THE SPATIAL DISTRIBUTION OF LEAD IN URBAN RESIDENTIAL SOIL AND CORRELATIONS WITH URBAN LAND COVER OF BALTIMORE, MARYLAND

By

KIRSTEN SCHWARZ

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Ecology and Evolution

written under the direction of

Dr. Steward T.A. Pickett and

Dr. Richard G. Lathrop, Jr.

and approved by

____

____

____

____

____

____

New Brunswick, New Jersey

January 2010

ABSTRACT OF THE DISSERTATION

The spatial distribution of lead in urban residential soil and correlations with urban land cover of Baltimore, Maryland

by KIRSTEN SCHWARZ

Dissertation Co-Directors:

Dr. Steward T.A. Pickett

Dr. Richard G. Lathrop, Jr.

Lead contamination of the urban environment is not a new phenomenon. A great deal of research has focused on the health effects of lead-based paint. Less attention, however, has been given to the potential problem of soil contaminated with lead from the past use of lead-containing products such as lead-based paint and leaded gasoline. Identifying areas of high contamination is necessary in order to prioritize soil remediation and public health efforts. This requires a comprehensive understanding of a highly heterogeneous and dynamic system.

This research addresses whether land *use* or land *cover* is a better predictor of lead concentrations in soil. Specifically, this research addresses whether landscape features, including trees, lawns, buildings, and roads, can be used to predict lead concentrations in soil. Through a method of rapid assessment of soil lead concentrations, I gathered spatially explicit data from urban residential yards to generate several models that predict the spatial distribution of lead in soil. Using the results of these models, potential inequities associated with the

modeled spatial distribution of lead in soil and socio-demographic features were explored.

The results of this study suggest that the distribution of lead in urban residential soils is more closely correlated with features of urban land cover compared to metrics of land use. Specifically, the spatial distribution of lead in urban residential soils is strongly influenced by three factors: housing age, distance to the major road networks, and distance to built structures. Through the comparison of various spatial models, this research demonstrates that a greater amount of variation in the data is explained by machine learning techniques compared to traditional modeling techniques. In addition, important correlations between the modeled distribution of lead in soil and socio-demographic features such as race and poverty have been identified. Specifically, a greater amount of soil contamination is predicted to be present in high poverty areas.

This research contributes to the growing field of urban ecology by advancing our knowledge of how spatial heterogeneity affects the distribution of a critical pollutant in urban systems. This work also tests the suitability of using land cover as a predictive ecological variable.

ACKNOWLEDGMENTS

Many thanks are owed to my advisors, Steward T.A. Pickett and Richard G. Lathrop. Their commitment to my education has been a gift, and I thank them both for being truly present throughout this process. Steward has taught me many things about ecology. Perhaps more importantly, Steward has introduced me to a larger, more wonderful world. In so many ways he embodies the word mentor. Rick instilled in me the confidence I was sometimes lacking which allowed me to continue pursuing my work. I am also thankful that Rick always ensured me that I had a home at Rutgers. I am very appreciative, and have greatly benefited from, his encyclopedic knowledge of all things spatial.

I am extremely grateful to the members of my committee for demonstrating incredible guidance, encouragement, and patience. Kathie Weathers took a chance on a freshly minted college graduate 10 years ago. Her wisdom, strength, character, and sincerity have been an inspiration every day since. Mary Cadenasso threw me into the wilds of South Africa and in the process showed me I was capable of this thing called ecology. I am extremely grateful for her enthusiastic support and thoughtful contributions. Rich Pouyat, being a tremendous advocate for graduate students, has always been extremely gracious with his time and knowledge of both soils and Baltimore. He not only made collaboration with the BES possible, he made it fun. Bonnie McCay always made it known that even though she was geographically distant; she was, in fact, nearby. Her guidance and words of wisdom have been greatly appreciated.

I am indebted to friends and colleagues at the Cary Institute of Ecosystem Studies, which has served as home for the past several years, especially David Lewis, Lisa Martel, Amy Schuler and Samuel Simkin.

I would also like to thank friends and colleagues at the Baltimore Ecosystem Study, which has been a welcoming environment to grow personally and professionally, especially Alan Berkowitz, Holly Beyar, Chris Boone, Peter Groffman, Morgan Grove, Jen Jenkins, Jarlath O'Neil-Dunne, Austin Troy, Jonathan Walsh, Mary Washington, Ian Yesilonis and Weiqi Zhou.

To friends and colleagues at Rutgers for making my stint in NJ pleasant and intellectually engaging, especially Kyle Bennett, Jen Freeman, Shannon Galbraith-Kent, Matt Kimball, Inga LaPuma, David LaPuma, Jeanmaire Molina, Marsha Morin, Matt Palmer, Kristen Ross, Ishanni Sen, and Emilie Stander.

I extend heartfelt thanks to Elizabeth Cook, a cherished friend and colleague, who has listened to more than her fair share of my dissertation woes. Thank you for always steering me back to reality and providing much needed perspective.

Most importantly, I extend very special thanks to my family. My parents, Otto and Joanne Schwarz, have supported me in every endeavor, wise or otherwise. My siblings Nancy, Rob and Ericka are a constant source of strength and support and my nieces, Taylor and Madison, inspire me every day to be a better person. Lastly, for sharing in the laughter and tears of this long process I offer the most sincere gratitude to my partner Josh Sonner, whose unwavering faith in my ability to succeed, although at times incomprehensible to me, was a lifeline.

DEDICATION

This dissertation is dedicated to my parents, Otto and Joanne Schwarz, who sacrificed much so that I could be afforded every opportunity they never had. I will always be grateful.

AND

To my grandfather, Otto Schwarz Sr., for introducing me to the beauty of the natural world through his love of gardening.

TABLE OF CONTENTS

INTRODUCTION

In many ways the story of lead poisoning in the United States is a public health success. The ban on lead containing products resulted in dramatic decreases in the blood lead levels of children. Therefore, many are surprised to learn that lead continues to be a current public health threat with children still affected by the legacy of lead in the environment. Lead's toxicity combined with its persistent nature and widespread use set the stage for a serious and long term public health issue.

Many sources of lead can contribute to elevated blood lead levels in humans. One important source that has not gained much attention is elevated levels of lead in urban soils. Lead from deteriorating lead-based paint and leaded gasoline has polluted many urban soils. While we know that lead concentrations in soil can be influenced by features in the landscape, we need more information regarding what features are most important to the spatial patterning of lead and how those features together with the larger landscape context form the patterns observed in the field. Ecologists can contribute to our understanding of lead contaminated soil by using the tools of landscape and ecosystem ecology to examine relationships between urban ecosystem structure and the ecosystem function of lead retention.

With the United States Department of Health and Human Services declaring the elimination of childhood lead poisoning a national health priority, many states have committed to eliminating childhood lead poisoning by 2010. In order to meet this goal, there is a need to focus on less concentrated and widely dispersed sources of lead in the environment. Models depicting the spatial distribution of lead in soil are necessary in order to identify areas of high contamination in the landscape and to protect vulnerable populations.

The overarching question of this research is: what features of the urban landscape contribute to the spatial patterning of lead in urban residential soils? I approach the question of lead contamination in soil by first examining correlations between lead in soil and urban land use and cover. Specifically, chapter one addresses whether land *cover*, in contrast to land *use*, can serve as a useful predictor of lead levels in soil. Chapter one also examines the importance of legacies to the spatial patterning of current concentrations of lead in soil by considering both present land cover and land cover during peak deposition of lead.

Next, both intensive and extensive sampling of urban soil was conducted to examine patterns between lead in soil and individual landscape features. Chapter two examines the contribution of both fine scale landscape features, such as trees, lawns, and buildings, and the larger landscape context, including road networks and housing age, to the spatial patterning of lead. While much is known regarding the effect of individual features and landscape scale patterns of lead deposition, little is known about the combined effect of fine scale and landscape scale patterns. Chapter two also tests the assumption that lead is concentrated at the surface of soil. Finally, chapter two examines the contribution of different lead sources to the urban environment.

Building on the results from chapter two, in chapter three I present three empirical models in a Geographic Information System and test their validity using an independent dataset. All three models use readily available data for urban areas to model the distribution of lead in residential soils.

Partnering the results from the empirical model with census tract data, chapter four examines potential correlations between the modeled distribution of lead in soil and social characteristics. Specifically, I examine race and poverty rates to address whether communities of color and economically disadvantaged communities are disproportionately exposed to lead pollution in soil.

CHAPTER 1. CORRELATIONS BETWEEN PRESENT AND HISTORIC LAND COVER AND LEAD CONCENTRATIONS IN URBAN SOILS OF BALTIMORE, MARYLAND USA

Abstract

The inextricable link between ecological structure and function is a fundamental principle of landscape ecology. However, structure and function describe different aspects of a landscape. Land cover, or structure, defines the landscape features that are present, while land use, or function, describes their utility. Although it is widely assumed that land use is a predictor of ecosystem services, common measures of land use often conflate structure and function. Thus, the assumed link between land use and services may be inherently flawed when ecosystem services are primarily dependent on ecological structure. Urban soils act as a sink for anthropogenic lead. Previous studies in Baltimore, MD have found that this ecosystem service of lead retention shows no relationship to land use as described by an Anderson Level II land use classification. I test the hypothesis that in contrast to land use, land cover describing ecological structure, may be a better predictor of lead in urban soils by partnering an urban land cover classification scheme called HERCULES (High Ecological Resolution Classification for Urban Landscapes and Environmental Systems) with soil lead data collected in 2000 for the Urban Forest Effects (UFORE) Model. This comparison was made using recent (2004) and historic (1968) imagery to test whether current soil lead patterns correlate more strongly with present land cover or land cover during peak deposition from automobiles which occurred in the late 1960s and early 1970s. A statistically significant positive correlation was found

between lead in soil and 2004 building ($p = 0.0287$, $r^2 = 0.0386$) and pavement ($p = 0.0078$, $r^2 = 0.0566$) cover. Bare soil, lawns, and trees showed no relationship. Historic land cover from 1968 showed a similar pattern with a statistically significant positive correlation between lead in soil and building ($p = 0.0004$, $r^2 = 0.197$) cover. Although not statistically significant, lead in soil and 1968 pavement cover were also related ($p = 0.0641$, $r^2 = 0.0589$). While the r-squared values are low, suggesting only a small percentage of the variation can be accounted for in correlations with one land cover type, these data support the idea that the distribution of lead in urban soils is dependent on the spatial pattern and density of certain landscape features. These data also highlight the importance of historic land cover in examining current patterns of a persistent environmental chemical.

Introduction

Lead, a naturally occurring element present in very small quantities in the Earth's crust, has found its way into modern society in numerous ways. The two main anthropogenic sources of lead are lead-based paints and leaded gasoline. The widespread use of these two products has resulted in contamination of the urban environment. Lead in the environment is a potential public health issue, given that lead, a heavy metal, is a potent neurotoxin (Silbergeld 1992). Urban soils act as a sink for anthropogenic sources of lead in the environment (Wong *et al.* 2006). Therefore, describing the spatial distribution of lead in urban soils is a goal of both the public health community and ecologists. A first step to describing the spatial distribution of lead in soil is testing the assumption that land use is a predictor of ecosystem services, specifically lead retention in urban soils. The following research addresses whether land *use* or land *cover* is a better predictor of lead concentrations in urban soil.

Lead's chemical and physical properties made it invaluable during the early 20th century. Thomas Midgley Jr., while working as a chemist for General Motors, discovered that the addition of Tetra-Ethyl lead eliminated engine knocking. This discovery spurred the formation of the Ethyl Corporation, which was followed by the worldwide use of leaded gasoline. Although the problem of engine knocking was solved, the public health and environmental trade-offs were severe with lead persisting in the environment more than 20 years after its widespread use in the United States. The addition of lead to gasoline from 1920-1986 resulted in the release of 4 to 5 million metric tons of lead into the atmosphere (ATSDR 1988). It is estimated that 75% of lead from gasoline was deposited along roadways via exhaust with the remaining 25% staying in the engine and/or oil (ATSDR 1988). Lead is found in large quantities in the soil adjacent to roadways (Lagerwerff and Specht 1970) with a precipitous decline observed with greater distances from the road (Motto *et al.* 1970, Wang *et al.* 2006). Since urban areas, including residential neighborhoods, exhibit both a higher density of roads and increased traffic volume, it is not surprising that lead is a common pollutant found in urban soils.

Beginning in the late 19th century, lead was also added to some paints as a pigment and to increase durability. The amount of lead in paint changed during the time period it was used (1884-1978), varying from 1-50% lead by dry weight (Browne and Laughnan 1953, Shannon 1996). Lead-based paints that contain more than 0.06% lead were banned for residential use in 1978 (United States Department of Housing and Urban Development, Legislative history of lead-based paint, hud.gov). The legacy of lead-based paint, however, remains, primarily due to the impressive pervasiveness of lead-based paints. An estimated 24 million dwellings built before 1978 contain some lead-based paint (United States Department of Housing and Urban Development, Lead, hud.gov). When interior and exterior paint deteriorate over time, especially on friction surfaces like windowsills and door jams, it chips and peels, finding its way into the soil surrounding homes.

In addition to its pervasiveness, lead is a known neurotoxin that is especially dangerous to certain populations, including children under the age of six and pregnant women. Lead contamination of the urban environment remains one of the largest public health concerns facing urban populations, with approximately 310,000 children aged 1-5 affected by lead poisoning in the US during the 1999-2002 survey period (CDC 2005). The Centers for Disease Control and Prevention (CDC) level of concern for blood lead levels (BLLs) has been lowered by 88% over the last 40 years (Miranda *et al.* 2002) and currently stands at 10 µg/dL (CDC, Standard surveillance definitions and classifications, cdc.gov). Some have argued for a reduction in the current standard from 10 to 2 µg/dL (Gilbert and Weiss 2006) citing studies that demonstrate BLLs below the current CDC level of concern can result in intellectual impairment (Lanphear *et al.* 2000, see also Koller *et al.* 2004). Children and pregnant women are not the only populations affected by elevated lead levels. Recent studies have highlighted the many adverse health effects experienced by adults and aging populations (Payton *et al.* 1998, Schwartz *et al.* 2000, Stewart *et al.* 2006). In addition, the burden is not equally distributed among socio-economic classes, with low income minority children living in older homes or urban areas more likely to be affected by elevated BLLs (CDC 1997, CDC 2005). Although residents of more rural areas are not immune to anthropogenic lead contamination, urban areas are disproportionately affected by consistently higher lead poisoning rates compared to national averages. For example, the incidences of lead poisoning in

Cleveland, Providence, Philadelphia, Buffalo, Milwaukee, Chicago, Detroit, St. Louis, and Baltimore all exceed the national average (Environmental Health Watch, Childhood lead poisoning, ehw.org). This is especially pertinent considering 80% of the US population (United States Census Bureau, Geographic Comparison Table, census.gov) and 50% of the global population (The United Nations Population Fund, Linking population, poverty, and development, unfpa.org) live in cities.

While soil serves as a sink for anthropogenic lead, it may also be a source of lead contamination to human populations under certain environmental conditions. However, research on lead and lead poisoning has largely focused on lead-based paint. While lead-based paint is indisputably a major contributor to childhood lead poisoning, soil cannot be overlooked. Lead in the form of both 1) deteriorating lead-based paint and 2) lead in soil and dust from leaded gasoline and lead-based paint can have direct effects on public health. Several studies have concluded that soil is an important contributor to children's lead burden (Duggan and Inskip 1985, Mielke 1997). Supporting this idea is the revealing pattern that cities with greater concrete coverage, like Manhattan, have lower percentages of children with elevated BLLs compared to cities with more exposed soil, for example, New Orleans and Philadelphia (Mielke 1999).

Lead in soil comes from multiple sources and cannot be explained by only one variable. For example, soil lead patterns in Baltimore cannot be explained by paint alone. In fact, in large cities, soil lead levels have been found to correlate strongly with traffic patterns (Kelly *et al.* 1996). This is especially important considering that lead is emitted from automobiles in the form of small particles. Smaller particles of lead pose a greater health threat since they are more easily absorbed in the gastrointestinal and pulmonary tracts (Miranda *et al.* 2002). Soil can also be an important *indoor* contaminant when it is tracked into the home (Caravanos *et al.* 2006).

The public health community has historically focused on children consuming lead-based paint chips as a pathway for childhood lead poisoning. However, lead can be transferred to humans by a variety of pathways. Auto and industrial emissions can be inhaled via ambient air (USEPA 2006). Soil lead from auto emissions and paint can be inhaled, ingested directly, incorporated into house dust that is ingested or inhaled, adsorbed to plants that are eaten by humans, or taken up by certain species of plants known as hyper-accumulators and then consumed by humans. Children, who exhibit pica, or cravings for non-food items, are especially vulnerable to soil lead exposure. Public awareness of leaded dust and soil as potential pathways for human exposure is lacking. A public health survey on caregivers' knowledge of childhood lead poisoning showed that while 61% of participants identified eating paint chips as a source of lead poisoning, only 15% identified lead paint dust and less than 3% identified soil (Mahon 1997). This is in stark contrast to studies that have identified soil as an important pathway of human exposure (Mielke and Reagan 1998).

Lead is a chemically "sticky" molecule. Given the right soil properties, lead can remain in the soil for a very long time with some estimates of residence time greater than 100 years (Reiners *et al.* 1975, Smith and Siccama 1981). Even in competitive situations where other metals are available in the soil, lead strongly adsorbs to soil (Gomes *et al.* 2001). Lead binds to organic matter and oxides of manganese and iron in the soil (Yesilonis *et al.* 2008). This is especially important in residential areas given that organic matter is generally higher, especially in older turf grass systems, compared to native soils and some other urban land uses (Pouyat *et al.* 2008). Soil pH is also an important factor. Neutral and alkaline soils have fewer hydrogen ions and thus more available sites for lead to bind. Acidic soils on the other hand have a greater concentration of hydrogen ions, which generally translates to lead being more mobile in the soil (Elless *et al*. 2007).

In order to facilitate remediation efforts, the spatial pattern of lead in soil must be identified. While geospatial mapping of contamination accomplishes this task it is time intensive and costly. Therefore, there is a need to identify strong predictive variables so that one can calculate the spatial distribution of lead contamination in soil. This would allow for remediation efforts to be concentrated in the most highly polluted areas. The question remains as to whether land use, a functional variable, or land cover, a structural variable, is a better predictor of lead levels in urban soils.

Previous work in Baltimore, Maryland has shown that no correlation exists between lead levels in soil and land use as defined by a well known land use classification scheme called Anderson Level II (Pouyat *et al.* 2007). Specifically, the authors examine whether various land use categories, including commercial or transportation, industrial or urban open, unmanaged forest, park or golf course, residential, and industrial were correlated with lead concentrations in surface soils (0-10 cm). An ANOVA demonstrated that none of the land use categories were statistically significant at the 0.05 probability level. The authors conclude that the lack of correlation between land use and heavy metal concentrations in soil suggests "that these elements are more related to site history and the spatial arrangement of contaminant sources in urban landscapes" (Pouyat *et al.* 2007, p. 1017). Commonly-used land use classifications are not at the appropriate categorical resolution to identify landscape features that may be important contaminant sources or sinks of anthropogenic lead. In addition, commonly-used land use classifications do not exhibit the fine scale resolution necessary to address the spatial arrangement of contaminant sources. Therefore, it is necessary to describe the landscape in such a way that accounts for 1) landscape features that may be important sources and/or sinks for anthropogenic lead and 2) fine scale heterogeneity that may be important to the spatial patterning of lead in soil.

Here I test the hypothesis that land cover, as defined by a new land cover classification scheme called HERCULES (High Ecological Resolution Classification for Urban Landscapes and Environmental Systems, Cadenasso *et al.* 2007), is a better predictor of lead concentrations in soil. Earlier studies support this hypothesis by demonstrating that individual landscape features are important to lead sequestration in soil. For example, "splash zones" close to heavily traveled roads exhibit elevated lead levels (Lagerwerff and Specht 1970). Lead adheres to the surface of buildings and other structures and can be rinsed to the soil near the foundation by rain, resulting in elevated lead levels at the foundations of buildings, even those with no history of lead-based paint use (Mielke 1999). Lead can also be scrubbed out of the atmosphere by vegetation. Weathers *et al.* (2000) found a difference between coniferous and deciduous vegetation, with high-elevation coniferous forests exhibiting higher soil lead concentrations, supporting the idea that conifers are more efficient at removing gases and particulates from the atmosphere. All of these landscape features - roads, soil, buildings, and trees - are present in the urban matrix and can affect lead sequestration. What remains unknown is the collective importance of these different landscape features and which structures are most important in predicting lead sequestration in urban soils. By partnering existing soil lead data from Baltimore, Maryland with urban land cover data, we can begin to address these questions.

Methods

In order to test the hypothesis that land cover, in contrast to land use, is a better predictor of lead concentrations in soil, I partnered existing soil lead data that was collected in 2000 as part of the Urban Forest Effects Model (United States Department of Agriculture Forest Service, Urban Forest Effects Model, fs.fed.us) with urban land cover data that was generated using the HERCULES (Cadenasso *et al.* 2007) land cover classification. Below the individual datasets, including soil lead, HERCULES, and the aerial photographs, are described in more detail.

Data - Soil Lead

Soil lead data were available from study plots (n = 125) that were established in 1999 as part of the Baltimore Ecosystem Study (BES). Data collected from the plots were originally used to calibrate the Urban Forest Effects (UFORE) Model. The UFORE model was developed by David Nowak of the United States Department of Agriculture (USDA) Forest Service, Northeastern Research Center to characterize the structure of urban forests. UFORE plots in Baltimore City were selected using a stratified random sampling scheme. Plots were stratified by land use with type corresponding roughly to Anderson Level II land cover classes (Anderson *et al*. 1976) and weighted by area. The distribution of plots by land-use type was: commercial (n = 2); industrial (n = 3); institutional (n = 10); transportation right-of-ways (n = 7); high density residential (n = 19); medium density residential (n = 33); golf course (n = 3); riparian (n=2); park (n = 10); urban open (n = 10); and forest (n = 26). With the intended goal of characterizing urban soils, composite soil samples were collected at the UFORE plots from the surface 10 cm in the summer of 2000 (Pouyat *et al*. 2007). The samples were air dried and passed through a 2 mm sieve. Samples were acid digested using a modified USEPA method 3050B (United States Environmental Protection Agency, Method 3050B, epa.gov) at the BES and University of Maryland, Baltimore County lab. The digested samples were filtered and sent to Cornell University Nutrient Analysis Laboratory where they were run on an Inductively Coupled Plasma (ICP) and analyzed for Al, Ca, Cd, Co, Cu, Cr, Fe, K, Mg, Mn, Mo, Na, Ni, P, Pb, S, Ti, V, and Zn (Pouyat *et al.* 2007).

Data - HERCULES

Most common land use classification schemes have not been able to capture the rich spatial heterogeneity that is characteristic of urban areas. Urban residential areas are often classified as simply "urban" in standard land classification schemes (Cadenasso *et al.* 2007). This indicates very little about the ability of that landscape to retain important environmental pollutants such as lead. Cadenasso *et al.* (2007) developed a new object-based land cover classification called HERCULES. HERCULES characterizes urban structure based on the composition of three landscape features: buildings, surface material, and vegetation (Figure 1). Object-based classifications have been proposed as alternatives better suited to capture the spatial heterogeneity of urban areas compared to traditional pixel-based techniques. Studies comparing traditional pixel-based techniques to object-based classifications have shown improved classification accuracy in urban areas (Felipe *et al.* 2003, Flanders *et al.* 2003, Tadesse *et al.* 2003, Moeller *et al.* 2004). HERCULES classifies land cover, keeping structure separate from function. This allows examination of individual landscape features and their contribution to lead retention in urban soil. Thus, this research moves beyond the traditional budgetary approach of studying what is moving in and out of a defined system, and instead, focuses on the specific features of the system that control the ecosystem service of lead retention.

Data - Aerial Photos

False-color infrared aerial photos of Baltimore City taken in August 2004 were used in this study. These photos were obtained in digital format and did not require any post processing. Historical black and white aerial photos of Baltimore City from March 1968 were acquired from the Baltimore City Department of Planning. These photos represent Baltimore’s land cover during peak deposition from leaded gasoline use in automobiles, which was estimated to occur during the late 1960s and early 1970s (Mielke 1999). The historical photos were scanned and georeferenced at the University of Vermont's Spatial Analysis Lab in 2009 using ArcGIS 9.3.

GIS Analysis

Within a Geographic Information System (ArcGIS 9.3), I spatially joined the UFORE soil data to the HERCULES land cover classification system for the 2004 photographs to examine possible correlations between individual landscape features and lead levels in soil. A 30 meter radius buffer was added to each of the 125 UFORE soil-sampling points. The distance of 30 meters was selected based on literature that has documented an increase in soil lead concentrations adjacent to major roadways (Motto *et al*. 1970, Ordonez *et al.* 2003, Wang *et al.* 2006). Within the 30 meter buffer, individual landscape features as defined by HERCULES (bare soil, coarse vegetation, fine vegetation, buildings, and roads) were classified using a heads-up digitizing technique (Figure 2). This method was also applied to the 1968 photographs; however, only a subset of the Baltimore City 1968 aerial photographs was available for analysis. Imagery from 1968 was available for 59 of the UFORE study plots. Area was calculated for the individual landscape features within each buffer using the Xtools Pro extension in ArcGIS (Data East 2003). Next, the total area for each land cover type was calculated within each 30 meter buffer and assigned the associated soil lead concentration from the UFORE plot. Single factor regression analyses on individual land cover types and lead concentrations were performed using JMP statistical software (SAS Institute 2007).

Results

For the 2004 aerial photos, a statistically significant positive relationship between lead concentrations in soil and building ($p = 0.0287$, $r^2 = 0.0386$) and pavement ($p = 0.0078$, $r^2 = 0.0566$) cover was found (Figure 3). Bare soil ($p = 0.6136$, $r^2 = 0.0021$), coarse vegetation ($p = 0.1370$, $r^2 = 0.0180$), and fine vegetation ($p = 0.2978$, $r^2 = 0.0089$) showed no relationship. For the 1968 aerial photos, a similar pattern was discovered. A statistically significant positive relationship between lead concentrations in soil and building ($p = 0.0004$, $r^2 = 0.197$) cover was found (Figure 4). There appears to be a relationship between lead in soil and pavement cover; however, it was not statistically significant ($p = 0.0641$, $r^2 = 0.0589$). Bare soil ($p = 0.4004$, $r^2 = 0.0124$), coarse vegetation ($p = 0.1410$, $r^2 = 0.0376$), and fine vegetation ($p = 0.4276$, $r^2 = 0.0111$) showed no relationship.

Discussion

By partnering existing soil lead data with land cover data, this research tests the hypothesis that land cover is a better predictor of lead concentrations in soil compared to land use. The results support the idea that certain landscape features, specifically pavement and building cover, are important to the spatial patterning of lead in soil. The relationship between lead in soil and building and pavement cover was expected given that buildings and roads represent the two main anthropogenic sources of lead to the environment: lead-based paint and leaded gasoline. However, the relationship between lead in soil and land cover features does not account for much of the variation in the data, supporting earlier claims that additional factors may be more important to the spatial patterning of lead in soil (Pouyat *et al.* 2007). Even though the relationship is not strong, evidenced by low r^2 values, these data support the idea that metrics of land cover that account for individual landscape features, in contrast to land use, may be better suited to mapping the spatial distribution of lead in soil.

The high variability of lead in soil may impact the relationship between soil lead levels and land cover features within a 30 meter buffer. High variability within a plot may result in lead levels better correlating with land cover directly where the samples were collected and not as strongly to the surrounding area. Soil samples collected from the UFORE plots were composite samples. By combining multiple samples into one, much of the inherent variability in lead concentrations may have been lost explaining weak relationships between lead concentrations in soil and metrics of both land use *and* cover. In addition, the soil lead data used in this analysis were not originally collected to examine patterns of soil lead in the landscape. Studies conducted with the specific goal of measuring lead in soil generally sample surface soils. Sampling the top 10 cm of soil could result in lower lead measurements since lead is generally concentrated in the topmost layer of soil.

A slightly stronger relationship between soil lead levels and 1968 building cover compared to 2004 building cover (Figures 3 and 4) was discovered. This highlights the importance of considering historic data when examining current patterns of a persistent environmental chemical. This could be a real pattern due to the 1968 land cover more accurately describing the landscape features on the ground during peak deposition from automobiles. However, it could also be an artifact of the photographs.

There are several land cover patterns that could explain the difference between lead levels and 1968 and 2004 building cover. Buildings that were constructed after 1968 would be included in the 2004 building cover estimates. However, these building would be less likely to contain lead-based paint and therefore not contribute to lead concentrations in soil. This in particular could account for the weaker relationship between lead in soil and 2004 building cover. Plant growth and the associated increase in canopy cover between 1968 and 2004 could result in greater coarse vegetation cover in 2004 and a reciprocal decrease in building cover as the rooflines become obscured by vegetation. This could also contribute to the weaker relationship between lead in soil and 2004 land cover.

Beyond temporal differences such as increased vegetation, there are differences in the actual photographs that may contribute to differences in the 1968 and 2004 analyses. The 1968 photographs are leaf-off compared to the 2004 photographs which are leaf-on. Therefore, building footprints are much more visible in the 1968 photographs (Figure 5). The differences between photograph series could result in an underestimation of coarse vegetation cover and an overestimation of building cover which could have important implications for the relationships observed. In addition, the 2004 photographs are false-color infrared making vegetation appear red and therefore easier to identify compared to the 1968 photographs. Although steps were taken to reduce potential biases due to difference among photograph series, these challenges are unavoidable when conducting historical photograph analyses.

Land *use* and land *cover* relationships with lead in soil yield contrasting results. The analyses presented here support the hypothesis that some individual features of land cover are better predictors of lead retention in urban soils than land use. While past research in Baltimore has shown no correlation with lead levels in soil and land use (Pouyat *et al.* 2007), these results suggest that lead levels are better correlated with some elements of land cover, such as building and impervious cover. These results have important implications to our understanding of the spatial distribution of lead in soil and can be used to map areas of high contamination in the landscape. An adequate understanding of the spatial distribution of lead in soil is needed in order to inform remediation efforts and protect vulnerable populations from potential exposure to a serious neurotoxin.

To gain a more complete understanding of the distribution of lead in soil, future work should focus on the contribution that individual landscape features make to the spatial patterning of lead in soil. This can be accomplished by conducting fine scale intensive soil sampling stratified by land cover features predicted to be important to the spatial patterning of lead, such as buildings and roads. In addition to fine scale patterns of lead in soil, a better understanding regarding the influence of the larger landscape context is also needed. For example, future studies, in addition to fine scale sampling, could also consider where the sampling location is in relation to the major road networks or the age of the neighborhood that the sample is collected from. Consideration of both fine scale patterns in partner with the larger landscape context should provide the most accurate information regarding the spatial distribution of lead in soil.

Literature Cited

Agency for Toxic Substances and Disease Registry (ATSDR). 1988. The nature and extent of lead poisoning in children in the United States: A report to Congress, July 1988.

Anderson J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. Land use and land cover classification systems for use with remote sensor data. USGS Professional Paper 964, 1-28.

Browne, F.L. and D.F. Laughnan. 1953. Effect of coating thickness on the performance of house paints under different programs of maintenance. Official Digest 137-159.

Cadenasso, M.L., S.T.A. Pickett, and K. Schwarz. 2007. Spatial heterogeneity in urban ecosystems: reconceptualizing land cover and a framework for classification. Frontiers in Ecology and the Environment 5(2): 80-87.

Carvanos, J., A.L. Weiss and R.J. Jaeger. 2006. An exterior and interior dust deposition survey in New York City: Results of a 2-year study. Environmental Research 100: 159-164.

Centers for Disease Control and Prevention. Standard surveillance definitions and classifications. Cdc.gov. 28 October 2009 <http://www.cdc.gov/nceh/lead/data/definitions.htm>

Centers for Disease Control and Prevention. 2005. Blood lead levels – United States, 1999-2002. Morbidity and Mortality Weekly Report 54(20): 513-516.

Centers for Disease Control and Prevention. 1997. Update: blood lead level – United States, 1991-1994. Morbidity and Mortality Weekly Report 46(07): 141-146.

Duggan, M.J., and M.J. Inskip. 1985. Childhood exposure to lead in surface dust and soil: a community health problem. Public Health Reviews 13: 1-54.

Elless, M.P., C.A. Bray and M.J. Blaylock. 2007. Chemical behavior of residential lead in urban yards in the United States. Environmental Pollution 148: 291-300.

Environmental Health Watch. Childhood lead poisoning. Ehw.org. 28 October 2009 <http://www.ehw.org/Lead/LEAD_home3.htm>

Felipe, L., G. Regu, and B. Kock. 2003. Automatic classification of land cover with high resolution data of the Rio De Janeiro City Brazil. In: Proceedings 2nd GRSS/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Areas. Berlin. 172-176.

Flanders, D., M. Hall-Beyer, and J. Pereverzoff. 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Canadian Journal of Remote Sensing 29(4):441-452.

Gilbert, S.G., B. Weiss. 2006. A rationale for lowering the blood lead action level from 10 to 2 μg/dL. NeuroToxicology 27: 693-701.

Gomes, P.C., M.P.F. Fontes, A.G. da Silva, E. de S. Mendonca, and A.R. Netto. 2001. Selectivity sequence and competitive adsorption of heavy metals by Brazilian soils. Soil Science Society of America Journal 65: 1115-1121.

JMP 7.0.2. 2007. SAS Institute Inc.

Kelly, J., I. Thornton and P.R. Simpson. 1996. Urban geochemistry: a study of the influence of anthropogenic activity on the heavy metal content of soils in traditionally industrial and non-industrial areas of Britain. Applied Geochemistry 11: 363-370.

Koller, K., T. Brown, A. Spurgeon, and L. Levy. 2004. Recent developments in low-level lead exposure and intellectual impairment in children. Environmental Health Perspectives 112(9): 987-994.

Lanphear , B.P., K. Dietrich, P. Auinger, and C. Cox. 2000. Cognitive deficits associated with blood lead concentrations <10 mg/dL in US children and adolescents. Public Health Report 115: 521–9.

Legerwerff, J.V. and A.W. Specht. 1970. Contamination of roadside soil and vegetation with cadmium, nickel, lead, and zinc. Environmental Science and Technology 4(7): 583-586.

Mahon, I. 1997. Caregivers' knowledge and perceptions of preventing childhood lead poisoning. Public Health Nursing 14(3): 169-182.

Mielke, H.W. 1999. Lead in the inner cities. American Scientist 87(1): 62-73.

Mielke, H.W. and P.L. Reagan. 1998. Soil is an important pathway of human lead exposure. Environmental Health Perspectives 106(1): 217-229.

Mielke, H.W. 1997. Leaded dust in urban soils shown to be greater source of childhood lead poisoning than leaded paint. Lead Perspectives. March/April 997:28-31.

Miranda, L.M., D.C. Dolinoy and M.A. Overstreet. 2002. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environmental Health Perspectives 110 (9): 947-953.

Moeller, M.S., W.L. Stefanov, and M. Netzband. 2004. Characterizing landcover changes in a rapidly growing metropolitan area using long term satellite imagery. In: ASPRS Annual Conference Proceedings. Denver, Colorado. May 2004.

Motto, H.L., R.H. Daines, D.M. Chilko, and C.K. Motto. 1970. Lead in soils and plants: its relationship to traffic volume and proximity to highways. Environmental Science and Technology 4(3): 231-237.

Ordonez, A., J. Loredo, E. DeMiguel, and S. Charlesworth. 2003. Distribution of heavy metals in the street dusts and soils of an industrial city in Northern Spain. Archives of Environmental Contamination and Toxicology 44(2): 160-170.

Payton, M., K.M. Riggs, A. Spiro III, S.T. Weiss and H.Hu. 1998 Relations of bone and blood lead to cognitive function: The VA normative aging study. Neorotoxicology and Teratology 20(1): 19-27.

Pouyat, R.V., I.D. Yesilonis, J. Russell-Anelli, and N.K. Neerchal. 2007. Soil chemical and physical properties that differentiate urban land-use and cover types. Soil Science Society of America Journal 71(3): 1010-1019.

Pouyat, R.V., I.D. Yesilonis, and N.E. Golubiewski. 2008. A comparison of soil organic stocks between residential turf grass and native soil. Urban Ecosystems 12: 45-62.

Reiners, W. A., R. H. Marks, and P. M. Vitousek. 1975. Heavy metals in subalpine and alpine soils of New Hampshire. Oikos 26: 264–275.

Schwartz, B.S., W.F. Stewart, K.I. Bolla, D. Simon, K, Bandeen-Roche, B. Gordon, J.M. Links, and A.C. Todd. 2000. Past adult lead exposure is associated with longitudinal decline in cognitive function. Neurology 55: 1144-1150.

Shannon, M.W. 1996. Etiology of childhood lead poisoning. *In* Lead poisoning in childhood. Edited by S.M. Pueschel, J.G. Linakis and A.C. Anderson. Brooks Publishing Co. pp. 37-58.

Silbergeld, E.K. 1992. Neurological perspective on lead toxicity. *In* Human Lead Exposure. Edited by H.L. Needleman. CRC Press, Boca Raton, Florida. pp. 89-103.

Smith, W.H. and T.G. Siccama. 1981. The Hubbard Brook Ecosystem Study:

biogeochemistry of lead in the northern hardwood forest. Journal of Environmental Quality 10(3): 323-333.

Stewart, W.F., B.S. Schwartz, C. Davatzikos, D. Shen, D. Liu, X. Wu, A.C. Todd, W. Shi, S. Bassett, and D. Youssem. 2006. Past adult lead exposure is linked to neurodegeneration measured by MRI. Neurology 66: 1476-1484.

Tadesse, W., T.L. Coleman, and T.D. Tsegaye. 2003. Improvement of land use and land cover classification of an urban area using image segmentation from Landsat ETM$^+$ data. In: Proceedings of the 30th International Symposium on Remote Sensing of the Environment. November 10-14, 2003. Honolulu, Hawaii.

United Nations Population Fund. Linking population, poverty, and development. Unfpa.org. 28 October 2009 <http://www.unfpa.org/pds/urbanization.htm>.

United State Census Bureau. Geographic comparison table. Census.gov. 28 October 2009 <http://factfinder.census.gov/servlet/GCTTable?_bm=y&-geo_id=01000US&-_box_head_nbr=GCT-P1&-ds_name=DEC_2000_SF1_U&-format=US-1>.

United States Department of Agriculture Forest Service. Urban Forest Effects Model – UFORE. Fs.fed.us. 28 October 2009 <http://www.nrs.fs.fed.us/tools/ufore/>.

United States Department of Housing and Urban Development. Legislative history of lead-based paint. Hud.gov. 28 Oct 2009 <http://www.hud.gov/offices/cpd/affordablehousing/training/web/leadsafe/ruleoverview/legislativehistory.pdf>.

United States Department of Housing and Urban Development. Lead. Hud.gov. 28 October 2009 <http://www.hud.gov/offices/lead/library/hhi/Lead.pdf>.

United States Environmental Protection Agency. 2006. Air quality criteria for lead. EPA/600/R-5/144aF. Research Triangle Park, North Carolina.

United States Environmental Protection Agency. Method 3050B. Epa.gov. 28 October 2009 <http://www.epa.gov/waste/hazard/testmethods/sw846/pdfs/3050b.pdf>.

Wang, J., H. Ren, J. Liu, J. Yu, and X. Zhang. 2006. Distribution of lead in urban soil and its potential risk in Shenyang City, China. Chinese Geographical Science 16(2): 127-132.

Weathers, K.C., G.M. Lovett, G.E. Likens, and R.Lathrop. 2000. The effect of landscape features on deposition to Hunter Mountain, Catskill Mountains, New York. Ecological Applications 10(2): 528-540.

Wong, C.S.C., X. Li, and I. Thornton. 2006. Urban environmental geochemistry of trace metals. Environmental Pollution 142(1): 1-16.

XTools Pro.v1. 2003. Data East LLC., Novosibirsk, Russia.

Yesilonis, I.D., B.R. James, R.V. Pouyat, B. Momen. 2008. Lead forms in urban turfgrass and forest soils as related to organic matter content and pH. Environmental Monitoring and Assessment 146: 1-17

Acknowledgements

Thank you to Dr. David Nowak of the USDA Forest Service, Northeastern Research Station for the development of the UFORE model. I greatly benefited from conversations with Ian Yesilonis regarding the UFORE soil data. I would also like to thank the Baltimore City Planning Department, especially Thor Nelson and Jamie Gerhart for providing the historic aerial photos for this research. Finally, thank you to the University of Vermont's Spatial Analysis Lab, especially Jarlath O'Neil Dunne, for georeferencing the historic aerial photos. This work was supported by National Science Foundation grants DEB 042376 and 0808418.

Figure Legends

Figure 1. Hierarchy of urban landscape structure as defined by HERCULES (Cadenasso *et al.* 2007). These landscape features may be important to lead retention.

Figure 2. Land cover features in the 30 meter buffer around a soil lead sample point (red point) in Baltimore, MD. The top panel shows the individual landscape features or elements: BD = Building, CV = Coarse Vegetation (trees and shrubs), FV = Fine Vegetation (grasses and herbs) and PV = Pavement. The bottom panel shows the same buffer with solid colors representing the individual landscape elements.

Figure 3. Soil lead levels (ppm) regressed against the area (m^2) of individual landscape features for the 2004 aerial photographs. A significant relationship was found between lead and buildings ($p = 0.0287$) and lead and pavement ($p = 0.0078$). Bare soil, coarse vegetation, and fine vegetation showed no relationship.

Figure 4. Soil lead levels (ppm) regressed against the area (m^2) of individual landscape features for the 1968 aerial photographs. A significant relationship was found between lead and buildings ($p = 0.0004$). Pavement, although not statistically significant, also showed a relationship with lead in soil ($p = 0.0641$). Bare soil, coarse vegetation, and fine vegetation showed no relationship.

Figure 5. Land cover features in the 30 meter buffer around a soil lead sample point in Baltimore, MD. The top panel shows the land cover for 1968 while the bottom panel shows the land cover for 2004. Of importance to the visual interpretation of these photos is that the 1968 photographs are leaf-off while the 2004 photographs are leaf-on. The 2004 photographs are also false color infrared and therefore vegetation is red.

Figure 1.

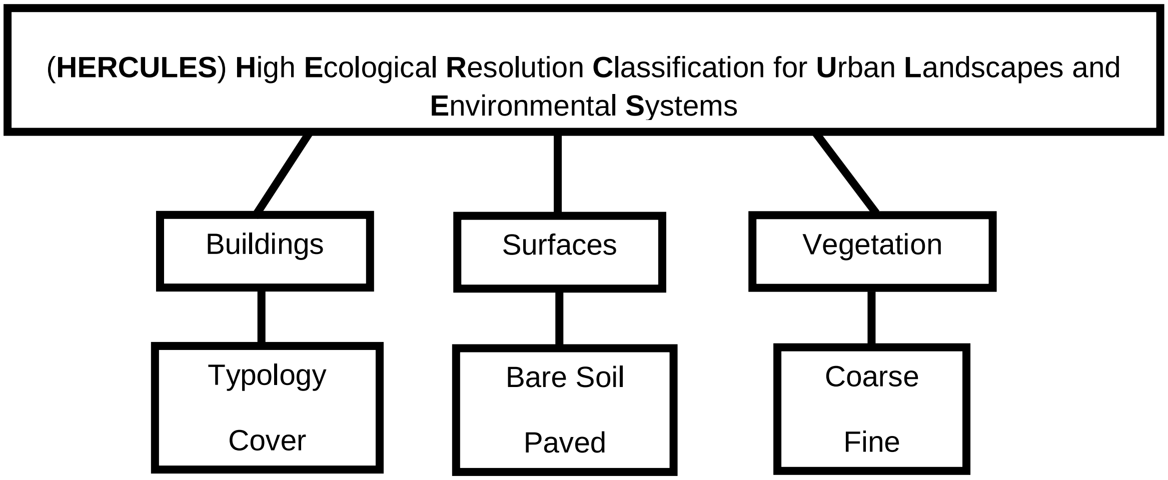

(**HERCULES**) **H**igh **E**cological **R**esolution **C**lassification for **U**rban **L**andscapes and **E**nvironmental **S**ystems
Buildings
Surfaces
Vegetation
Typology
Cover
Bare Soil
Paved
Coarse
Fine

Figure 2.

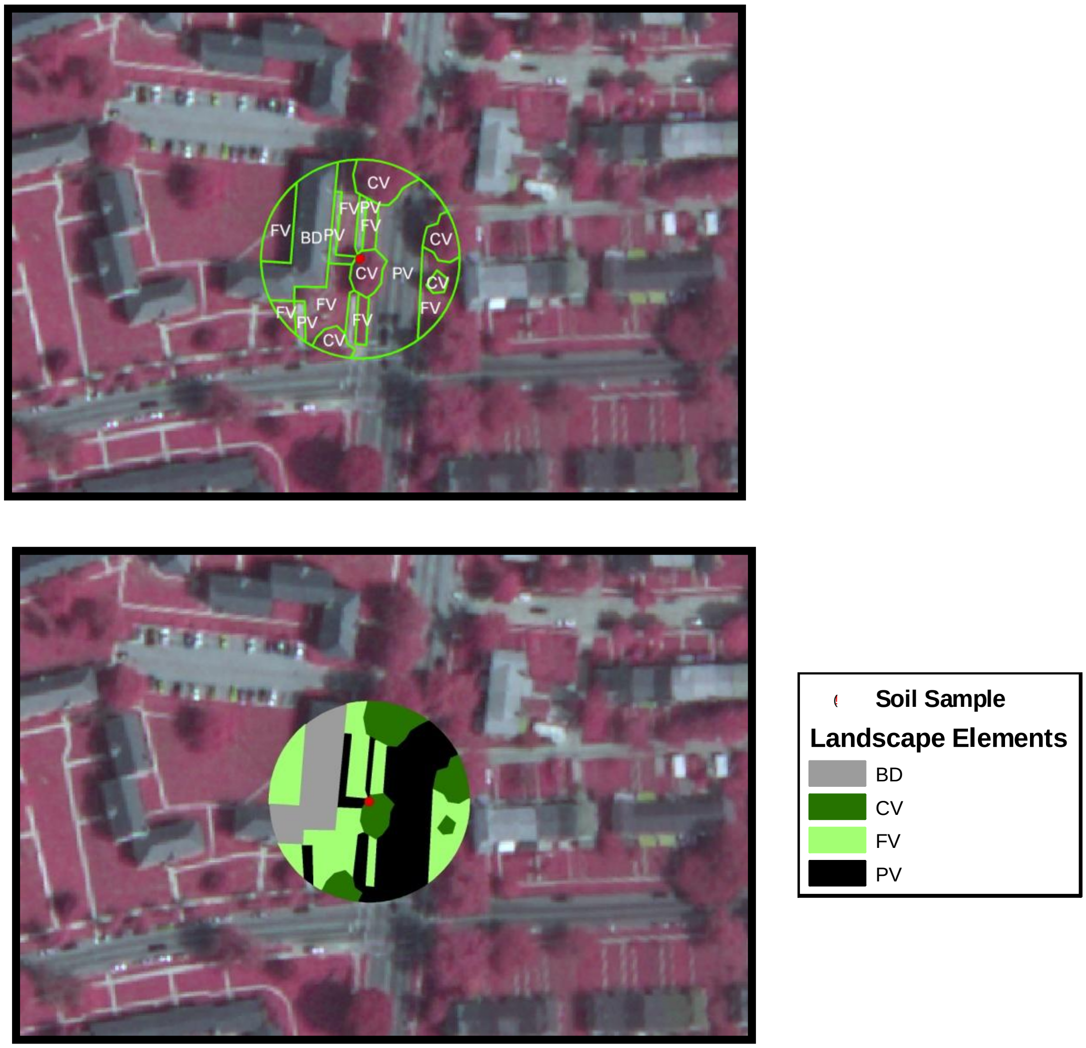
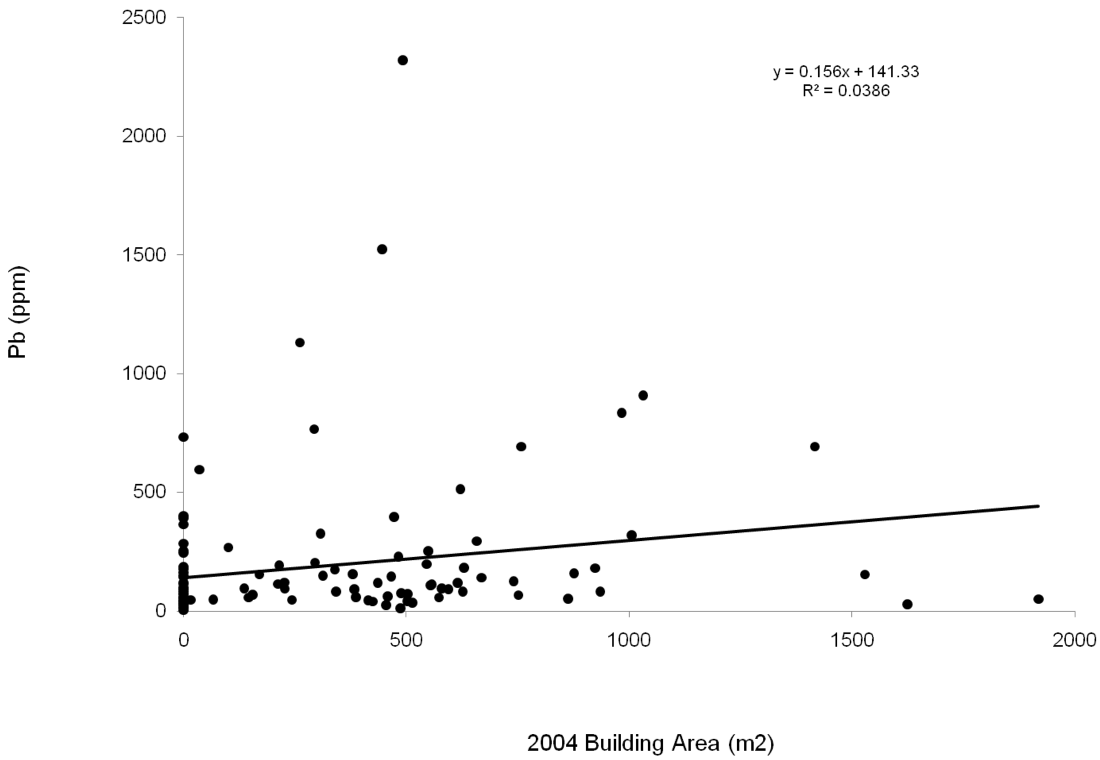


Figure 3.


y = 0.156x + 141.33
R² = 0.0386
Pb (ppm)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2004 Building Area (m2)

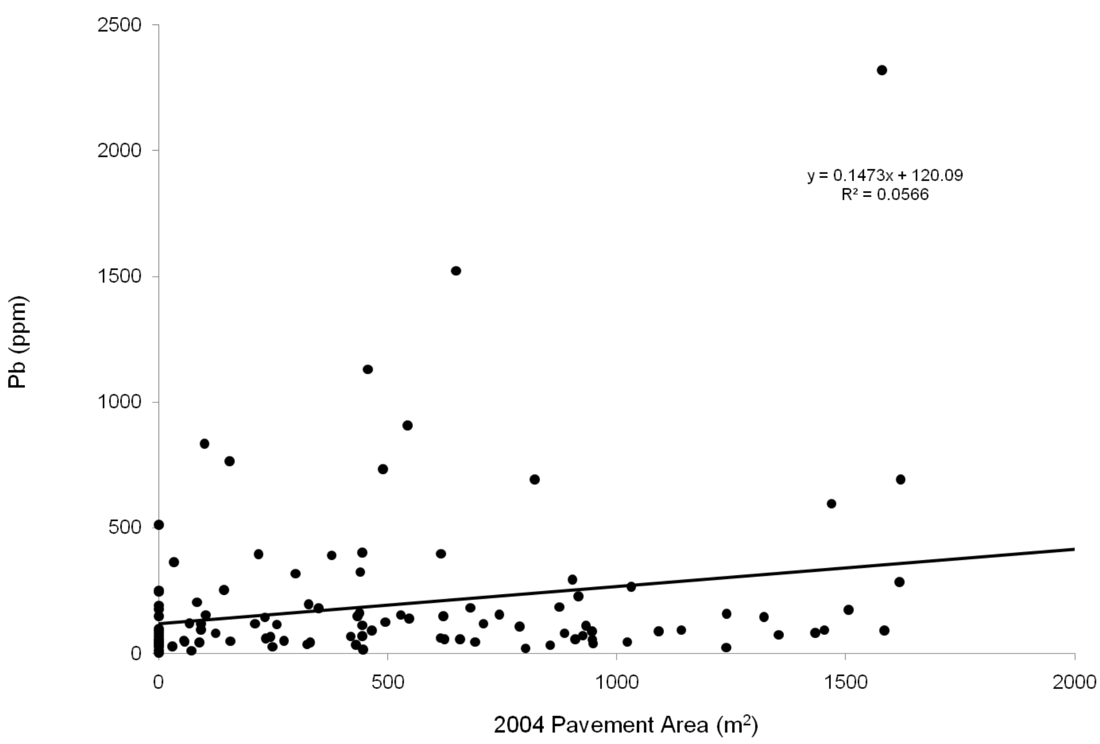
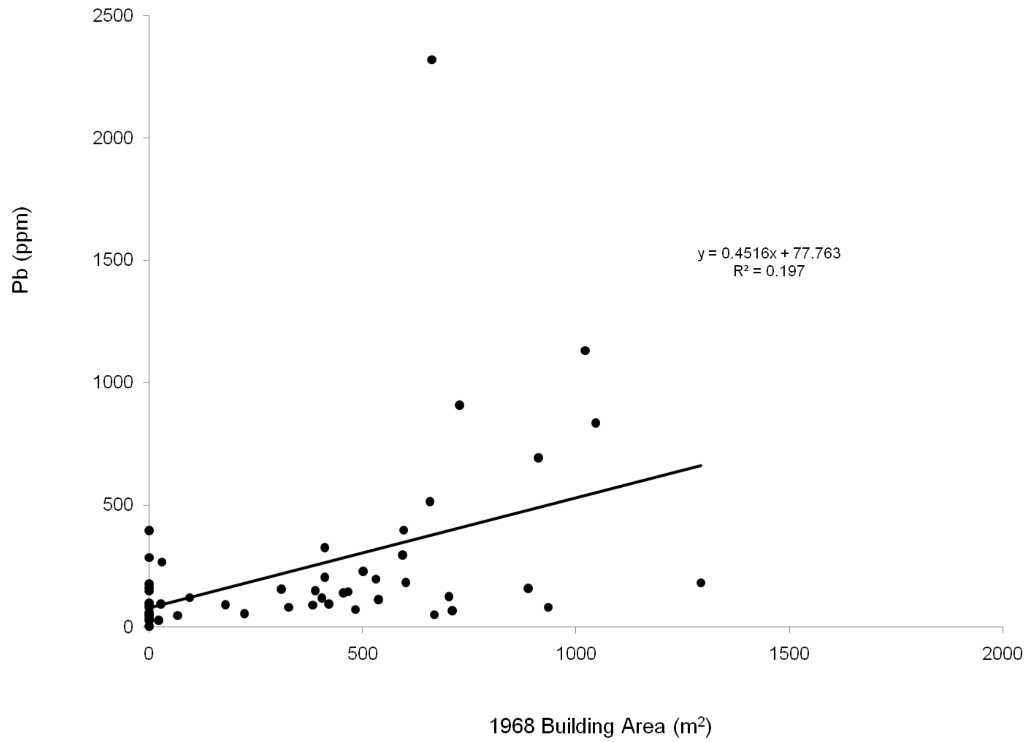

y = 0.1473x + 120.09
R² = 0.0566
Pb (ppm)
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2004 Pavement Area (m²)

Figure 4.


2500
2000
1500
1000
500
0
Pb (ppm)
y = 0.4516x + 77.763
R² = 0.197
0
500
1000
1500
2000
1968 Building Area (m2)

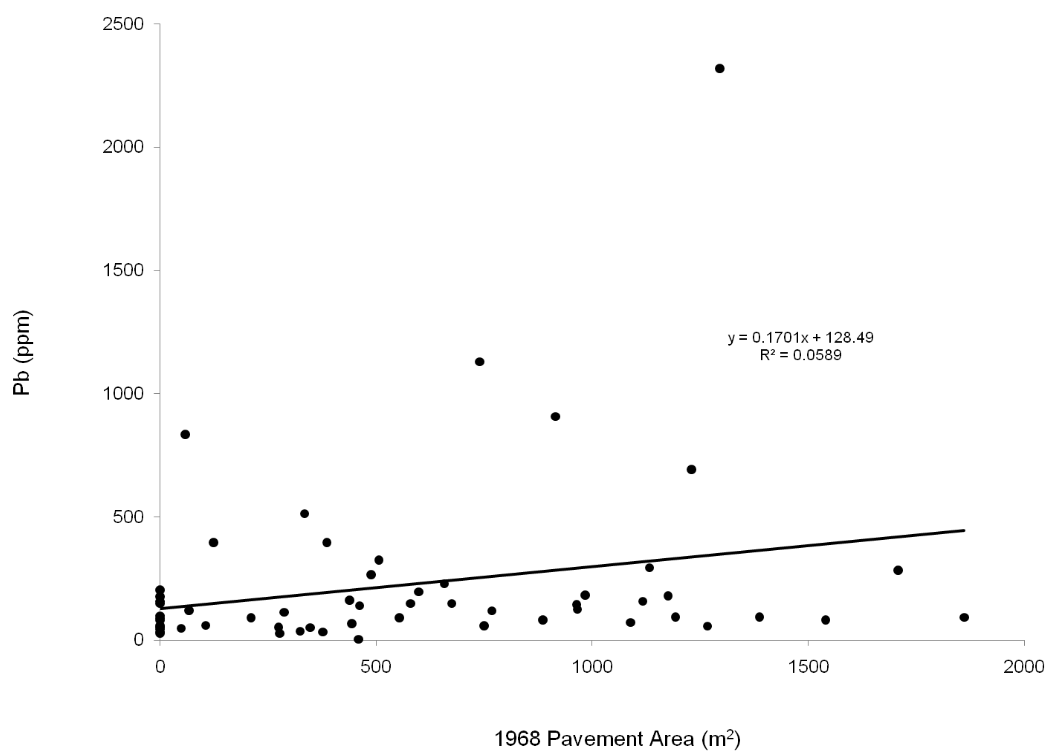

2500
2000
1500
1000
500
0
Pb (ppm)
y = 0.1701x + 128.49
R² = 0.0589
0
500
1000
1500
2000
1968 Pavement Area (m2)

Figure 5.

CHAPTER 2. ASSSESSING THE IMPORTANCE OF LANDSCAPE CONTEXT AND INDIVIUDAL LANDSCAPE FEATURES TO THE SPATIAL DISTRIBUTION OF LEAD IN URBAN RESIDENTIAL SOILS OF BALTIMORE, MARYLAND USA.

Abstract

The ecosystem concept has traditionally been used to understand the pools, fluxes and cycling of nutrients in a defined system. I apply the ecosystem concept to an urban system to examine the internal controls that regulate the spatial distribution of a critical environmental pollutant, lead, in residential soils. Lead contamination of urban residential soils is a public health concern. While most are aware of the dangers associated with lead based paint, fewer are cognizant of the legacy that lead containing products, including paint and leaded gasoline, have left in soil. Lead contamination of soil is widely dispersed and highly variable. As a result, there is as a great need to identify hotspots in the landscape in order to facilitate successful remediation of lead contaminated soil. Ecologists often assume that land use is a predictor of environmental contamination. However in the case of soil lead concentrations, previous research in Baltimore, Maryland has shown that *land use* is not a useful proxy. Based on earlier work that showed a correlation between lead concentrations in soil and some elements of land cover, I propose that *land cover* is a better predictor of lead contamination in the urban environment. Thus, I have conducted fine scale intensive sampling of soil lead on 61 residential properties in Baltimore Maryland using a field portable x-ray fluorescence (XRF) multi-element analyzer to examine which elements of land cover are most important to the spatial patterning of lead in soil. Sampling was stratified by landscape

features that are predicted to affect lead levels including housing age, distance to major road networks, and housing material. Thirty percent of the properties sampled had average XRF lead concentrations that exceed the United States Environmental Protection Agency (USEPA) reportable limit of 400 ppm and 53% had at least one reading that exceeded the USEPA reportable limit. The mean value for all field measurements within the detection limit of the XRF was 362 ppm and the median value was 123 ppm. Field results were confirmed using Atomic Absorption Spectroscopy (AAS) and showed a very strong correlation with field results ($p = <0.0001$, $r^2 = 0.82$). XRF field results were also confirmed using an XRF lab technique ($p = <0.0001$, $r^2 = 0.87$). Fine scale measurements of lead in residential soils of Baltimore City highlight the prevalence of a persistent environmental pollutant in residential soils which could potentially pose a public health risk, especially to vulnerable populations. These data will be used in future work to construct an empirically based model of soil lead concentrations in Baltimore City residential soils.

Introduction

Lead (Pb), a naturally occurring metal found in very small quantities of the Earth's crust, has become a common and widely dispersed environmental pollutant. Urban areas, in particular, have been significantly affected by greater amounts of lead in the environment. Background levels of lead in unpolluted agricultural soils typically average 10 ppm (Holmgren *et al.* 1993); however, urban areas commonly exhibit background levels that are much higher. Wong *et al.* (2006) found an order of magnitude difference between mean urban soil Pb concentrations in Hong Kong (89.9 ppm) and rural soil Pb concentrations (8.66 ppm). Lead contamination in soil stems primarily from two important anthropogenic sources: lead-based paints and leaded gasoline. Even though these products are generally no longer used in the United States, old lead-based paint and particulates from the combustion of leaded gasoline have become incorporated into the soil. Approximately 24 million dwellings built before 1978, contain lead-based paint (United States Department of Housing and Urban Development, Lead, hud.gov) and an estimated 4 to 5 million metric tons of lead was released into the atmosphere from the combustion of leaded gasoline (ATSDR 1988).

Lead contaminated soil may contribute to childhood blood lead levels. In recognition of the dangers associated with lead in soil, the United States Environmental Protection Agency (USEPA) has set two separate guidelines for lead in soil: 400 parts per million (ppm) for children's play areas and 1,200 ppm for all other areas of the yard (USEPA, Federal register 40 CFR Part 745 Lead;

Identification of dangerous levels of lead; Final Rule, epa.gov). Reducing the amount of lead in soil may reduce children's blood lead levels. For example, a study addressing the effect that soil lead abatement had on children's blood lead levels found that a reduction of 2060 ppm lead in soil translated to a 2.25 to 2.70 µg/dL decrease in blood lead levels (Aschengrau *et al.* 1994). Although children are not the only population affected by elevated soil lead, they are particularly vulnerable to the effects of lead due to 1) increased hand-to-mouth activity, 2) greater absorption of lead compared to adults, and 3) greater sensitivity to the effects of lead (USEPA, Lead in paint, dust, and soil, epa.gov). Historically the focus surrounding lead and public health has been on children ingesting leaded paint chips. However, it is now widely accepted that lead dust and lead in soil pose a significant public health threat to children (Duggan and Inskip 1985, Mielke 1997). The issue of exposure is further complicated by lead speciation (Ge *et al.* 2000, Ryan *et al.* 2004) and particle size (Miranda *et al.* 2002), which can affect the bioavailability and metal sorption intensity of lead. For example, concentrations of lead have been found to increase with decreasing particle size due to an increase in surface area (Thuy *et al.* 2000). In addition, smaller lead particles can absorb more easily into the gastrointestinal and pulmonary tracts (Miranda *et al.* 2002).

It is clear that lead is both widely dispersed in the urban environment and a potential health threat, especially to children. Therefore, it is essential that we establish a conceptual framework in which to study the factors impacting the spatial distribution of this critical environmental pollutant. The ecosystem concept has been used by ecologists to understand the pool, fluxes, and cycling of nutrients and energy through a defined system. Earlier explorations of the ecosystem concept measured inputs and outputs to an ecosystem. Using this information, researchers were able to quantify the amount of nutrients or energy remaining in the system. As the field of ecosystem science progressed, the ecosystem concept evolved from simply measuring system inputs and outputs to defining the controls within the system. We can apply the ecosystem approach to urban systems as well, asking not only what enters and leaves a system but what internal processes and characteristics regulate the pools and fluxes of environmental pollutants as well as nutrients.

Scientists have previously used the ecosystem approach to understand lead dynamics in a forested ecosystem (Siccama and Smith 1978, Smith and Siccama 1981). Studies conducted at the Hubbard Brook long term ecological research site in New Hampshire have shown that while there is a significant input of atmospheric lead to a watershed, low levels of lead are present in water draining from that watershed (Siccama and Smith 1978, Bormann and Likens 1979). This work illustrates the ability of forest soils to sequester lead. The ecosystem function of lead sequestration is influenced by ecosystem structure including forest type and elevation (Weathers *et al.* 2000), soil organic matter, oxides of manganese and iron (Yesilonis *et al.* 2008a), and soil pH (Elless *et al.* 2007). In addition to forested systems, the ecosystem approach has also been applied to urban systems (Pickett *el al.* 1997). Here I extend the ecosystem approach to study soil lead dynamics in an urban system.

In addition to serving as the “brown infrastructure” of cities (Pouyat *et al.* submitted), human altered soils often provide ecosystem services comparable to those of unaltered soils, including reduced bioavailability of pollutants (Effland and Pouyat 1997). If urban soils, like forested soils, serve as a sink for anthropogenic sources of lead, a logical hypothesis is that the ecosystem function of lead sequestration remains correlated with ecosystem structure. However, structure in an urban setting is very different than that of a forested system and can be defined by three main components: surface material, buildings, and vegetation (Ridd 1995). The different components that describe urban structure could affect lead retention in soil by 1) serving as a source of lead to the environment in the case of surface material and buildings and/or 2) serving as a surface for lead particles to adhere to, in the case of buildings and vegetation.

An ecosystem approach to studying the lead dynamics in an urban system promotes a broader conceptual understanding of the issue of lead contaminated soil. Specifically, this approach accounts for the controls within the system, such as the distribution and density of landscape features that may affect the ecosystem function of lead retention. Identifying and understanding controls within the system is the first step towards forming a predictive model, which can help identify potential hotspots of lead contamination in the landscape thus informing remediation efforts and protecting human populations from potential exposure. In order to address this, I examined 1) what individual landscape features have the greatest influence on the spatial distribution of lead and 2) how the larger landscape context influences the distribution of lead in soil (Figure 1). In this research, the study system is Baltimore City. I applied intensive soil lead sampling at a spatially extensive scale to characterize the distribution of lead in soil.

Methods

Sampling Scheme

Residential parcels were selected using a stratified sampling scheme. The focus on residential soil was intentional since it is arguably the most important in regards to public health. Properties were stratified by 1) housing age (1986-present and pre-1978), 2) distance to major road networks (0-30 meters, 30+ meters) and 3) housing material (brick and wood frame). Newer houses constructed in 1986 or later were built after the 1978 ban on lead-based paint and the 1986 ban on leaded gasoline. Housing age was determined using the Maryland Property View Dataset (Maryland Department of Planning, Maryland property data, state.md.us). The contribution of lead from automobiles can be assessed both spatially and temporally. Parcels within 0-30 meters of a road may be more heavily affected by historic leaded gasoline deposition than parcels 30 meters or farther from a road. The distance of 30 meters was selected based on several sources that have documented an increase in soil lead concentrations adjacent to major roadways (Motto *et al*. 1970, Ordonez *et al.* 2003, Wang *et al.* 2006). In the absence of reliable traffic density data, road size was used as a surrogate variable. Major roads were identified using the Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) database (United States Census Bureau, Topologically Integrated Geographic Encoding and Referencing (TIGER) database, census.gov). Roads described as "primary highways with limited access," "primary roads without limited access," and "secondary and connecting roads" were classified as major road networks[1]. Both brick and wood frame homes were sampled. This contrast in housing material was aimed at distinguishing between the different sources of lead contamination (lead-based paint and leaded gasoline). This assumes that brick homes contain little or no external lead-based paint.

Our sampling scheme consisted of four sampling groups (Table 1). Sampling groups 1 and 2 represent older homes close to the major road networks that differ in housing material with group 1 consisting of brick buildings and group 2 consisting of wood frame buildings. Sampling groups 1 and 3 represent older brick homes that differ in their distance to the major road networks. Properties in group 1 are within 30 meters of the major road networks while properties in group

---

[1] Primary highways with limited access (A1) are described in the TIGER classification: "Interstate highways and some toll highways are in this category (A1) and are distinguished by the presence of interchanges. These highways are accessed by way of ramps and have multiple lanes of traffic. The opposing traffic lanes are divided by a median strip. The TIGER/Line files may depict these opposing traffic lanes as two distinct lines in which case, the road is separated". Primary roads without limited access (A2) are described in the TIGER classification: "This category (A2) includes nationally and regionally important highways that do not have limited access as required by category A1. It consists mainly of US highways, but may include some state highways and county highways that connect cities and larger towns. A road in this category must be hard-surface (concrete or asphalt). It has intersections with other roads, may be divided or undivided, and have multi-lane or single-lane characteristics". Secondary and connecting roads are described in the TIGER classification: "This category (A3) includes mostly state highways, but may include some county highways that connect smaller towns, subdivisions, and neighborhoods. The roads in this category generally are smaller than roads in category A2, must be hard-surface (concrete or asphalt), and are usually undivided with single-lane characteristics. These roads usually have a local name along with a route number and intersect with many other roads and driveways".

3 are farther than 30 meters from the major road networks. Sampling groups 1 and 4 correspond to brick homes close to the major road networks that differ in housing age with group 1 representing older homes and group 4 representing newer houses. The contrasts achieved through this design allows us test the importance of landscape context by examining the contribution of housing age, housing material, and distance to major roadways to the lead burden of residential soils. Sixty one properties were sampled because a minimum of 60 sampling locations (parcels) is recommended for subsequent spatial analysis (Fortin and Dale 2005).

Site Selection

Sampling was limited to owner occupied housing to avoid potential tenant landlord conflicts. Ownership status was identified using the Maryland Property View Dataset and confirmed using the Maryland Department of Assessment and Taxation Real Property Data Search (Maryland Department of Assessment and Taxation, Real property data search, resiusa.org). Within a Geographic Information System (GIS), potential properties were identified using the query function. In addition to selecting homeowners based on the sampling criteria, potential sites were also selected based on their proximity to one another and geographic distribution across Baltimore City. Letters were sent to homeowners to solicit participation in the project. Following initial letters, reminder post cards were sent and follow-up phone calls were placed. When phone calls were not possible, home visits were conducted. The overall response rate for homeowner recruitment was approximately 8%.

Sampling

Residential parcels in Baltimore City were sampled November 2007 through September 2008. Soil lead measurements were made *in situ* using a USEPA approved x-ray fluorescence (XRF) multi-element spectrum analyzer, which allows for efficient in-field soil sampling for lead concentrations. A Niton XLt 700 series multiple element analyzer was used for the measurements. Soil metal content to a depth of approximately 2 mm was evaluated with *in situ* surface measurement. Following USEPA method 6200 (2007), Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment, a minimum of 5% of *in situ* samples were confirmed via laboratory analyses. Samples were analyzed for lead concentration at an independent USEPA recognized lab (BTS laboratories, Richmond VA) using Atomic Absorption Spectroscopy (AAS) analyses.

XRF technology has been used extensively by the USEPA and others (Clark *et al.* 1999, Carr *et al.* 2008). In x-ray fluorescence, an atom is hit with a high energy photon. The source of the high energy photon comes from a radio isotope or miniature x-ray tube. The high energy photon results in an electron being ejected from the K or L shell of an atom. The ejected electron is replaced by an electron from the L or M shell. The replacement electron drops down to a shell of lower energy releasing energy in the form of an x-ray. The x-ray that is emitted is unique for each element and is measured by the XRF analyzer. This information is converted into a spectrum, which qualitatively evaluates whether or not that element is present. Based on what medium is being tested a mode calibration (thin, thick, or paint) is selected. The mode calibration converts the spectra into a quantitative output. The thick/bulk mode is used for soil analysis. This assumes that a sample is infinitely 'thick'. This is problematic in that obstructions in the soil (rock etc.) prevent the x-ray from reaching the detector, which may affect the results. In addition, soil moisture is known to affect results. Water in the soil absorbs x-rays preventing them from reaching the detector resulting in decreased lead concentration readings. Even with the restraints inherent in XRF field measurements, this method allowed us to gather the density of samples necessary to examine spatial patterns of lead in urban residential soil. The limit of detection (LOD) for lead is dependent on the type of sample being tested and the other elements that are present in the sample; however, for our purposes, lead concentrations were high enough to be detected except in very few cases where readings were below the LOD of the instrument.

Within each parcel, soil samples were collected in transects extending from the house. At every site an attempt was made to take a reading as close to the structure as possible. Moving away from the structure a minimum of 3 samples were collected. At each sampling location, the corresponding landscape feature was recorded as 1) open lawn with no tree canopy above 2) under tree canopy,

3) adjacent, i.e. within 1 meter, to building, 4) next to the road, meaning the measurement that was closest to the major road network, or 5) landscaped (flower beds, etc.). All the landscape features that applied per location were recorded. The number of samples taken per parcel was dependent on the size of the parcel.

In addition to surface samples, three soil cores were collected at each site using a soil probe with a ¾ inch diameter. When possible, core sampling locations were restricted to the backyard to reduce any visible disturbance to the yard. Sample locations were selected randomly within the backyard but generally restricted to the open lawn if possible. If the backyard was inaccessible or there was no soil present in the backyard, samples were collected in the front yard. Each core was collected to a depth of 12 cm. In the field, the sample was divided into four sub-samples representing different depths (0-3 cm, 3-6 cm, 6-9 cm, and 9-12 cm) for a total of 12 samples per site. A total of 732 core samples were collected.

Sample preparation was completed at the Cary Institute of Ecosystem Studies in Millbrook, NY. Samples were dried at 60°C until a consistent weight was maintained. Wet and dry weights were recorded and gravimetric water content was calculated using the following equation: (wet weight – dry weight)/dry weight. Samples were passed through a 2 mm sieve (Fisher Scientific Company) and ground in a ball mill (Kleco Model 4200). Samples were analyzed for total Pb content at BTS laboratories in Richmond, VA (www.btslabs.com) using USEPA method SW-846, 7420 (lead, atomic absorption, direct aspiration).

In contrast to the XRF field technique, where no sample preparation was involved, our XRF lab technique involved sample preparation by drying, sieving, and grinding, as described above. At least 10 grams of soil was then placed in an XRF cup, which consists of a plastic ring and cup, mylar sheet, whatman filter paper, and polyester fiberfill. The samples were run on a Niton XLt 700 series multi-spectrum analyzer at the Cary Institute of Ecosystem Studies. Core samples were also prepared and analyzed via XRF at the Cary Institute.

Results

Lab Results

Following USEPA method 6200 (2007) for field portable XRF, a minimum of 5% of field samples were confirmed in a laboratory setting by comparing field results to AAS results. Regression analysis showed a very strong correlation between XRF field results and AAS lab results ($p = <0.0001$, $r^2 = 0.82$, Figure 2). The slope of the line (y = **0.7637**x + 15.765) indicates that our field measurements underestimate the amount of lead in soil compared to AAS lab results. I also compared the field results to a XRF lab technique and found a very similar relationship ($p = <0.0001$, $r^2 = 0.87$). The slope of the line (y = **0.6715**x + 21.8) again indicates that our field measurements underestimate the amount of lead in soil compared to XRF lab results. In comparing the two lab techniques, regression analysis showed a very strong correlation between the XRF lab technique and the AAS lab technique ($p = <0.0001$, $r^2 = 0.97$, Figure 3). The slope of the line (y = **1.1549**x - 25.102) reveals a slight overestimation using the XRF lab technique. This is comparable to results reported in the literature. Griffith *et al.* (2008) reported a similar relationship between lab XRF and ICP measurements ($r^2 = 0.972$).

Field Results

The USEPA has set two separate guidelines for lead in soil: 400 parts per million (ppm) for children's play areas and 1,200 ppm for all other areas of the yard (USEPA, Federal register 40 CFR Part 745 Lead; Identification of dangerous levels of lead; Final Rule, epa.gov). Many publications compare results to the more conservative guideline of 400 ppm. I will do the same here. The mean value for all measurements taken within the detection limit of the instrument (n = 1121) was 363 ppm. The standard deviation was 794 ppm and the standard error was 24 ppm. The median value was 124 ppm. The lowest reading within the level of detection was collected from a landscaped portion of the yard of an older brick house far from a major road (7.4 ppm). The highest reading (9151 ppm) within the upper limits of the XRF (10,000 ppm) was collected adjacent to a painted porch of an older brick building far from a major road. Thirty percent of the properties sampled had average Pb values that exceeded 400 ppm and 53% had at least one reading somewhere on the property that exceeded 400 ppm. Median lead values were higher next to buildings compared to other areas in the yard that were sampled (Figure 4). This pattern was consistent for both brick and wood frame homes. When examining data from all 61 properties together, an ANOVA indicates mean lead concentrations are significantly different among landscape features ($F = 63.491$, $df = 3$, $p < 0001$). The lead concentration data was log transformed in order to meet the assumption of a normal distribution. Using Tukey's multiple comparison test, I found significant differences between classes in which soil lead concentrations near buildings were significantly greater than all other landscape features and concentrations near major roads were significantly greater than those near lawns or under trees (Figure 4). The significance of these results, however, may be confounded by the block sampling design, in which multiple XRF readings were taken within each yard. None of the properties built after 1986, after the ban on leaded gasoline and lead-based paint, exhibited any field readings that exceed the USEPA reportable limit of 400 ppm (Figure 5).

When examined for all sampling groups, the soil core data showed very little change in lead concentration with depth. The median Pb value for the 0-3 cm samples was 182 ppm and only increased slightly with depth: 185 ppm for the 3-6 cm samples, 186 ppm for the 6-9 cm samples, and 186 ppm for the 9-12 cm samples. Examining the core data by sampling group provides more information (Figure 6). Median lead values for all sampling depths are higher for the older brick and frame homes that are close to a major road compared to other sampling groups.

Field lead concentration results were categorized as 0-400 ppm (low), 400-2000 ppm (moderately high), 2000-5000 ppm (high) and greater than 5000 ppm (very high). These categories were based on the USEPA lead safe yard project (USEPA, Lead safe yard project, epa.gov). The field results were distributed to homeowners (Figure 7a) through color coded maps. In addition, homeowners were also provided with a USEPA fact sheet regarding lead poisoning and a Maryland Cooperative Extension fact sheet on lead in garden soils. XRF field data was added to a GIS in order to calculate distance to roadways and distance to buildings for every field measurement taken (Figure 7b). Lead levels decrease with increasing distance from major roads and buildings.

A Note on Sources

Determining the sources of lead in soil is beyond the scope of this study; however, the XRF provides information on many elements and some interesting patterns have been identified that are worth mentioning. There are several patterns in the data that 1) point to multiple sources of Pb in Baltimore's soil and 2) highlight the importance of leaded-gasoline, in addition to leaded paint, to soil lead pollution in Baltimore.

Previous studies have looked at element ratios, specifically Lead/Titanium (Pb/Ti), to determine the source of lead in soil (Clark *et al.* 2006). Since Ti is a component of both lead-based paint (Pb titanate) and lead-free paint (titanium dioxide), Ti is an indicator of total paint products in soil. Soils contaminated with lead-based paint exhibit enrichment of both Pb and Ti, while soils contaminated with leaded-gasoline exhibit higher Pb levels compared to Ti (Clark *et al.* 2006). In our Baltimore study soils, Pb and Ti in soil do not appear to be enriched in the same way, as evidenced by the lack of a strong correlation between the two elements ($p = <0.0001$, $r^2 = 0.067$, Figure 8). If paint was the only source of lead to soil, one would expect a stronger correlation (higher r^2 value) between the two elements. In addition, other studies have shown that Pb/Ti ratios from non-urban paint sources are typically less than 1 (Clark *et al.* 2006). Samples collected in this study exhibit Pb/Ti ratios greater than 1 for samples collected next to both brick and wood frame buildings indicating Ti concentrations are lower than Pb concentrations (Figure 9). This suggests multiples sources of lead, such as lead-based paint and leaded gasoline, are at play. Finally, examination of the small percentage of samples that exceeded 1000 ppm Pb indicates the possibility of contamination from both lead-based and lead-free paint (Figure 10). Paint is generally assumed to contribute to lead levels that are this high. Soil samples with high Pb values and lower Ti values are likely contaminated with *lead-based* paint while samples with high Ti values and lower Pb levels are likely contaminated with *lead-free paint*. The fact that soil may be contaminated from both types of paint makes pinpointing sources even more challenging.

Discussion

It is imperative to understand the effect that urban characteristics, including physical, biological, and chemical, have on the distribution of trace metals (Wong *et al.* 2006). Previous studies, such as the USEPA lead safe yard project, have examined fine scale patterns of lead on residential properties, but their focus was not on relating the distribution of lead to the distribution of landscape features. In addition, these studies did not consider the larger landscape context in which residential parcels are embedded. Consideration of fine scale patterns in partner with the larger landscape context, is necessary if the goal is to predict lead distribution in soil at the city scale. Spatial prediction at this scale requires knowledge regarding both the influence of landscape context, such as road networks and neighborhood age, and spatial patterns at the scale of the individual parcel, such as the influence of buildings, trees, and lawns.

The results of this study illustrate the influence of both landscape context and landscape features. A graph of the field measurements plotted by distance to the major road networks shows a sharp decrease in lead concentration with increasing distance away from the road network (Figure 11a). A noticeable exception to this pattern occurs at approximately 1,500 meters. This anomaly is likely due to a lead-based paint source at a particular sampling location. When measurements that were taken directly adjacent to buildings were eliminated and the pattern re-examined, the overall pattern of decreasing lead with increasing distance from the road remained while the anomaly disappeared (Figure 11b). Housing age is also an important predictor of lead concentrations in soil.

Although Baltimore City was more progressive regarding its stance on lead-based paint by banning interior lead-based paint in 1951 and requiring warning labels on leaded paint in 1958 (Mushak and Crocetti 1990), federal regulations did not ban the sale of paint containing more that 0.06% lead until 1978. The fact that Baltimore City banned interior lead paint and required warning labels on all leaded paint in the 1950s might speak to a potentially important regional influence. Examining all field measurements by housing age reveals that the highest lead levels are present in homes built during the 1920s and 1930s (Figure 5). This corresponds to the time period when lead-based paint was heavily used. Measurements taken from houses built after the ban on lead-based paint and leaded gasoline do not exceed the USEPA reportable limit of 400 ppm. This is especially important considering that previous work has demonstrated the re-suspension of contaminated soil in urban soils (Clark *et al.* 2008, Laidlaw and Filippelli 2008). Re-suspension of fine particulates, which are easily inhaled, can have important health implications (Wong *et al*. 2006). If re-suspension of contaminated soil is occurring on the properties that were built after 1986, it is not at a level considered toxic by the USEPA.

In addition to the influence of landscape context, I also examined the importance of individual landscape features. In comparing measurements taken from four locations 1) immediately adjacent to a built structure, 2) next to a major road, 3) in the lawn but away from buildings and the road, and 4) under tree canopy but away from buildings and the road, a difference in lead concentrations relative to sampling location is revealed (Figure 4). This is consistent with earlier work done by Mielke *et al.* (2004) which reported a difference among sampling locations and median Pb levels in New Orlean's soil with samples near the foundation of buildings exhibiting the highest lead levels, followed by busy streets, residential streets, and open areas. Results from this study show that measurements taken close to buildings and roads are higher compared to measurements taken in the lawn or under tree canopy. In addition to having a higher mean and median value, measurements taken directly adjacent to a building, regardless of building material, also exhibit a much wider range. Since measurements taken next to buildings may be influenced by a paint source, this variability could be explained by both paint condition and the variable amounts of lead used in lead-based paints.

The patterns revealed by the results are not surprising considering that both buildings and roads represent a source of lead to the environment. These sources, however, seem to be operating at different spatial scales. Lead contamination of the soil associated with built structures seems to be very localized operating at a very fine spatial scale (Figure 12) whereas roads appear to have a much more extensive influence (Figure 11). This is most likely due to differences in the mode of distribution and particle size. Mielke (1999) has suggested that particles from the combustion of leaded gasoline accumulate next to roadways as well as travel and adhere to buildings resulting in elevated lead next to buildings. Others (Cook and Ni 2007) have proposed a modification of this hypothesis (modified aerosol hypothesis), suggesting that only particles of intermediate mass have an affinity for buildings. Smaller particles, they suggest, become blanket contamination while larger particles are deposited close to the source. This is consistent with the patterns found in this study where paint sources seem to contribute to higher lead concentrations that are localized and gasoline contributes to more spatially disperse lower levels of contamination.

The patterns revealed by the lead and titanium data are important because they highlight the complexity of lead sources that are found in urban systems. The lack of a strong correlation between lead and titanium suggest that soil lead contamination in Baltimore is a combination of multiple sources and cannot be accounted for by paint alone. Therefore, hotspots next to buildings are likely to occur in the absence of paint and hotspots next to painted surfaces may be the result of a combination of lead sources. In addition, there may be contamination of Ti from both lead-based and lead-free paint, which further complicates modeling efforts. The results for Pb:Ti ratios in Baltimore are very different than what has been documented in Boston. In a study with a similar number of soil samples collected from urban gardens in Boston using the same XRF techniques, much higher Ti numbers are observed highlighting the importance of paint to soil lead contamination (Clark *et al.* 2006, Clark *et al.* 2008). This further strengthens the argument that leaded gasoline is an important source in Baltimore.

The patterns revealed by the soil core data are also noteworthy. Lead concentrations have been shown to be higher at the surface of the soil profile compared to lower in the soil profile (Wang *et al.* 2006); however, this study examines the vertical distribution of lead in finer increments. The results indicate that in urban systems, where disturbance of the soil profile is common, lead may not be concentrated in the top of the soil profile. Looking at all samples together, there was very little difference in lead concentration with sampling depth. Upon closer examination of the results by sampling group, differences exist between groups, but there was still no distinct pattern with depth. Interestingly, none of the sampling groups show the highest median values at the surface 0-3 cm. This could be due to burial of lead in the soil profile, leaching of lead in the soil profile, or physical disturbance of the soil profile. New brick homes close to a major road exhibit the highest median lead values at depths of 9-12 cm. This might be explained by clean fill that is brought in during construction possibly covering up older lead contamination. The soil core data suggest that assumptions regarding lead concentration and depth may not be true for complex urban systems where disturbance of the soil profile is common.

While interpreting the results of this study, it is important to keep in mind that this work addresses total lead concentrations. Although studies have shown that total lead concentrations correlate with the amount of bioavailable lead (Clark *et al.* 2006), total lead concentrations do not represent the amount of lead in the soil that is readily available. The amount of bioavailable lead in soil is dependent on speciation (Ge *et al.* 2000) and particle size (Miranda *et al.* 2002). It is also important to recognize that 400 ppm is a federal standard. Some states have set guidelines for lead in soil that are much lower than this; for example, Minnesota has a bare soil standard of 100 ppm. Reagan and Silbergeld (1989) have advocated for a national lead soil standard of 100 ppm based on protecting vulnerable populations. Other countries, as well, have set lower standards for soil. For example, the Netherlands have a target bare soil lead standard of 85 ppm (Yesilonis *et al.* 2008b). Using the Netherlands target value of 85 ppm, 95% of properties sampled in this study would exhibit lead contamination above the target level.

This research highlights the spatially heterogeneous distribution of lead in a complex ecosystem type. Residential landscapes are a rapidly expanding important component of the human dominated urban landscape. I used the ecosystem concept to address the spatial distribution of a critical environmental pollutant in complex urban ecosystems, and specifically within residential areas. Overall, I found both individual landscape features and the larger landscape context influence the spatial distribution of lead in urban residential soils. Lead concentrations varied by sampling location in which samples collected next to a building exhibited higher lead values compared to all other areas sampled. In addition, lead values decreased with increasing distance from the road networks, thus highlighting the importance of landscape context. Additionally, the field sampling technique used in this study, which strongly correlated with accepted laboratory analyses, allowed me to collect a large amount of data that will be used in future modeling efforts. Through the ecosystem approach utilized here, we can start to understand how fine scale individual landscape features partner with the larger landscape context to drive the spatial distribution of lead in soil. This knowledge is essential for modeling the spatial distribution of lead in urban residential soils, predicting lead hotspots, and targeting remediation efforts to minimize human exposure.

Literature Cited

Aschengrau, A., A. Beiser, D. Bellinger, D. Copenhafer, and M. Weitzman. 1994. The impact of soil lead abatement on urban children's blood lead levels: phase II results from the Boston lead-in-soil demonstration project. Environmental Research 67: 125-148.

Agency for Toxic Substances and Disease Registry (ATSDR). 1988. The nature and extent of lead poisoning in children in the United States: A report to Congress, July 1988.

Bormann, F.H. and G.E. Likens. 1979. Patterns and processes in a forested ecosystem. New York, Wiley and Sons.

Carr, R., C. Zhang, N. Moles, and M. Harder. 2008. Identification and mapping of heavy metal pollution in soils of a sports ground in Galway City, Ireland, using a portable XRF analyser and GIS. Environmental geochemistry and health 30(1): 45-52.

Clark, H.F., D. Hausladen, and D.J. Brabander. 2008. Urban gardens: Lead exposure, recontamination mechanisms, and implications for remediation design. Environmental Research 107(3): 312-319.

Clark, H.F., D.J. Brabander, and R.M. Erdil. 2006. Sources, sinks, and exposure pathways of lead in urban garden soil. Journal of Environmental Quality 35: 2066-2074.

Clark, S., W. Menrath, M. Chen, S. Roda and P. Succop. 1999. Use of a field portable X-Ray fluorescence analyzer to determine the concentration of lead and other metals in soil samples. Annals of Agricultural and Environmental Medicine 6(1): 27-32.

Cook, R. and L. Ni. 2007. Elevated soil lead: Statistical modeling and apportionment of contributions from lead-based paint and leaded gasoline. The Annals of Applied Statistics 1(1): 130-151.

Duggan, M.J. and M.J. Inskip. 1985. Childhood exposure to lead in surface dust and soil: a community health problem. Public Health Review 13: 1-54.

Effland, W.R. and R.V. Pouyat. 1997. The genesis, classification, and mapping of soils in urban areas. Urban Ecosystems 1:217-228.

Elless, M.P., C.A. Bray and M.J. Blaylock. 2007. Chemical behavior of residential lead in urban yards in the United States. Environmental Pollution 148: 291-300.

Fortin, M-J., and M. Dale. 2005. Spatial analysis: A guide for ecologists. Cambridge University Press. New York. 365pp.

Ge, Y., P. Murray, and W.H. Hendershot. 2000. Trace metal speciation and bioavailability in urban soils. Environmental Pollution 107(1): 137-144.

Griffith, D.A., D.L. Johnson, and A. Hunt. 2008. The geographic distribution of metals in urban soils: the case of Syracuse, NY. GeoJournal: 1-17.

Holmgren, G.G.S., M.W. Meyer, R.L. Chaney, and R.B. Daniels. 1993. Cadmium, lead, zinc, copper, and nickel in agricultural soils of the United States of America. Journal of Environmental Quality 22(2): 335–348.

Laidlaw, M. and G. Filippelli. 2008. Resuspension of urban soils as a persistent source of lead poisoning in children: A review and new directions. Applied Geochemistry 23(8): 2021-2039.

Maryland Department of Planning. Maryland property data. State.md.us. 28 October 2009 <http://www.mdp.state.md.us/data/index.htm>.

Maryland Department of Assessment and Taxation. Real property data search. Resiusa.org. 28 October 2009 < http://sdatcert3.resiusa.org/rp_rewrite/>.

Mielke, H.W., G. Wang, C.R. Gonzales, E.T. Powell, B. Le and V.N. Quach. 2004. PAHs and metals in the soils of inner-city and suburban New Orleans, Louisiana, USA. Environmental Toxicology and Pharmacology 18(3): 243-247.

Mielke, H.W. 1997. Leaded dust in urban soil shown to be greater source of childhood lead poisoning than leaded paint. Lead Perspectives (March/April): 28-31.

Mielke, H.W. 1999. Lead in the inner cities. American Scientist 87(1): 62-73.

Miranda, L.M., D.C. Dolinoy and M.A. Overstreet. 2002. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environmental Health Perspectives 110 (9): 947-953.

Motto, H.L., R.H. Daines, D.M. Chilko, and C.K. Motto. 1970. Lead in soils and plants: its relationship to traffic volume and proximity to highways. Environmental Science and Technology 4(3): 231-237.

Mushak, P. and A. Crocetti. 1990. Methods for reducing lead exposure in young children and other risk groups: an integrated summary of a report to the US Congress on childhood lead poisoning. Environmental Health Perspectives 89: 125.

Ordonez, A., J. Loredo, E. DeMiguel, and S. Charlesworth. 2003. Distribution of heavy metals in the street dusts and soils of an industrial city in Northern Spain. Archives of Environmental Contamination and Toxicology 44(2): 160-170.

Pickett, S.T.A., W.R. Burch Jr., S.E. Dalton, T.W. Foresman, J.M. Grove, and R. Rowntree. 1997. A conceptual framework for the study of human ecosystems in urban areas. Urban Ecosystems 1: 185-199.

Pouyat, R.V., K. Szlavecz, I.D. Yesilonis, P.M. Groffman, and K. Schwarz. Chemical, Physical, and Biological Characteristics of Urban Soils. Submitted to Soil Science Society of America Monograph.

Reagan, P. and E. Silbergeld. 1989. Establishing a health based standard for lead in residential soils. Trace Substances in Environmental Health (CR Cothern, ed). Suppl Environ Geochem Health 23: 199-238.

Ridd M.K. 1995. Exploring a V–I–S (vegetation–impervious surface–soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. International Journal of Remote Sensing 16: 2165–85.

Ryan, J.A., K.G. Scheckel, W.R. Berti, S.L. Brown, S.W. Casteel, R.L. Chaney, J. Hallfrisch, M. Doolan, P. Grevall, M. Maddaloni, and D. Mosby. 2004. Reducing children's risk from lead in soil. Environmental Science and Technology 38 (1): 18A-24A.

Siccama, T.G. and W.H. Smith. 1978. Lead accumulation in a northern hardwood forest. Environmental Science & Technology 12(5): 593-594.

Smith, W.H. and T.G. Siccama. 1981. The Hubbard Brook Ecosystem Study: biogeochemistry of lead in the northern hardwood forest. Journal of Environmental Quality 10(3): 323-333.

Thuy, H.T.T., H.J. Tobschall, and P.V. An. 2000. Distribution of heavy metals in urban soils - a case study of Danang-Hoian area (Vietnam). Environmental Geology 39 (6): 603-610.

United States Census Bureau. Topologically Integrated Geographic Encoding and Referencing (TIGER) database. Census.gov. 28 October 2009. <http://www.census.gov/geo/www/tiger/>.

United States Department of Housing and Urban Development. Lead. Hud.gov. 28 October 2009 <http://www.hud.gov/offices/lead/library/hhi/Lead.pdf>.

United States Environmental Protection Agency. Lead safe yard project. Epa.gov. 28 October 2009 <http://www.epa.gov/ne/leadsafe/index.html>.

United States Environmental Protection Agency. Lead in paint, dust, and soil. Epa.gov. 28 October 2009 <http://www.epa.gov/lead/pubs/leadinfo.htm#facts>.

United States Environmental Protection Agency. Federal register 40 CFR Part 745, Lead; Identification of dangerous levels of lead; Final rule. Epa.gov. 28 October 2009 <http://www.epa.gov/EPA-TOX/2001/January/Day-05/t84.pdf>.

United States Environmental Protection Agency. 2007. Method 6200: Field portable x-ray fluorescence spectrometry for the determination of elemental concentrations in soil and sediment, Revision 0. United States Environmental Protection Agency, Washington D. C., February 2007.

Wang, J., H. Ren, J. Liu, J. Yu, and X. Zhang. 2006. Distribution of lead in urban soil and its potential risk in Shenyang City, China. Chinese Geographical Science 16(2): 127-132.

Weathers, K.C., G.M. Lovett, G.E. Likens, and R.Lathrop. 2000. The effect of landscape features on deposition to Hunter Mountain, Catskill Mountains, New York. Ecological Applications 10(2): 528-540.

Wong, C.S.C., X. Li, and I. Thornton. 2006. Urban environmental geochemistry of trace metals. Environmental Pollution 142(1): 1-16.

Yesilonis, I.D., B.R. James, R.V. Pouyat, B. Momen. 2008a. Lead forms in urban turfgrass and forest soils as related to organic matter content and pH. Environmental Monitoring and Assessment 146: 1-17

Yesilonis, I., R. Pouyat, and N.K. Neerchal. 2008b. Spatial distribution of metals in soils in Baltimore, Maryland: role of native parent material, proximity to major roads, housing age and screening guidelines. Environmental Pollution 156(3): 723-731

Acknowledgements

I am especially grateful to the homeowners for access to their property and to the University of California, Davis for use of the XRF. Thank you to Mike Khavul, Kemi Osunsan, Joshua Sonner, and Tanya Watts for their assistance in the field and William Davies and I'Asia Deyo for assistance in the lab. Heather Clark and Dr. Daniel Brabander were extremely generous in sharing their knowledge of XRF technology, soil lead contamination, and element ratios. This work is a contribution to the Long Term Ecological Research Program (a program of the National Science Foundation) and the Cary Institute of Ecosystem Studies and was supported by NSF grants DEB 042376 and 0808418.

Figure Legends

Figure 1. Conceptual model depicting the sampling strategy. Individual parcels serve as the unit of study (panel b). In order to understand the spatial distribution of lead in soil the larger urban matrix in which the parcel is embedded (panel a) must be considered. In addition, collecting measurements at a fine spatial scale allows exploration of the contribution each landscape feature makes to lead retention (panel c).

Figure 2. Regression analysis of XRF field results and AAS lab results showing a strong correlation between field and lab results. Dashed line represents a 1:1 ratio; solid line represents the relationship between field and lab results.

Figure 3. Regression analysis of XRF lab results and AAS lab results showing a very strong correlation between both lab techniques. Dashed line represents a 1:1 ratio; solid line represents the relationship between lab techniques.

Figure 4. Lead concentrations (ppm) by sampling location. The “lawn” classification represents areas of the lawn not adjacent to a major roadway or building. The “under tree” classification represents samples collected directly under a tree canopy, not adjacent to a major roadway or building. The “near major road” classification represents samples closest to a major roadway defined as primary and secondary roads in the TIGER classification. The “near building” classification represents samples collected directly adjacent to a built structure. Error bars represent the lowest and highest values falling within 1.5 times the interquartile range. Median values are represented by the change in bar color from grey to black. Mean values are represented by the white diamonds; different letters represent a significant difference between means. Observations outside the error bars are not shown; however, they were not removed from the dataset.

Figure 5. XRF field measurements and housing age. The dotted black line represents the time period when leaded-paint was used (1884-1978). The solid black line represents the time period when leaded gasoline was used (1929-1986). The USEPA reportable limit for lead in soil (400 ppm) is represented by the soil gray line.

Figure 6. Vertical distribution of lead in the soil profile by sampling group. Older brick and wood frame houses close to a major road exhibit the highest lead levels, although none of the groups show a strong pattern with increasing depth.

Figure 7. An example of the color coded map that was sent to homeowners (a) and the corresponding GIS dataset with aerial photos of Baltimore City (b).

Figure 8. Lead and Titanium concentrations do not show a strong relationship. If paint was the only source of lead to soil, a stronger correlation between the two elements would be expected.

Figure 9. Lead/titanium ratios and distance to building. Ratios are higher close to the building and decrease with distance. Ratios greater than one highlight the possibility of multiple sources of lead in the soil.

Figure 10. Relationship between lead and titanium for samples with high (> 1000 ppm) lead concentrations. The sample circled in green exhibits much higher titanium levels compared to lead. This could be an example of lead-free paint contamination. In contrast, the sample circled in red exhibits high lead levels in relation to titanium and could be an example of lead-based paint contamination.

Figure 11. XRF field measurements and distance to the major roadways, defined as primary and secondary roads in the TIGER classification. The samples circled in red on the top graph are likely contaminated by a paint source. As shown in the bottom graph, they disappear when samples that were collected adjacent to buildings are eliminated.

Figure 12. XRF field measurements and distance to built structures.

Table 1. Contrasts achieved through the sampling scheme. Sites were stratified by housing age, distance from road, and housing material.

Group	***Housing Age***	***Distance to Road***	***Housing Material***	***Number of Plots***
1	>29 (1978 and older)	**0-30m**	**Brick**	**17**
2	>29 (1978 and older)	**0-30m**	**Wood**	**11**
3	>29 (1978 and older)	**>30m**	**Brick**	**25**
4	<21 (1986-2007)	**0-30m**	**Brick**	**7**

Contrast Achieved in Sampling Scheme	
Housing Material	Group 1 vs. Group 2
Distance from Major Road	Group 1 vs. Group 3
Housing Age	Group 1 vs. Group 4

Figure 1.

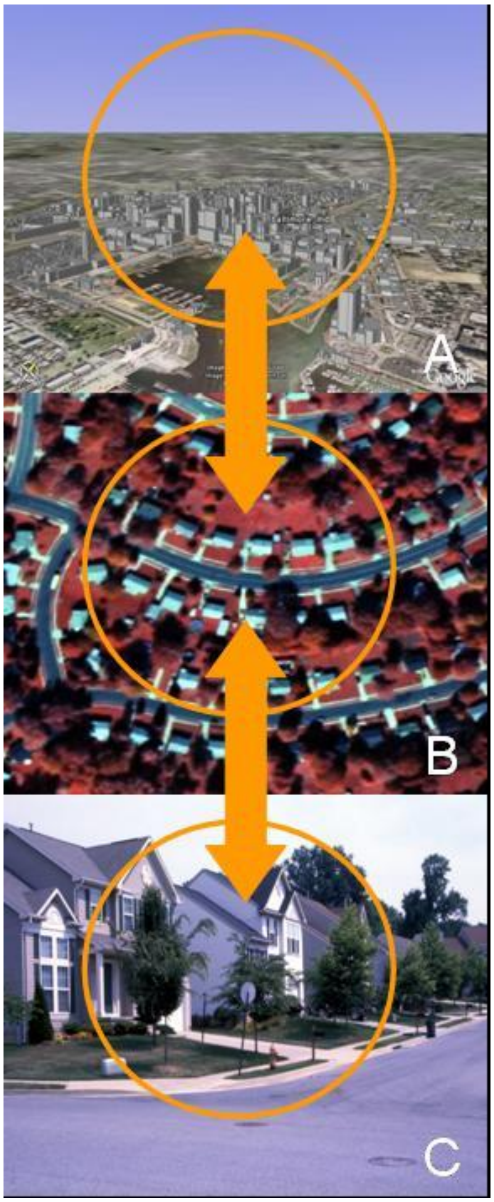


Figure 2.

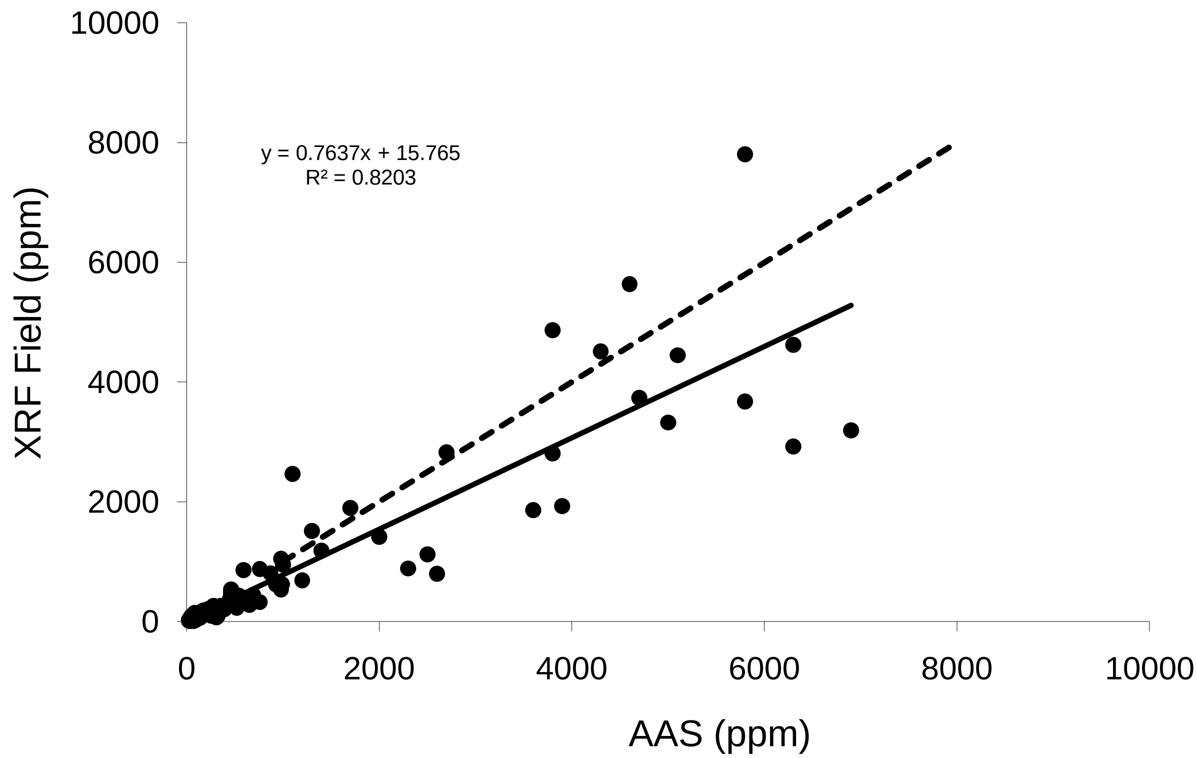

10000
8000
6000
4000
2000
0
y = 0.7637x + 15.765
R² = 0.8203
XRF Field (ppm)
0
2000
4000
6000
8000
10000
AAS (ppm)

Figure 3.

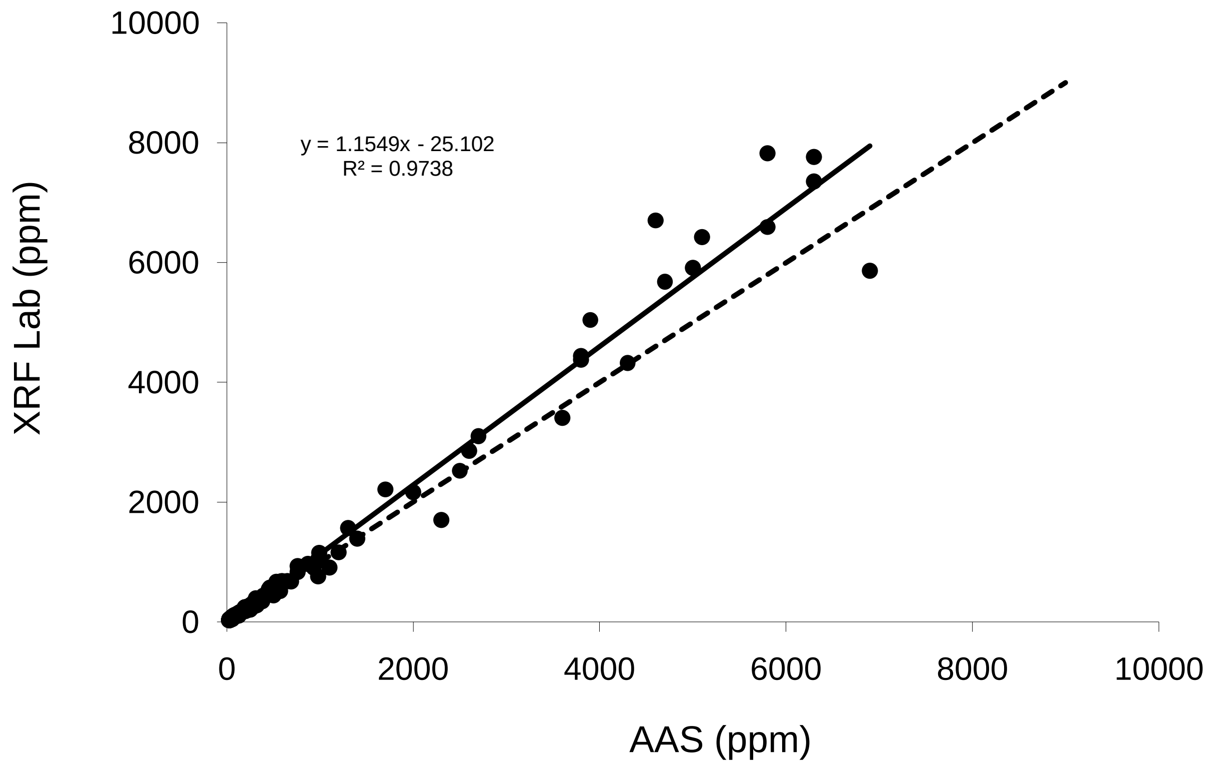

y = 1.1549x - 25.102
R² = 0.9738
XRF Lab (ppm)
AAS (ppm)
0
2000
4000
6000
8000
10000

Figure 4.

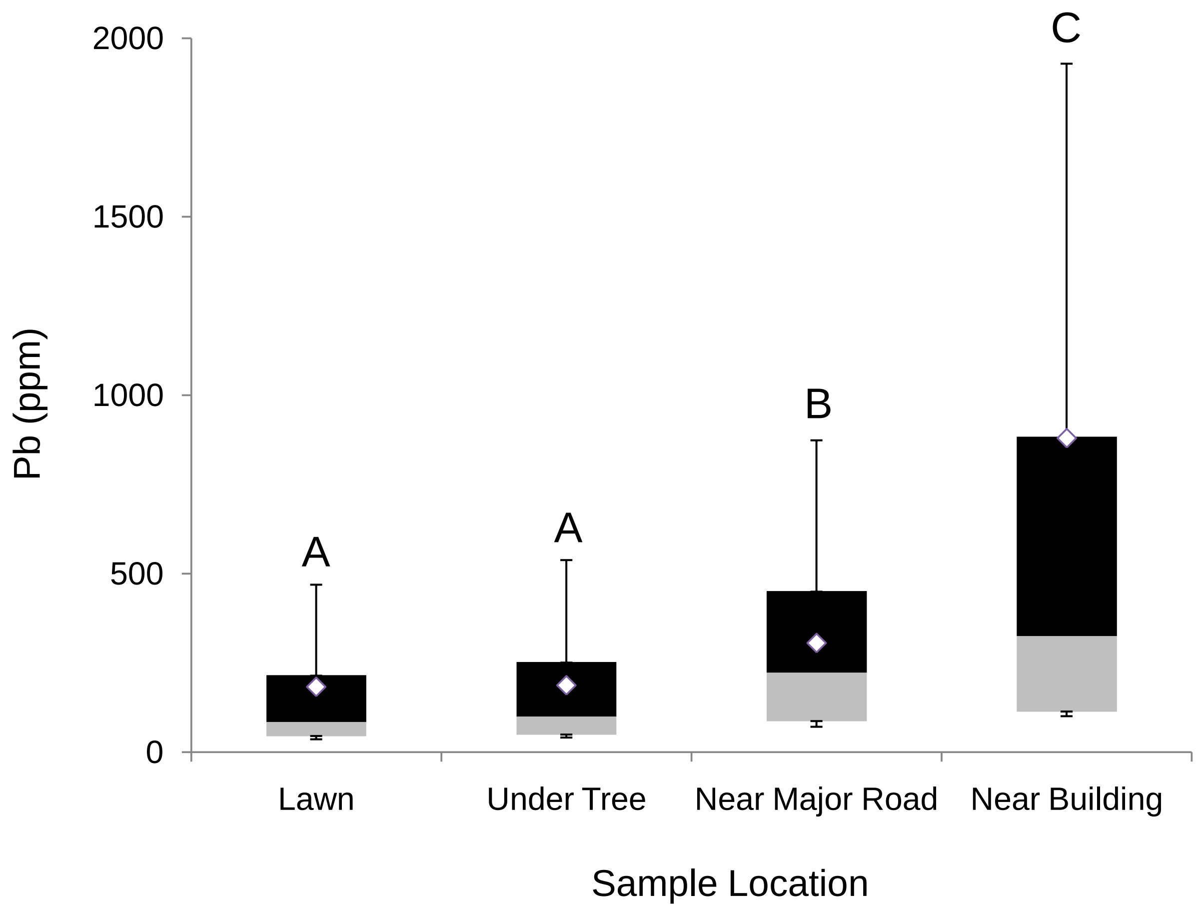

2000
1500
1000
500
0
Pb (ppm)
A
A
B
C
Lawn
Under Tree
Near Major Road
Near Building
Sample Location

Figure 5.


10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Pb (ppm)
1848
1868
1888
1908
1928
1948
1968
1988
2008
Housing Age

Figure 6.

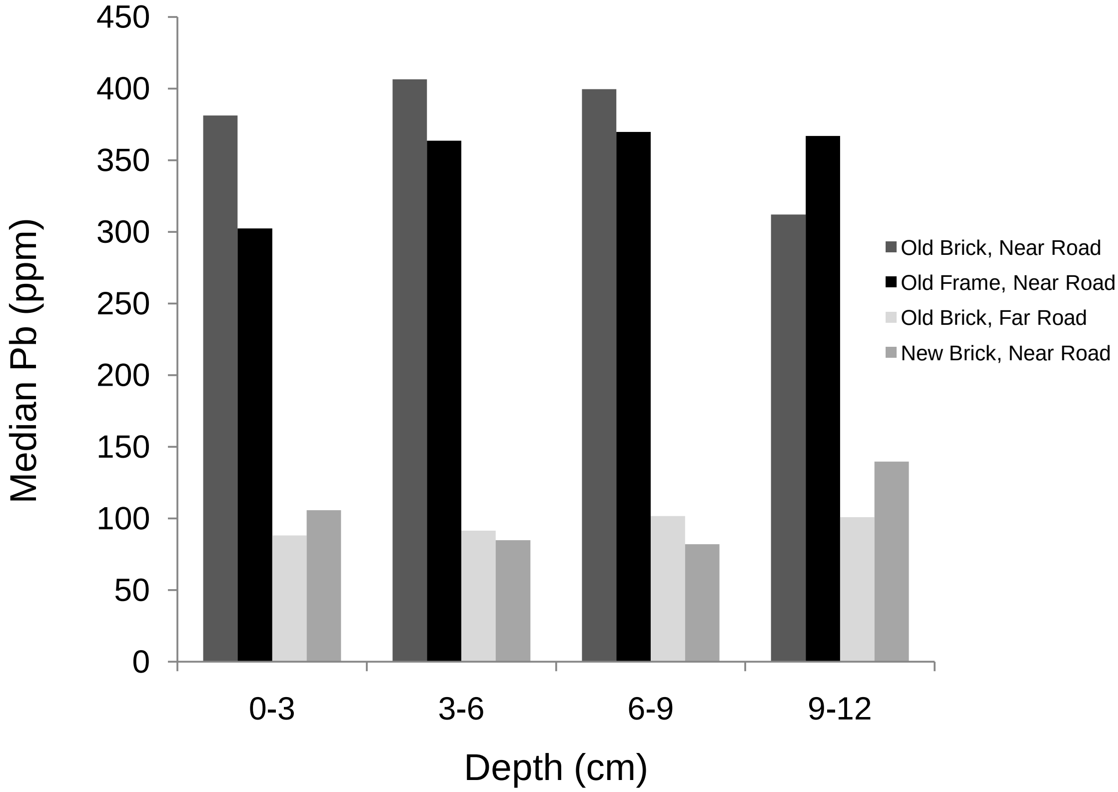

450
400
350
300
250
200
150
100
50
0
Median Pb (ppm)
0-3
3-6
6-9
9-12
Depth (cm)
Old Brick, Near Road
Old Frame, Near Road
Old Brick, Far Road
New Brick, Near Road

Figure 7.

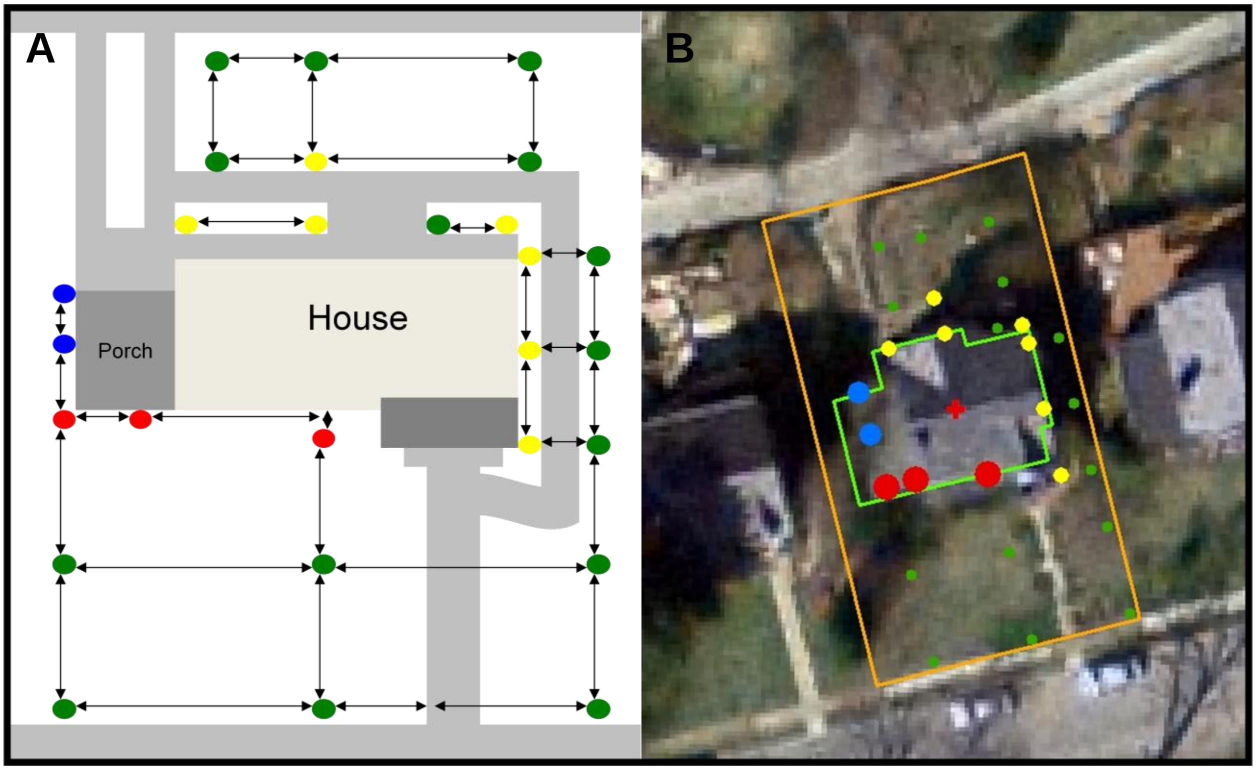


Figure 8.

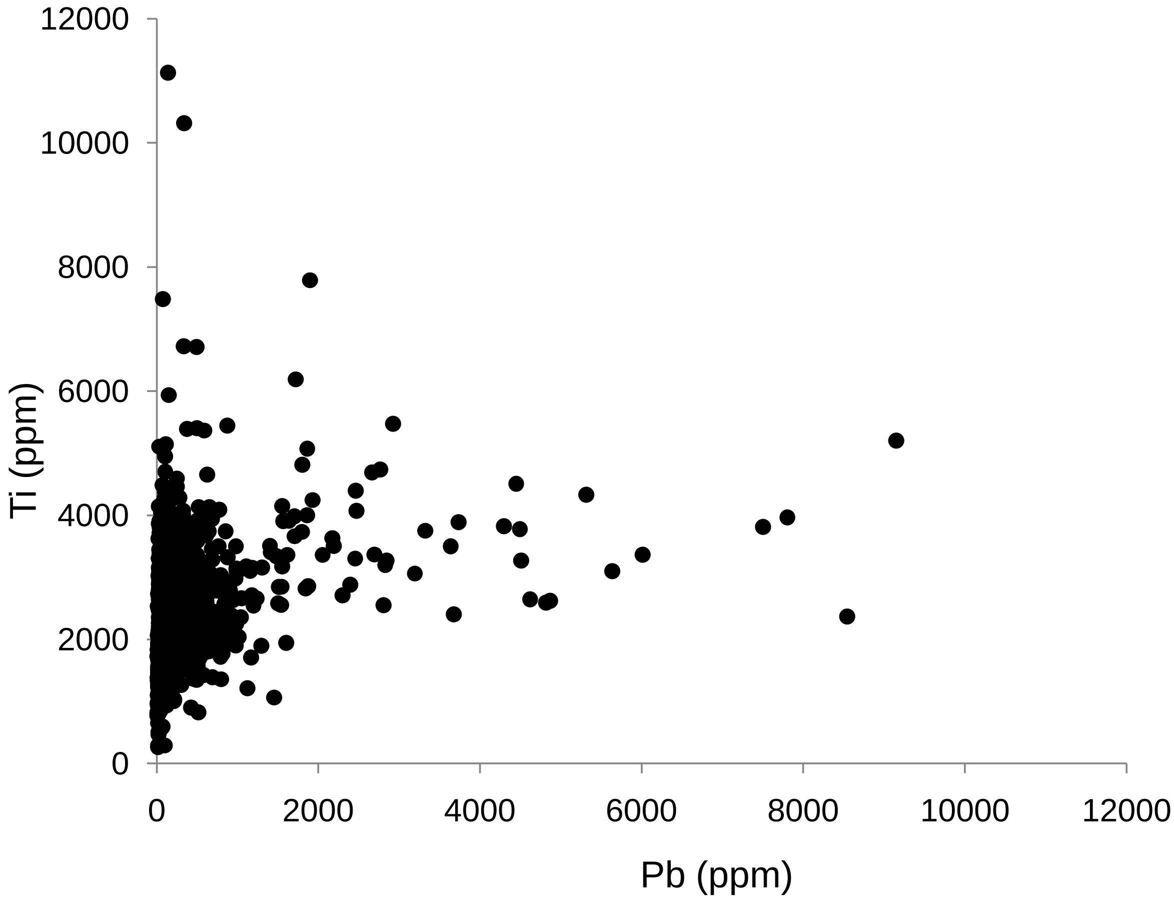

12000
10000
8000
6000
4000
2000
0
Ti (ppm)
0
2000
4000
6000
8000
10000
12000
Pb (ppm)

Figure 9.

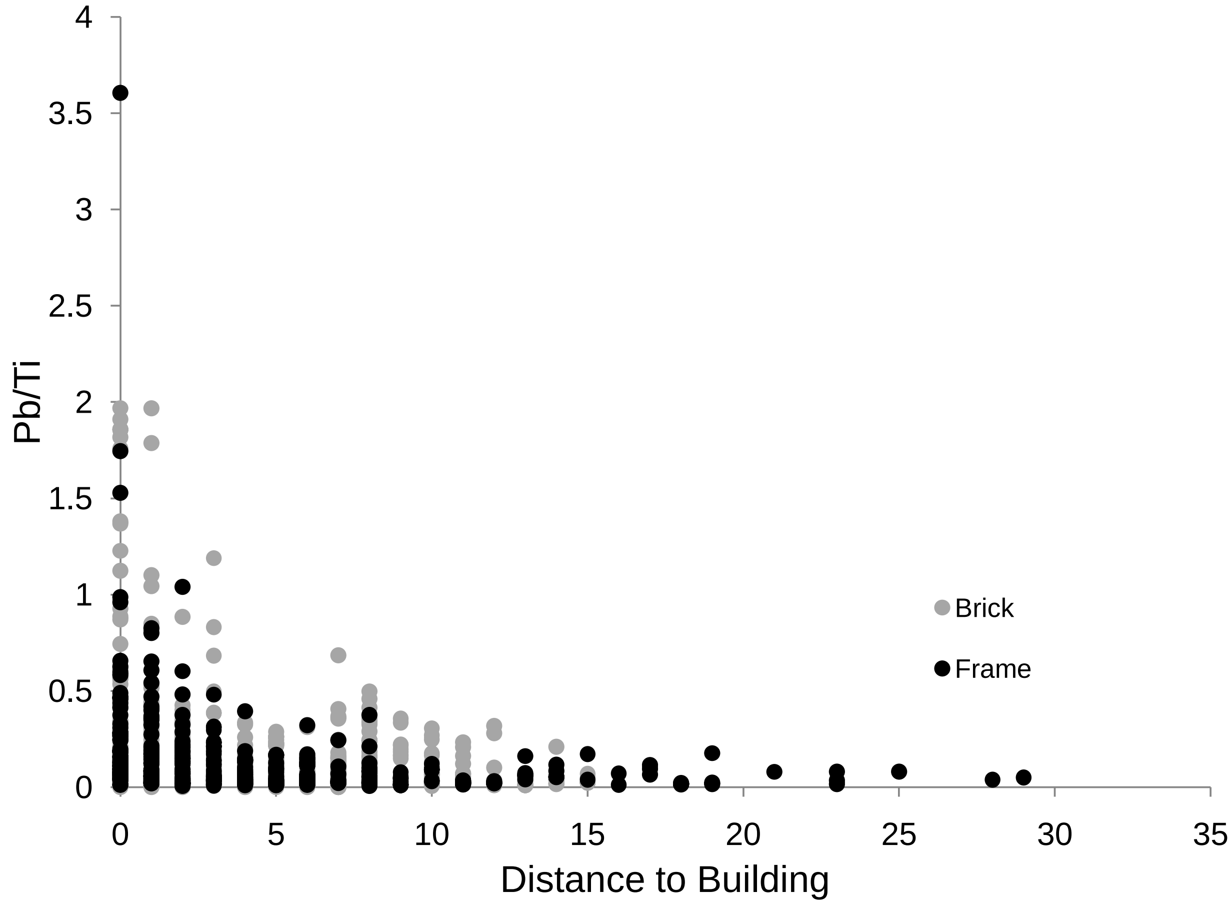

4
3.5
3
2.5
2
1.5
1
0.5
0
Pb/Ti
0
5
10
15
20
25
30
35
Distance to Building
Brick
Frame

Figure 10.

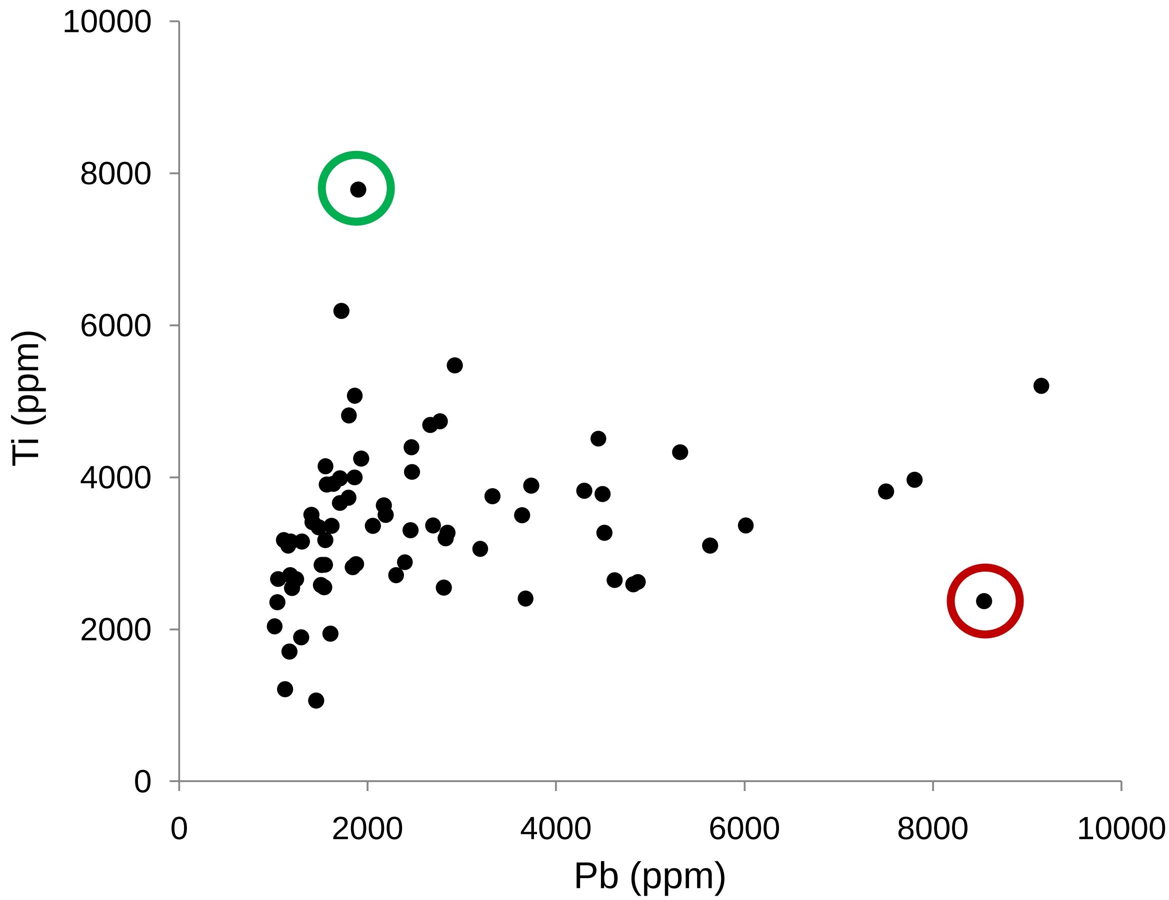

10000
8000
6000
4000
2000
0
Ti (ppm)
0
2000
4000
6000
8000
10000
Pb (ppm)

Figure 11a.

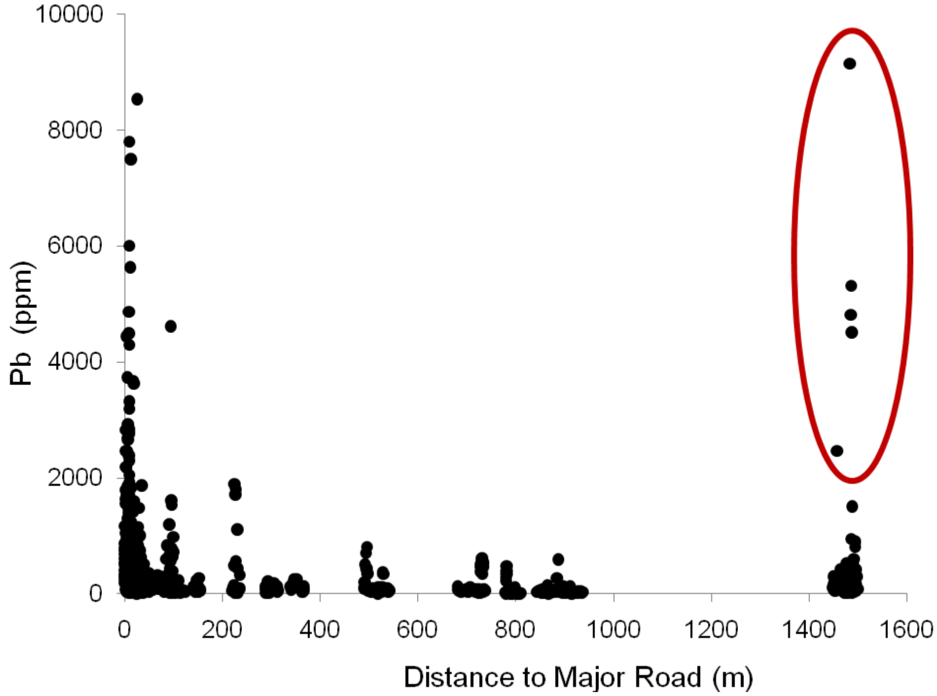


Figure 11b.

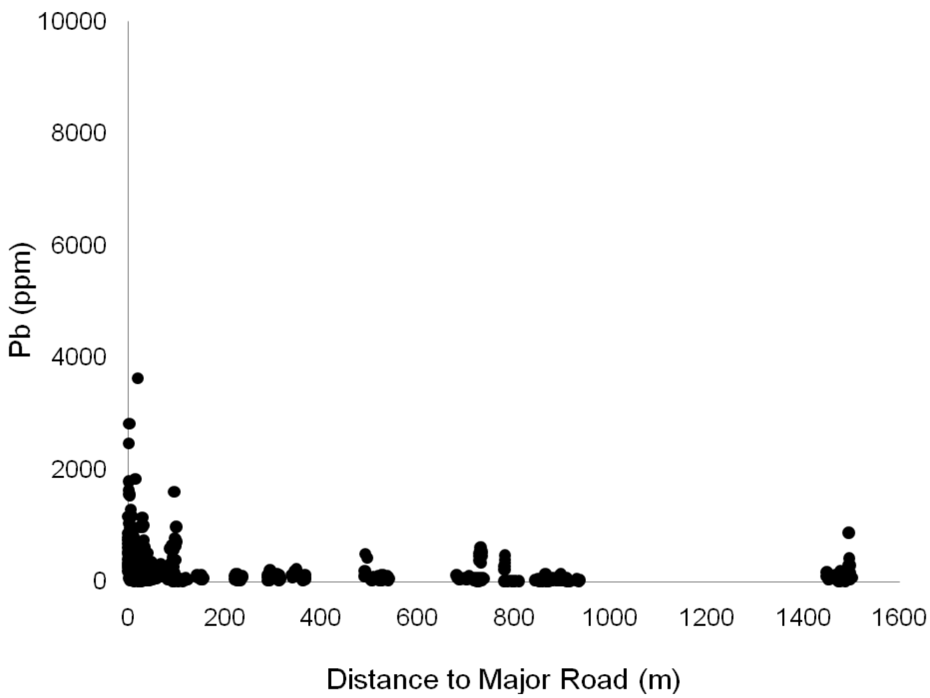
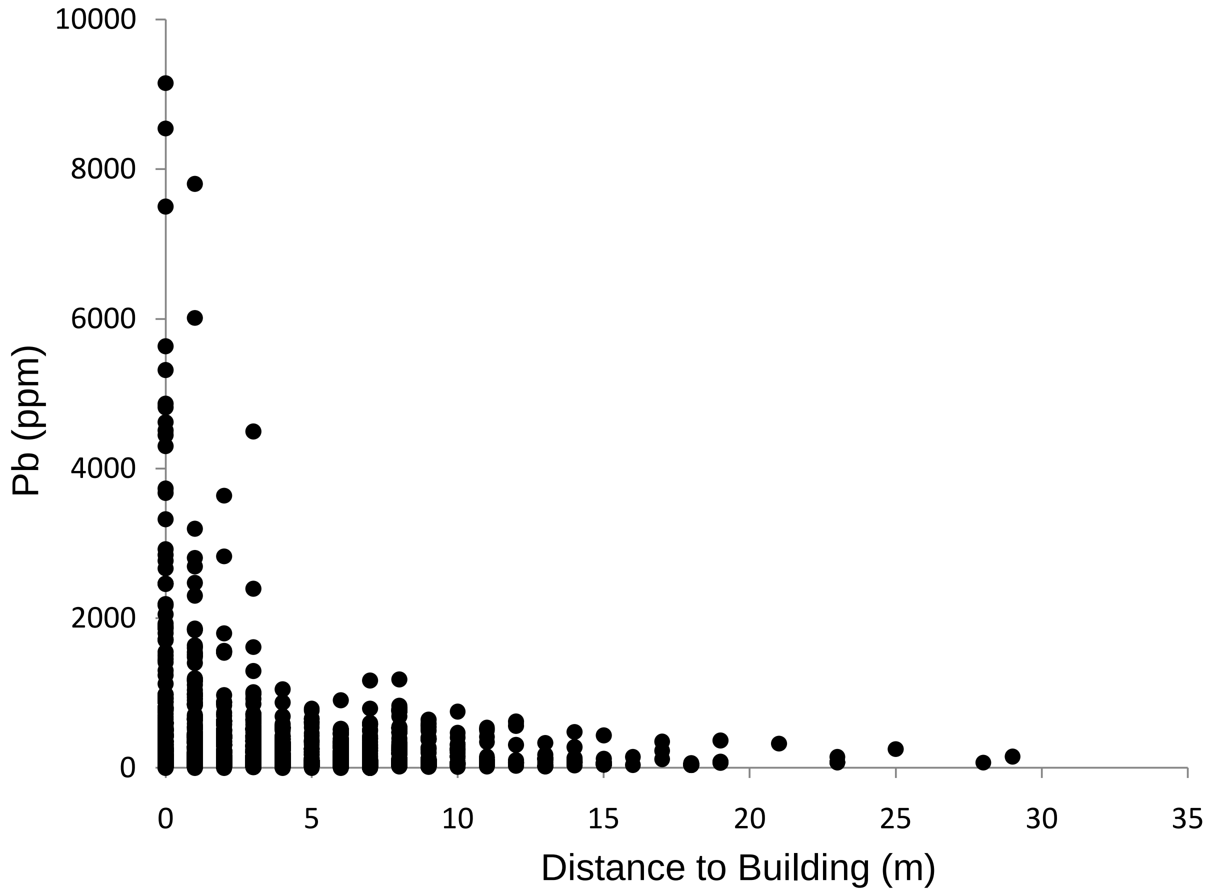


Figure 12.


10000
8000
6000
4000
2000
0
Pb (ppm)
0
5
10
15
20
25
30
35
Distance to Building (m)

CHAPTER 3. AN EMPIRICALLY-BASED GIS MODEL OF RESIDENTIAL SOIL LEAD CONCENTRATIONS IN BALTIMORE, MARYLAND USA.

Abstract

Lead (Pb) contamination of Baltimore's residential soil is widespread, highly variable, and a potential public health concern. With the inception of Maryland's plan to eliminate childhood lead poisoning by 2010, attention has been focused on old lead-based paint sources. Lead-based paint is an important factor in childhood lead exposure; however, it is not the only source of lead in the environment. Soil contaminated with lead from past use of leaded gasoline, deteriorating lead-based paint and industrial sources is also an important source of lead in the environment. Intensive sampling of 61 residential properties in Baltimore City revealed that 53% had soil Pb levels that exceeded the United States Environmental Protection Agency (USEPA) reportable limit of 400 ppm. These data were used as the input to several models to predict the spatial distribution of lead in urban residential soils throughout Baltimore City. Here I examine three different modeling approaches within a geographic information systems (GIS) environment: a traditional general linear model (GLM), and two machine learning techniques: Classification and Regression Trees (CART) and Random Forests (RF). The GLM revealed that housing age, distance to road, distance to building, and the interactions between distance to road and housing age, and distance to building and distance to road explained 38% of the variation in the data. The CART model confirmed the importance of these variables, with housing age, distance to building, and distance to major road networks

determining the terminal nodes of the CART model. Using the same three predictor variables, the RF model was able to explain 42% of the variation in the data. An independent dataset was used to evaluate the accuracy of the models. The overall accuracy, which is a measure of agreement between the model and an independent dataset, was 89.66% for the GLM model, 82.76% for the CART model, and 72.41% for the RF model. The producer's and user's accuracy, which indicate errors of omission and commission, was greater for the low classification of lead concentrations (Pb < 400 ppm) compared to the high classification (Pb > 400 ppm) for all three models. This research highlights the usefulness of empirical models to predict the spatial distribution of lead in urban residential soils. Empirically-based GIS models have the potential to assist public health officials and city agencies in focusing efforts on contaminated soil removal and remediation.

Introduction

Baltimore City has several defining characteristics that indicate the presence of multiple environmental lead (Pb) sources. It's an **older urban** area with a **centralized transportation network** that at one time had a strong **industrial presence**. Urban areas often have higher soil Pb concentrations compared to their rural counterparts. For example, Wong *et al*. (2006) reported an order of magnitude difference between mean urban soil Pb concentrations and rural soil Pb concentrations. Zhai *et al.* (2003) found soil Pb levels in the capital city of Gaborone, Botswana to be 5.7 times those of rural soils. A similar pattern was discovered in New Orleans by Mielke *et al.* (2004) where the median soil Pb levels for inner-city soils were highly enriched (656 ppm) compared to median soil Pb levels for suburban soils (12 ppm). The median soil Pb levels for New Orleans' suburban soils fell within the range of Pb naturally occurring in soils as reported by Holmgren *et al.* (1993).

Baltimore city, like many other U.S. cities, experienced increased growth in the early twentieth century. The increase in growth translated to a large portion of the Baltimore housing stock being built during a time when lead-based paint and leaded gasoline were widely used. Baltimore has approximately 368,000 homes that were built before 1950 and 897,000 homes that were built between 1950 and 1978 (Maryland Department of the Environment 2007). Any home built before the 1978 ban on lead-based paint is more likely to contain lead-based paint. Seventy five percent of homes built between 1950 and 1978 likely have lead-based paint, while 95% of homes built before 1950 likely contain lead-based paint (Maryland Department of the Environment 2007).

The year a house was built not only indicates the potential for the presence of lead-based paint, but it also indicates how long a house may have been accumulating lead from the exhaust of cars burning leaded gasoline (Mielke 1999). Previous studies have shown a correlation between elevated soil Pb levels and the presence of road networks (Facchinelli *et al.* 2001, Imperato *et al.* 2003, Ordonez *et al.* 2003, Li *et al.* 2004, Wang *et al.* 2006, Yesilonis *et al*. 2008). In addition, research has also shown a decrease in soil Pb levels with increasing distance away from road networks (Motto *et al.* 1970, Wang *et al.* 2006). This highlights the fact that within Baltimore City, the centralized transportation network could be an important historic source of lead to the environment and will likely influence the spatial distribution of lead in soil. In fact, other studies have documented differences in soil Pb concentrations at different sampling locations *within* cities. Chirenje *et al.* (2004) found higher levels of lead in soil in residential and commercial areas compared to urban parks. Mielke *et al.* (2004) reported a difference among sampling locations and median Pb levels in New Orleans soil with samples near the foundation of buildings exhibiting the highest lead levels, followed by busy streets, residential streets, and open areas. Finally, industrial land use has been correlated with an increase in soil Pb levels in multiple cities (Thuy *et al.* 2000, Facchinelli *et al.* 2001, Imperato *et al.* 2003, Wang *et al.* 2006). These studies support the idea that urban soil lead concentrations are greatest near buildings, specifically those built before 1978, major roads and industrial areas.

Baltimore's older housing stock, extensive transportation network and the presence of heavy industry are all likely sources of lead to the environment and have the potential to contribute to elevated lead levels in Baltimore's soil. This poses a potential public health concern, particularly in older residential neighborhoods. Elevated soil Pb concentrations are thought to contribute to elevated blood lead levels (BLLs) in children (Duggan and Inskip 1985, Aschengrau *et al.* 1994, Mielke *et al.* 1997). Specifically, Mielke *et al.* (1997) showed that soil lead concentrations were more closely related to children's BLLs than housing age. Elevated BLLs in children can result in devastating and irreversible health effects. Perhaps even more alarming is the fact that adverse health effects have been associated with BLLs below the current threshold (Lanphear *et al.* 2000, Koller *et al.* 2004), which is defined by the Centers for Disease Control and Prevention and the World Health Organization as 10 µg/dL.

Given that elevated soil Pb levels pose a potential public health concern, it is necessary to describe the spatial distribution of soil Pb levels. Although many studies have reported on soil Pb levels, fewer have described the spatial distribution of lead in soil (Markus and McBratney 2001). Studies that have examined the spatial distribution of lead in soil have often used geostatistical techniques, such as kriging, to estimate lead concentrations (Shinn *et al.* 2000, Cattle *et al.* 2002). Kriging, which creates predictions based on interpolating neighboring values, has served as a useful tool in predicting soil Pb values. Other studies that have used portable x-ray florescence (XRF) and geographic information systems (GIS) to map lead contamination (Carr *et al.* 2008) have employed a regular grid design and intensively sampled one location, which makes interpolation between samples possible. In contrast, our data, which was collected in residential areas of Baltimore City, consists of 61 intensively sampled areas which are very distant from one another resulting in several areas with an abundance of lead data and vast stretches between that contain no data. This data structure is not optimal for kriging techniques especially given that the variability of lead in residential soils has been shown to be very high (Machemer and Hosick 2004). Instead I have created three different empirically-based GIS models, one that uses a traditional general linear model or GLM (Weathers *et al.* 2006), and two machine learning techniques: classification and regression trees or CART (Breiman 1984) and Random Forests or RF (Cutler *et al.* 2007). The development of three different models allowed me to evaluate which model was the most useful in predicting the spatial pattern of lead in soil.

Methods

Detailed soil Pb data were collected from 61 residential parcels in Baltimore, Maryland from November 2007 through September 2008. Measurements were collected using a handheld portable XRF multi-element analyzer (n = 1121). See chapter 2 for more details regarding the sampling scheme. Sampling locations were added to a GIS and soil Pb data were joined to the sampling points using the “join field” function in ArcGIS 9.3. The primary and secondary roads from the TIGER classification (United States Census Bureau, Topologically Integrated Geographic Encoding and Referencing (TIGER) database, census.gov) were digitized to create a layer that represented the major road networks. Building footprints were obtained from the City of Baltimore and used with permission under a license agreement through the City. The building footprints dataset represents photogrammetrically captured building footprints that exceed 100 square feet. The dataset was created by the Sanborn Map Company and reflects ground conditions during 2003. Finally, distance of each soil lead sample to the major road networks and building footprints were calculated using the “near” function in ArcGIS 9.3. The data were then exported from ArcGIS for use in additional software packages.

Model Construction - General Linear Model (GLM)

Following a method developed by Weathers *et al.* (2006), I created an empirical model that predicts total soil Pb concentrations in residential areas as a function of landscape features. First, values in the dataset that exceeded the range of the sampling XRF instrument, which included lead values lower than the level of detection and values that exceeded 10,000 ppm, were eliminated from the dataset (n = 9). In testing for normality, the original data failed the Shapiro-Wilk W test for normality ($W=0.403684$, $prob<W = 0.0000$). After log transforming the

data, the data still failed the Shapiro-Wilk W test for normality (W=0.980644, prob<W = 0.0001); however, the distribution of the data more closely resembled that of a normal distribution (Figure 1). I then constructed a general linear model in JMP (v.7.0.2, SAS Institute 2007) using the log transformed data. Soil Pb concentration served as the dependent variable for the GLM while housing age, distance to major road networks, distance to building, and the interactions served as the independent variables. After removing variables that were not statistically significant, the GLM parameter estimates were recalculated (Table 1). The GLM revealed that housing age, distance to road, distance to building, and the interactions between distance to road and housing age, and distance to building and distance to road explained 38% of the variation in the data

In order to apply the parameter estimates across the city, GIS layers with continuous data for all of the predictor variables are needed. Within a GIS, a raster for distance to road and distance to building was created using the "Euclidean distance" tool in the spatial analyst extension of ArcGIS. A $1m^2$ cell size was used in the construction of both raster layers. In order to create a raster of housing age, a parcel boundary layer obtained from the City of Baltimore was joined to the Maryland Property View Assessors and Taxation point file which contains housing age information using the "spatial join" function in ArcGIS 9.3. Only parcels described as residential were retained in the dataset. Building footprints were extracted from the residential parcels using the "erase" function in ArcGIS 9.3 to eliminate areas were no soil is present. Finally, the parcel polygon layer was converted to a raster using the “polygon to raster” conversion tool in ArcGIS 9.3 with housing age as the value field and a 1m^2 cell size.

The following equation described in Weathers *et al.* (2006) was used to create a GIS based model of soil lead concentration as a function of landscape features.

Model = [ß$_0$ (at x,y)] + [ß$_1$ x (Z$_1$ at x,y)]

+ [ß$_2$ x (Z$_2$ at x,y)]

+ [ß$_3$ x (Z$_1$ at x,y) x (Z$_2$ at x,y)]

+...+ [ß$_n$ x (Z$_n$ at x,y)]

Where Z$_1$, Z$_2$, Z$_n$ = independent variables (i.e. Distance to major road networks, distance to buildings, and housing age),

ß$_0$ = intercept,

ß$_1$ = landscape variable Z$_1$ coefficient,

ß$_2$ = landscape variable Z$_2$ coefficient, and

ß$_3$ = interaction

Using the parameter estimates from the GLM (Table 1), the following expression was entered into the raster calculator, which is part of the spatial analyst extension in ArcGIS:

Pb = (26.506363 - (0.000272 * [EucDistRoad]) - (0.034337 * [EucDistBuild]) - (0.012447 * [YearBuilt]) - (0.00001637 * (([EucDistRoad] - 284.694) * ([YearBuilt] - 1936.61))) - (0.000019 * (([EucDistBuild] - 4.19982) * ([EucDistRoad] - 284.694))))

The raster calculator performs mathematical calculations on raster datasets which represent the independent variables. The resulting raster based model reflects spatially explicit log transformed lead values. I used the following equation in the raster calculator to create a model that reflected actual values: Exp10([LogPbModel]). Finally, the resulting GIS model was reclassified using the “reclassify” tool in ArcGIS so that cells would reflect a binary model of values that were greater or less than the USEPA’s reportable limit of 400 ppm (Figure 2). The conversion to a binary model was made for two reasons, 1) to reflect metrics relevant to management and 2) to compare with the CART model, which does not generate continuous lead values.

Model Construction – Classification and Regression Trees (CART)

Classification and Regression Trees (CART) were used to construct an alternative model to the one based on a GLM. CART is a non-parametric, machine learning, statistical method used in ecology for exploration, description, and prediction of grouped data (De’Ath and Fabricius 2000, Golubiewski 2006). CART models are an approach to classification that does not assume data normality (Sutton 2005). Classification and regression trees produce a hierarchy of decision rules displayed in the form of a binary tree (Sutton 2005). They are often used in place of multiple regressions. Compared to linear models, tree-based models are easier to interpret, can handle missing data, capture non-additive behavior (De'Ath and Fabricius 2000) and are resistant to outliers

(Sutton 2005). Gahegan (2003) cautions that while machine learning techniques offer the capability to improve predictive power with complex data there are only as good as "(1) the data are representative, and (2) the methods are capable of learning the trend contained therein" (p. 86). Compared to more traditional statistical techniques, machine learning techniques differ in the amount of *a priori* knowledge, in other words they rely on patterns in the data rather than any underlying theory (Gahegan 2003). One potential advantage to this approach is the ability to discover novel and unexpected patterns (Gahegan 2000). CART modeling was chosen for all of these reasons, specifically because it allows for use of non-normal, complex data.

Lead values for each sampling location were categorized as either low (0-400 ppm) or high (> 400 ppm) and used as the dependent variable in a classification tree. The classification tree was constructed in S-Plus (version 6.1, Insightful Corporation 2005). Several variables were evaluated as independent or predictor variables including: 1) sample location, 2) sample group, 3) housing age, 4) housing material, 5) distance to major road networks, and 6) distance to building. The algorithm evaluates which variables to include in the actual tree construction. Only housing age, distance to major road networks, and distance to building were included. The terminology used to interpret classification trees parallels the terminology used to describe actual trees. For example, the top node of the tree is referred to as the "root". Terminal nodes are referred to as "leaves" and a split is a rule that results in the formation of new "branches."

One issue in building (i.e. growing) classifications trees is deciding which trees are significant. A common problem with CART analysis is growing the tree too long, thus making the tree overly complex. This results in the model fitting the original dataset very well, but limits its applicability to more general situations, in other words, overfitting the data. In order to evaluate which trees were significant I examined the misclassification error rate in relation to the complexity of the tree or number of terminal nodes (Figure 3). With increasing tree complexity there was a decrease in the misclassification error rate, which was expected. However, after 6 nodes an increase in tree complexity did not result in a lower misclassification error rate. I therefore examined classification trees with 5 and 6 nodes and determined that the 5 node tree was most reasonable. The rule associated with the 6^{th} node appeared to be an artifact of the sampling design. Specifically, the 6^{th} node predicted lead contamination was restricted to within 14.5 meters of the major road networks. This prediction was most likely a result of the fact that more than half of the properties sampled were within 30 meters of a major roadway by design. The rule associated with the 6^{th} node does not account for contamination on parcels that are farther than 30 meters from the road, a pattern that was observed in 28% of the properties sampled.

The predications from the 5 node classification tree (Figure 4) were mapped in ArcGIS 9.3 using Model Builder. The rules determined by the CART model were translated into a series of conditional statements that were constructed in Model

Builder using the “single output map algebra” tool (Figure 5). The resultant map (Figure 6) shows the areas were soil Pb levels are predicted to exceed the USEPA reportable limit of 400 ppm.

Model Construction – Random Forest (RF)

In contrast to CART, which produces a single tree, Random Forests (RF) combines many trees (Breiman 2001). Each tree is constructed using a subset of the training data (i.e. field data). The remaining data are said to be “out-of-bag” (Cutler *et al.* 2007). The multiple trees generated in RF, each constructed from a subset of data, are then used to predict the withheld or "out-of-bag" observations as a means of accuracy assessment (Cutler *et al.* 2007). RF models are ideally suited for ecological data. RF models can handle complex interactions, missing data and exhibit high classification accuracy (Cutler *et al.* 2007).

Models developed using RF have been shown to be more robust compared to other modeling techniques. For example, Peters *et al.* (2007) compared two different models, a multiple logistic regression in a GLM framework and RF, to predict vegetation type occurrence based on various habitat descriptors. In comparing the two techniques, they concluded that the RF model could lead to better predictive ecohydrological models (Peters *et al.* 2007). In addition, Cutler *et al.* (2007) used RF models to classify presence-absence data and found increased accuracy with RF models compared to other classifiers.

In R, I used "ModelMap" (Freeman and Frecino 2009), which uses a RF package (Liaw and Wiener 2009), to create a RF model based on the training data. Then the model is used to predict on new data with the same predictor variables as used in the original model. The field data collected with the XRF served as the training data. A composite raster containing the three predictor variables (housing age, distance to road, distance to building) was created using the "Composite Bands" function in ArcGIS 9.3 and served as the new data. A RF regression model consisting of the default value of 500 trees explained 42 % of the variation in the data. Models developed using RF consistently show that distance to building and housing age are more important predictors compared to distance to major road networks. Moreover, distance to building and housing age are roughly equal in terms of variable importance. The output from the ModelMap package is an ASCII grid file that was converted to a raster using the "ASCII to raster" function in ArcGIS 9.3. The raster file was reclassified resulting in a map (Figure 7) that shows areas were soil Pb levels are predicted to exceed the USEPA reportable limit of 400 ppm.

Model Validation

In order to test the accuracy of the models, soil Pb data that was collected as part of the Urban Forest Effects Model (United States Department of Agriculture Forest Service, Urban Forest Effects Model – UFORE, fs.fed.us) was overlaid with the resulting models to evaluate the amount of agreement between the model predictions and an independently collected dataset. Only those UFORE soil samples that were collected in residential areas that intersected the model were used in the accuracy assessment (n = 29). For the GLM model 26 of the 29 UFORE soil samples were correctly classified as lead concentrations (< or >) 400 ppm based on the model predictions, resulting in an overall accuracy of 89.66% (Table 2). The producer's accuracy, representing errors of omission, for samples classified as low (0-400 ppm) is 96% and for high (over 400 ppm) is 50% (Table 2). The user's accuracy, representing errors of commission, for the low category (0-400 ppm) is 92.31% and for the high category (over 400 ppm) is 66.66% (Table 2).

For the CART model 24 of the 29 UFORE soil samples were correctly classified based on the model predictions resulting in an overall accuracy of 82.76% (Table 3). The producer's accuracy for samples classified as low (0-400 ppm) is 92% and for high (over 400 ppm) is 25% (Table 3). The user's accuracy for the low category (0-400 ppm) is 88.46% and for the high category (over 400 ppm) is 33.33% (Table 3). The RF model classified 21 of the 29 UFORE soil samples correctly, resulting in an overall accuracy of 72.41% (Table 4). The producer's accuracy for samples classified as low (0-400 ppm) is 76% and for high (over 400 ppm) is 50% (Table 4). The user's accuracy for the low category (0-400 ppm) is 90.48% and for the high category (over 400 ppm) is 25% (Table 4).

The models more accurately predicted samples in the low category (<400 ppm) than in the high lead concentration (>400 ppm) category. This could be an artifact of the field data that was used as the input to the model since a greater number of measurements fell into the low category compared to the high category. Lower accuracy in the high lead concentration category could also be due to the limited number of samples in the independent UFORE dataset used to validate the models that were part of the high category (n=4). In addition, the soil samples collected as part of the UFORE project were sampled to a depth of 10 cm. The samples collected via XRF, which were used as the input to the model, are surface measurements that roughly correspond to the upper 2 mm of surface soil. It is assumed that soil Pb is concentrated in surface soils and decreases with depth (Wang *et al*. 2006, Griffith *et al*. 2008); therefore the UFORE samples may represent more dilute Pb levels compared to XRF readings.

Results and Discussion

Visual examination of the predictions of the three models reveals a similar pattern with contamination concentrated in the city center and along the major road networks (Figures 2, 6, and 7). This pattern is consistent with other studies in Baltimore City, which have found elevated soil Pb in older parts of the city and areas with higher road densities (Yesilonis *et al.* 2008). These results are also consistent with earlier work that has found a correlation between elevated soil lead and roads (Facchinelli *et al.* 2001, Imperato *et al.* 2003, Ordonez *et al.* 2003, Li *et al.* 2004, Wang *et al.* 2006) and work that has found a spatial and temporal component to the distribution of lead in soils (Yesilonis *et al.* 2008).

The area predicted to be above the USEPA reportable limit of 400 ppm differed among models. The GLM model predicted an area of 1,215,611 square meters would exhibit soil Pb levels above 400 ppm, while the CART model predicted an area of 4,077,071 square meters would be above 400 ppm. The RF model predicted the greatest amount of area, 10,575,386 square meters, to be above 400 ppm. Close examination of a detailed section of the mapped lead levels highlights the difference in areas predicted to be above the USEPA limit among the three models (Figure 8). One potential reason that the CART model predicts more than three times the area predicted by the GLM to be contaminated could be the fact that it applies a blanket statement that elevated soil Pb levels will occur within 1.5 meters of any building built before 1934 regardless of housing material. This prediction, not surprisingly, is consistent with patterns that I observed in the XRF data in which elevated levels of lead were found next to both brick and wood frame homes. This prediction is also consistent with other studies which have documented elevated soil Pb in the absence of frame homes (Shinn *et al.* 2000). The RF Model predicts even greater amounts of contamination compared to the CART model. The difference between the two models can be explained by two patterns observed in the RF model predictions: 1) contamination is not limited to homes built before 1934 as in the CART model and 2) in contrast to CART model predictions that state contamination will occur within 1.5 meters of a building, the RF model predicts contamination to be more widespread, in some cases occurring throughout the entire parcel. The patterns observed in the RF model predictions are also consistent with the XRF field data which exhibited soil contamination in homes built after 1934 and occasionally revealed contamination throughout a parcel.

It is important to keep in mind that even though the model predictions differ in the extent of contamination, an independent dataset that was used to validate the models resulted in similar numbers for overall accuracy. Overall accuracy is an important indicator of how well a model describes actual conditions; however, the data used in the validation have their limitations as well. Therefore, it is also important to keep in mind that although the Random Forest displayed the lowest overall accuracy (72.41%), the model explained the greatest amount of variation (42%) in the data. This is consistent with the idea that classification techniques are better suited at describing complex, non-normal data compared to general linear models.

Many public health programs that focus on lead remediation remain mitigative as opposed to preventive (Miranda *et al.* 2002). There is a great need to shift to more preventative measures (Griffith *et al.* 2008). With the realization that elevated soil Pb is a potential public health issue, there is also a great need to accurately predict hotspots of elevated soil Pb. GIS has been proposed as a technique that can help to explain the spatial structure of environmental data

(Getis *et al.* 2004). With increasing GIS capacity and an abundance of spatial statistic techniques, researchers now have the ability to create empirically-based models that can predict hotspots of lead contamination in urban residential soils. These new techniques hold great potential and can increase our understanding of environmental health issues such as childhood Pb poisoning (Miranda and Dolinoy 2005). In addition, using GIS-based models could alleviate the need for costly and time-intensive clinical trials (Griffith *et al.* 2008).

When examining spatial patterns of ecological data, the use of several different techniques is encouraged (Perry *et al*. 2002). Here I compared the use of empirically-based GIS models using a traditional GLM technique and two machine learning techniques, CART and RF. Although the GLM model out-performed the CART and RF model in terms of overall, producer and user's accuracy, all three models exhibited overall accuracies above 70%. The resulting models predict the spatial distribution of lead in soil for residential areas of Baltimore City. The *methods* used in this work could be applied to other urban areas; however, the resulting *models* may be regionally constrained. Urban centers in the southeastern United States are less affected by lead contamination compared to their rural counterparts because the growth phase of those cities occurred after the ban on products containing lead (Miranda *et al.* 2002). The models described above would most likely be applicable to cities in the Northeastern United States with similar housing stock and transportation networks.

This research highlights the usefulness of empirical models to predict the spatial distribution of lead in urban residential soils and improves our knowledge of the effect that various landscape features have on lead concentrations in soil. Empirically-based GIS models have the potential to improve upon current soil remediation efforts by pinpointing areas of high contamination, thus reducing the cost of intensive soil sampling. Accurate characterization of soil lead concentrations can also assist the public health community in focusing on a widely dispersed source of lead in the environment.

The multiple models created provide the public health community with several screening tool options. The more restricted GLM and CART model predictions provide a reasonable option if resources for further study are limited. However, if resources are available for comprehensive follow-up studies, the RF model predictions are less likely to exclude possible lead contamination by erring on the side of commission rather than omission. Therefore, public health policy based on the RF model predictions may be more protective of human health.

Literature Cited

Aschengrau, A., A. Beiser, D. Bellinger, D. Copenhafer, and M. Weitzman. 1994. The impact of soil lead abatement on urban children's blood lead levels: phase II results from the Boston lead-in-soil demonstration project. Environmental Research 67: 125-148.

Breiman, L., J. Friedman, C.J. Stone, and R.A. Olshen. 1984. Classification and regression trees. Chapman and Hall, London.

Breiman, L. 2001. Random forests. Machine learning 45(1): 5-32.

Carr, R., C. Zhang, N. Moles, and M. Harder. 2008. Identification and mapping of heavy metal pollution in soils of a sports ground in Galway City, Ireland, using a portable XRF analyser and GIS. Environmental geochemistry and health 30(1): 45-52.

Cattle, J.A., A.B. McBratney, and B. Minasny. 2002. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. Journal of Environmental Quality 31 1576-1588.

Chirenje, T., L.Q. Ma, M. Reeves, and M. Szulczewski. 2004. Lead distribution in near-surface soils of Florida cities: Gainesville and Miami. Geoderma 119: 113-120.

Cutler, D.R., T.C. Edwards Jr, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler. 2007. Random forests for classification in ecology. Ecology 88(11): 2783-2792.

De'ath, G. and K.E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81(11): 3178-3192.

Duggan, M.J. and M.J. Inskip. 1985. Childhood exposure to lead in surface dust and soil: a community health problem. Public Health Review 13: 1-54.

Facchinelli, A., E. Sacchi, and L. Mallen. 2001. Multivariate statistical and GIS-based approach to identify heavy metal sources in soil. Environmental Pollution 114 313-324.

Freeman, E. and T. Frescino. 2009. ModelMap. Methods in R: 18.

Gahegan, M. 2000. The case for inductive and visual techniques in the analysis of spatial data. Journal of Geographical Systems 2(1): 77-83.

Gahegan, M. 2003. Is inductive machine learning just another wild goose or might it lay the golden egg? International Journal of Geographical Information Science 17(1): 69-92.

Getis, A., L. Anselin, A. Lea, M. Ferguson, and H. Miller. 2004. Spatial analysis and modeling in a GIS environment. A research agenda for geographic information science. Boca Raton, FL, CRC Press: 157-196.

Golubiewski, N. 2006. Urbanization increases grassland carbon pools: Effects of landscaping in Colorado's front range. Ecological Applications 16(2): 555-571.

Griffith, D.A., D.L. Johnson, and A. Hunt. 2008. The geographic distribution of metals in urban soils: the case of Syracuse, NY. GeoJournal: 1-17.

Holmgren, G.G.S., M.W. Meyer, R.L. Chaney, and R.B. Daniels. 1993. Cadmium, lead, zinc, copper, and nickel in agricultural soils of the United States of America. Journal of Environmental Quality 22(2): 335–348.

Imperato, M., P. Adamo, D. Naimo, M. Arienzo, D. Stanzione, and P, Violante. 2003. Spatial distribution of heavy metals in urban soils of Naples city (Italy). Environmental Pollution 124(2): 247-256.

JMP 7.0.2. 2007. SAS Institute Inc.

Koller, K., T. Brown, A. Spurgeon, and L. Levy. 2004. Recent developments in low-level lead exposure and intellectual impairment in children. Environmental Health Perspectives 112(9): 987-994.

Lanphear , B.P., K. Dietrich, P. Auinger, and C. Cox. 2000. Cognitive deficits associated with blood lead concentrations <10 mg/dL in US children and adolescents. Public Health Report 115: 521–9.

Li, X., S. Lee, S. Wong, W. Shi, and I. Thornton. 2004. The study of metal contamination in urban soils of Hong Kong using a GIS-based approach. Environmental Pollution 129(1): 113-124.

Liaw, A. and M. Wiener. 2009. Classification and regression by randomForest. Resampling Methods in R: The boot Package: 18.

Machemer, S. and T. Hosick. 2004. Determination of soil lead variability in residential soil for remediation decision making. Water, Air, & Soil Pollution 151(1): 305-322.

Markus, J. and A.B. McBratney. 2001. A review of the contamination of soil with lead II. Spatial distribution and risk assessment of soil lead. Environment international 27 399-411.

Maryland Department of the Environment. 2007. Maryland Plan to Eliminate Childhood Lead Poisoning by 2010. Baltimore: 110.

Mielke, H.W., G. Wang, C.R. Gonzales, E.T. Powell, B. Le and V.N. Quach. 2004. PAHs and metals in the soils of inner-city and suburban New Orleans, Louisiana, USA. Environmental Toxicology and Pharmacology 18(3): 243-247.

Mielke, H.W. 1999. Lead in the inner cities. American Scientist 87(1): 62-73.

Mielke, H.W., D. Dugas, P.W. Mielke Jr., K.S. Smith, and C.R. Gonzales. 1997. Associations between soil lead and childhood blood lead in urban New Orleans and rural Lafourche Parish, Louisiana. Environmental Health Perspectives 105(9): 950-954.

Miranda, M. and D. Dolinoy. 2005. Using GIS-based approaches to support research on neurotoxicants and other children's environmental health threats. Neurotoxicology 26(2): 223-228.

Miranda, L.M., D.C. Dolinoy and M.A. Overstreet. 2002. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environmental Health Perspectives 110 (9): 947-953.

Motto, H.L., R.H. Daines, D.M. Chilko, and C.K. Motto. 1970. Lead in soils and plants: its relationship to traffic volume and proximity to highways. Environmental Science and Technology 4(3): 231-237.

Ordonez, A., J. Loredo, E. DeMiguel, and S. Charlesworth. 2003. Distribution of heavy metals in the street dusts and soils of an industrial city in Northern Spain. Archives of Environmental Contamination and Toxicology 44(2): 160-170.

Perry, J.N., A.M. Liebhold, M.S. Rosenberg, J. Dungan, M. Miriti, A. Jakomulska, and S. Citron-Pousty. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography 25 (5): 578-600.

Peters, J., B. De Baets, N.E.C. Verhoest, R. Samson, S. Degroeve, P. De Becker, and W. Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling 207(2-4): 304-318.

S-Plus 6.1 for Windows. 2002. Insightful Corporation.

Shinn, N.J., J. Bing-Canar, M. Cailas, N. Peneff, and H.J. Binns. 2000.

Determination of spatial continuity of soil lead levels in an urban residential neighborhood. Environmental Research 82(1): 46-52.

Sutton, C. D. 2005. Classification and regression trees, bagging, and boosting. Handbook of Statistics 24: 303-329.

Thuy, H.T.T., H.J. Tobschall, and P.V. An. 2000. Distribution of heavy metals in urban soils - a case study of Danang-Hoian area (Vietnam). Environmental Geology 39 (6): 603-610.

United States Census Bureau. Topologically Integrated Geographic Encoding and Referencing (TIGER) database. Census.gov. 28 October 2009. <http://www.census.gov/geo/www/tiger/>.

United States Department of Agriculture Forest Service. Urban Forest Effects Model – UFORE. Fs.fed.us. 28 October 2009 <http://www.nrs.fs.fed.us/tools/ufore/>.

Wang, J., H. Ren, J. Liu, J. Yu, and X. Zhang. 2006. Distribution of lead in urban soil and its potential risk in Shenyang City, China. Chinese Geographical Science 16(2): 127-132.

Weathers, K.C., G.M. Lovett, G.E. Likens, and R.Lathrop. 2000. The effect of landscape features on deposition to Hunter Mountain, Catskill Mountains, New York. Ecological Applications 10(2): 528-540.

Wong, C.S.C., X. Li, and I. Thornton. 2006. Urban environmental geochemistry of trace metals. Environmental Pollution 142(1): 1-16.

Yesilonis, I., R. Pouyat, and N.K. Neerchal. 2008. Spatial distribution of metals in soils in Baltimore, Maryland: role of native parent material, proximity to major roads, housing age and screening guidelines. Environmental Pollution 156(3): 723-731

Zhai, M., H.A.B. Kampunzu, M.P. Modisi, and O. Totolo. 2003. Distribution of heavy metals in Gaborone urban soils (Botswana) and its relationship to soil pollution and bedrock composition. Environmental Geology 45(2): 171-180.

Acknowledgements

I am especially grateful to the homeowners for access to their property and to the University of California, Davis for use of the XRF. My understanding of GIS and spatial statistics greatly benefited from conversations with John Bognar, Dr. Adele Cutler, Amanda Elliot, Dr. Elizabeth Freeman, Scott Haag, David Lewis, Dr. Zewei Maio, Samuel Simkin, Jim Trimble, and Dr. Weiqi Zhou. This work is a contribution to the Long Term Ecological Research Program (a program of the National Science Foundation) and the Cary Institute of Ecosystem Studies and was supported by NSF grants DEB 042376 and 0808418.

Figures Legends

Figure 1. Distribution of soil Pb data before and after log transformation. Both distributions failed the Shapiro-Wilk W test for goodness of fit; however, the log transformed data more closely resembles a normal distribution.

Figure 2. A map illustrating the predictions of the GLM model. Areas in green are predicted to fall below the USEPA reportable limit of 400 ppm while areas in red are predicted to be higher than 400 ppm. No data are available for areas where only the aerial photo is shown.

Figure 3. A graph of the CART misclassification error rates by number of terminal nodes on the tree. As the tree gains complexity the misclassification error rate decreases; however after 6 nodes an increase in complexity does not result in a lower misclassification error rate.

Figure 4. Classification tree from S-Plus showing the 5 terminal nodes and corresponding thresholds. The length of the branch is proportional to the variance explained with longer branches explaining more variance.

Figure 5. A series of conditional statements constructed in ArcGIS Model Builder. These rules are derived from the predictions of the 5 node classification tree.

Figure 6. A map illustrating the predictions of the CART model. Areas in green are predicted to fall below the USEPA reportable limit of 400 ppm while areas in red are predicted to be higher than 400 ppm. No data are available for areas where only the aerial photo is shown.

Figure 7. A map illustrating the predictions of the Random Forest model. Areas in green are predicted to fall below the USEPA reportable limit of 400 ppm while areas in red are predicted to be higher than 400 ppm. No data are available for areas where only the aerial photo is shown.

Figure 8. Area of detail showing model predictions from the three different models. The GLM model predictions are shown in panel A, the CART model predictions in panel B, and the Random Forest predictions in panel C.

Table 1. GLM created in JMP using log transformed data with Pb concentration as the dependent variable and distance to major road networks, distance to buildings, housing age, and interactions as the independent variables. As indicated by the R square value, the model describes 38% of the variation in the data.

Response Log10Pb
Summary of Fit

RSquare	0.380777
RSquare Adj	0.378
Root Mean Square Error	0.428757
Mean of Response	2.163011
Observations (or Sum Wgts)	1121

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	26.506363	1.125501	23.55	<.0001
Housing_Ag	-0.012447	0.00058	-21.45	<.0001
Dist_Road	-0.000272	3.023e-5	-8.99	<.0001
Dist_Build	-0.034337	0.003151	-10.90	<.0001
(Dist_Road-284.694)*(Housing_Ag-1936.61)	-1.637e-5	2.123e-6	-7.71	<.0001
(Dist_Build-4.19982)*(Dist_Road-284.694)	-0.000019	7.822e-6	-2.44	0.0150

Table 2. Error matrix of the GLM model and UFORE soil lead data.

Model Predictions Pb (ppm)	**UFORE (Reference Data) Pb (ppm)**	
	0-400 ppm (low)	*Over 400 ppm (high)*
0-400 ppm (low)	**24**	2
Over 400 ppm (high)	1	**2**
Overall Accuracy = 26/29 = **89.66%**		

Producer's Accuracy (measure of omission error)	**User's Accuracy** (measure of commission error)
0-400 ppm (low) = 24/25 = **96%** 4% omission error	0-400 ppm (low) = 24/26 = **92.31%** 7.69 % commission error
over 400 ppm (high) = 2/4 = **50%** 50% omission error	over 400 ppm (high) = 2/3 = **66.66%** 33.33% commission error

Table 3. Error matrix of the CART model and UFORE soil lead data.

Model Predictions Pb (ppm)	**UFORE (Reference Data) Pb (ppm)**	
	0-400 ppm (low)	*Over 400 ppm (high)*
0-400 ppm (low)	**23**	3
Over 400 ppm (high)	2	**1**
Overall Accuracy = 24/29 = **82.76%**		

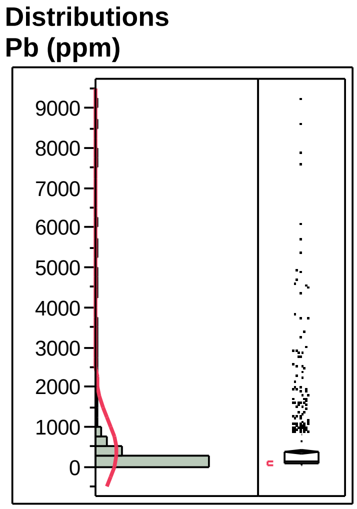
Producer's Accuracy (measure of omission error)	**User's Accuracy** (measure of commission error)
0-400 ppm (low) = 23/25 = **92%** 8% omission error	0-400 ppm (low) = 23/26 = **88.46%** 11.54 % commission error
over 400 ppm (high) = 1/4 = **25%** 75% omission error	over 400 ppm (high) = 1/3 = **33.33%** 66.66% commission error

Table 4. Error matrix of the Random Forest model and UFORE soil lead data.

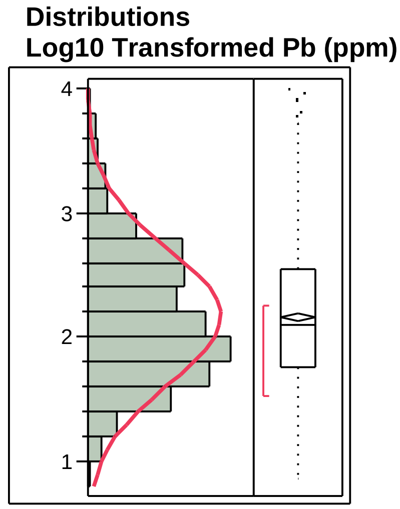
Model Predictions Pb (ppm)	**UFORE (Reference Data) Pb (ppm)**	
	0-400 ppm (low)	*Over 400 ppm (high)*
0-400 ppm (low)	**19**	2
Over 400 ppm (high)	6	**2**
Overall Accuracy = 21/29 = **72.41%**		

Producer's Accuracy (measure of omission error)	**User's Accuracy** (measure of commission error)
0-400 ppm (low) = 19/25 = **76%** 24% omission error	0-400 ppm (low) = 19/21 = **90.48%** 9.52 % commission error
over 400 ppm (high) = 2/4 = **50%** 50% omission error	over 400 ppm (high) = 2/8 = **25%** 75% commission error

Figure 1.



Distributions
Pb (ppm)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0



Distributions
Log10 Transformed Pb (ppm)
4
3
2
1

Figure 2.

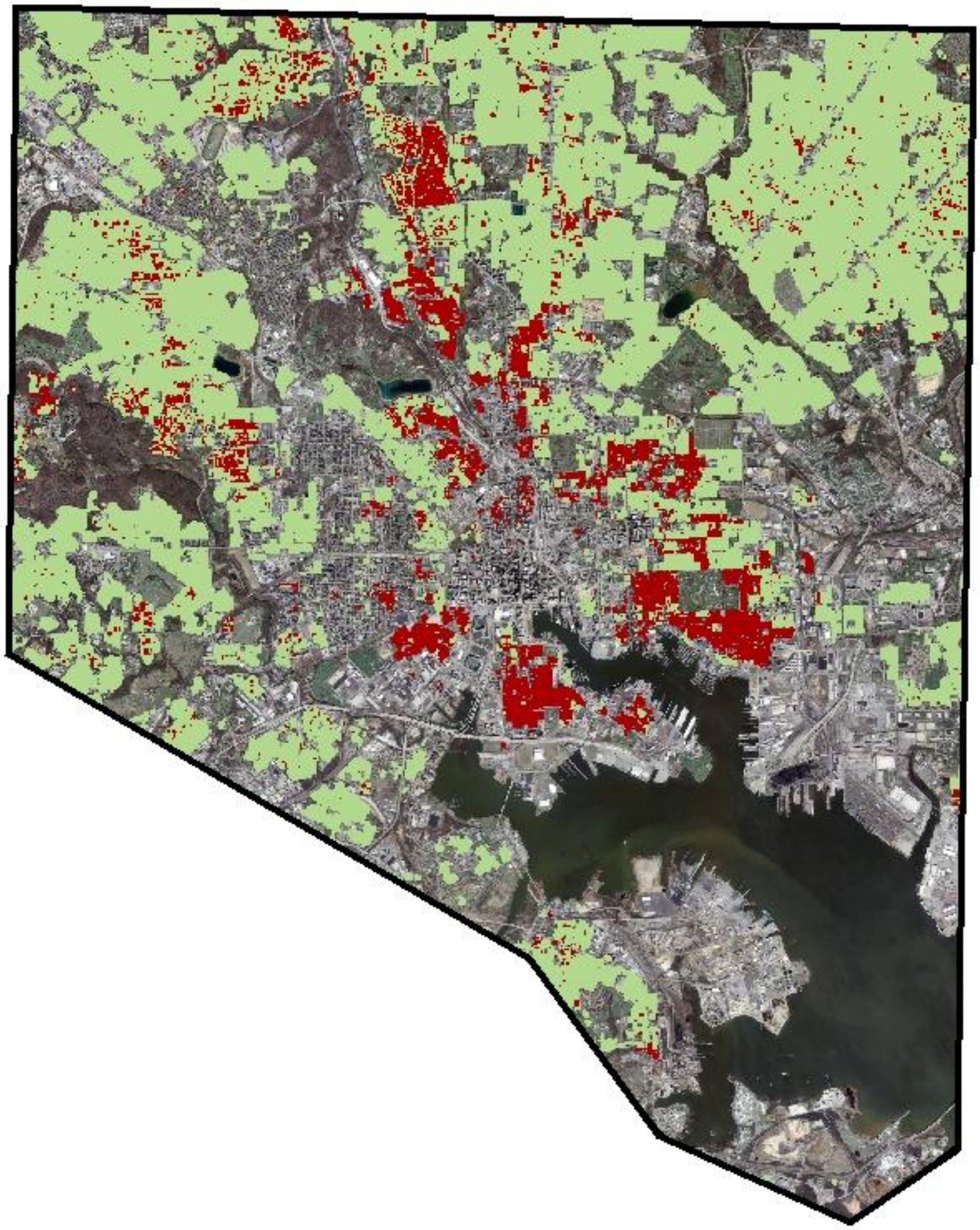


Figure 3.

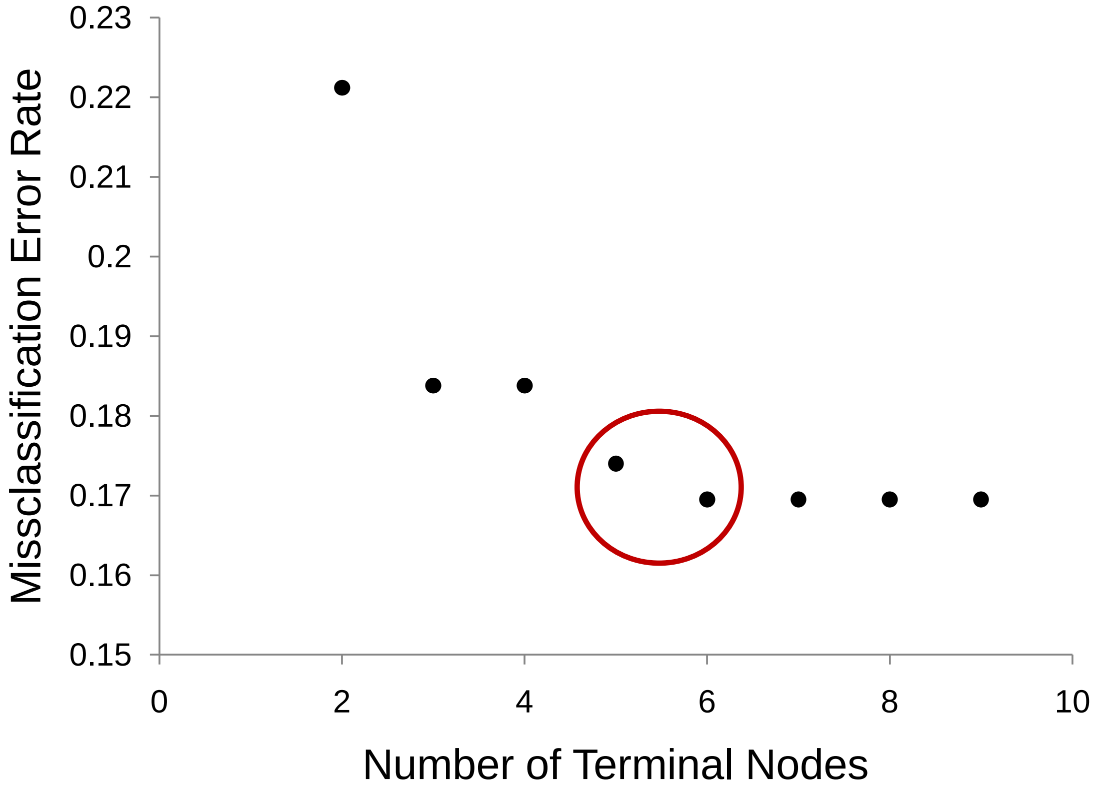

0.23
0.22
0.21
0.2
0.19
0.18
0.17
0.16
0.15
Missclassification Error Rate
0
2
4
6
8
10
Number of Terminal Nodes

Figure 4.

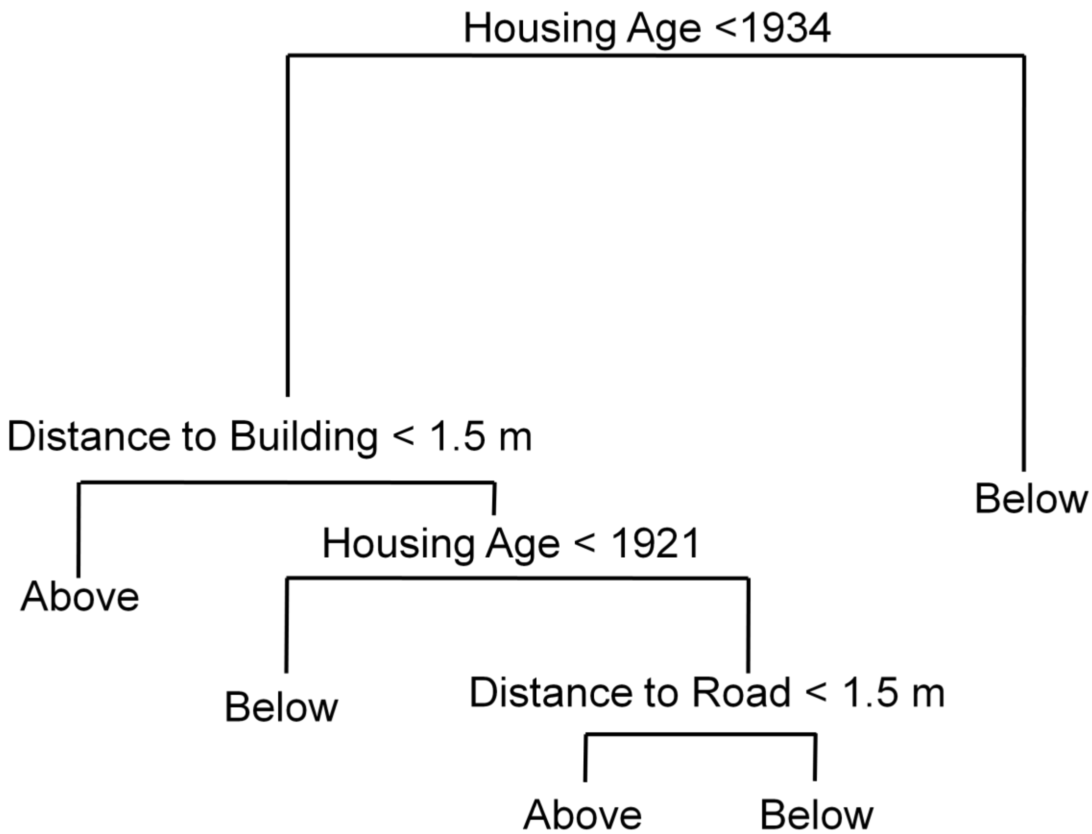

Housing Age <1934
Distance to Building < 1.5 m
Below
Above
Housing Age < 1921
Below
Distance to Road < 1.5 m
Above
Below

Figure 5.

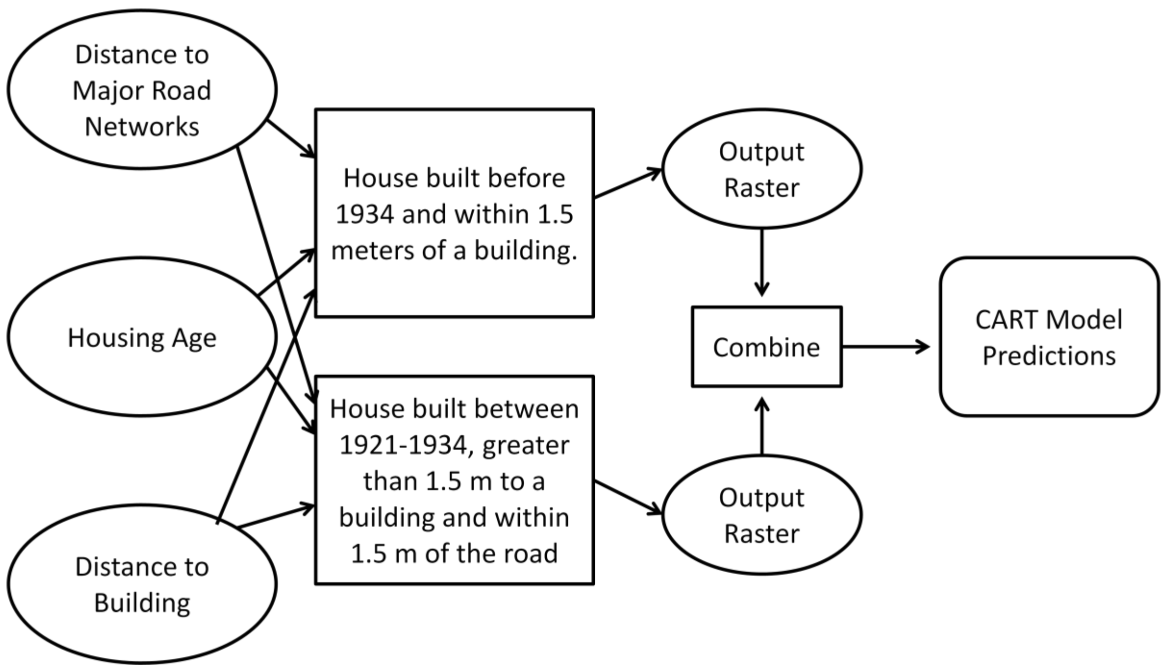

Distance to Major Road Networks
Housing Age
Distance to Building
House built before 1934 and within 1.5 meters of a building.
House built between 1921-1934, greater than 1.5 m to a building and within 1.5 m of the road
Output Raster
Output Raster
Combine
CART Model Predictions

Figure 6.

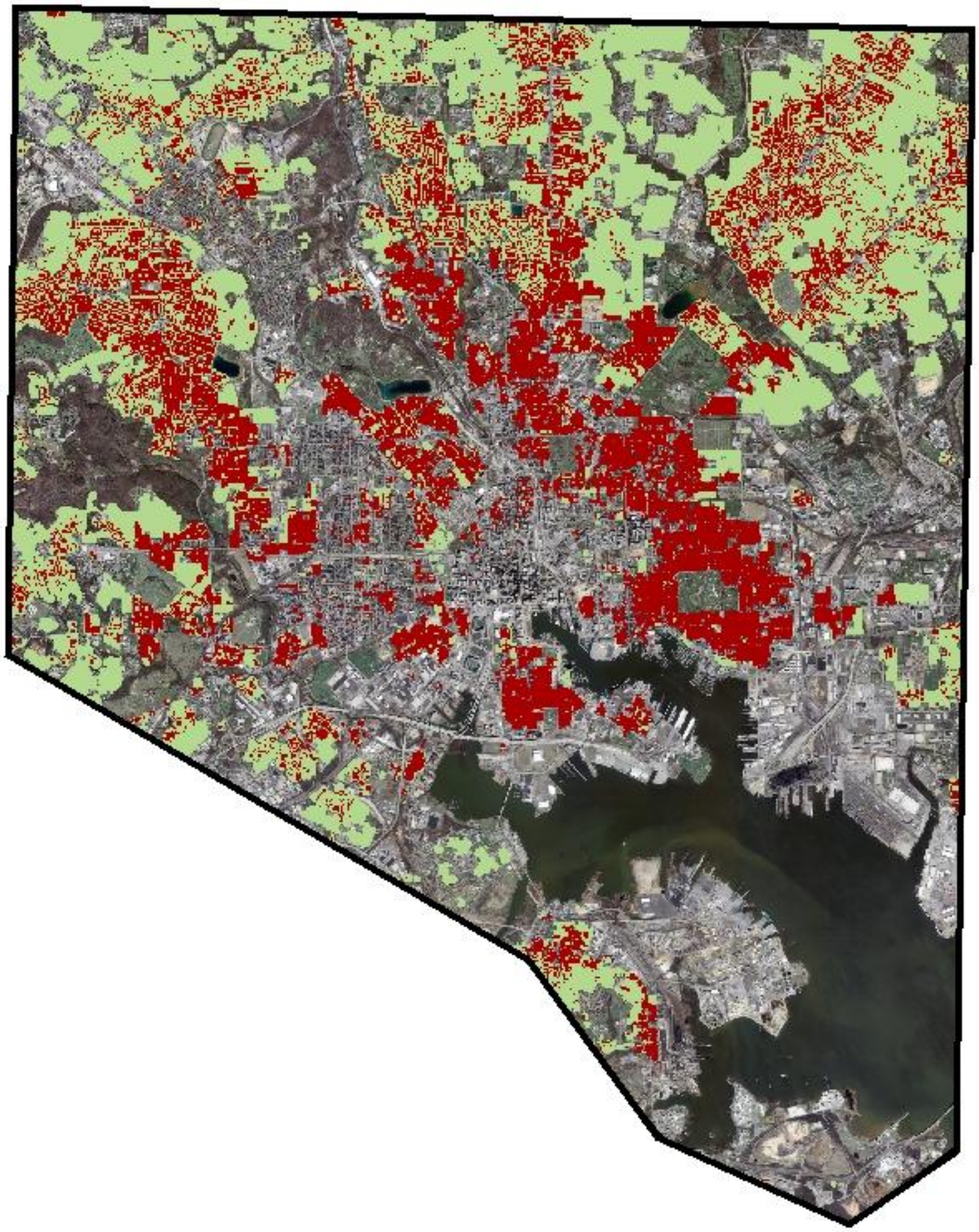


Figure 7.

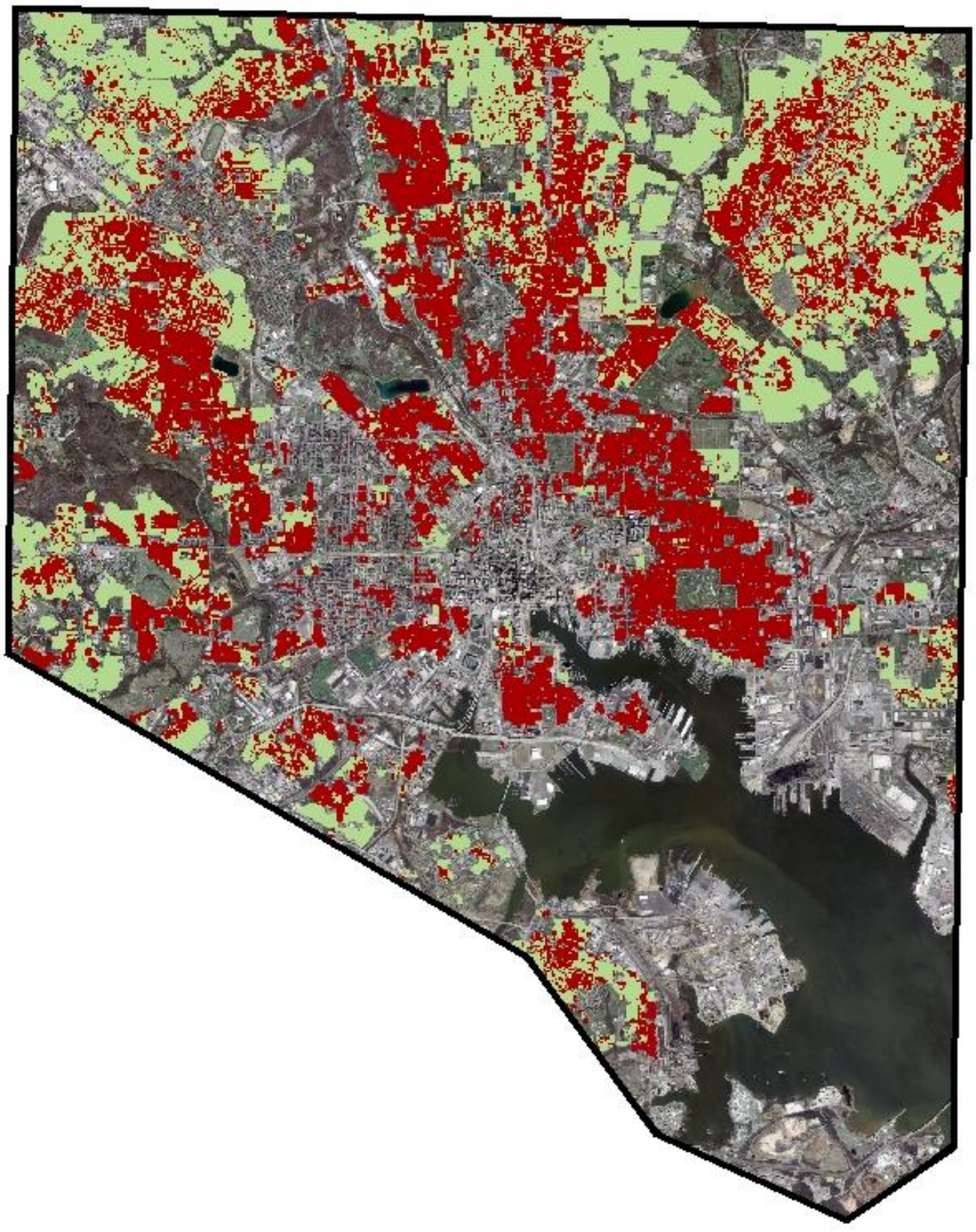
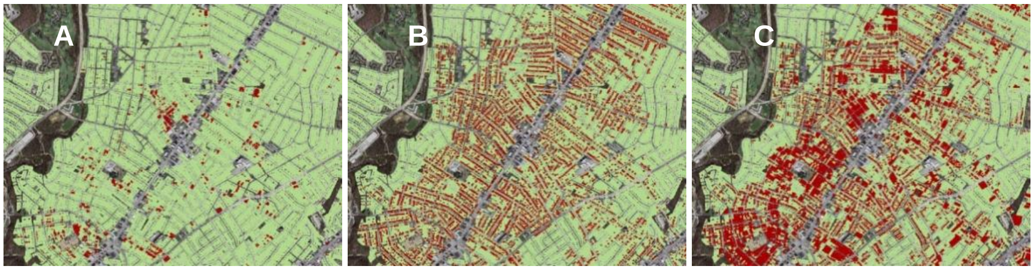


Figure 8.



CHAPTER 4. ENVIRONMENTAL INEQUITIES ASSOCIATED WITH THE SPATIAL DISTRIBUTION OF LEAD IN SOIL IN RESIDENTIAL AREAS OF BALTIMORE, MARYLAND USA.

Abstract

Valued for its unique chemical properties of corrosion resistance and low melting point, lead was historically used in several commercial products, most notably lead-based paint and leaded gasoline. Through these products, lead has made its way into our homes, our environment, and our bodies. Lead poisoning is an important public health concern because there is no safe level of lead in the body, and the effects of lead poisoning are devastating and irreversible. Lead poisoning is an entirely preventable disease. Yet, every year children who are most vulnerable to the effects of lead poisoning due to their unique physiology and increased hand-to-mouth activity continue to be poisoned. In addition, the incredible burden of lead poisoning does not appear to be equally distributed among populations. Poor, African American children living in large metropolitan areas and older homes are more likely to have elevated blood lead levels (BLLs). Many sources of lead can contribute to elevated BLLs, including lead-based paint, atmospheric lead, and lead in water, food, and soil. Here I partner an empirical model for residential soil lead concentrations in Baltimore, Maryland with census tract data to examine potential associations between demographic data and areas predicted to have soil lead concentrations above the United Stated Environmental Protection Agency's (USEPA) reportable limit of 400 ppm. In order to examine potential relationships between race and elevated soil lead concentrations, model predictions were overlaid with majority black (>50%)

census block groups and majority white (>50%) census block groups. A greater percent of area in the majority black census block groups is predicted to have elevated soil lead concentrations (11.57%) compared to the majority white census block groups (10.82%); however, this difference is small. In order to examine potential relationships between socio-economic variables and elevated soil lead concentrations, model predictions were overlaid with census blocks that exhibited poverty rates above the national average (12%) and those below. The model predicts that census blocks with poverty rates above the national average have a greater percent of soil with lead levels in excess of 400 ppm (18.35%) compared to census blocks with poverty rates below the national average (10.70%). Based on empirical model predictions, the data suggest that important associations exist between demographic data and the spatial patterning of lead in soil in Baltimore, Maryland. These results support earlier research conducted in Baltimore that demonstrated environmental inequities are not limited to communities of color, but rather affect all economically disadvantaged groups.

Introduction

The health impacts of lead poisoning are severe. Each year children continue to be exposed to harmful amounts of a known potent neurotoxin, lead (Pb), and the results of excessive exposure are devastating and irreversible (Rogan and Ware 2003). Lead exposure in children has been found to have profound negative effects on cognitive and behavioral function (Wakefield 2002) with some studies documenting adverse health effects below the current Centers for Disease Control's (CDC) level of concern of 10 µg/dL (Lanphear *et al.* 2000, Koller *et al.* 2004). These findings highlight the fact that there is no "safe" level of lead in the body. Adverse health effects of lead exposure are many and range from intellectual impairment and hyperactivity to coma, convulsions and death (ATSDR 1999). In addition, some scientists argue that lead exposure is associated with criminal activity or elevated risk for adjudicated delinquency (Needleman *et al.* 2002, Nevin 2007, Wright *et al.* 2008, Nevin *et al.* 2009).

Lead is especially dangerous to children. Children's bodies absorb more lead, an estimated 30-50%, compared to adults who only absorb 10-20% (Shannon 1996). Additionally, children are more sensitive to the effects of lead (USEPA, Lead in paint, dust, and soil, epa.gov). Once lead enters the body through ingestion or inhalation, short-term storage occurs in the blood, while longer term storage occurs in bone. Lead not stored in bones is eventually excreted in urine and feces. However, the amount excreted differs between adults and children. Adults excrete an estimated 99% while children only excrete about 32% (ATSDR

1999). Severe childhood lead poisoning can be treated with chelation therapy; however, studies have shown that even after undergoing treatment, cognitive loss does not appear to be restored (Rogan *et al.* 2001).

The toxic properties of lead have been known for a very long time. J. Lockhart Gibson was one of the first people to link childhood lead poisoning with lead-based paint in 1904 (as cited in Markowitz and Rosner 2000), but global lead pollution dates back to ancient times (Nriagu 1996). As evidence became clear of the toxic properties of lead-based paint in the 1920s and 1930s, many countries banned or restricted the use of interior lead-based paint. France, Belgium, Austria, Tunisia, Greece, Czechoslovakia, Great Britain, Sweden, Belgium and Poland all banned or placed restrictions on interior lead-based paint in the 1920s (as cited in Markowitz and Rosner 2000). The United States was slow to follow, only banning the use of lead-based paint in 1978. The sluggish response in the United States was potentially related to a trade group that represented lead pigment manufacturers called the Lead Industries Association (LIA) (Markowitz and Rosner 2000). The LIA actively tried to counter claims by the public health sector with an aggressive advertising campaign. It was also suggested that LIA marketing was directed toward urban populations. In a review of this ad campaign and the LIA's role in promoting a dangerous product, it is noted that "the LIA specifically targeted markets in urban areas" (Markowitz and Rosner 2000, p.41).

The burden of lead poisoning is not equally distributed. Even though great progress has been made in reducing the use of lead, lead poisoning is still a major public health concern, especially for children in urban areas. Although residents of more rural areas are not immune to anthropogenic lead contamination, urban populations are often more affected by lead poisoning (Laidlaw and Filippelli 2008). For example, the rates of lead poisoning in some cities are as high as 15-20% (Laidlaw and Filippelli 2008), much higher than the 2006 national average of 1.2% (CDC, CDC's national surveillance data, cdc.gov). Earlier studies have shown that elevated BLLs are significantly associated with age, African American race, low income, and older homes (Pirkle *et al.* 1998). It is well-documented that low-income minority communities have disproportionately higher rates of elevated BLLs (Kraft and Scheberle 1995). Direct evidence from Baltimore suggests that African Americans experience disproportionately higher lifetime lead doses as evidenced by higher tibia lead concentrations compared to Whites (Theppeang *et al.* 2008). Differences in socio-economic status did not explain the differences in tibia lead concentrations between African Americans and Whites. The authors argue differences are due to cumulative environmental lead exposures (Theppeang *et al.* 2008).

These data, along with the literature suggest that lead poisoning is an important environmental justice issue in urban areas (Kraft and Scheberle 1995). However, previous studies in Baltimore have shown that the accepted patterns of environmental racism do not hold true for some *point* sources of pollution. This is evidenced by higher concentrations of toxic release inventory (TRI) facilities in white working-class neighborhoods (Boone 2002). The pattern may be different for *non-point* sources of pollution such as elevated concentrations of lead in soil.

The societal costs of lead poisoning are also enormous. Those who balk at the potential high cost of remediation efforts fail to recognize that the annual cost of poisoning from environmental lead in the United States is $43.3 billion (Landrigan *et al.* 2002). That high number might be difficult to put in context. Consider that the same study estimates the annual cost of pediatric asthma of environmental origin to be $2 billion and the cost of childhood cancer of environmental origin to be $0.3 billion; it becomes very clear that the battle over lead poisoning has yet to be won. Interestingly, a study that examined the benefits associated with replacing windows in pre-1960 housing, a common source of childhood lead poisoning, estimated that there would be a net benefit, which considered energy savings, cost, market value benefit, and lifetime earning benefit, of $67 billion (Nevin *et al.* 2008). Given the fact that lead poisoning is 1) entirely preventable, 2) carries devastating health consequences that cannot be reversed, and 3) is costing the United States billions of dollars every year, it is clear that the public health focus needs to be on prevention (Rogan and Ware 2003).

Identifying the source of lead is a challenging task. Soil is an important sink for environmental lead and is thought to contribute to elevated BLLs (Duggan and Inskip 1985, Aschengrau *et al.* 1994, Mielke 1997); however, soil is one of many lead sources with which children may come into contact. To further complicate matters, some argue that pinpointing a single lead source, i.e. paint, is more difficult when lead poisoning is low (10-20 µg/dL) because multiple sources may be at play (Shannon 1996). In addition, at low BLLs, ingesting paint chips is probably not the source of poisoning. If lead paint chips were the source of lead poisoning BLLs would likely be much higher (Shannon 1996). In addition to the issue of multiple sources, other factors are also important to total lead body burdens. Nutrition has been found to be an important factor in lead absorption in the body, such that both calcium and iron deficiencies have been associated with greater lead absorption (Shannon 1996, Ryan *et al.* 2004). In addition, diets high in fat have also been associated with increased lead absorption (Shannon 1996). Pica, or ingestion of non-food items, is also thought to be a risk factor. Identifying the multiple sources of lead in the environment and assigning risk to those sources is not the aim of this study. Here I focus on one potential source of lead in the environment and the possible relationships between that potential source and certain demographic features.

In this study I examine potential associations between demographic data and soil lead concentrations in order to address possible environmental justice inequities in Baltimore Maryland. Specifically, I partner an empirical model of residential soil lead concentrations above and below the USEPA reportable limit of 400 ppm in Baltimore, Maryland with census tract data. This study, therefore, examines geographic equity defined by Bullard (1994) as "the location and spatial configuration of communities and their proximity to environmental hazards and locally unwanted land uses" (p. 13). Social equity, which considers environmental decision-making, is not addressed in this study. A more complete understanding of which communities are predicted to experience higher levels of lead in soil could aid soil remediation efforts.

Methods

Within a GIS, a map of the predicted soil lead values derived from an empirical model was overlaid with census tract data (Figures 1 and 2). The empirical model is based on soil lead concentrations that were collected in Baltimore City, Maryland in 2007-2008 using field portable x-ray fluorescence. The model was developed using classification and regression trees (CART) and predicts soil lead concentrations in two categories: those predicted to be below the USEPA guideline of 400 ppm (low) and those predicted to be above the USEPA guideline (high). In an earlier study, I compared several different models that predicted the spatial patterning of lead in soil as a function of landscape features. One model utilized a traditional general linear model (GLM) while another model used Classification and Regression Trees (CART), a machine learning technique. I used the model developed with CART for this analysis because CART models have been shown to better capture complex patterns that linear models sometimes miss (De'Ath and Fabricius 2000). CART models may therefore be more adept at predicting soil lead concentrations in highly heterogeneous environments, particularly urban areas. In addition, the predictions of the CART model more closely resemble spatial patterns revealed by previous research on the distribution of heavy metals in Baltimore City (Yesilonis *et al.* 2008). The CART model also captures important hotspots in the landscape detected in the field data that was used to construct the model. By combining existing census tract data with predicted soil lead concentrations I was able to examine potential associations between demographic factors and the spatial distribution of lead in soil.

I first examined relationships between race and areas predicted to have soil lead concentrations greater than 400 ppm. I chose to examine majority black and majority white census block groups because they are the predominant races represented in Baltimore City; approximately 95.9% of Baltimoreans self-identified as either black (64.3%) or white (31.6%) on the last United States Census (United States Census Bureau, Baltimore City Quickfacts, census.gov). Areas predicted to exceed 400 ppm by the CART model were converted from raster to polygon format. Using the census tract data, I calculated the percent of black and white residents at the census block level and created two new layers. One layer consisted of the majority black census block groups and the other consisted of the majority white census block groups. I defined "majority" as greater than fifty percent. Using the "dissolve" function in ArcGIS 9.3, the majority black census block groups were converted to one continuous polygon. The same was done for the majority white census block groups. Next, I used the Hawths Tools Extension "Polygon in Polygon Analysis" to calculate the area predicted to have high lead concentrations (> 400 ppm) in the majority black and majority white census block layers. I then repeated this process for areas predicted to have low lead concentrations (< 400 ppm). This allowed the total modeled area that was contained within the different census block groups to be calculated. The total modeled area includes only residential parcels that contain housing age data. In addition, the model does not include the area occupied by building footprints. The calculations are:

(Area Predicted > 400 ppm for the Majority Black Census Blocks) / (Total Modeled Area) = 1,957,194 m^2 / 16,919,219 m^2 *100 = **11.57%**

(Area Predicted > 400 ppm for the Majority White Census Blocks) / (Total Modeled Area) = 1,979,896 m^2 / 18,303,631 m^2 *100 = **10.82%**

A similar analysis was done to examine correlations between soil lead levels and the percent of residents living below the poverty line. The percentage of people living below the poverty line in the United States is 12% (Central Intelligence Agency, The world factbook (2004 estimate), cia.gov). Using 12% as a threshold, I divided the census block groups of Baltimore City into those with poverty rates below the national average and those with poverty rates above the national average. Again I used the Hawths Tools Extension “Polygon in Polygon Analysis” to calculate the amount of area predicted to have lead levels greater and less than 400 ppm in each of the census block groups. The calculations are below:

(Area Predicted > 400 ppm for Census Blocks with >12% of Population Living Below the Poverty Line) / (Total Modeled Area) = 408,195 m^2 / 2,224,797 m^2 *100 = **18.35%**

(Area Predicted > 400 ppm for Census Blocks with <12% of Population Living Below the Poverty Line) / (Total Modeled Area) = 3,416,401 m^2 / 31,933,595 m^2 *100 = **10.70%**

Results and Discussion

A greater percentage of area in the majority black census block groups is predicted to have elevated soil lead concentrations (11.57%) compared to that in the majority white census block groups (10.82%); however, this difference is slight. A greater percentage of area in census blocks with poverty rates above the national average is predicted to have elevated soil lead concentrations (18.35%) compared to areas with poverty rates below the national average (10.70%).

It is necessary to address important caveats to this study. This study does not address elevated BLLs. Instead, this study addresses whether or not associations exist between demographic features and the spatial distribution of lead in soil. Lead in soil is only one of many important sources of lead in the environment that may contribute to elevated BLLs. Therefore, this study does not address the association of soil contaminated with lead to elevated BLLs or the associated risk. In addition, the model used in this study is a physical environmental model and does not take into account socio-economic variables that may be important to predicting the spatial patterning of lead.

By partnering an empirical model that predicts lead concentrations in soil with census tract data I was able to identify associations between certain demographic features and the spatial distribution of lead in soil. The results indicate that areas with poverty rates above the national average may be disproportionately burdened with a greater amount of area predicted to have elevated soil lead concentrations. These results parallel earlier research conducted in Baltimore that showed that environmental inequities are not limited to communities of color but rather affect all economically disadvantaged groups (Boone 2002, Pickett *et al.* 2008). Further research is needed to explain why these inequities exist and what role, if any, history, racial biases, and politics played in the resultant pattern.

Literature Cited

Aschengrau, A., A. Beiser, D. Bellinger, D. Copenhafer, and M. Weitzman. 1994. The impact of soil lead abatement on urban children’s blood lead levels: phase II results from the Boston lead-in-soil demonstration project. Environmental Research 67: 125-148.

Agency for Toxic Substances and Disease Registry. 1999. Toxicological profile for lead. ATSDR. Atlanta, Georgia.

Boone, C.G. 2002. An assessment and explanation of environmental inequity in Baltimore. Urban Geography 23 (6): 581-595.

Bullard, R. 1994. Overcoming racism in environmental decision making. Environment 36(4): 10-20.

Centers for Disease Control and Prevention. CDC’s national surveillance data. Cdc.gov. 28 October 2009 <http://www.cdc.gov/nceh/lead/data/national.htm>.

Central Intelligence Agency. The world factbook (2004 estimate). Cia.gov. 28 October 2009 < https://www.cia.gov/library/publications/the-world-factbook/fields/2046.html?countryName=UnitedStates&countryCode=us&regionCode=na&#us>.

De'ath, G. and K.E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81(11): 3178-3192.

Duggan, M.J., and M.J. Inskip. 1985. Childhood exposure to lead in surface dust and soil: a community health problem. Public Health Reviews 13: 1-54.

Koller, K., T. Brown, A. Spurgeon, and L. Levy. 2004. Recent developments in low-level lead exposure and intellectual impairment in children. Environmental Health Perspectives 112(9): 987-994.

Kraft, M. and D. Scheberle. 1995. Environmental justice and the allocation of risk: the case of lead and public health. Policy Studies Journal 23(1).

Laidlaw, M. and G. Filippelli. 2008. Resuspension of urban soils as a persistent source of lead poisoning in children: A review and new directions. Applied Geochemistry 23(8): 2021-2039.

Landrigan, P.J., C.B. Schechter, J.M. Lipton, M.C. Fahs, and J. Schwartz. 2002.

Environmental pollutants and disease in American children: estimates of morbidity, mortality, and costs for lead poisoning, asthma, cancer, and developmental disabilities. Environmental Health Perspectives 110(7): 721-728.

Lanphear , B.P., K. Dietrich, P. Auinger, and C. Cox. 2000. Cognitive deficits associated with blood lead concentrations <10 mg/dL in US children and adolescents. Public Health Report 115: 521–9.

Markowitz, G. and D. Rosner. 2000. "Cater to the children": the role of the lead industry in a public health tragedy, 1900-1955. American Journal of Public Health 90 (1): 36-46.

Mielke, H.W. 1997. Leaded dust in urban soils shown to be greater source of childhood lead poisoning than leaded paint. Lead Perspectives. March/April 997:28-31.

Needleman, H.L., C. McFarland, R.B. Ness, S.E. Fienberg, and M.J. Tobin. 2002. Bone lead levels in adjudicated delinquents A case control study. Neurotoxicology and Teratology 24(6): 711-717.

Nevin, R. 2007. Understanding international crime trends: the legacy of preschool lead exposure. Environmental Research 104(3): 315-336.

Nevin, R. 2009. Trends in preschool lead exposure, mental retardation, and scholastic achievement: Association or causation? Environmental Research.

Nevin, R., D.E. Jacobs, M. Berg, and J. Cohen. 2008. Monetary benefits of preventing childhood lead poisoning with lead-safe window replacement. Environmental Research 106(3): 410-419.

Nriagu, J.O., 1996. A history of global metal pollution. Science 272: 223-224.

Pickett, S.T.A., M.L. Cadenasso, J.M. Grove, P.M. Groffman, L.E. band, C.G. Boone, W.R. Burch Jr., C.S.B. Grimmond, J. Hom, J.C. Jenkins, N.L. Law, C.H. Nilon, R.V. Pouyat, K. Szlavecz, P.S. Warren, and M. A. Wilson. 2008. Beyond urban legends: An emerging framework of urban ecology, as illustrated by the Baltimore Ecosystem Study. BioScience 58(2): 139-150.

Pirkle, J.L., R.B. Kaufmann, D.J. Brody, T. Hickman, E.W. Gunter, and D.C. Paschal. 1998. Exposure of the US population to lead, 1991-1994. Environmental Health Perspectives: 745-750.

Rogan, W.J., K.N. Dietrich, J.H. Ware, D.W. Dockery, M. Salganik, J. Radcliffe, R.L. Jones, N.B. Ragan, J.J. Chisolm and G.G. Rhoads. 2001. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. New England Journal of Medicine 344(19): 1421-1426.

Rogan, W.J. and J.H. Ware. 2003. Exposure to lead in children-how low is low enough? 348: 1515-1516.

Ryan, J.A., K.G. Scheckel, W.R. Berti, S.L. Brown, S.W. Casteel, R.L. Chaney, J. Hallfrisch, M. Doolan, P. Grevall, M. Maddaloni, and D. Mosby. 2004. Reducing children's risk from lead in soil. Environmental Science and Technology 38 (1): 18A-24A.

Shannon, M.W. 1996. Etiology of childhood lead poisoning. *In* lead poisoning in childhood. Pueschel, S.M., J.G. Linakis, and A.C. Anderson (eds). Paul H. Brooks Publishing, Baltimore MD.

Theppeang, K., T.A. Glass, K. Bandeen-Roche, A.C. Todd, C.A. Rohde, and B.S. Schwartz. 2008. Gender and race/ethnicity differences in lead dose biomarkers. American Journal of Public Health 98(7): 1248.

United States Census Bureau. Baltimore City Quickfacts. Census.gov. 28 October 2009 < http://quickfacts.census.gov/qfd/states/24/2404000.html>.

United States Environmental Protection Agency. Lead in paint, dust, and soil. Epa.gov. 28 October 2009 <http://www.epa.gov/lead/pubs/leadinfo.htm#facts>.

Wakefield, J. 2002. The lead effect? Environmental Health Perspectives 110(10): A574.

Wright, J.P., K.N. Dietrich, M.D. Ris, R.W. Hornung, S.D. Wessel, B.P. Lanphear, M. Ho, and M.N. Rae. 2008. Association of prenatal and childhood blood lead concentrations with criminal arrests in early adulthood. PLoS Med 5(5): e101.

Yesilonis, I., R. Pouyat, and N.K. Neerchal. 2008. Spatial distribution of metals in soils in Baltimore, Maryland: role of native parent material, proximity to major roads, housing age and screening guidelines. Environmental Pollution 156(3): 723-731

Acknowledgements

I am especially grateful to the homeowners for access to their property and to the University of California, Davis for use of the XRF. My understanding of GIS and spatial statistics greatly benefited from conversations with John Bognar, Dr. Adele Cutler, Amanda Elliot, Dr. Elizabeth Freeman, Scott Haag, David Lewis, Dr. Zewei Maio, Samuel Simkin, Jim Trimble, and Dr. Weiqi Zhou. This work is a contribution to the Long Term Ecological Research Program (a program of the National Science Foundation) and the Cary Institute of Ecosystem Studies and was supported by NSF grants DEB 042376 and 0808418.

Figure Legends

Figure 1. The green areas represent the majority black (>50%) census block groups. The blue areas represent the majority white (>50%) census block areas. Areas that are predicted to have soil lead values above the USEPA reportable limit of 400 ppm are depicted in red.

Figure 2. The orange areas represent census block groups with >12% of the population living below the poverty line. The purple areas represent census block groups with <12% of the population living below the poverty line. Areas that are predicted to have soil lead values above the USEPA reportable limit of 400 ppm are depicted in red.

Figure 1.

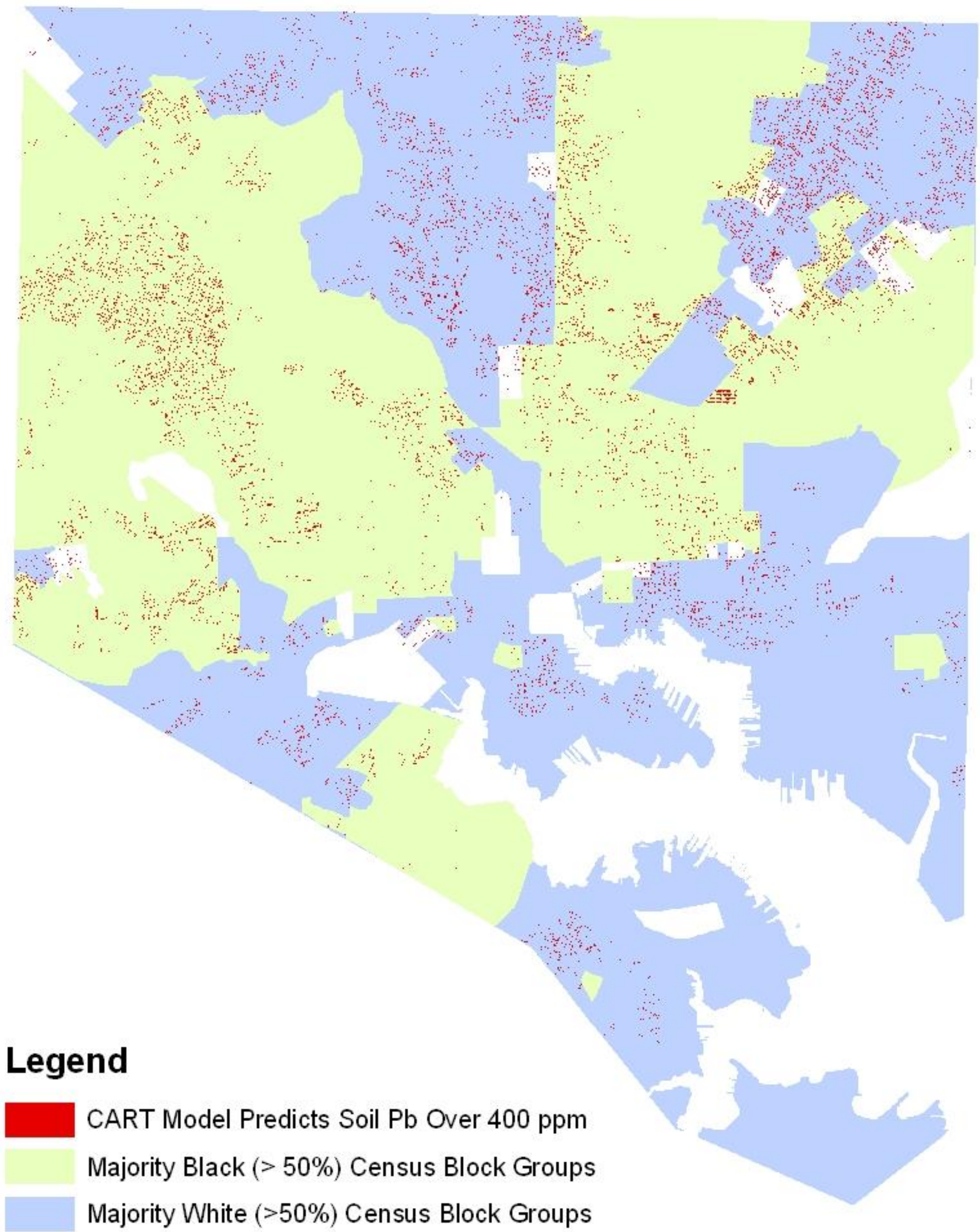
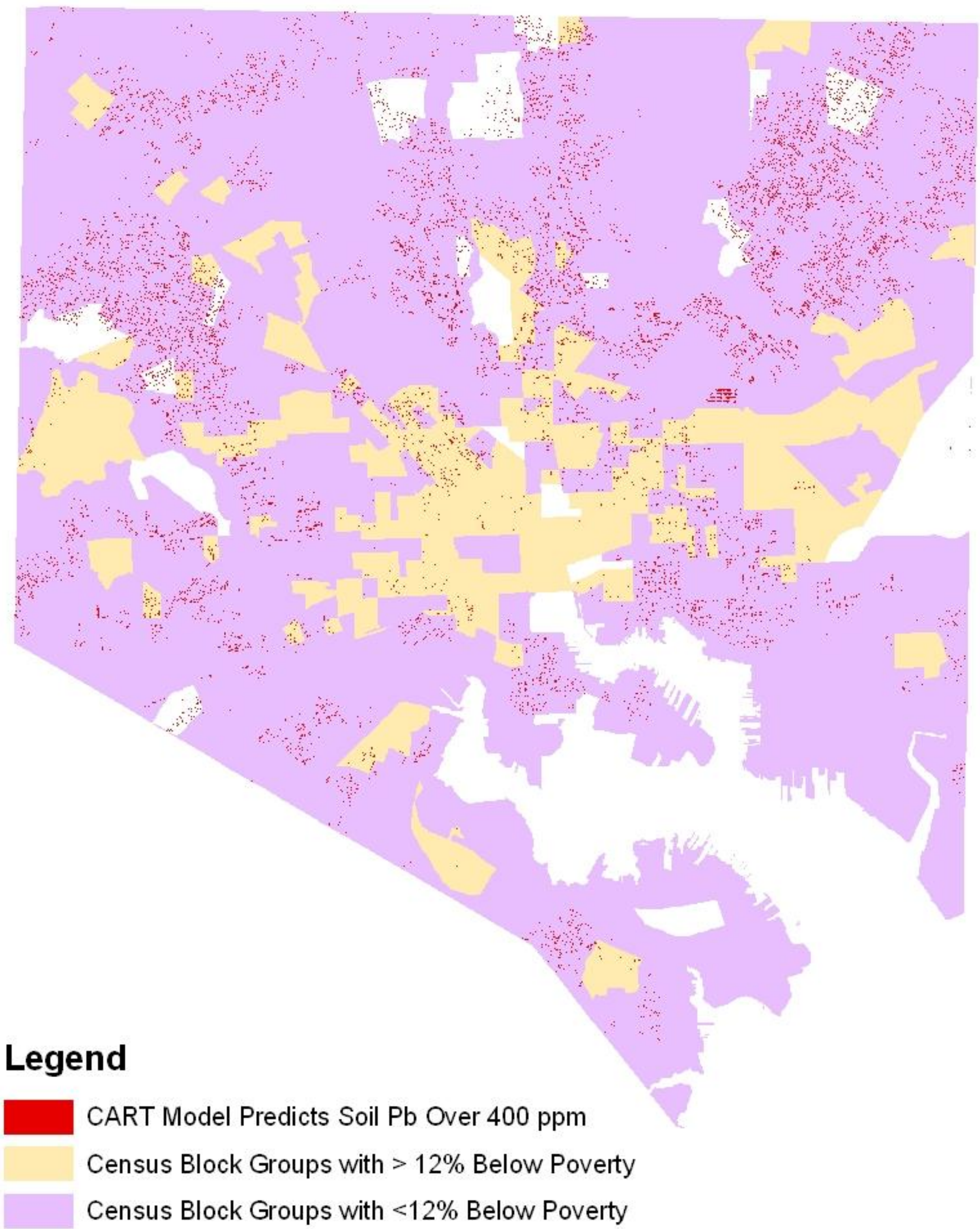

Legend
CART Model Predicts Soil Pb Over 400 ppm
Majority Black (> 50%) Census Block Groups
Majority White (>50%) Census Block Groups

Figure 2.


Legend
CART Model Predicts Soil Pb Over 400 ppm
Census Block Groups with > 12% Below Poverty
Census Block Groups with <12% Below Poverty

CONCLUSIONS

Lead contamination of the urban residential environment in Baltimore, Maryland is highly variable and pervasive. This research has demonstrated that the spatial distribution of lead in urban residential soils is influenced by land cover, legacies of lead-containing products, and spatial heterogeneity and has important implications for environmental inequity in some communities. This research has shown that in contrast to land use, land cover is a predictor of lead concentrations in soil. In addition, historical land cover is also an important predictor of lead in soil. Certain features of land cover are more important than others. Housing age, distance to roads, and distance to building strongly influence the spatial distribution of lead in soil and explain up to 42% of the variation in the data. The various spatial models in this dissertation demonstrate similar overarching patterns of contamination; however, the models predict differing extents of contamination. Further work is needed in order to determine which model is most appropriate. Finally, by partnering the modeled soil lead predictions with demographic features, I have found that a larger amount of contamination is predicted to occur in high poverty areas. This finding supports earlier work in Baltimore that suggested environmental inequities are not limited to communities of color and instead affect all economically disadvantaged groups.

There are several implications of this work. The first implication is that land cover, in contrast to land use, may be a better predictor of lead concentrations in urban residential soils. The importance of land cover as a predictive ecological variable can likely be applied to other ecosystem functions that may be based on spatially differentiated source and sink patches, and networks of connection. This suggests that for spatially complex situations where sources and sinks are interspersed, metrics describing ecosystem structure, such as land cover, may be more appropriate to ecological studies than classifications that focus on coarse categories of land use. The second implication is that legacies of lead in the environment are important to the current spatial distribution of lead in soil. The third implication is the necessity to use intensive sampling to capture inherent variability in urban residential soil lead concentrations. Therefore, a sample collected from the middle of the yard would likely exclude areas of high contamination. In addition, composite samples can reduce variability, missing hotspots in the landscape. This information should be considered when conducting geochemical mapping of the urban residential environment. The fourth implication is the utility of spatial models in describing the spatial distribution of lead in soil. Although the modeling work in this dissertation is focused on Baltimore, Maryland, this approach is applicable to other urban areas because the features that were found to be significant to the spatial distribution of lead in this study, including housing age, distance to road, and distance to building are likely important in other cities of similar structure, age, and history. Finally, the information from this dissertation can be used in remediation efforts to target areas of high contamination in the yard, for example soil adjacent to buildings, thereby reducing cost and the challenge of contaminated soil disposal. Knowledge regarding the association between impoverished communities and high soil lead concentrations can inform the public health community and help protect vulnerable communities from potential exposure to an important environmental contaminant.

VITA

Kirsten Schwarz

1996 – 2000	B.A. in Human Ecology, College of the Atlantic, Bar Harbor, Maine
2000 – 2003	Research Assistant, Cary Institute of Ecosystem Studies, Millbrook, New York
2003 – 2006	Teaching Assistant, Department of Life Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey
2004	Simkin, S.M., D.N. Lewis, K.C. Weathers, G.M. Lovett, and K. Schwarz. Determination of Sulfate, Nitrate, and Chloride in Throughfall using Ion-Exchange Resins. *Water, Air and Soil Pollution*. 153: 343-354.
2006	Grove, J.M., M.L. Cadenasso, W.R. Burch, Jr., S.T.A. Pickett, K. Schwarz, J. O'Neil-Dunne, M. Wilson, A. Troy, and C.G. Boone. Data and Methods Comparing Social Structure and Vegetation Structure of Urban Neighborhoods in Baltimore, Maryland. *Society and Natural Resources*. 19: 117-136.
2006 – present	Research Assistant, Cary Institute of Ecosystem Studies, Millbrook, New York
2007	Cadenasso, M.L., S.T.A. Pickett, and K. Schwarz. Spatial Heterogeneity in Urban Ecosystems: reconceptualizing land cover and a framework for classification. *Frontiers in Ecology and the Environment*. 5(2): 80-87.
2008	Cadenasso, M.L., S.T.A. Pickett, and K. Schwarz. Measuring change - Reply. *Frontiers in Ecology and the Environment*. 6(2):67-68.
2009	Zhou, W., K. Schwarz, and M.L. Cadenasso. Mapping Urban Landscape Heterogeneity: Agreement between Visual Interpretation and Digital Classification Approaches. Landscape Ecology. DOI: 10.1007/s10980-009-9427-8.

2003 – 2010	Ph.D. in Ecology and Evolution, Rutgers, The State University of New Jersey, New Brunswick, New Jersey