td>
<td>0.47</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
<td>26186</td>
<td>23795</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05  **p<0.01

For each explanatory variable, the last three columns of Table 5.1 categorize the t-values produced by GWR for each Florida tract by the level of statistical significance (p<.10) and direction of the association. Because of the large sample size (n=3,151), standard t-values of -1.645 and 1.645 are used to represent the 90 percent level threshold for negative and positive significance, respectively. In this example, proportion Black is positively and significantly related to cancer risk from all sources in 93 percent of tracts and is not significant in 7 percent of tracts.

Although the Hispanic and Asian proportions were positively and significantly related to cancer risk from all sources in the global regression model, the GWR results in Table 5.1 indicate that these variable coefficients include both negative and positive values. For both variables, the statistical association with cumulative cancer risk is not significant in more than one-fourth of census tracts in Florida. Across a small percentage
of tracts, the t-values are negatively significant - the opposite relationship than what was indicated by the global regression model. For proportion below poverty, 60 percent of tracts indicate non-significant t-values, matching the global regression model result. However, the global model does not account for the fact that cancer risk has a positively significant relationship with poverty rate in more than ten percent of Florida tracts, which is consistent with the environmental injustice expectations. GWR model parameters for population density are divided almost equally between positive and negative significance, while the global regression model produced a positively significant coefficient for the entire study area. These variations in not only the significance (t-values), but also direction (coefficient signs), point to the need to assess statistical relationships for environmental justice analysis locally, rather than globally. The Akaike Information Criterion (AIC) is the recommended measure for comparing a global regression model with a GWR model (Fotheringham et al. 2002). As shown in Table 5.1, the GWR model yields a lower AIC score than the global model, indicating a superior fit or improvement in overall model performance.

To find evidence of spatial nonstationarity in the relationship between cumulative cancer risk and each explanatory variable, the interquartile range of an individual GWR parameter can be compared to twice the standard error of the global regression parameter (Fotheringham et al. 2002). For cumulative cancer risk from all sources, the interquartile ranges for all six explanatory variables from the GWR model were found to be larger than their respective standard error obtained from conventional regression, thus providing an informal confirmation of the presence of spatial nonstationarity for their estimated parameters across Florida.
To assess spatial differences in overall model fit between the GWR and global regression models, estimates of local model performance for each census tract in the form of R-squared values are depicted as a choropleth map in Figure 5.1. The red line in the map legend denotes the adjusted R squared value from the original global model, with two classes above and below this value. This map can be used to visualize and identify locations within Florida where GWR produces an improvement or decline in overall model fit. The local R-squared values improve upon the global R-squared value of 0.52 in a majority of tracts in Florida, with the GWR models performing the best in northern Florida and parts of central and southern Florida.

The final method of assessing spatially varying relationships are choropleth maps for each explanatory variable that represent the geographic distribution of their t-values across Florida, with respect to cumulative cancer risk (Figures 5.2 to 5.7). The choropleth classification is based on a manual distribution that separates both positive and negative t-values at the 90 percent and 95 percent level of significance. Because of the large sample size (n=3,151), standard t-values of +1.96 and +1.645 are used to denote 95 percent and 90 percent levels of significance, respectively. On all these maps, the red color is used to depict positive significance, the blue color is used to display negative significance, and the grey color is used to represent non-significant t-values.
Figure 5.1 Local R-Squared from GWR Model of Cancer Risk from All Sources
Figure 5.2 Distribution of t-values for Proportion Black and Cancer Risk from All Sources

Figure 5.3 Distribution of t-values for Proportion Hispanic and Cancer Risk from All Sources
Figure 5.4 Distribution of t-values for Proportion Asian and Cancer Risk from All Sources

Figure 5.5 Distribution of t-values for Home Ownership Rate and Cancer Risk from All Sources
**Figure 5.6** Distribution of t-values for Poverty Rate and Cancer Risk from All Sources

**Figure 5.7** Distribution of t-values for Population Density and Cancer Risk from All Sources
Proportion Black demonstrates a positive relationship with cancer risk from all sources in almost 93 percent of tracts throughout Florida, with the exception of the panhandle (Figure 5.2). This corresponds to the global regression result. However, a number of tracts show a negative relationship between the Hispanic and Asian proportions and cumulative cancer risk from all sources, contrary to the global regression results. While proportion Hispanic is positively related to cancer risk across much of the state, it is negatively associated with risk in the northeastern portion of Florida (Figure 5.3). Proportion Asian is also positively associated with cancer risk from all sources across most of the state, but is negatively significant near the Miami metropolitan area (Figure 5.4).

Home ownership rate was not significant in the global model, yet is negatively associated with cancer risk from all sources in tracts located in north, central, and southwest Florida (Figure 5.5). The poverty rate variable behaves inconsistent with the expectations of environmental injustice across most of the state, exhibiting a negative relationship in most tracts. However, the Tampa Bay metropolitan area is the only location in Florida where census tracts have an expected positive relationship with cancer risk (Figure 5.6). Tracts are sharply split between having a significantly positive and negative relationship between cancer risk and population density, with positive associations located in the panhandle and through Tampa Bay and rural central Florida and negative associations existing in north-central and south Florida (Figure 5.7). The results associated with GWR analysis of cancer risk from four known sources of air toxics are summarized and described in the remaining sections of this chapter.
5.2 Lifetime Cancer Risk from Major Point Sources

When cancer risks from major point sources of air toxics such as manufacturing facilities and power plants are examined, the GWR parameter estimates for all explanatory variables do not show substantial variation when compared to the global regression model parameters (Table 5.2). The GWR results for the Black proportion, for example, are reasonably consistent with the global regression model with 93 percent of tracts yielding non-significant t-values. The Hispanic proportion was positively related to cancer risk in the global model, but was found to be positively significant in less than 50 percent of Florida tracts according to GWR. The Asian proportion was not deemed significant in the global model, but is found to be negatively significant in 36 percent of tracts based on the GWR output.

In terms of socioeconomic characteristics, home ownership rate was also not significant in the multiple regression model, but GWR results indicate negative significance in 18 percent of tracts and positive significance in 34 percent of tracts. Most tracts mirrored the global coefficient for poverty rate, while in the case of population density, the global estimate does not fall within the GWR range and no tracts were found to be significant. The AIC score for the GWR model is marginally lower than the same statistic for the global regression model, indicating a slight improvement in overall model performance. For cancer risk from major point sources of air toxics, the comparison between the interquartile ranges from GWR model parameters and standard errors from the global regression model parameters indicate that the Black proportion and home ownership rate are the only variables suggesting spatial nonstationarity based on this measure.
Table 5.2 GWR Results for Cancer Risk from Major Point Sources

<table>
<thead>
<tr>
<th></th>
<th>Global Regression Coefficient</th>
<th>GWR Coefficients</th>
<th>Statistical Significance of t-values Across Census Tracts in Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Min</td>
<td>Med</td>
</tr>
<tr>
<td>Proportion Black</td>
<td>0.11</td>
<td>-0.80</td>
<td>-0.02</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>0.57**</td>
<td>-0.53</td>
<td>0.55</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>-0.62</td>
<td>-7.88</td>
<td>-3.46</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>0.10</td>
<td>-1.25</td>
<td>-0.16</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>1.01**</td>
<td>-1.11</td>
<td>1.32</td>
</tr>
<tr>
<td>Population Density (Ln)</td>
<td>-0.04*</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Adjusted R squared</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
<td>12602</td>
<td>12554</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05   **p<0.01

Local R-squared values associated with the regression model for cancer risks from major point sources are very low with a median value of only 0.03 and a maximum of 0.06 (Table 5.2). However, the global model’s R-squared value of 0.01 is exceeded in most census tracts in Florida. The spatial distribution of R-squared values show slightly better model performance at the extreme north and south portions of the state with the poorest performance in central Florida (Figure 5.8). Choropleth maps depicting the geographic distribution of explanatory variables whose significant t-values vary across space are provided in Figures 5.9 to 5.13. On all these maps, the red color is used to depict positive significance, the blue color is used to display negative significance, and the grey color is used to represent non-significant t-values.
Figure 5.8 Local R-Squared of Cancer Risk from Major Point Sources

Figure 5.9 Distribution of t-values for Proportion Black and Cancer Risk from Major Point Sources
Figure 5.10 Distribution of t-values for Proportion Hispanic and Cancer Risk from Major Point Sources

Figure 5.11 Distribution of t-values for Proportion Asian and Cancer Risk from Major Point Sources
Figure 5.12 Distribution of t-values for Home Ownership Rate and Cancer Risk from Major Point Sources

Figure 5.13 Distribution of t-values for Poverty Rate and Cancer Risk from Major Point Sources
Although no census tracts show a positive relationship between proportion Black and cancer risk from major point sources, the 7 percent of tracts that are negatively associated with this type of risk are located in the east-central portion of the state (Figure 5.9) where the Black proportion is relatively smaller. Tracts showing a positive association between cancer risk from major point sources and Hispanic proportion can be found mainly in the southern Florida (Figure 5.10). While the Asian proportion is not positively associated with cancer risk from major point sources in any tract, 36 percent of tracts indicate a negative relationship. These tracts are located in the northeast and east central areas of Florida (Figure 5.11).

Although a negative association was expected between home ownership rate and cancer risk from major point sources, only 18 percent of tracts exhibit this relationship. These tracts are located in northern Florida (Figure 5.12), whereas those with a positive relationship (34 percent) are located in southern Florida. According to the environmental justice framework, the proportion below poverty is anticipated to have a positive relationship with risk, and indeed 76 percent of tracts confirm this association. These tracts are located throughout central and southern Florida (Figure 5.13). A few tracts showing a negative relationship between poverty rate and cancer risk from major point sources are located north of Jacksonville on the Georgia border.

5.3 Lifetime Cancer Risk from Other Point Sources

When cancer risks from other point sources or small emitters such as auto body shops and dry cleaning facilities are examined, the GWR parameter estimates for the explanatory variables exhibit some variation compared to those from the global regression model (Table 5.3). The GWR coefficients for the Black proportion indicate
that less than three-fourth of tracts in Florida have a significantly positive relationship with cancer risk which matches the global regression result, but the remaining tracts yield non-significant t-values. The coefficient for the Hispanic proportion also demonstrates the same relationship as the global regression parameter in a majority of tracts, but the same variable is either non-significant or negatively associated with cancer risk from this source in the remaining 30 percent of tracts in the state. For the proportion of Asians, 60 percent of tracts are positively associated with cancer risk from major point sources, also matching the global regression coefficient.

Table 5.3 GWR Results For Cancer Risk From Minor Point Sources

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Global Regression Coefficient</th>
<th>GWR Coefficients</th>
<th>Statistical Significance of t-values Across Census Tracts in Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min Med Max</td>
<td>Negative 90% level</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Proportion</td>
<td>Black</td>
<td>3.64**</td>
<td>-1.08</td>
</tr>
<tr>
<td>Proportion</td>
<td>Hispanic</td>
<td>3.73**</td>
<td>-19.85</td>
</tr>
<tr>
<td>Proportion</td>
<td>Asian</td>
<td>21.94**</td>
<td>-21.67</td>
</tr>
<tr>
<td>Proportion</td>
<td>Owner Occupied</td>
<td>0.44</td>
<td>-7.28</td>
</tr>
<tr>
<td>Proportion</td>
<td>Below Poverty</td>
<td>3.85**</td>
<td>-14.36</td>
</tr>
<tr>
<td>Population</td>
<td>Density (Ln)</td>
<td>0.36**</td>
<td>0.00</td>
</tr>
<tr>
<td>Adjusted R squared</td>
<td>0.18</td>
<td>0.07</td>
<td>0.38</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
<td>18586</td>
<td>16948</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05  **p<0.01

Tracts exhibiting a negative relationship between cancer risk from other point sources and home ownership rate comprise 26 percent, while 5 percent show a positive relationship. The global model, however, did not find home ownership to have a significant effect on cancer risk from this source. Reflecting the global coefficient, 28 percent of tracts in Florida indicate a positive relationship between risk and poverty rate,
while 12 percent show the opposite relationship. While the global regression parameter found population density to be significant across Florida, GWR finds 41 percent of tracts to have the opposite relationship. As shown in Table 5.3, the GWR model yields a substantially lower AIC score than the global model, indicating a considerable improvement in overall model performance. For cancer risk from other point sources of air toxics, the comparison between the interquartile ranges from the GWR model parameters and standard errors from the global regression model parameters indicated spatial nonstationarity for all explanatory variables.

Local R-squared values associated with the multivariate regression model for cancer risks from other point sources are generally higher and suggest a much better model performance compared to cancer risk from major point sources, based on a median value of 0.38 and a maximum of 0.62 (Table 5.3). The global model’s R-squared value of 0.18 is exceeded in a majority of census tracts in Florida. The spatial distribution of R-squared values shows that tracts with the best model performance are located in Jacksonville, Tampa, and West Palm Beach (Figure 5.14).

Choropleth maps depicting the geographic distribution of significant t-values for explanatory variables are provided in Figures 5.15 to 5.20. On all these maps, the red color is used to depict positive significance, the blue color is used to display negative significance, and the grey color is used to represent non-significant t-values.
**Figure 5.14** Local R-Squared of Cancer Risk from Other Point Sources

**Figure 5.15** Distribution of t-values for Proportion Black and Cancer Risk from Other Point Sources
Figure 5.16 Distribution of t-values for Proportion Hispanic and Cancer Risk from Other Point Sources

Figure 5.17 Distribution of t-values for Proportion Asian and Cancer Risk from Other Point Sources
**Figure 5.18** Distribution of t-values for Home Ownership Rate and Cancer Risk from Other Point Sources

**Figure 5.19** Distribution of t-values for Poverty Rate and Cancer Risk from Other Point Sources
The Black proportion is positively related to cancer risk from other point sources throughout northeastern Florida, south Florida, and the west-central coast (Figure 5.15). Tracts that are positively associated with proportion Hispanic are located in south Florida and west Florida, while tracts negatively associated with cancer risk are located in northeast Florida (Figure 5.16). Cancer risk is positively related to the Asian proportion in northeast Florida, on the east coast, and around the Tampa Bay metropolitan area (Figure 5.17). The socioeconomic variables show both positive and negative associations with cancer risk from other sources. Home ownership rate is negatively significant with tracts in the northeast and north-central portions of the state, spreading down to the Tampa Bay metropolitan area (Figure 5.18). A handful of tracts show a positive relationship in areas of West Palm Beach and the Florida Keys. While poverty rate has a
significantly positive effect on cancer risk among tracts located in Tampa Bay and Daytona Beach (Figure 5.19), tracts in southwest Florida and north-central Florida depict a negative relationship between poverty and cancer risk - the opposite relationship with respect to the global model parameter. Population density is expected to have a positive association with cancer risk from other point sources and tracts located in the panhandle and central to south Florida reflects this relationship (Figure 5.20). However, tracts in north-central and south Florida show the opposite relationship.

5.4 Lifetime Cancer Risk from On-Road Mobile Sources

When cancer risks from on-road mobile sources such as cars, trucks, and other vehicles found on public roadways are examined, the GWR parameter estimates for the explanatory variables suggest minimal spatial nonstationarity (Table 5.4). The t-values generated by the GWR model for all three racial/ethnic variables are positively related to cancer risk in all Florida tracts, an exact match with the global regression result. The t-values for home ownership rate also reflect the global coefficient, with 100 percent of tracts classified as negatively significant. The only variable indicating any variation in results is poverty rate. While the global regression model did not find this variable to be significant, its coefficient is negatively significant in 97 percent of tracts according to the GWR model. Population density also showed no spatial variation in the significance of its t-values, with all tracts indicating a significantly positive relationship that is identical to the global regression findings. The lack of variation across the study area in model parameters could point to a problem in the adaptive bandwidth selection for this particular dependent variable. The AIC score for the GWR model associated with cancer risk from on-road mobile sources is marginally lower than the same statistic for the
global model, indicating only a minor improvement in model performance. As can be expected based on the GWR results summarized in Table 5.4, the comparison between the interquartile ranges from the GWR model parameters and standard errors from the global regression model did not indicate spatial nonstationarity for any of the explanatory variables.

Table 5.4 GWR Results For Cancer Risk From On-Road Mobile Sources

<table>
<thead>
<tr>
<th>Regression Coefficient</th>
<th>GWR Coefficients</th>
<th>Statistical Significance of t-values Across Census Tracts in Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Med</td>
</tr>
<tr>
<td>Proportion Black</td>
<td>4.44**</td>
<td>6.26</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>8.12**</td>
<td>6.45</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>26.12**</td>
<td>56.98</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>-1.27*</td>
<td>-6.06</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>0.63</td>
<td>-11.76</td>
</tr>
<tr>
<td>Population Density (Ln)</td>
<td>1.44**</td>
<td>0.00</td>
</tr>
<tr>
<td>Adjusted R squared</td>
<td>0.40</td>
<td>0.19</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
<td>20797</td>
<td>20577</td>
</tr>
</tbody>
</table>

* p<0.05   **p<0.01

The local R-squared values indicate a median value of 0.28 and a maximum of 0.33, with no tracts improving upon the global model performance of 0.40 (Figure 5.21) for on-road mobile sources of ambient air toxics. The GWR model performed the best in the Florida panhandle and South Florida (Figure 5.21). Because of the lack of spatial variation in the significant t-values, only the poverty rate choropleth map is provided (Figure 5.22). Unlike the rest of the state, tracts in the southern portion of Florida do not show a significant relationship between poverty rate and cancer risk from on-road mobile sources. On all these maps, the red color is used to depict positive significance, the blue...
color is used to display negative significance, and the grey color is used to represent non-significant t-values.

**Figure 5.21** Local R-Squared of Cancer Risk from On-Road Mobile Sources

**Figure 5.22** Distribution of t-values for Poverty Rate and Cancer Risk from On-Road Mobile Sources
5.5 Lifetime Cancer Risk from Non-Road Mobile Sources

When cancer risks from non-road mobile sources of air toxics (e.g., airplanes, lawnmowers, and ships) are examined, the GWR parameter estimates for the explanatory variables suggest a greater amount of spatial nonstationarity (Table 5.4) compared to the GWR model for on-road sources. This model provides fewer tracts with significant t-values than those from the other dependent variables, with all six variables indicating that at least 50 percent tracts are classified as non-significant. The t-values for the Black proportion show that this variable is not significant in 87 percent of tracts, yet has a significantly positive relationship in the global model. For this variable, only 3 percent of tracts are positively associated with risk, while 10 percent are negatively related. The Hispanic variable behaves similarly, with 19 percent of tracts reflecting the positively significant global coefficient, while 15 percent of tracts yielding a negative association with cancer risk. The proportion of Asians was not a significant indicator of risk in the global model, but 23 percent of tracts in the GWR model indicate a significantly negative relationship between this variable and cancer risk from non-road mobile sources.

Home ownership rate was negatively related to cancer risk in the global model, and maintained this association in 41 percent of tracts in the GWR model. Poverty rate only matched the negative global coefficient in 7 percent of tracts, while 13 percent of tracts exhibited the opposite relationship. Population density was split between negatively (20 percent) and positively (30 percent) significant t-values in the GWR model, while the global model was positively related to cancer risk from non-road mobile sources. As seen in Table 5.5, the GWR model produces a substantially lower AIC score than the global model, indicating a considerable improvement in overall model performance over the
global model for cancer risks from non-road mobile sources. The comparison between
the interquartile ranges for all GWR model parameters and the standard error of the
global regression parameters provided strong evidence of nonstationarity for all six
explanatory variables.

**Table 5.5** GWR Results For Cancer Risk From Non-Road Mobile Sources

<table>
<thead>
<tr>
<th>Regression Coefficient</th>
<th>GWR Coefficients</th>
<th>Statistical Significance of t-values Across Census Tracts in Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Med</td>
</tr>
<tr>
<td>Proportion Black</td>
<td>0.56**</td>
<td>-3.82</td>
</tr>
<tr>
<td>Proportion Hispanic</td>
<td>2.47**</td>
<td>-16.06</td>
</tr>
<tr>
<td>Proportion Asian</td>
<td>0.23</td>
<td>-41.13</td>
</tr>
<tr>
<td>Proportion Owner Occupied</td>
<td>-1.39**</td>
<td>-5.45</td>
</tr>
<tr>
<td>Proportion Below Poverty</td>
<td>-1.24*</td>
<td>-6.62</td>
</tr>
<tr>
<td>Population Density (Ln)</td>
<td>0.30**</td>
<td>-0.91</td>
</tr>
<tr>
<td>Adjusted R squared</td>
<td>0.21</td>
<td>0.06</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
<td>13508</td>
<td>12739</td>
</tr>
</tbody>
</table>

* p<0.05   **p<0.01

The local R-squared values generated by the GWR model for cancer risks from
non-road mobile sources are generally higher than global regression model’s R-squared
value of 0.21, with a median value of 0.39 and a maximum value of 0.81. The tracts with
the best model performance are located through the panhandle and central Florida (Figure
5.23). The geographic distribution of significant t-values for all explanatory variables are
shown in Figures 5.23 to 5.29. On all these maps, the red color is used to depict positive
significance, the blue color is used to display negative significance, and the grey color is
used to represent non-significant t-values.
Figure 5.23 Local R-Squared of Cancer Risk from Non-Road Mobile Sources

Figure 5.24 Distribution of t-values for Proportion Black and Cancer Risk from Non-Road Mobile Sources
Figure 5.25 Distribution of t-values for Proportion Hispanic and Cancer Risk from Non-Road Mobile Sources

Figure 5.26 Distribution of t-values for Proportion Asian and Cancer Risk from Non-Road Mobile Sources
**Figure 5.27** Distribution of t-values for Home Ownership Rate and Cancer Risk from Non-Road Mobile Sources

**Figure 5.28** Distribution of t-values for Poverty Rate and Cancer Risk from Non-Road Mobile Sources
The Black proportion has a negative relationship with cancer risk from non-road mobile sources in tracts located in the Miami metropolitan area and central Florida (Figure 5.24), while 3 percent of tracts have a positive relationship with this type of risk in southwest Florida, the Tampa Bay metropolitan area, Fort Lauderdale, and central Florida. The Hispanic proportion demonstrates a negative relationship with this dependent variable in tracts located in the Miami and Orlando metropolitan areas as well as along the Alabama border (Figure 5.25), while a positive association is observed around the Tampa Bay and Palm Bay metropolitan areas and in south Florida. In Orlando, Miami, and West Palm Beach, tracts indicate a negative relationship between cancer risk from non-road mobile sources and the Asian proportion (Figure 5.26), while
tracts throughout central Florida and southwest Florida demonstrate the expected positive relationship with risk.

Tracts that show a significant and negative association between home ownership rate and cancer risk from non-road mobile sources are located along the Alabama border, through the Jacksonville, Orlando, Tampa, Fort Myers/ Cape Coral, Port St. Lucie, and Miami metropolitan areas, along with all of southwest Florida (Figure 5.27). The expected positive relationship exists between this dependent variable and poverty rate among tracts in the Miami, West Palm Beach, Jacksonville, Tampa, and Palm Bay metropolitan areas, while a negative relationship is found in south Florida (Figure 5.28).

Population density is positively associated with cancer risk from non-sources of air toxics throughout north Florida, parts of central Florida, and throughout southwest Florida. Pockets of positive relationships can be seen around Orlando and throughout south Florida (Figure 5.29).

5.6 Summary of GWR Results

Satisfactory model performance among the five sources of risk varies across Florida, but certain areas of the state consistently demonstrate better model fit when compared with surrounding tracts. GWR model performance improves upon traditional multivariate regression in north Florida’s panhandle area and north of the Tampa Bay region for most sources of cancer risk. However, it becomes difficult to compare performance between models due to the method used to determine the optimal number of census tracts necessary to conduct GWR analysis. The criteria for bandwidth selection were consistent for all models, but the adaptive kernel method (as described in section 3.3) results in variation in the sample size produced by minimization of the AIC. Out of a
total xxx census tracts within and adjacent to Florida, 719 tracts were used to determine local parameters in the GWR models for both cumulative cancer risk and minor point sources. Larger sample sizes of 2,470 and 2,966 tracts were utilized in the GWR models for major point sources and on-road mobile sources, respectively. These sample sizes explain the relative homogeneity in the choropleth maps for local model parameters associated with these two variables. The GWR model for non–road mobile sources used the smallest sample size of only 276 tracts. This low number is reflected in the comparatively nuanced choropleth maps that show greater spatial variability in model parameters compared to the other GWR models.

Even with this variation in sample size, certain geographic patterns are consistently displayed in the relationship between the dependent and explanatory variables. Proportion Black indicates a positive relationship with multiple sources of cancer risk throughout peninsular Florida. Proportion Hispanic has a positive association with cancer risk from various sources throughout south Florida, and also exhibits a negative relationship in many models in north Florida. This uniformity of patterns in both the expected positive relationships between racial variables and risk as well as the unanticipated negative associations can also be seen among the distributions of the Asian GWR results. For the most part, relationships remain positive across much of the state, but negative associations can be seen in the Miami metropolitan area and in central Florida, in multiple models. Home ownership rate has pockets of unexpected positive relationships in south Florida, while the poverty rate variable shows the opposite expected associations with risk throughout the east coast and Tampa Bay. Population density demonstrates a positive relationship with multiple sources of cancer risk in the
panhandle and central Florida, while the opposite is observed in north central and south Florida.
CHAPTER 6: CONCLUSIONS

Human geography has been concerned with how the industrial and commercial developments of the last 150 years have adversely impacted different sectors of society, particularly the under-privileged and under-represented. The emergence of the environmental justice movement in the 1980s led to an increased focus on how racial/ethnic minorities and socioeconomically disadvantaged residents were impacted by human-induced hazards and related health risks. Geographers became interested in environmental justice’s inherently spatial issues of distributional equity and social justice (Walker and Bulkeley 2006). A variety of quantitative methodologies have been utilized in the environmental justice research literature to evaluate whether technological hazards and risks are distributed equitably with respect to different racial, ethnic, and socioeconomic groups.

Previous quantitative research on environmental justice, however, has been impeded by three specific limitations in the types of data utilized and methodologies employed. First, a large number of studies have measured disproportionate burdens based on distance to pollution sources or the quantity or toxicity of emitted pollutants, thus failing to account for the adverse health impacts of exposure to these chemicals. Second, past studies have focused primarily on inequities associated with major stationary sources of pollution such as industrial manufacturing facilities, ignoring mobile emission sources and smaller emitters that also pollute the local environment and contribute substantially to environmental health risks. Finally, almost all previous studies using conventional
statistical methods such as multivariate regression have assumed that relationships between dependent and explanatory variables are stationary across space in a given study area. Since traditional regression analysis cannot uncover local variations, it leads to broad generalizations about the entire study region, thus ignoring the notion that specific places within the same study area might differ from each other with regard to the nature and extent of environmental injustice.

This thesis extends quantitative research on environmental justice by addressing these three problems, through a case study of estimated cancer risks from inhalation exposure to air toxics in Florida. The 1999 National-Scale Air Toxics Assessment (NATA) data used for this study provides modeled estimates of cancer and non-cancer risks from air toxics at the census tract level. This is an improvement over previous studies because the risk assessment methodology incorporates chemical fate and transport, weather patterns, and human inhalation data. The NATA also provides estimates of adverse health risks for multiple types of air toxics emission sources, allowing for less readily acknowledged sources of pollution to be accounted for and incorporated into environmental justice analyses. In order to address the limitations of conventional regression, this study uses a spatially sensitive statistical technique that allows for a unique regression model to be produced at each data point. Instead of generating one set of model parameters and diagnostics for the entire study area, geographically weighted regression gives more influence to data points closer to the origin point for which the model is being built, thus allowing local variations in statistical relationships to be examined.
The first stage of this study utilizes traditional statistical methods such as linear correlation and multiple regression to examine the environmental justice implications of lifetime cancer risk for entire study area. Bivariate correlation analysis provides strong evidence of racial/ethnic and socioeconomic inequities in the distribution of cancer risks from all emissions sources. The multivariate regression results provide evidence that race and ethnicity, particularly Black and Hispanic proportions, are consistent predictors of the presence of cumulative cancer risk, as well as cancer risk from the four individual sources of air toxics, even after controlling for socioeconomic characteristics and population density. Increased Asian proportions are indicative of greater cancer risk from other point sources and on-road mobile sources. The socioeconomic variables indicate mixed results, with lower home ownership rate being related to increased cancer risk from only mobile sources of air toxics. An increased poverty rate is associated with greater cancer risk for point sources, but not mobile sources of air toxics. Higher population density was a positive indicator of cancer risk for both types of mobile sources, minor point sources, and cumulative cancer risk from all sources.

In contrast to the conventional correlation and multivariate regression analysis, the geographically weighted regression (GWR) analysis indicated that many of the observed statistical associations between cancer risk and specific explanatory variables are not uniform across Florida, but are more significant in some areas and not evident in other locations. Despite the differences in regression methodology, the GWR results consistently indicate the pervasive effect of Black and Hispanic proportions in explaining the geographic distribution of cancer risks from air toxics in Florida, even when controlling for other explanatory factors. A higher proportion of Black residents in a
census tract suggests greater cancer risk from all sources throughout much of the state. More specifically, the Black proportion is positively related to cancer risk from other point sources among tracts in Jacksonville and other areas in north Florida, around Tampa Bay, and throughout south Florida, while cancer risk from mobile sources display the same relationship northwest of Orlando and Tampa, southeast of Naples, in Miami Beach, and West Palm Beach.

Although cancer risks are also distributed inequitably with respect to the Hispanic population, the GWR analysis reveals that locations which show a significant and positive association for the Hispanic proportion are different from those which depict a similar relationship for the Black proportion. The presence of Hispanic residents indicates greater cancer risks from all sources throughout the Tampa Bay metropolitan area and along the east coast of Florida. More specifically, higher proportions of Hispanic residents suggest increased cancer risk from point sources across south Florida, and from non-road mobile sources in Tampa Bay, Cape Canaveral, Miami, and Fort Lauderdale.

While GWR confirms where these expected relationships exist, it also highlights areas where the opposite is observed. While the Hispanic proportion is positively associated with cancer risk across much of Florida, it seems to negatively relate to cancer risk from other point sources around Jacksonville. Higher home ownership rates indicate increased risk from other point sources in West Palm Beach and the Florida Keys. Lower poverty rates point to increased cumulative cancer risk from all sources in most Florida tracts, but the exception occurs in the Tampa Bay metropolitan area where cancer risks are disproportionately distributed with respect to people in poverty. These local variations in the global positive relationship between cancer risk and ethnicity or socioeconomic
status are obscured by conventional multivariate regression, but are revealed by the use of GWR in conjunction with choropleth mapping. The results of this study clearly suggest the need for researchers to recognize the usefulness of GWR as an exploratory data analysis tool for environmental justice assessment.

It is important to discuss, however, the caveats that bound the improvements made by the data and methodology used for this research. There are specific limitations associated with EPA’s NATA data set used for this research. The 1999 NATA estimates cancer risk only from inhalation exposure to air toxics and does not account for exposure through other pathways such as ingestion and skin contact. The NATA information is also not a substitute for actual health outcomes data, and only represent modeled estimates of cancer risk based on EPA’s risk assessment guidelines. The 1999 NATA also assesses the cancer risk from 133 out of 188 air toxics, and thus fail to provide a complete assessment of all toxic air pollutants of concern (EPA 2008d). Also, this study is a cross-sectional analysis of adverse health risks at a specific point in time (1999), and should not be used to elucidate causal relationships between race/ethnicity and exposure to pollution or deduce the chronological order of events leading to the current disparities.

Geographically weighted regression also presents challenges in its application. A relatively new method of analysis, GWR is ripe for critique and assessment of its capabilities. The choice of adaptive or fixed bandwidth for the selection of neighboring census tracts is a critical decision that could greatly influence the results. The adaptive or variable bandwidth approach was utilized in this study to account for the spatial variation in the size of the tracts or the density of tract centroids. However, a fixed kernel bandwidth which uses distance rather than a specific number of neighbors to determine
the local sample size could be a more appropriate method for analysis in certain cases where resolution is more uniform (Mennis and Jordan 2005). It is also important to consider that GWR also does not always improve the overall performance or fit of a multivariate regression model. As seen in the results for cancer risk from major point sources, both the global regression and GWR models resulted in low R-squared values and point to a potential problem with the choice of explanatory variables.

This thesis has several important implications for public policy. The results of the study indicate that environmental injustice is occurring in specific places within Florida and also identify the air pollution sources responsible for the disproportionate health impacts. This information can assist local advocacy groups, grassroots organizations, and minority populations in their attempts to obtain increased oversight of hazardous facilities or highways near areas populated by minority or low-income residents, in addition to providing evidence of unequal exposure to hazards for possible litigation. Since this study uses health risk assessments from the EPA to evaluate spatial inequities, the results can lend legitimacy to the environmental health burdens imposed on community members and thus advance the aims of environmental justice. GWR can be particularly useful because it has the capacity to provide specific information about locations that are disproportionally impacted by hazardous air pollutants. These locational differences in the unique relationships between the explanatory variables and different types of pollution sources can assist policy formation and regulation by highlighting the unique issues faced by a city or region and represent a starting point for qualitative analysis of the various economic, historical, and political processes at work in the study area. Several variables such as the proportion of Black and Hispanic populations are consistently and
positively associated with cancer risk from all sources in all parts of the state. This inequitable pattern points to the need to assess the environmental justice implications of current and future zoning and siting regulations in the state of Florida. Along with enforcement of environmental regulations, disproportionately exposed minority and low-income residents must be included in the decision-making process in order to combat unintended discrimination and inequitable exposure to the negative by-products of modern society.

In conclusion, this thesis provides strong evidence of racial/ethnic and socioeconomic disparities in the adverse health effects of exposure to outdoor air toxics from both point and mobile air pollution sources in Florida. The results demonstrate that conventional multivariate regression for environmental justice analysis can hide important local variations in the relationships between cancer risk and relevant explanatory variables such as race, ethnicity, and socioeconomic status. Further research is needed to characterize the various socioeconomic, political, and spatial processes that are causing cancer risk inequities in specific areas of Florida. Meanwhile, these results raise new challenges for both policy makers and environmental justice advocates in terms of developing regulatory and pollution prevention strategies that address both stationary and mobile emission sources.
REFERENCES


Daniels, Glynis and Samantha Friedman. 1999. Spatial inequality and the distribution of industrial toxic releases: Evidence from the 1990 TRI. *Social Science Quarterly* 80: 244-62.


EPA (U.S. Environmental Protection Agency). 2008c. The four steps of the assessment.  

EPA (U.S. Environmental Protection Agency). 2008d. Frequent questions.  


Pastor, Manuel, Rachel Morello-Frosch, and James L. Sadd. 2005. The air is always cleaner on the other side: Race, space, and ambient air toxics exposures in California. *Journal of Urban Affairs* 27:127-48.


Zhuang, Duan. 2006. Spatial dependence and neighborhood effects in mortgage lending: A geographically weighted regression approach. LUSK Center for Real Estate, USC.
Appendices
Appendix A: Frequency Histograms

**Figure A-1** Frequency Histogram of Cumulative Cancer Risk per Million People

**Figure A-2** Frequency Histogram of Major Point Sources of Cancer Risk per Million People
Appendix A (continued)

**Figure A-3** Frequency Histogram of Other Point Sources of Cancer Risk per Million People

**Figure A-4** Frequency Histogram of On-Road Mobile Sources of Cancer Risk per Million People
Appendix A (continued)

Figure A-5 Frequency Histogram of Non-Road Mobile Sources of Cancer Risk per Million People

Figure A-6 Frequency Histogram of Proportion Black by Census Tract, 2000
Appendix A (continued)

**Figure A-7** Frequency Histogram of Proportion Hispanic by Census Tract, 2000

**Figure A-8** Frequency Histogram of Proportion Asian by Census Tract, 2000
Appendix A (continued)

Figure A-9 Frequency Histogram of Poverty Rate by Census Tract, 2000

Figure A-10 Frequency Histogram of Home Ownership Rate by Census Tract, 2000
Appendix A (continued)

Figure A-11 Frequency Histogram of Population Density by Census Tract, 2000

Figure A-12 Frequency Histogram of the Natural Log of Population Density by Census Tract, 2000