1-1-2014

Evaluating Spatial Outliers And Integrating Temporal Data In Air Pollution Models For The Detroit-Windsor Airshed

Brendan Francis O'leary
Wayne State University,

EVALUATING SPATIAL OUTLIERS AND INTEGRATING TEMPORAL DATA IN AIR POLLUTION MODELS FOR THE DETROIT-WINDSOR AIRSHED

by

BRENDAN F. O'LEARY

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

In partial fulfillment of the requirements

For the degree of

MASTER OF SCIENCE

2014

MAJOR: Geology

Approved By:

_____

Advisor: Date

DEDICATION

To my parents,

John and Karen O'Leary

For their self-sacrifice, love, and support in all of my academic endeavors**.**

ACKNOWLEDGMENTS

Dr. Lawrence Lemke provided consistent and invaluable support and direction to my graduate school experience at Wayne State University. His commitment to my success in completing this thesis highlighted a broader pattern of effective academic advising, supportive direction, and constructive feedback. He was particularly adept in teaching effective communication skills, correcting my spelling, developing a cohesive research laboratory team, and providing me with an exceptional role model for my professional career.

My committee, Dr. John Reiners Jr. and Dr. Mark Baskaran, supplied invaluable feedback and constructive criticism on my thesis project. In particular, I appreciated Dr. Reiners' personal commitment to my educational development by providing editorial corrections and suggestions on the drafts of my manuscript and grant proposal. In addition, Dr. Reiners along with Dr. Lemke, provided me with an opportunity to be involved with the GeoDHOC project. This was a unique interdisciplinary educational experience that broadened my interdisciplinary perspectives in ways that I did not expect. The GeoDHOC research team was generous in their collegial inclusion. In particular Gianluca Sperone provided perspective with his sense of humor during this research project and his assistance with GIS. Pierre Goovaerts was helpful with the outliers analysis and SpaceStat. The W.C. Kellogg Foundation and Wayne State University provided the funding for this research.

Lastly, I wish to thank my lab partners, Lauren Bugdalski and Amanda Pruehs, for their supportive comradery and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.0. Introduction

In response to poor urban air quality in many of the United States' cities, the U.S. Congress established the Clean Air Act in 1970 "to protect public health and welfare from different types of air pollution caused by a diverse array of pollution sources" (Chow et al., 2007). The Clean Air Act established the air quality management framework that is currently in place today. This framework measures regional ambient air pollutant concentrations but does not adequately account for the heterogeneous nature of urban regions.

The heterogeneity of urban airsheds results from the interplay of spatially and temporally complex systems (e.g. Kim et al., 2005; Pinto et al., 2004). It complicates human exposure estimates in urban areas and creates the need for accurate, highly detailed spatiotemporal air contaminant models. Distributed and prolonged air quality measurements are resource intensive, however, and study designs frequently balance tradeoffs between spatial and temporal resolution (Beevers et al., 2013). Increased density of spatial information and/or increased frequency of temporal information also increases the potential for erroneous measurements within air quality datasets. Consequently, there is a growing need to develop practical methods to 1) evaluate the presence and influence of anomalous air quality measurements and 2) integrate detailed spatial and temporal air quality data from multiple sources (e.g. Mayer, 1999; Ross et al., 2013; Wilson et al., 2005).

Currently, air sampling networks in the United States provide regional air pollution estimates from central-site monitors. Although regulatory monitoring networks offer valuable time series measurements, they commonly lack the spatial resolution needed to estimate neighborhood-scale exposure (e.g. Baxter et al., 2013; Ozkaynak et al., 2013; Sarnat et al., 2013;

Wilson et al., 2005). This can result in greater uncertainty (Luginaah et al., 2006) coupled with corresponding underestimation of variability (Baxter et al., 2013) in large population exposure estimates.

Alternatively, temporary networks of active or passive air samplers can collect a higher spatial density of measurements over limited, discontinuous periods of time (e.g., Miller et al., 2010; Ross et al., 2013). Such, short-term monitoring networks can be logistically difficult to implement and expensive to repeat (Cocheo et al., 2008). Nevertheless, their measurements are readily incorporated into land use regression (LUR) and geostatistical interpolation (i.e., kriging) algorithms to generate pollutant concentration models at increased spatial resolution (e.g., Hoek et al., 2008; Jerrett et al., 2005; Künzli et al., 2004; Sampson et al., 2011). LUR and kriging models share similar limitations but have different strengths For example, both require a large number of sampling sites and are not readily adaptable to changing meteorological conditions (Isakov et al., 2011). For example, LUR models can reproduce small scale features such as roadway configurations that contribute to mobile source pollutants (Mercer et al., 2011) whereas kriging smooths concentration estimates. Moreover, kriged models generate measures of uncertainty using estimation of error variance throughout the model domain (Vicedo-Cabrera et al., 2013) whereas LUR models do not yield corresponding spatial uncertainty estimates.

Accurate datasets, without corrupt information, are essential for creating models that promote informed decision making. Data outliers can greatly influence finely resolved spatial models, like kriged or LUR models, at the neighborhood level (Chang-Tien et al., 2003). Identifying outliers can lead to useful information and unexpected outcomes such as severe air pollutant exposure for specific geographical zones (Chang-Tien et al., 2003; Torres et al., 2011) or the identification of air pollution adversely affecting socio-economic groups (Zou et al.,

2014). Therefore, along with measuring uncertainty, an evaluation of the accuracy and influence of individual data points is essential when making interpretations at highly resolved spatial scales.

Temporal data in conjunction with spatial data are needed to address chronic exposure across air pollution gradients in urban areas. Exposure estimates over time are required to study certain health conditions, such as those associated with pregnancy, which have a specific gestational period. Therefore, spatiotemporal models, which can provide individual exposure estimates and time-base exposure estimates, are needed for epidemiological studies (Brauer et al., 2003).

The combination of spatial modeling with temporal data adds definition to the heterogeneous nature of air pollution and delivers arguably better exposure estimates (Mölter et al., 2010; Romanowicz et al., 2006). This thesis contributes to the development of emerging modeling approaches by presenting practical methods to refine spatial models and assimilate detailed temporal data with high spatial resolution models of urban air pollutants. Specifically this study investigates the spatial and temporal patterns of nitrogen dioxide (NO_2), total BTEX (benzene, toluene, ethylbenzene, and xylene), volatile organic compounds (VOCs), particulate mass less than 2.5 microns in aerodynamic diameter ($PM_{2.5}$), particulate mass less than 10 microns (PM_{10}), and polycyclic aromatic hydrocarbons (PAH) in the cities of Detroit and Windsor.

1.1. Background on the Detroit-Windsor Airshed

This study investigates the contiguous airshed of Detroit, Michigan (U.S.A.) and Windsor, Ontario (Canada). Detroit is located on the north side of the Detroit River in

southeastern Michigan and Windsor is situated on the south side of the Detroit River in southwestern Ontario. Windsor is also the southernmost metropolitan city in Canada. Both of these cities are identified as high air pollutant zones resulting from industrial and transportation emissions (Health Canada, 2000; MDEQ, 2008; Simon et al., 2005; U.S. EPA, 2009).

Detroit and Windsor are connected by the Ambassador Bridge, which is the busiest commercial international border crossing in North America (Figure 1.1). The Ambassador Bridge and Detroit-Windsor Tunnel concentrate traffic in both cities and create focused areas of air pollution (Wheeler et al., 2008). Municipal and medical waste incinerators, automobile manufacturing plants, steel mills, the Detroit Edison Rouge River coal-fired power plant, and the Marathon refinery are examples of major emission sources for Detroit and its surrounding communities. In Windsor, industrial, municipal, and transportation are the major sources of air pollution (Wheeler et al., 2008). Beyond industry, traffic induced emissions are a large source of air pollution for Detroit and Windsor. Major interstate highways in Detroit include I-75, I-94, and I-96, and Detroit's major state highways include M-1, M-10, and M-39 (Molaroni, 2010). Major highways in Windsor include Huron Church Rd (3), E.C. Row Expressway, and Macdonald Cartier Freeway (401).

1.2 Air Pollutant Sources and Monitoring

Air pollutants come from a range of sources in Detroit and Windsor that include stationary and on road sources. A majority of NO_2 is derived from on road sources (MDEQ, 2008). In Detroit, VOC compositions and trends indicate the dominance of vehicular sources over the many industrial sources with the possible exceptions of styrene and several chlorinated

VOCs (Batterman et al., 2002). According to the US Environmental Protection Agency (USEPA), in Michigan, 34% of PM_{10} comes from particles that originate from point sources, such as power plants, and various manufacturing and industrial processes; 32% comes from "area" sources that do not originate from any specific point; and another 34% comes from vehicle emissions. $PM_{2.5}$ from vehicle emissions comprises 50% of the ambient $PM_{2.5}$ in Michigan. Area sources make up 37% and point sources contribute 13% (U.S. EPA, 2009).

Air sampling networks established to monitor compliance with National Ambient Air Quality Standards (NAAQS) and the National Air Pollution Surveillance (NAPS) program are important sources of outdoor air quality information in Detroit and Windsor. The Michigan Air Sampling Network (MASN) uses strategically placed monitors to assess air pollutant levels throughout the state (MDEQ, 2013). The NAPS program provides similar data with two locations in Windsor. Comparable air sampling networks in other states and countries provide long-term air quality measurements that may be used to estimate exposure for surrounding communities (e.g. Dockery et al., 1993; Pope et al., 2009; Samet et al., 2000; Zanobetti et al., 2003).

Currently, Detroit and Windsor have a limited number of continuous monitoring locations. Given the need for accurate models to predict and understand the spatial and temporal variability of air pollutants within heterogeneous urban airsheds, air pollutant modeling is required to estimate exposure at a spatially resolved level that complements temporally resolved measurements from the long term air sampling networks.

1.3. Previous Detroit and Windsor Air Quality Studies

Multiple studies in the Detroit and Windsor airshed have focused on correlating air pollution and human exposure. Recent investigations include the Detroit Exposure and Aerosol Research Study (DEARS), the Windsor, Ontario Exposure Assessment Study (WOEAS) and the Geospatial Determinants of Health Outcomes Consortium (GeoDHOC) studies.

The objective of the DEARS project was to determine if air pollutant data from centralized monitoring stations could be used to estimate exposure in neighborhoods from multisource air pollutants. It was conducted between 2004 and 2007 and took a number of approaches to compare central monitoring station measurements to alternative exposure measurements including indoor, outdoor, and personal monitors (Williams et al., 2009). The DEARS study confirmed that air pollutants vary at the neighborhood scale and are significantly affected by weather.

The WOEAS investigation assessed the contribution of ambient air pollutants to personal and indoor exposures of adults and asthmatic children living in Windsor. The variability of air pollution, particularly around the Ambassador Bridge, was analyzed using a combination of land use regression (LUR) modeling (Luginaah et al., 2006; Wheeler et al., 2008) and personal monitors. The study found that VOC indoor concentrations are a good predictor of personal exposure (Stocco et al., 2008). The WOEAS study also concluded that using central monitoring locations for personal exposure in epidemiological studies creates error because of the spatial variability and wide range of home infiltration factors associated with particulate matter (Kearney et al., 2011).

The initial GeoDHOC study was designed to examine spatial correlations between ambient air pollution concentrations and asthma exacerbations in Detroit and Windsor at a high

level of spatial resolution. The underlying premise of the study was that correlations among mappable environmental attributes and health indicators can be used to better understand and manage urban community health (Miller et al., 2010). The original aims included: 1) simultaneous collection and modeling of air pollutant concentrations in Detroit and Windsor, 2) collection and evaluation of concurrent asthma morbidity data using asthma-related ambulatory care encounters, emergency department visits, and hospital discharge records in Detroit and Windsor, and 3) integration of the environmental and asthma data into a geographic information system (GIS) and modeling framework (Lemke et al., 2013).

The initial GeoDHOC analysis demonstrated spatial variability in air pollutant distributions between and, more importantly, within the cities of Detroit and Windsor at neighborhood scales (Miller et al., 2010; Miller et al., 2012a), as well as statistically significant correlations between the rate of asthma events and concentrations of specific air pollutants averaged over postal code scales (i.e., zip codes in the USA and forward sortation areas in Canada) (Lemke et al., 2013). The GeoDHOC study has subsequently been expanded to investigate the relationship between spatially distributed airborne environmental contaminants and adverse birth outcomes in Detroit. In a pilot project funded by the W.K. Kellogg Foundation ("Geospatial Analysis of Air Pollution and Adverse Birth Outcomes in Detroit", John Reiners, Jr., PI), the GeoDHOC team is testing a hypothesized geospatial association between concentrations of specific pollutants (i.e., NO_2, PM, and VOCs) in the Detroit airshed and adverse birth outcomes (i.e., low birth weight and premature delivery) for pregnant women living in Detroit (Reiners et al., 2014). To achieve these correlations, a pragmatic method for integrating high spatial resolution measurements from a temporary monitoring network with time

series measurements from fixed regulatory monitoring stations was implemented (O'Leary and Lemke, 2014) and this work forms a significant portion of the foundation for this thesis.

1.4 Spatial Outliers

Often air quality data needs refinement to account for anomalous measurements (Torres et al., 2011). These irregular measurements need investigation to minimize potential risks that can occur from inaccurate air pollutant models. Frequently, models are used to provide a more robust analysis of air pollution (e.g., O'Leary and Lemke, 2014) but outliers within the dataset can substantially skew model output. Therefore, accurate measurements, proper assumptions, and appropriate corrections are needed to ensure air pollutant datasets constitute a relevant tool for informed decision making.

Generically, an outlier in statistics is one that appears to deviate markedly from other members of the sample group (Barnett and Lewis, 1994). Outliers are identified by comparing the values in question to the rest of the distribution or a subset of the distribution (Torres et al., 2011). Two types of spatial outliers are recognized by differences in magnitude relative to surrounding points (**Figure 1.1**):

Global spatial outliers are statistical outliers if their value is unlike the values in the entire distribution within a geographically defined study area. These data points may be higher or lower than the other data points.

Local spatial outliers differ from the data points immediately surrounding them. These points may not be extremes within the entire distribution but stand out when compared to other data within their local geographic neighborhood.

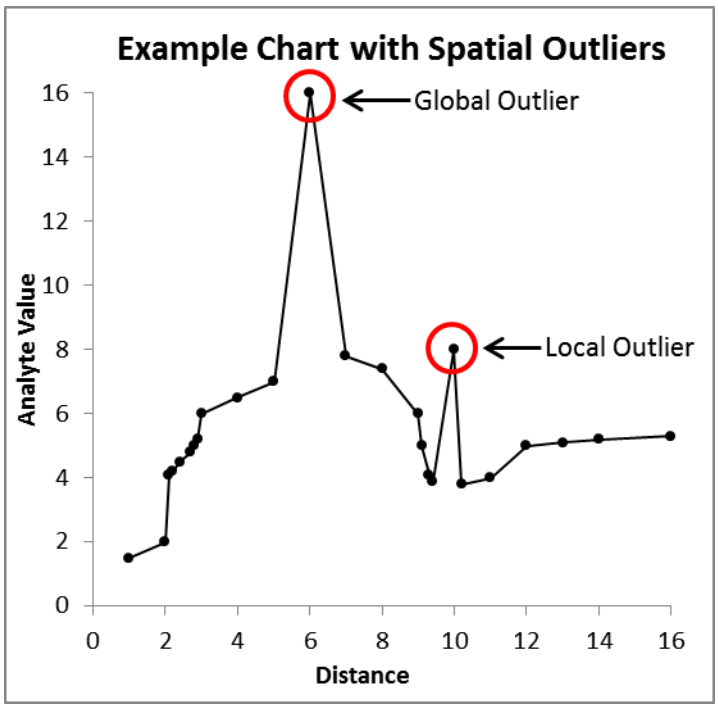


Figure 1.1. Example chart of measurements made along a linear profile illustrating a global outlier and a local spatial outlier.

For the purposes of this thesis, all spatial outliers, including global and local spatial outliers, are referred to simply as outliers.

With new tools and technology, spatial data are more assessable and manageable leading to an increase in spatial outlier research. At the Joint Research Center (JRC) in Europe, geostatistical tools have been used to correct for air pollutant outliers and identify air quality indicators (Kracht et al., 2013). The JRC includes over 6,000 air monitoring sites, and outliers are identified using a low pass filter to remove high concentrations while preserving low frequencies. The Moran's I is used frequently to identify outliers (e.g., McGrath and Zhang, 2003; Walker et al., 2013; Zhang et al., 2008; Zhang and McGrath, 2004; Zou et al., 2014). In Ireland, a number of studies used the Moran's I method to identify spatial outliers when mapping the soil organic carbon (McGrath et al., 2003, Zhang and McGrath, 2004, and Zhang et al.,

2008). Walker et al. (2013) also used the Moran's I to detect spatial outliers in geomorphic changes to sand dunes in Western Canada. Clougherty et al. (2013) took a different approach in New York City and identified outliers as air quality samplers that were ± 3 standard deviations away from the mean.

1.5 Study Motivation and Objectives

The motivation for this thesis stems from a desire to better quantify correlations between ambient air pollution concentrations and asthma exacerbations and other health outcomes in Detroit and Windsor. Previously, high spatial resolution correlations were established in the GeoDHOC pilot study (Lemke et al., 2013). This thesis investigation expands the assessment between ambient air pollution and asthma exacerbations by 1) reassessing the GeoDHOC spatial models for outliers and 2) incorporating temporal data into the original and subsequent models.

Specifically, the first objective is to determine if outliers are present in the datasets and, if so, quantify the magnitude of their impact on modeled spatial pollution distributions. Spatial outliers, both high and low in magnitude, can substantially influence ordinary kriged models and lead to potentially inaccurate pollutant concentration estimates across portions of the airshed.

The second objective is to incorporate temporal data into the modeling of the air pollutant maps. In the initial GeoDHOC study, Lemke et al. (2013) did not include temporal information in their analysis, relying instead on a two-week air pollution model to estimate exposures uniformly throughout 2008 in Detroit and Windsor. This thesis applies the methodology of O'Leary and Lemke (2014) to extrapolate the September 2008 and June 2009 GeoDHOC datasets and create a time series of monthly concentration maps for 2008.

The final objective of this study is to determine how a) the correction for outliers and/or b) increased temporal model resolution influence 2008 asthma associations. The associations were evaluated on the same spatial scale as Lemke et al. (2013) and used the same linear regression technique to evaluate asthma correlations with the newly created models from objectives 1 and 2.

1.6. Hypotheses

The heterogeneous nature of urban air complicates human exposure estimates and creates a need for accurate, highly detailed spatiotemporal air contaminant models. This investigation evaluated two hypotheses while developing improved models of the Detroit-Windsor airshed by identifying spatial data outliers and incorporating temporal trends to better define neighborhood level air contaminant concentrations.

Hypothesis #1:

Prior GeoDHOC models indicate areas of high concentrations that are inconsistent with regional concentration trends. The 2008 and 2009 GeoDHOC datasets and the MASN Detroit datasets were re-evaluated using statistical software including ArcGIS, Surfer, SGeMS, and SpaceStat to identify global and local spatial outliers. It was hypothesized that spatial models with outliers removed will improve health correlations for asthma exacerbations.

Hypothesis #2:

The GeoDHOC asthma study relied upon a two-week sampling period in September 2008 to estimate annual air pollutant concentration levels. This thesis utilized MASN data along with the GeoDHOC June 2009 data to temporally resolve air pollutants throughout 2008 and 2009. It

was hypothesized that pollutant models with increased temporal resolution will also improve health correlations for asthma exacerbations.

1.7. Thesis Structure

This thesis comprises five chapters. Chapter 1 introduced the need to spatial data outlier identification and spatiotemporal models along with previous GeoDHOC studies.

The datasets used for this study are described in Chapter 2. Chapter 3 documents the methodology used for identifying spatial data outliers, temporal scaling, and asthma correlations. Results are reported in Chapter 4. A discussion of the results, limitations, conclusions, and recommendations of future research are presented in Chapter 5.

CHAPTER 2. DATA

2.0 Introduction

Air pollution, meteorological data, and health information for this study were derived from a number of sources. Air pollution data came from two sources. The first was a spatially-distributed set of air pollution measurements developed by the Geospatial Determinants of Health Outcomes Consortium (GeoDHOC) (Lemke et al., 2013; Miller et al., 2010). The second dataset consisted of time series measurements of air pollutants from the Michigan Air Sampling Network (MASN) which is operated by the Air Quality Division of the Michigan Department of Environmental Quality (MDEQ). Meteorological data were obtained from the National Oceanic and Atmospheric Administration (NOAA). Asthma information was collated from the Henry Ford Health System (HFHS) in Detroit and the Canadian Institute for Health Information (CIHI) in Windsor (Lemke et al., 2013). Each of these datasets is described in more detail below.

2.1 GeoDHOC

The GeoDHOC conducted two, two-week air sampling campaigns in Detroit, Michigan and Windsor, Ontario between September 5-20, 2008 and May 29–June 13, 2009. A total of 100 passive samplers and 50 active samplers were deployed in 2008. An identical sampling plan was implemented in 2009 with additional passive samplers totaling 133 passive samplers and 50 active samplers (**Figure 2.1**).

Sites with a combination of active and passive samplers are designated by codes including the letter 'A' and the station number (e.g., D-A-21 or W-A-4 in Detroit and Windsor, respectively) (**Figure 2.2**). Passive samplers measured NO_2, SO_2, and volatile organic compound (VOC) concentrations at an approximate spatial density of 5 km^2 per sample (**Figure**

2.2). BTEX compounds comprised 64% and 72% of total VOCs measured in 2008 and 2009, respectively. Active samplers measured polycyclic aromatic hydrocarbons (PAHs) and particulate matter (PM) in three size fractions (PM_1, $PM_{1\text{-}2.5}$, and $PM_{2.5\text{-}10}$) at an approximate spatial density of 10 km^2 per sampler throughout both cities.

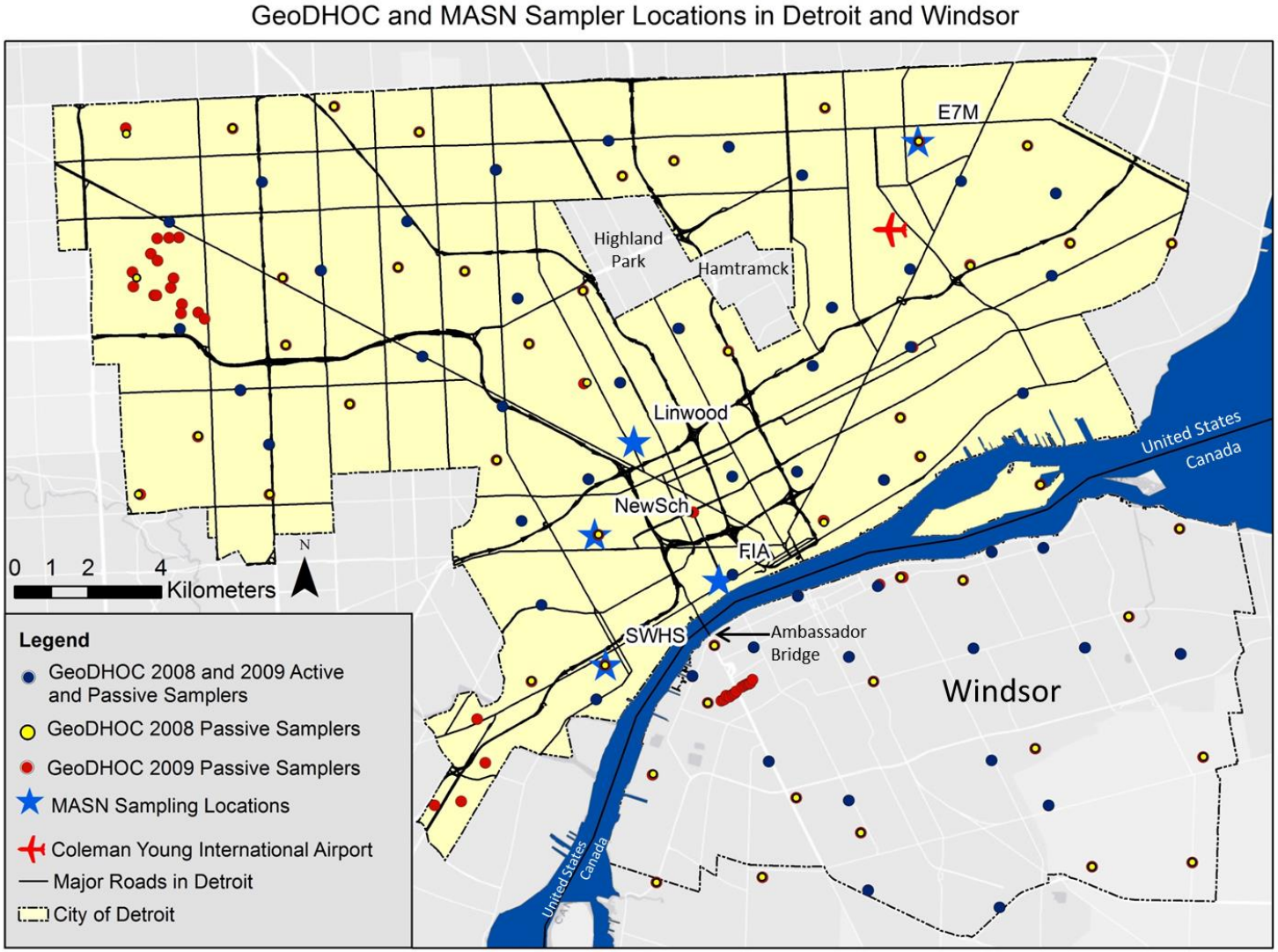


Figure 2.1. GeoDHOC study air sample locations in Detroit and Windsor. MASN site name abbreviations given in Table 2.1.

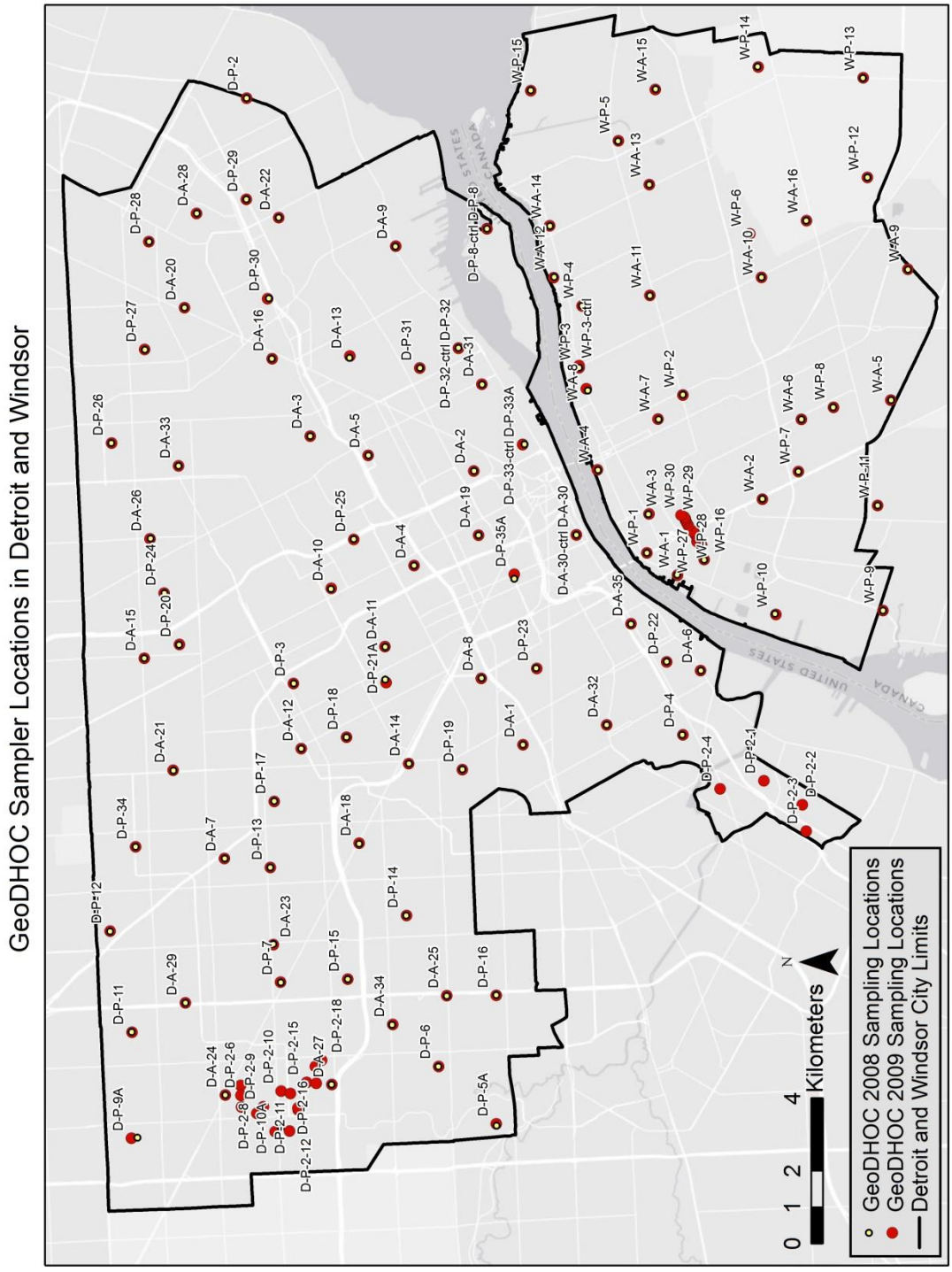


Figure 2.2. GeoDHOC sampling locations for Detroit and Windsor with station designations.

The pollutant measurement datasets differed slightly in the number of points associated with each analyte. NO_2 had 97 sampling points in 2008 and 122 sampling points in 2009. BTEX and VOCs had 98 and 133 sampling points in 2008 and 2009, respectively. The greater number of samplers for these analytes in 2009 arose from the inclusion of additional sampling points in northwest and southwest Detroit, as well as the inclusion of a profile perpendicular to Huron-Church road on the approach to the Ambassador Bridge in Windsor. $PM_{2.5}$, PM_{10}, and PAHs had 38 sampling points in 2008 and 37 sampling points in 2009. Implementation of quality control/quality assurance procedures (Miller et al., 2010), resulted in the exclusion of different points from the 2008 and 2009 datasets for all of the analytes listed above.

Pollutant distribution models were created using ordinary kriging with a 300m x 300m grid spacing. Details of sampling, QA/QC and mapping methods for the GeoDHOC data set are given by Miller et al. (2010). These models demonstrated neighborhood-scale spatial variability of air pollutants within the Detroit-Windsor airshed during the 2008 (Miller et al., 2010) and 2009 (O'Leary and Lemke, 2014) sampling periods. Kriging variance maps illustrating the spatial distribution of estimation uncertainty in the Miller et al. (2010) and O'Leary and Lemke (2014) maps are provided in the appendices (**Figure A.1**).

2.2 MASN

The second air pollution data set consisted of time series measurements at five MASN locations within the city of Detroit from 2008-2010 (**Figure 2.1**). Not all analytes were measured at each location (**Table 2.1**). Measurements at two nearby MASN locations outside the city (Allen Park and Dearborn) were excluded from the study because $PM_{2.5}$ and PM_{10} measurements at these sites did not differ materially in temporal trends from the Detroit station

measurements during the period examined. Hence, only MASN samplers located in the city of Detroit were included in the study. Measurements at two National Air Pollution Surveillance (NAPS) monitoring stations in Windsor were also excluded because they are outside Detroit, the refined study area for temporal analyses described in Section 3.2 of this thesis.

The Detroit MASN data set includes single sampling locations for NO_2, VOCs including individual BTEX components, and PM_{10} (**Table 2.1**). The East 7 Mile location was the only active NO_2 sampling location in the study area during the study period. At this location, NO_2 was sampled continuously using automated chemiluminescence (Federal Reference Method (FRM) RFNA-0179-035) (U.S. EPA, 2013) and hourly concentrations were reported. BTEX concentrations at Southwestern High School were derived from air samples collected over a 24 hour period every 12 days using SUMMA canisters. These samples were analyzed for VOCs using gas chromatography/mass spectrometry following EPA method TO-15 (U.S. EPA, 1999). PM_{10} concentrations were measured at the Southwestern High School site over a 24 hour period every six days using a High-Volume Air Sampler (FRM RFPS-1287-064) (U.S. EPA, 2013).

Table 2.1. Michigan Air Sampling Network (MASN) monitoring sites in Detroit.

Site Name	Abbreviation	Location	Analyte	Method	Sampling Frequency	Sample/report Duration
East Seven Mile	E7Mile	northeast Detroit	NO_2	FRM 35	Continuous	1 hour
			$PM_{2.5}$	FRM 118	3 days	24 hours
Linwood	Linwood	central Detroit	$PM_{2.5}$	FRM 118	1 to 3 days	24 hours
Newberry School	NewSch	south central Detroit	$PM_{2.5}$	FRM 118	3 days	24 hours
FIA / Lafayette St.	FIA	south central Detroit	$PM_{2.5}$	FRM 118	3 days	24 hours
Southwestern High School	SWHS	southwest Detroit	$PM_{2.5}$	FRM 118	3 days	24 hours
			PM_{10}	FRM 64	6 days	24 hours
			VOCs	EPA TO-15	12 days	24 hours

$PM_{2.5}$ was measured at five Detroit MASN sampling locations during the 2008-2010 study period (**Table 2.1**). $PM_{2.5}$ was measured over a 24 hour period using a $PM_{2.5}$ Sequential Air Sampler (Rupprecht & Patashnick Company, Incorporated Partisol®-Plus Model 2025, FRM RFPS-0498-118) (U.S. EPA, 2013). $PM_{2.5}$ was sampled every 3 days at each site, with the exception of the FIA/West Lafayette Street site where daily samples were available after October 1, 2009.

2.3 Meteorological Data

Weather data from the National Oceanic and Atmospheric Administration (NOAA) were obtained for the Coleman A. Young International Airport, located in northeast Detroit (**Figure 2.1**). These data provided hourly observations of wind speed, wind direction, temperature, and precipitation for 2008 (National Climatic Data Center, 2014). Monthly average wind rose plots (**Figure 2.2**) indicate prevailing westerly winds throughout much of the year, although greater variability in wind direction and speed was generally present in the spring (March through May) and late summer (August through September) **(Figure 2.2)**.

2.4 Asthma Data

Asthma data for this study were derived directly from the GeoDHOC analysis reported by Lemke et al. (2013). In that study, asthma events were defined as hospital admissions and emergency department visits with a primary diagnosis of asthma. In 2008, approximately 2,800 asthma hospitalizations and emergency room visits were reported in the Henry Ford Health System (HFHS) patient database for Detroit. During the same year, approximately 650 comparable asthma events were reported in the Discharge Abstract Database (DAD) and the

National Ambulatory Care Reporting System (NACRS) in Windsor (Lemke et al., 2013). These counts, normalized to number of asthma events per 1000 patients in each zip code tabulation area in Detroit and each postal forward sortation area in Windsor, constitute the spatially distributed asthma frequencies used in this thesis.

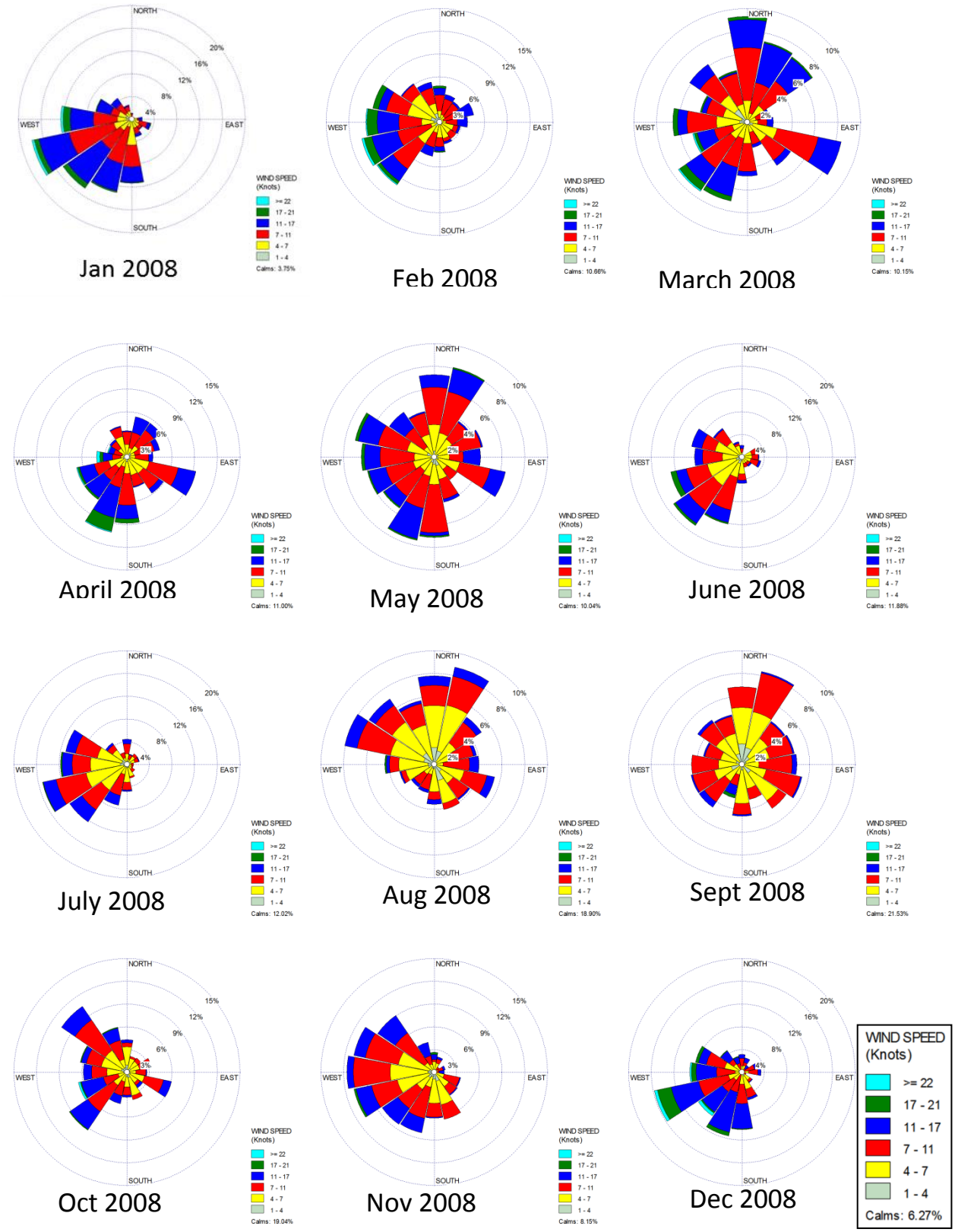


Figure 2.3. Monthly windrose graphs for the Coleman Young International Airport in Detroit for 2008. The scale is consistent for each month.

CHAPTER 3. METHODS

3.0 Introduction

This study consists of two major investigations, spatial outlier analysis (hypothesis #1) and temporal scaling (hypothesis #2), to address the relationship between asthma exacerbations and air pollution in Detroit and Windsor. The spatial outlier analysis utilized a multistep process to assess outliers in the Geospatial Determinants of Health Outcomes Consortium (GeoDHOC) datasets for September 2008 and June 2009. Potential outliers were initially identified using four different outlier identification methods and then refined to a set of spatial data outliers based on a convergence of these methods. New ordinary kriged spatial models were subsequently derived from the resulting air pollution datasets with spatial data outliers removed. The second effort incorporated Michigan Air Sampling Network (MASN) time series air pollutant concentration measurements. The original GeoDHOC models and the revised models with outliers removed were temporally scaled by incorporating the MASN time-based dataset. The resulting air pollutant concentration models were then correlated with acute asthma events in Detroit and Windsor to reassess the impact of air pollution on asthma and test the two hypotheses of this study (Section 1.6).

3.1 Outliers

Four independent methods were used to identify potential outliers that were subsequently reassessed using a combination of the methods to select a final set of spatial outliers. These include box plots, difference maps, variogram clouds, and the Moran's I, each of which takes into account different aspects of individual pollutant measurements relative to surrounding measurements. The box plot and Moran's I approach are quantitative means of assessing spatial data outliers whereas the difference maps and variogram clouds are more qualitative

assessments. Analytes examined for outliers included nitrogen dioxide (NO_2), total benzene, toluene, ethylbenzene, and xylene (BTEX), total volatile organic compounds (VOCs), mass of particulate matter less that 2.5 microns in aerodynamic diameter ($PM_{2.5}$), mass of particulate matter less than 10 microns (PM_{10}), and polycyclic aromatic hydrocarbon (PAH).

Both global and local outliers (Section 1.4) were considered. Global outliers, which are values unlike the rest of the entire distribution, were determined using box plots and difference maps. In contrast, local outliers, which are different from nearby and adjacent points, were identified using difference maps and Moran's I and confirmed using variogram clouds. Because combinations of global and local outlier identification methods were used, the outliers were all grouped together.

SpaceStat (Biomedware, Inc.) was used to create boxplots, variogram clouds, and Moran's I analysis charts. SpaceStat links each of its charts and plots in a common dataset that enables direct graphical comparisons among the different methods employed in this study. The difference maps were generated using Surfer 11 (Golden Software). SpaceStat was used for the variography and to create the ordinary kriged models with the outliers removed. The final map versions of the difference maps and new models were created in ArcMap 10.0 (ESRI).

3.1.1 Box Plots

Box plots are aspatial representations illustrating the spread of concentration distributions. The box plots created for this study are a statistical visualization of the range of each dataset relative to the median and interquartile range (IQR) **(Figure 3.1)**. The black box in the center of the graph represents the IQR and the line in the middle is the median. The "whiskers" (the

horizontal lines above and below the box), represent +/- 1.5 times the IQR beyond the box boundaries **(Figure 3.1)**.

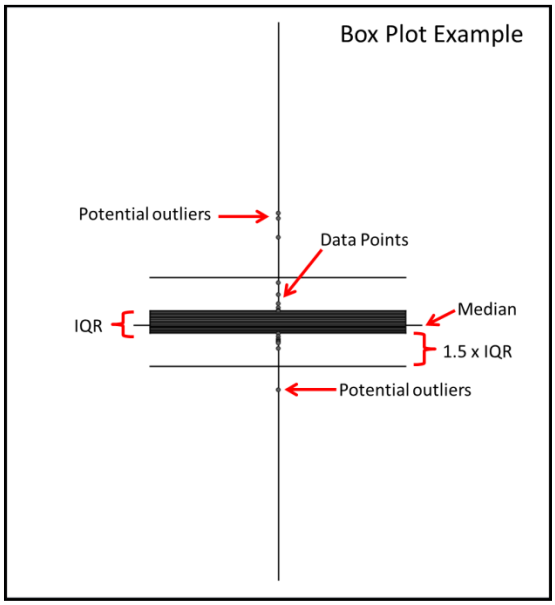


Figure 3.1. Example Box Plot with labels.

This study treated points located outside of the whiskers as potential outliers. Histogram charts were generated to confirm that these points were located on the extreme ends of the distribution for each dataset.

3.1.2 Variogram Clouds

Variogram clouds are exploratory plots of dataset spatial variability, in which each point represents the dissimilarity (square root of the absolute difference) between any two measurement locations as a function of their Euclidean distance. Each variogram cloud takes the entire dataset into account and is useful for identifying local variability. Points that fall in the upper left corner demonstrate a high degree of dissimilarity across short separation distances **(Figure 3.2)**. Conversely, points in the bottom right show little dissimilarity over long distances.

Points that are close together are expected to display more similarity than points separated by a greater geographical distance (Isaaks and Srivastava, 1989). Consequently, pairs of points displaying high dissimilarity over a geographically small distance were initially considered potential spatial data outliers for this study. However, selecting all of the points in the top left was found to be too arbitrary. The box plot (aspatial technique) was therefore used to identify extreme values in conjunction with the variogram clouds (spatial technique) to identify outliers. If the resulting highlighted points identified through the box plot analysis showed a large degree of dissimilarity in a short geographical distance on the variogram cloud **(Figure 3.2)**, the sampling location was considered a potential spatial outlier because it registered as both a global (box plot) and local (variogram cloud) anomaly.

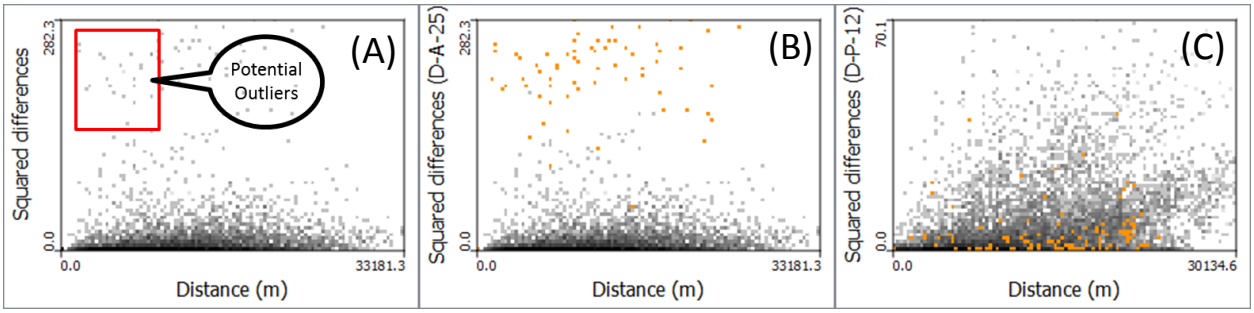


Figure 3.2. Example variogram clouds. Graph A shows an example variogram cloud with a box representing a region with potential outliers. Graph B highlights points associated with a single sampler location (identified on a corresponding box plot for the same dataset) that is considered to be a potential outlier. Graph C highlights points associated with a single sampler location that is not considered to be a potential outlier.

3.1.3 Difference Maps

Difference maps quantify the change in pollutant measurements between the two GeoDHOC sampling periods in September 2008 and June 2009. They were generated using grid arithmetic by subtracting the September 2008 GeoDHOC ordinary kriged model grids from the June 2009 ordinary kriged model grids for each analyte**:**

$$Difference\ Map = (GeoDHOC\ June\ 2009\ OK\ model) - (GeoDHOC\ Sept.\ 2008\ OK\ model) \quad \text{(Eqn. 3.1)}$$

Isoconcentration lines were added to each difference map to help highlight large concentration changes over short distances.

Two different approaches were used when identifying potential outliers based on the difference maps. The first approach assessed the concentration difference, both positive and negative, between the two sampling periods. Major differences typically generated 'bulls eye' shaped isoconcentration lines centered on a single sampler location **(Figure 3.3)**. The September 2008 sampling locations and June 2009 sampling locations were assessed individually because the sampling locations varied from year to year (Section 2.1). Sampling points that demonstrated a large difference between the two sampling years were considered potential outliers.

The second approach assessed abnormally high or low model values adjacent to, but offset from, individual sampling locations. Some modeled grids included high or low concentration values that were markedly different from nearby sampling locations. These features can result as artifacts from gridding and contouring algorithms when closely spaced control points have markedly different values. The resulting 'bulls eye,' which is shifted away from an actual sampling location, constitutes a potentially inaccurate model estimate. For example, in **Figure 3.3**, the lower red box shows a zone in Windsor where the sampler location, black dot, is located

to the left of the center contour circle. The locations near the sampling point were determined to be inaccurately modeled and the associated sampler was identified as a potential spatial data outlier.

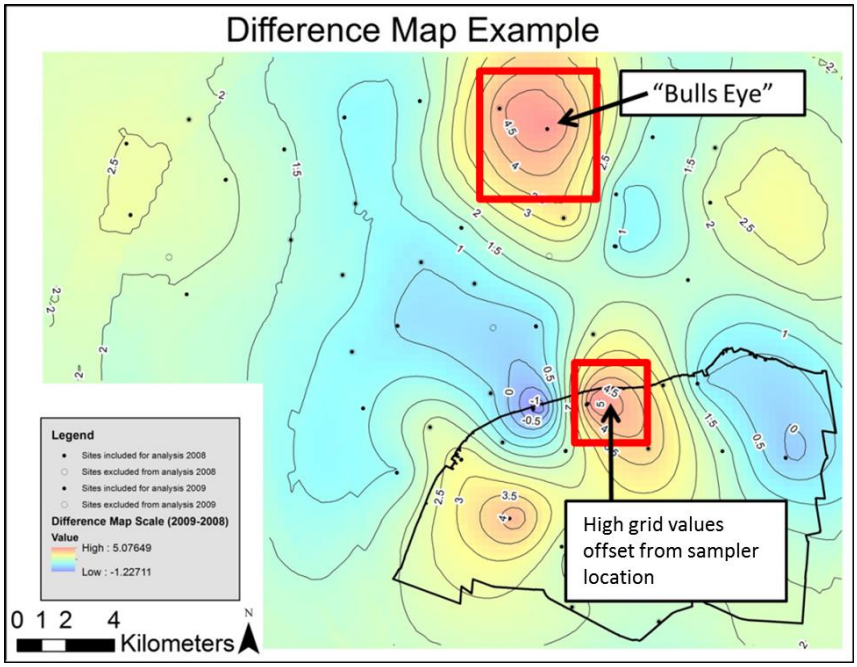


Figure 3.3. Example difference map illustrating different approaches to spatial outlier identification.

3.1.4 Local Moran's I

The Local Moran's I is a weighted correlation coefficient that quantifies spatial randomness for each location in the dataset. Spatial patterns are evaluated by the Moran algorithm for each point measurement using a set number of surrounding points. In this study, eight surrounding points were utilized. Locations that deviate from spatial randomness demonstrate a specific spatial pattern. Spatial patterns include clustering of high or low values,

or, alternatively, individual high or low values surrounded by samplers with consistent, but contrasting (low or high) values, respectively (**Figure 3.4**).

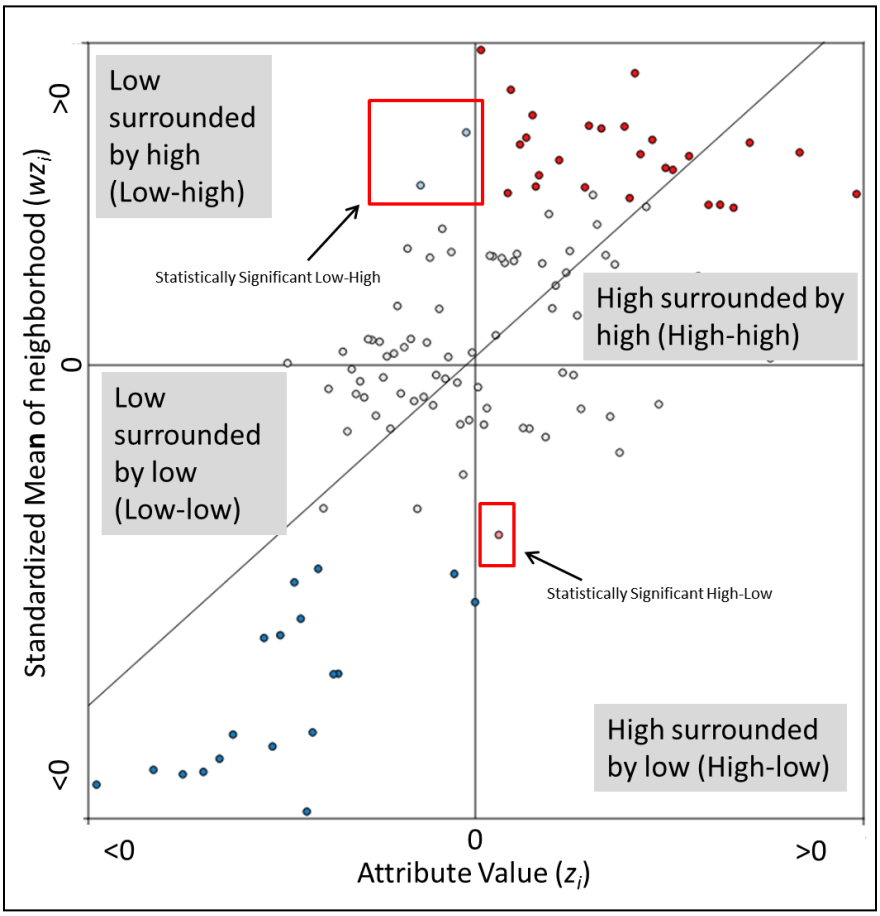


Figure 3.4. **Local Moran's I scatter plot example with high-low and low-high points highlighted. Positive I values are found in the upper right-hand and lower left-hand quadrants of this chart. Potential outliers, indicated by statistically significant negative I values, appear as light blue or pink dots in the upper left and lower right quadrants and are highlighted by the boxes on this example chart.**

The Local Moran's I function assumes a normal population to evaluate spatial distributions with spatial randomness representing the null hypothesis. The Jarque-Bera test was used to test for normality (Kiefer and Salmon, 1983) by examining the GeoDHOC dataset for each analyte in Sept. 2008 and June 2009. If the dataset generated a p-value equal to or less than 0.05, it was determined to be non-normal and, consequently, a normal score transform was performed. Subsequently, the Moran's I was calculated in SpaceStat using the following equation:

$$I_i = z_i \sum_{j=1}^{n} w_{i,j} z_j \, , i \neq j \qquad \text{(Eqn. 3.2)}$$

where *I* is the Moran's I coefficient, *z* is the z score, *w* is a weighting parameter applied to the neighbors, *i* identifies the sampler point being evaluated, *j* is an index value corresponding to each of the nearby neighbors, and *n* is the number of neighbors (Anselin, 1995). A value of zero indicates no spatial autocorrelation. Positive I_i values indicate that there is either a cluster of low or high values. Negative I_i values indicate that high and low values are clustered together (**Figure 3.4**). Consequently, potential outliers were identified as sampling points with a negative I_i value along with a p-value less than or equal to 0.05. These represent either sampling points with higher values than the surrounding sampling locations (high-low) or lower values that the surrounding sampling location (low-high).

Parameters of the Local Moran's I include number of randomizations, point adjacency method, neighbor weight method, and Simes Correction. This study used 999 Monte Carlo simulations to derive a p-value for each sampler location. The point adjacency method was set to the nearest 8 neighbors. This ensured that I_i only reflected the immediate surrounding sampler locations. The neighbor weight method was standardized to neighbor count.

The Simes correction was used to adjust p-values for the Local Moran's I calculations. Local clustering values are not independent because the Moran's I test statistics are derived from repeat trials drawn from the same dataset, and neighboring areas overlap, which makes the test statistics correlated. The Simes adjustment is defined as:

$$p_i' = (n + 1 - a)p_i \qquad \text{(Eqn. 3.3)}$$

with n as the number of p-values (number of neighbors + central location) and a as the index value starting at 1 and describing the location in the vector of p-values (number of neighbors + central location).

3.1.5 Outlier Determination

A three step process was implemented to select spatial data outliers. 1) Potential outliers were identified from the global dataset using each of the four methods described above. 2) Potential outliers identified only by a single method were disregarded. 3) Each remaining potential outlier was then individually assessed using a combination of the four methods to determine a final set of outliers for subsequent analysis in this study. Variogram clouds were plotted in conjunction with points identified on box plots to evaluate pairs of points associated with extreme values. Sampler locations identified with box plot analysis and Moran's I were plotted on the difference maps to facilitate visual comparison of those three identification methods. Sampler locations identified using Moran's I were also evaluated using variogram clouds to assess agreement for local outliers.

3.1.6 Kriging

After spatial data outliers were identified for each analyte, they were removed from the Miller et al. (2010) and O'Leary and Lemke (2014) datasets and remodeled using ordinary kriging. Kriging is a geostatistical method that uses a weighted linear regression to interpolate values at unsampled locations. There are numerous types of kriging that include ordinary kriging, simple kriging, universal kriging, and indicator kriging (Webster and Oliver, 2007). Ordinary kriging was employed, which, like the other kriging methods, is considered a 'best linear unbiased estimator' (B.L.U.E.) (Isaaks and Srivastava, 1989). It is 'linear' because of its use of weighted linear regressions. It is 'unbiased' by setting the mean residual (error) as close to zero as possible. And it is 'best' by minimizing the estimation error variance. Ordinary kriging distinguishes itself from other forms of kriging by assuming a constant local mean in the search neighborhood of the estimation point (Isaaks and Srivastava, 1989). This offers greater flexibility to incorporate spatial variability than other kriging techniques such as simple kriging which assumes a constant global mean (Isaaks and Srivastava, 1989) or kriging with a trend, which assumes an *a priori* model of spatial variation (Goovaerts, 1997). Kriging requires a covariance model to quantify variance as a function of separation distance for each variable of interest. Variogram models are generated through a process called variography.

A variogram (or, more strictly, a semi-variogram) measures variance or covariance between one or more values measured at distinct locations as a function of the distance between the locations. When generating an initial model for kriging, an 'experimental variogram' is used. For this study, experimental semi-variograms were defined as half the mean squared difference between values separated by a defined distance or 'lag' (Isaaks and Srivastava, 1989):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)} (v_i - v_j)^2 \qquad \text{(Eqn. 3.4)}$$

where γ is the variance at lag increment *h*, *N(h)* is the total number of data pairs separated by lag *h*, and v_i and v_j are the attribute values of a single pair at locations *i* and *j*.

In contrast to the variogram cloud introduced in section 3.1.3, in which every pair of points is shown, experimental variograms display the average variance summed over multiple pairs of points at specified separation distances. Defining a lag distance with an appropriate tolerance to group pairs reduces the noise (contained in the variogram cloud) and enables a covariance structure to emerge. A continuous variogram model is then fit to the experimental variogram to define the covariance structure. Two common types of variogram models were used in this study:

Spherical $$\gamma(h) = c\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right] \qquad \text{(Eqn. 3.5)}$$

Exponential $$\gamma(h) = c\left[1 - \exp\left(-\frac{3h}{a}\right)\right] \qquad \text{(Eqn. 3.6)}$$

where γ is the variance at lag increment *h*, *c* represents the sill contribution, and *a* represents the range of the spherical model or the effective range (distance where γ reaches 95% of the sill contribution) in the exponential model (Goovaerts, 1997). These variogram models were used in conjunction with a nugget effect accounting for measurement uncertainty to interpolate over the entire study area with kriging estimates.

3.1.7 Kriging Parameters

Ordinary kriging was used to model the newly created GeoDHOC air pollution datasets with outliers removed for the Detroit and Windsor airshed. Models were created for the September 2008 and June 2009 sampling periods. The original September 2008 (Miller et al.,

2010) and June 2009 (O'Leary and Lemke, 2014) ordinary kriged GeoDHOC models are henceforth referred to as Model 1 **(Table 3.1).** The subsequent ordinary kriged maps with outliers removed are referred to as Model 2 **(Table 3.1)**.

In formulating new variogram models and kriging parameters, an attempt was made to adhere to the parameters of Model 1 (original models) as much as possible (**Table 3.2**). However, several parameters required changes. The lag distance and lag tolerance were kept the same as previous models (1000 meters and 500 meters, respectively). The lag count was increased to 20. The nugget effect, or discontinuity at the origin, was kept the same for NO_2, BTEX, and VOCs. The ordinary kriging search radius was set to 10,000 meters for passive sampler analytes (NO_2, BTEX, and VOCs). For active sampler analytes (PM and PAHs), the search radius was set to 20,000 meters because of the smaller number of samplers distributed throughout the same study area. Sill contributions were adjusted from Model 1 to Model 2 to account for the decreased variance observed when extreme values identified during the outliers analysis were excluded, subsequently reducing the range of observed values for each analyte.

The original particulate matter models (Miller et al., 2010) were developed in the mass increments in which they were measured during GeoDHOC sampling: PM_1, $PM_{1\text{-}2.5}$, and $PM_{2.5\text{-}10}$. Subsquent PM mass concentration models reported by Lemke et al. (2013) included the sumation of incremental particulate matter to derive mass concentration distributions for $PM_{2.5}$ and PM_{10} , two regulated PM fractions in the US and Canada. To facilitate comparison of PM-asthma correlations, PM_1, $PM_{2.5}$, and PM_{10} summations were used in this thesis. Consequently, the variogram models generated for $PM_{2.5}$ and PM_{10} in Model 2 were distinctly different from Model 1 (**Table 3.2**).

Table 3.1. Summary of the four models used in this study.

Model	Description	Source
Model 1	Ordinary kriged models for the Detroit-Windsor Airshed in September 2008 and June 2009 using the original GeoDHOC dataset.	Miller et al. (2010) and O'Leary and Lemke (2014)
Model 2	Ordinary kriged models for the Detroit-Windsor Airshed in September 2008 and June 2009 using the GeoDHOC dataset with spatial outliers removed.	Master's Thesis
Model 3	Model 1 with temporal scaling for 12 months in 2008. This model is only applicable to Detroit.	O'Leary and Lemke (2014)
Model 4	Model 2 with temporal scaling for 12 months in 2008. This model is only applicable for Detroit.	Master's Thesis

Table 3.2. Variogram and ordinary kriging parameters.

	Dataset Information			Experimental Variogram Parameters							Variogram Model and Kriging Parameters				
	Pollutant	Year	Sites	Lag Count	Lag (m)	Tolerance (m)	Angle Count	Start Angle	Angle Lag	Angle Tolerance	Model	Nugget	Sill Contribution	Range (m)	Search Radius (m)
Model 1 Parameters	Parameters for the original ordinary kriged GeoDHOC models in 2008														
	NO2	2008	all	16	1000	500	1	0	180	90	sph	0.75	9.6	15200	30000
	BTEX	2008	all	16	1000	500	1	0	180	90	sph	0.57	15	10000	30000
	VOC	2008	all	16	1000	500	1	0	180	90	exp	2.05	42	30000	30000
	PM1	2008	all	16	1000	500	1	0	180	90	sph	0.1	0.57	10000	30000
	PM 1-2.5	2008	all	16	1000	500	1	0	180	90	sph	1	1.2	6000	30000
	PM 2.5-10	2008	all	16	1000	500	1	0	180	90	sph	1	9	27000	30000
	Parameters for the original ordinary kriged GeoDHOC models in 2009														
	NO2	2009	all	16	1000	500	1	0	180	90	sph	1.3	12.6	12000	48000
	BTEX	2009	all	16	1000	500	1	0	180	90	sph	0.049	4.8	18500	55500
	VOC	2009	all	16	1000	500	1	0	180	90	sph	0.24	7.85	16500	49500
	PAH	2009	all	16	1000	500	1	0	180	90	sph	207	477	8500	25500
	PM1	2009	all	16	1000	500	1	0	180	90	sph	0.15	0.85	8000	40800
	PM 1-2.5	2009	all	16	1000	500	1	0	180	90	sph	0.13	0.12	8000	27000
	Pm 2.5-10	2009	all	16	1000	500	1	0	180	90	sph	0.23	2.09	11000	30000
Model 2 Parameters	Parameters for the new ordinary kriged GeoDHOC models with spatial data outliers removed for both the 2008 and 2009 datasets.														
	NO2	2008	All	20	1000	500	1	55	180	90	exp	0.75	8	15000	10000
	NO2	2009	All	20	1000	500	1	60	180	90	sph	1.3	11	16000	10000
	BTEX	2008	All	20	1000	500	1	35	180	90	exp	0.57	11	11500	10000
	VOC	2008	All	20	1000	500	1	38	180	90	exp	2.1	21	16000	10000
	PAH	2009	All	20	1000	500	1	50	180	90	sph	25	200	10000	20000
	PM2.5	2008	All	20	1000	500	1	53	180	90	sph	0.2	1.35	11000	20000
	PM2.5	2009	All	20	1000	500	1	55	180	90	sph	0.5	0.5	7000	20000
	PM10	2008	All	20	1000	500	1	50	180	90	sph	0.03	11.7	13000	20000
	PM10	2009	All	20	1000	500	1	50	180	90	sph	1.5	4.5	10000	20000

3.2 Temporal Scaling

Both the original Miller et al. (2010) and O'Leary and Lemke (2014) ordinary kriged GeoDHOC models (Model 1), as well as the subsequent models with outliers removed (Model 2) were temporally scaled to generate a series of monthly air pollutant maps for 2008 following the approach described by O'Leary and Lemke (2014). This process integrated the spatially-rich GeoDHOC datasets with the temporally-rich MASN datasets for NO_2, total BTEX, $PM_{2.5}$, and PM_{10}. The temporally scaled models inclusive of outliers are collectively referred to as Model 3; whereas the temporally scaled models exclusive of outliers are collectively referred to as Model 4 **(Table 3.1)**. Published asthma data were available only for 2008. Consequently the effects of temporal scaling process could be assessed only for the twelve months during the 2008 time period.

The process of integrating MASN and GeoDHOC data sets began with a comparison of their respective measurements and model estimates. Cross validation was initially applied to assess the point-wise accuracy of modeled estimates for both Model 1 and Model 2. Subsequently, a two-step process was employed to develop a set of spatially and temporally interpolated concentration maps for NO_2, total BTEX, $PM_{2.5}$ and PM_{10} across the study area. Temporal data were restricted to Detroit in the O'Leary and Lemke (2014) study. Consequently, temporal scaling of the GeoDHOC spatial models only included areas inside the municipal boundaries of Detroit and the enclosed municipalities of Hamtramck and Highland Park.

Initially, existing GeoDHOC maps for each pollutant were combined using a weighted average scheme to produce a series of spatially interpolated maps for twelve consecutive months. Subsequently, temporal trends derived from MASN time series measurements were superimposed on the monthly maps to generate a series of 12 monthly models spanning the

period from January 2008 through December 2008 for each pollutant. Originally, a 36-month period (2008-2010) was modeled in support of the perinatology study (O'Leary and Lemke, 2014) but this thesis is limited to 2008 to coincide with the period of asthma data availability. Computations were performed using Surfer 11.4, ArcMAP 10.0, and SpaceStat 3.2.20 software.

3.2.1 Comparison of Measurements and Model Estimates

GeoDHOC samplers were either collocated or placed in close proximity to three of the MASN samplers. These included the East 7 Mile location, Southwestern High School, and West Lafayette Street/FIA sites (**Figure 2.1**). Concentrations measured by the individual GeoDHOC samplers at these three sites were compared to MASN measurements from the corresponding time periods to assess agreement between the two data sets. In addition, the spatial and temporal variability of the GeoDHOC and MASN concentration data sets were compared.

In the cross validation process, individual measured values for each pollutant were removed and re-estimated with ordinary kriging using the remaining observations. The distribution of estimation error at sampled locations was examined for magnitude, bias, and independence.

3.2.2 Spatial Modeling

Model 1 and Model 2 served as the anchor points for each pollutant during spatial interpolation. These maps, which model the spatial variability of measurements integrated over continuous two-week sampling periods across the Detroit-Windsor airshed, were assumed to be representative of the spatial distribution during the month in which samples were collected (i.e., September 2008 and June 2009). A weighted average was applied to construct spatially

distributed concentration models for the eight months between September 2008 and June 2009. Weighting factors for each month were assigned using fractions of the nine months separating the anchor months based on proximity in time to each anchor month. For example, the October map was blended with an 8/9 weighting for September plus 1/9 weighting for June at each point in the 300 m x 300 m model grid. Similarly, November combined 7/9 September with 2/9 June, and so on. This procedure was implemented independently for each of the four pollutants mapped.

July 2009 and August 2009 spatial models were also constructed using an analogous procedure to complete a 12-month series. This procedure required the additional assumption that the modeled spatial distribution of air pollutants in September 2008 can be used as a proxy for the spatial distributions in September 2009. July and August spatial distributions were calculated using a 1/3 and 2/3 weighting to combine the June 2009 and September 2009 models for each pollutant, with higher weights assigned to the temporally more proximal month. The completed year-long series of monthly GeoDHOC models (September 2008 through August 2009) served as a template of spatially variable models that subsequently were refined using available MASN time series measurements.

3.2.3 Temporal Modeling

Implementation of the temporal scaling of the monthly GeoDHOC spatial models for both Model 1 and Model 2 occurred in four steps. First, monthly averages of MASN measurements were calculated for each pollutant. This process was straightforward for NO_2, total BTEX, and PM_{10}, which were each measured at a single MASN station so that an arithmetic average of the time series measurements made during each month could be used. In contrast,

$PM_{2.5}$ was measured at five MASN sites with an irregular spatial distribution throughout the city (**Figure 2.1**). Consequently, the four $PM_{2.5}$ stations clustered in south central Detroit were first averaged together. The resulting south central Detroit value was then averaged with the East 7 Mile site in northeast Detroit to derive a crudely declustered average $PM_{2.5}$ concentration incorporating information from all five MASN sites.

Second, individual monthly GeoDHOC models (Section 3.2) for each of the four pollutants analyzed were spatially averaged over the city of Detroit, inclusive of the embedded municipalities of Hamtramck and Highland Park (**Figure 2.1**). This process averaged the estimated values at each point in the 300 m x 300 m grid for each model.

Third, monthly spatial averages were compared to the corresponding monthly average of the MASN time series measurements. September 2008 and June 2009 are the only months in this study when direct comparisons between MASN measurements and unadjusted GeoDHOC spatial models for the City of Detroit are possible. Therefore, the mean of the September 2008 and June 2009 differences between GeoDHOC and MASN monthly averages for each pollutant was adopted as a target adjustment factor (TAF):

$$TAF_i = \frac{(GeoDHOC_{Sept\ 08} - MASN_{Sept\ 08}) + (GeoDHOC_{June\ 09} - MASN_{June\ 09})}{2} \qquad \text{(Eqn. 3.7)}$$

where subscript *i* represents each of the four pollutants considered.

Fourth, a spatially uniform bulk shift was calculated for each month and pollutant using the target adjustment factor and the difference between the GeoDHOC spatial model average and the corresponding MASN average:

$$Bulk\ Shift_{i,j} = TAF_i - \left[GeoDHOC_{i,j} - MASN_{i,j}\right] \quad \text{(Eqn. 3.8)}$$

where subscript j represents each of the months considered. When [$GeoDHOC_{i,j}$ – $MASN_{i,j}$] is positive and smaller than the target adjustment factor, positive bulk shift values are needed to increase the monthly difference up to the fixed target adjustment factor. Alternatively, if [$GeoDHOC_{i,j}$ – $MASN_{i,j}$] is positive but greater than the target adjustment factor, negative bulk shift values are needed to decrease the monthly difference down to the target adjustment factor. The converse occurs when [$GeoDHOC_{i,j}$ – $MASN_{i,j}$] is negative. The resulting bulk shift was subsequently used to adjust each monthly GeoDHOC spatial model:

$$Adjusted\ Monthly\ Model_{i,j} = (GeoDHOC\ Spatial\ Model_{i,j}) + (Bulk\ Shift_{i,j}) \quad \text{(Eqn. 3.9)}$$

The end product was two series (Models 3 and 4) of spatially and temporally variable concentration models for each of the four pollutants during each of the 12 months of 2008.

3.3 Asthma correlations

Linear regression was employed to generate correlations between the asthma dataset and air pollutant Models 2, 3, and 4. Model 1 asthma correlations are published in Lemke et al. (2013). Asthma data were aggregated at the zip code scale in Detroit and the equivalent forward sortation area scale in Windsor (Lemke et al., 2013). Consequently, air pollutant model concentrations were averaged over each of the postal regions for this study. Lemke et al. (2013) also reported cumulative asthma data for 2008. Therefore, the twelve individual months in

Model 3 and Model 4 were averaged to create an annual concentration for Model 3 and Model 4 in each postal zone for 2008. The SpaceStat aspatial linear regression tool was used for regression analysis defined by:

$$y = b_0 + b_1 x \qquad \text{(Eqn. 3.10)}$$

with b_1 as the slope, and b_0 as the y-intercept The asthma exacerbation rate was assigned as the dependent variable, *y*, and the air pollutant concentration was set as the independent variable, *x*.

Statistically significant correlations had a p-value of 0.05 or less. SpaceStat also returns model r^2 values which measure the strength of the association between the independent and dependent variables. The model r^2 values are defined as one minus the residual sum of squares (RSS) over the total sum of squares (TSS):

$$r^2 = 1 - \frac{RSS}{TSS}. \qquad \text{(Eqn. 3.11)}$$

CHAPTER 4. RESULTS

4.0 Introduction

Results from the spatial outlier analysis (Section 3.1) and temporal scaling (Section 3.2) are presented in this chapter in Sections 4.1 and 4.2, respectively. Spatial outlier identification and temporal scaling, used individually and in combination, generated three new air pollution models for the study area. In total, four different models (**Table 3.1**) provided a means to assess relationships between asthma rates to air pollution in the Detroit-Windsor international airshed (Section 3.3) and test the study hypotheses (Section 1.6). Results of these regressions are also presented below in Section 4.3.

4.1 Spatial Data Outliers Identification

Box plots, variogram clouds, difference maps, and Moran's I (Section 3.1) were used to identify potential spatial outliers in the GeoDHOC dataset. No single method, among the four used, emerged to definitively determine spatial outliers for each analyte. The set of potential outliers, identified using the four methods individually, was initially refined by requiring concurrence of at least two of the four methods to identify the same sampling location as a potential outlier. Each remaining potential outlier was then individually assessed using appropriate combinations of four methods.

4.1.1 Outlier Identification

The box plot provided useful exploratory statistics to evaluate the distribution of measurements for each pollutant. Box plot graphs showed the spread of each dataset and the relative magnitude of potential outliers indicated by the bottom and top whiskers **(Table 4.1)**. In

a number of instances, the bottom whisker (1.5 times the interquartile range below the first quartile) fell below zero (**Table 4.1, Figure 4.1**). This occurred for BTEX 2009, VOC 2008 and 2009, and PAH 2008 and 2009. In total, the box plots (**Figure 4.1**) displayed 24 potential outliers that exceeded ±1.5 times the interquartile range (**Table 4.2**). Unusally high values at or above the whisker were observed at 21 sampling locations and low values at or below the whisker were found at 3 locations. Histograms confirmed that these points were located on the extreme ends of the dataset (**Figure 4.2**).

Table 4.1. Summary table of GeoDHOC measured values for 2008 and 2009 and box plot values for each analyte.

Year	Analyte	Unit	# of data	min	max	mean	std	first quartile	median	third quartile	IQR	IQR x1.5	Bottom Whisker	Top Whisker
2008	NO_2	ppb	97	7.3	25.2	15.3	3.0	13.2	15.7	17.0	3.8	5.7	7.5	22.7
2009	NO_2	ppb	122	7.7	27.1	15.4	3.5	13.3	15.3	17.3	4.1	6.1	7.1	23.5
2008	BTEX	$\mu g/m^3$	98	1.0	30.9	8.9	4.0	6.4	8.8	10.5	4.2	6.3	0.1	16.8
2009	BTEX	$\mu g/m^3$	133	1.8	11.7	5.6	2.1	4.0	5.6	6.9	2.9	4.4	-0.4	11.3
2008	VOC	$\mu g/m^3$	98	3.8	46.6	13.8	5.7	9.8	13.7	16.4	6.7	10.0	-0.2	26.4
2009	VOC	$\mu g/m^3$	133	2.7	16.7	7.8	2.8	5.7	7.7	9.6	3.9	5.9	-0.2	15.4
2008	$PM_{2.5}$	$\mu g/m^3$	38	3.9	13.4	7.7	1.6	6.9	7.6	8.2	1.3	1.9	5.0	10.1
2009	$PM_{2.5}$	$\mu g/m^3$	37	7.8	17.7	9.5	1.4	8.5	9.2	10.0	1.5	2.2	6.3	12.2
2008	PM_{10}	$\mu g/m^3$	38	5.9	23.0	12.8	3.2	11.1	12.1	13.5	2.4	3.6	7.4	17.1
2009	PM_{10}	$\mu g/m^3$	37	11.8	21.9	15.8	2.6	13.9	15.4	16.8	2.9	4.3	9.7	21.1
2008	PAH	$\mu g/m^3$	38	3.3	53.5	20.1	13.7	6.1	18.3	28.5	22.4	33.6	-27.5	62.1
2009	PAH	$\mu g/m^3$	37	7.9	149.8	33.5	27.3	16.8	28.3	37.5	20.7	31.1	-14.3	68.6

Table 4.2. Summary table of potential spatial data outliers identified in the GeoDHOC 2008 and 2009 datasets. A check mark indicates that the sampling location was identified as a potential outlier by that method. Shading indicates that the sample location was determined to be a spatial data outlier. IQR is interquartile range and Mes. Diff. is measured difference (June 2009-Sept. 2008).

Year	Analyte	Unit	Site ID	Value	x 1.5 IQR	Box Plot Location	Box Plot	Variogram Cloud	Difference Map	Mes. Diff.
2008	NO2	ppb	D-A-35	25.2	22.7	top	✓	✓	✓	-4.5
2008	NO2	ppb	W-P-3	7.3	7.5	bottom	✓	✓	✓	7.9
2009	NO2	ppb	D-P-25	27.1	23.5	top	✓	✓	✓	-7.9
2009	NO2	ppb	D-A-14	25.2	23.5	top	✓	-	-	3.8
2009	NO2	ppb	W-P-23	25.0	23.5	top	✓	✓	-	-
2008	BTEX	$\mu g/m^3$	D-A-25	30.9	16.8	top	✓	✓	✓	-19.3
2008	BTEX	$\mu g/m^3$	D-A-5	20.5	16.8	top	✓	-	-	-8.8
2009	BTEX	$\mu g/m^3$	D-A-25	11.6	11.3	top	✓	-	-	-19.3
2009	BTEX	$\mu g/m^3$	D-A-5	11.7	11.3	top	✓	✓	-	-8.8
2009	BTEX	$\mu g/m^3$	D-A-33	9.8	11.3	-	-	-	✓	8.8
2008	VOC	$\mu g/m^3$	D-A-25	46.6	26.4	top	✓	✓	✓	-30.9
2008	VOC	$\mu g/m^3$	D-A-5	27.6	26.4	top	✓	-	-	-12.1
2009	VOC	$\mu g/m^3$	D-A-25	15.7	15.4	top	✓	-	-	-30.9
2009	VOC	$\mu g/m^3$	D-A-33	13.0	15.4	-	-	-	✓	8.7
2008	PM2.5	$\mu g/m^3$	W-A-2	3.9	5.00	bottom	✓	-	✓	5.7
2008	PM2.5	$\mu g/m^3$	D-A-6	10.5	10.1	top	✓	-	-	-0.2
2008	PM2.5	$\mu g/m^3$	W-A-4	13.4	10.1	top	✓	✓	-	-5.5
2009	PM2.5	$\mu g/m^3$	W-A-8	14.7	12.2	top	✓	✓	✓	7.3
2009	PM2.5	$\mu g/m^3$	D-A-33	12.9	12.2	top	✓	-	✓	3.8
2008	PM10	$\mu g/m^3$	W-A-2	5.9	7.4	bottom	✓	-	✓	9.9
2008	PM10	$\mu g/m^3$	D-A-32	23.0	17.1	top	✓	✓	-	-1.9
2008	PM10	$\mu g/m^3$	D-A-6	22.5	17.1	top	✓	✓	-	-3.1
2008	PM10	$\mu g/m^3$	W-A-4	20.6	17.1	top	✓	-	-	-8.3
2009	PM10	$\mu g/m^3$	W-A-8	21.9	21.1	top	✓	✓	✓	10.3
2009	PM10	$\mu g/m^3$	D-A-33	20.5	21.1	-	-	-	✓	5.4
2009	PAH	$\mu g/m^3$	D-A-32	98.7	68.6	top	✓	-	✓	73.7
2009	PAH	$\mu g/m^3$	W-A-3	149.8	68.6	top	✓	✓	✓	-

Each point identified using a box plot was subsequently evaluated on a variogram cloud to assess the relationship between spatial and concentration differences (**Figures 4.3 to 4.13)**. Of the 24 points identified on box plots, variogram clouds showed 13 sampling locations with a high degree of dissimilarity in a short geographical distance (**Table 4.2**). All but one of these points were sampling locations with concentrations plotting above the top whisker in the corresponding box plot.

The difference maps displayed a number of sampling points with large concentration differences between the September 2008 and June 2009 sampling periods (**Figure 4.14**). Potential outliers identified by the box plot method and the difference maps showed good agreement. In total, 12 of the 24 sampling points indicated by the box plot were in areas that demonstrated a distinct difference between the sampling periods (**Table 4.2**).

The Local Moran's I test revealed 22 statistically significant points (p-value equal or less than 0.05) that showed spatial patterns of outliers **(Table 4.3)**. Of these, 18 points were classified as low-high, meaning the sampling points had a lower concentration than the surrounding points. The remaining four points were classified as high-low points, indicating the sampling points had a higher concentration than the surrounding samplers.

Table 4.3. Statistically significant Local Moran's I results.

Analyte	Year	Unit	Site ID	Analyte Conc.	normal scores values	z-value	Mean neighbor value	Moran's I values	High/Low	P value
NO2	2008	ppb	D-A-10	15.1	-0.18	-0.18	1.08	-0.20	Low-high	0.002
NO2	2008	ppb	D-P-19	15.0	-0.23	-0.24	0.89	-0.21	Low-high	<0.001
NO2	2009	ppb	W-A-4	15.5	0.15	0.16	-0.66	-0.10	High-low	0.048
NO2	2009	ppb	D-P-23	15.1	-0.07	-0.07	0.92	-0.07	Low-high	0.016
NO2	2009	ppb	D-P-5A	14.5	-0.39	-0.39	0.72	-0.28	Low-high	0.026
BTEX	2008	µg/m³	W-P-7	9.6	0.64	0.30	-0.76	-0.23	High-low	0.022
BTEX	2008	µg/m³	D-A-34	8.6	1.24	-0.17	0.94	-0.16	Low-high	0.024
BTEX	2008	µg/m³	D-P-2	7.6	0.96	-0.38	0.96	-0.37	Low-high	0.008
BTEX	2009	µg/m³	W-P-2	5.6	0.04	0.04	-0.80	-0.03	High-low	0.024
BTEX	2009	µg/m³	D-P-27	5.1	-0.19	-0.19	0.78	-0.15	Low-high	0.04
BTEX	2009	µg/m³	D-A-8	5.0	-0.25	-0.25	0.74	-0.19	Low-high	0.022
BTEX	2009	µg/m³	D-A-10	5.0	-0.27	-0.27	0.97	-0.26	Low-high	0.006
BTEX	2009	µg/m³	D-P-12	4.6	-0.35	-0.35	0.89	-0.31	Low-high	0.018
VOC	2008	µg/m³	D-A-34	13.5	-0.14	-0.14	0.92	-0.13	Low-high	0.004
VOC	2008	µg/m³	D-P-2	12.0	-0.41	-0.41	0.94	-0.38	Low-high	0.012
VOC	2008	µg/m³	D-P-25	12.9	-0.25	-0.25	0.69	-0.17	Low-high	0.034
VOC	2009	µg/m³	D-A-10	6.9	-0.21	-0.21	0.92	-0.19	Low-high	0.002
VOC	2009	µg/m³	D-P-12	6.7	-0.23	-0.23	0.89	-0.21	Low-high	0.012
VOC	2009	µg/m³	D-P-27	6.6	-0.33	-0.33	0.75	-0.25	Low-high	0.018
PM2.5	2008	µg/m³	W-A-4	13.4	2.22	2.26	-0.64	-1.44	High-low	0.046
PAH	2008	µg/m³	D-A-15	18.3	-0.03	-0.03	0.99	-0.03	Low-high	<0.001
PAH	2009	µg/m³	D-A-24	16.8	-0.74	-0.75	0.81	-0.61	Low-high	0.014

Results from the Moran's I did not agree well with the the other data outlier methods. With the exception of W-A-4 for $PM_{2.5}$ in 2008, the location of potential Moran's I outliers did not correspond to the spatial location of potential outliers on the difference maps **(Figure 4.14).** Moreover, with the same exception of W-A-4 for $PM_{2.5}$, potential Moran's I outliers were not identifiable as potential outliers on variogram clouds (**Figure 4.15**). This was surprising because, in theory, both the Local Moran's I and vairogram clouds are used to identify local spatial outliers. The Moran's I results were also not global outliers because they plotted in the middle of the histograms for each dataset with the exeption of W-A-4 (**Figure 4.16**). Because of the poor association of the Moran's I results with the other methods, the Moran's I was excluded from further consideration as a spatial-outlier identification method in this study.

After the initial analysis using the three remaining outlier detection methods (excluding the Moran's I), ten sampling locations identified by only one method were dropped as potential outliers (**Table 4.2**). Based on the spatial position and magnitude of each of the remaining 17 potential outliers, as reflected in the combination of their corresponding box plot, variogram cloud, and difference map, a total of 13 outliers were identified **(Table 4.2)**. One or two outliers were found for each analyte execpt for BTEX and VOCs which lacked outliers in 2009 and PAH which lacked outliers in 2008. Two outliers were identified for NO_2 and $PM_{2.5}$ in 2008 and for PAH and $PM_{2.5}$ in 2009 **(Table 4.2)**.

4.1.2 Variogram Modeling and Kriging with Outliers Removed

Variogram models and ordinary kriged grids were computed from the datasets with the outliers removed to generate Model 2 of this study **(Figure 4.17)**. Although both Models 1 and 2

interpolate over a larger area than the Detroit and Windsor municipal borders, the models were compared using only the values for grid nodes located in Detroit (including Hamtramck and Highland Park) and Windsor.

Model 2 employed different variogram and kriging parameters than Model 1. The still contribution was lower for Model 2 compared to Model 1 for NO_2, BTEX, VOC, and PAH **(Table 3.2)**. As a result, standard deviations for NO_2, BTEX, VOC, and $PM_{2.5}$ (**Table 4.4**) were reduced in Model 2 compared to Model 1. Conversely, standard deviations increased markedly for PM_{10} and slightly for PAHs (**Table 4.4**) in Model 2 compared to Model 1. Different versions of the PM datasets were used for Models 1 and 2 (Section 3.1.6), however, so direct comparisons between these models are problematic.

Table 4.4. Comparison of original GeoDHOC ordinary kriged models (Model 1) to the revised ordinary kriged models with spatial data outliers removed (Model 2).

Year	Analyte	Unit	Model #	Mean	Min	Max	Range	STD	% Difference of STD
2008	NO_2	ppb	Model 1	15.2	9.1	22.6	13.4	2.42	-4.23
2008	NO_2	ppb	Model 2	15.2	9.1	20.0	10.9	2.32	
2009	NO_2	ppb	Model 1	15.1	8.2	23.6	15.4	3.16	-4.94
2009	NO2	ppb	Model 2	14.9	8.1	22.6	14.4	3.01	
2008	BTEX	$\mu g/m^3$	Model 1	8.8	1.9	26.4	24.5	2.96	-14.4
2008	BTEX	$\mu g/m^3$	Model 2	8.7	1.9	18.1	16.2	2.56	
2008	VOC	$\mu g/m^3$	Model 1	13.7	4.1	37.0	33.0	4.12	-13.1
2008	VOC	$\mu g/m^3$	Model 2	13.5	4.1	23.5	19.4	3.61	
2008	$PM_{2.5}$	$\mu g/m^3$	Model 1	7.7	5.1	10.1	5.0	0.67	-9.93
2008	$PM_{2.5}$	$\mu g/m^3$	Model 2	7.8	6.0	10.0	4.0	0.74	
2009	$PM_{2.5}$	$\mu g/m^3$	Model 1	9.6	7.7	13.0	5.2	0.77	-73.1
2009	$PM_{2.5}$	$\mu g/m^3$	Model 2	9.2	8.3	10.3	2.0	0.36	
2008	PM_{10}	$\mu g/m^3$	Model 1	12.7	8.8	19.6	10.8	1.89	22.1
2008	PM_{10}	$\mu g/m^3$	Model 2	13.1	10.0	22.9	13.0	2.36	
2009	PM_{10}	$\mu g/m^3$	Model 1	15.9	12.6	19.2	6.6	1.31	21.3
2009	PM_{10}	$\mu g/m^3$	Model 2	15.6	12.5	19.9	7.4	1.62	
2009	PAH	$\mu g/m^3$	Model 1	31.8	13.3	79.3	66.1	11.0	4.30
2009	PAH	$\mu g/m^3$	Model 2	28.4	10.0	65.1	55.1	11.5	

4.2 Temporal Scaling

4.2.0 Introduction

MASN time series measurements NO_2, total BTEX, $PM_{2.5}$, and PM_{10} were combined with Model 1 (the original GeoDHOC models) and Model 2 (with outliers removed) to generate a 12 month series of maps for 2008 using a process that included a comparison of proximal GeoDHOC and MASN measurements followed by spatial and temporal modeling (Section 3.2). The resulting temporally scaled datasets derived from Models 1 and 2 are called Model 3 and Model 4, respectively (**Table 3.1**). Results from each step in the derivation of these maps are presented in this section.

4.2.1 Comparison of Measurements and Model Estimates

NO_2, total BTEX, $PM_{2.5}$, and PM_{10} measurements at GeoDHOC sampling locations differed from MASN values measured at collocated or nearby locations in September 2008 and June 2009. Collocated GeoDHOC and MASN measurements for NO_2 and total BTEX agreed within 25%, with the exception of June 2009 BTEX, which varied by 81% (**Table 4.5**). Although collocated PM samplers were not available in this study, GeoDHOC measurements collected within 1 km of the Southwestern High School and FIA/Lafayette St. sites agreed within 13% (**Table 4.6**).

The spatial and temporal variability of air pollutant measurements are summarized in **Table 4.7**. The magnitudes of spatial and temporal variability are comparable for NO_2 and BTEX, while the observed temporal variability of $PM_{2.5}$ and PM_{10} is slightly greater than the sampled and modeled spatial variability.

Table 4.5. GeoDHOC and MASN collocated sampler concentrations

Analyte	Location	Distance from MASN sampler	Sept. 2008	June 2009
NO_2 (ppb)	MASN (East 7 Mile)		12.1	10.3
	GeoDHOC collocated (D-P-27)	< 1m	15.5	12.5
	% difference		24.7	19.4
BTEX ($\mu g/m^3$)	MASN (SWHS)		7.7	1.8
	GeoDHOC collocated (D-P-22)	< 1m	8.8	4.4
	% difference		13.0	80.9

Table 4.6. GeoDHOC and MASN nearby sampler concentrations

Analyte	Location	Distance from MASN sampler	Sept. 2008	June 2009
$PM_{2.5}$ ($\mu g/m^3$)	MASN (SWHS)		11.3	9.4
	GeoDHOC (D-A-6)	990m	10.5	10.3
	% difference		6.9	9.1
$PM_{2.5}$ ($\mu g/m^3$)	MASN (FIA)		11.2	9.0
	GeoDHOC (D-A-30)	395m	n/a	10.0
	% difference		n/a	9.8
PM_{10} ($\mu g/m^3$)	MASN (SWHS)		20.4	22.0
	GeoDHOC (D-A-6)	990m	22.5	19. 4
	% difference		9.6	12.8

Table 4.7. Comparison of Detroit values for the MASN temporal dataset, the GeoDHOC observed (sampled) dataset, Model 1, and Model 2. Model 2 BTEX 2009 was not created due to lack of outliers.

Dataset	Analyte	Unit	Year	Mean	Standard Deviation	Sample Variance	Range	Minimum	Maximum
MASN	NO_2	ppb	2008-2010	12.5	2.79	7.80	12.0	7.6	19.5
Sampled Dataset	NO_2	ppb	2008	15.3	2.95	8.69	17.9	7.3	25.2
Sampled Dataset	NO_2	ppb	2009	15.4	3.5	12.2	19.4	7.7	27.1
Model 1	NO_2	ppb	2008	16.2	1.82	3.33	11.1	11.5	22.6
Model 1	NO_2	ppb	2009	16.5	2.27	5.16	13.3	10.3	23.6
Model 2	NO2	ppb	2008	16.1	1.67	2.79	8.2	11.8	20.0
Model 2	NO2	ppb	2009	16.3	2.07	4.28	11.8	10.8	22.6
MASN	BTEX	$\mu g/m^3$	2008-2010	4.6	2.2	4.83	8.2	1.1	9.4
Sampled Dataset	BTEX	$\mu g/m^3$	2008	8.9	3.95	15.6	30.0	1.0	30.9
Sampled Dataset	BTEX	$\mu g/m^3$	2009	5.6	2.07	4.30	10.0	1.8	11.7
Model 1	BTEX	$\mu g/m^3$	2008	10.0	2.47	6.12	24.5	1.9	26.4
Model 1	BTEX	$\mu g/m^3$	2009	6.7	1.3	1.69	8.4	2.8	11.1
Model 2	BTEX	$\mu g/m^3$	2008	9.8	1.91	3.65	16.2	1.9	18.1
Model 2	BTEX	$\mu g/m^3$	2009	N/A	N/A	N/A	N/A	N/A	N/A
MASN	$PM_{2.5}$	$\mu g/m^3$	2008-2010	10.9	3.02	9.11	11.5	5.8	17.3
Sampled Dataset	$PM_{2.5}$	$\mu g/m^3$	2008	7.7	1.62	2.62	9.5	3.9	13.4
Sampled Dataset	$PM_{2.5}$	$\mu g/m^3$	2009	9.5	1.44	2.07	6.9	7.8	14.7
Model 1	$PM_{2.5}$	$\mu g/m^3$	2008	7.9	0.53	0.29	3.0	6.9	9.8
Model 1	$PM_{2.5}$	$\mu g/m^3$	2009	9.7	0.72	0.52	4.3	8.1	12.4
Model 2	$PM_{2.5}$	$\mu g/m^3$	2008	7.9	0.58	0.34	3.0	7.0	10.0
Model 2	$PM_{2.5}$	$\mu g/m^3$	2009	9.4	0.3	0.09	1.7	8.6	10.3
MASN	PM_{10}	$\mu g/m^3$	2008-2010	20.2	5.59	31.2	20.8	11.7	32.5
Sampled Dataset	PM_{10}	$\mu g/m^3$	2008	12.8	3.21	10.29	17.0	5.9	23.0
Sampled Dataset	PM_{10}	$\mu g/m^3$	2009	15.8	2.57	6.58	10.1	11.8	21.9
Model 1	PM_{10}	$\mu g/m^3$	2008	13.2	1.81	3.29	8.9	10.7	19.6
Model 1	PM_{10}	$\mu g/m^3$	2009	16.2	1.16	1.34	5.3	13.9	19.2
Model 2	PM_{10}	$\mu g/m^3$	2008	13.5	2.4	5.76	12.3	10.6	22.9
Model 2	PM_{10}	$\mu g/m^3$	2009	16.2	1.34	1.80	6.5	13.4	19.9

For the September 2008 and June 2009 Models 1 and 2, cross validation of predicted (kriged) vs. observed concentrations demonstrates that kriging modeled estimates agree well with observed values (**Table A1)**. Small mean errors (close to zero), small standard deviations, lack of trend in the spatial distribution of estimation errors, and the absence of conditional bias on scatter plots of error versus estimated values indicate a lack of bias in kriged model estimations. Kriging variance maps for Model 1 and Model 2 exhibit low estimation variance throughout the majority of the study area for each pollutant (**Figure A1**).

4.2.2 Spatial Modeling

The temporally unadjusted 12-month series of GeoDHOC Models 1 and 2 for NO_2, total BTEX, $PM_{2.5}$, and PM_{10} are presented in **Figure 4.18**. Model 1 spatially averaged unadjusted mean values for September 2008 and June 2009 concentrations in the City of Detroit differed by 2 to 40% **(Table 4.8)**. Model 2 spatially averaged unadjusted mean values for September 2008 and June 2009 concentration in the City of Detroit differed by 1 to 58% **(Table 4.8)**. Spatially averaged mean concentrations for the 12-month series models (**Table A2**) vary progressively between the September 2008 and June 2009 anchor months as a consequence of the averaging method employed to construct them. Spatial variability is retained in each of the monthly models, although, in some cases, it is slightly attenuated as indicated by coefficient of variation values for individual months that are lower than the two anchor months, September 2008 and June 2009 (**Table A2**).

Table 4.8. Model 1 and Model 2 Detroit September 2008 and June 2009 spatial model mean, standard deviation (SD) and coefficient of variation (CV).

Model 1					
		Values for Sept. 2008 Detroit	Values for June 2009 Detroit	Difference (Sept. 2008 – June 2009)	% Difference
NO_2	Mean (ppb)	16.2	16.5	-0.3	2.1
	SD (ppb)	1.8	2.3	-0.5	
	CV (%)	11.3	13.7	-2.4	
BTEX	Mean (µg/m^3)	10	6.7	3.3	39.9
	SD (µg/m^3)	2.5	1.3	1.2	
	CV (%)	24.8	19.5	5.3	
$PM_{2.5}$	Mean (µg/m^3)	7.9	9.7	1.8	20.2
	SD (µg/m^3)	0.5	0.7	-0.2	
	CV (%)	6.7	7.4	-0.7	
PM_{10}	Mean (µg/m^3)	13.2	16.2	3	20.9
	SD (µg/m^3)	1.82	1.2	0.7	
	CV (%)	13.8	7.1	6.7	
Model 2					
		Values for Sept. 2008 Detroit	Values for June 2009 Detroit	Difference (Sept. 2008 – June 2009)	% Difference
NO_2	Mean (ppb)	16.1	16.3	-0.2	1.1
	SD (ppb)	1.7	2.1	-0.4	
	CV (%)	10.3	12.7	-2.4	
BTEX	Mean (µg/m^3)	10.0	17.0	-7.1	52.3
	SD (µg/m^3)	2.5	2.4	0.1	
	CV (%)	24.8	14.2	10.6	
$PM_{2.5}$	Mean (µg/m^3)	7.9	9.4	-1.5	16.9
	SD (µg/m^3)	0.5	0.3	0.2	
	CV (%)	6.7	3.2	3.5	
PM_{10}	Mean (µg/m^3)	13.5	16.2	-2.7	18.4
	SD (µg/m^3)	2.4	1.3	1.1	
	CV (%)	17.8	8.3	9.6	

4.2.3 Temporal Modeling

Although differences were observed in $PM_{2.5}$ measurements among the five MASN locations, temporal variation tracked consistently from station to station during 2008 through

2010 (**Figure 4.19**). Averaging the south central $PM_{2.5}$ mean value of 11.3 µg/m^3 with the East 7 Mile value of 8.5 µg/m^3 yielded a declustered Detroit average $PM_{2.5}$ value of 9.9 µg/m^3 for September 2008. Similar averaging of the south central $PM_{2.5}$ mean value of 9.1 µg/m^3 with an East 7 Mile value of 8.7 µg/m^3 yielded a value of 8.9 µg/m^3 for June 2009.

Averaged city of Detroit model concentrations for NO_2, total BTEX, $PM_{2.5}$, and PM_{10} (Model 1 and Model 2) also varied in relation to temporally averaged MASN measurements for the months of September 2008 and June 2009 (**Table 4.9**). Averaging of monthly differences between GeoDHOC and MASN values for September 2008 and June 2009 in Model 1 (**Equation 3.7**) resulted in positive target adjustment factors of 5.2 ppb and 3.5 µg/m^3 for NO_2 and total BTEX, and negative target adjustment factors of -0.6 µg/m^3 and -6.5 µg/m^3 for $PM_{2.5}$ and PM_{10}, respectively (**Table 4.9**). For Model 2, the NO_2 and total BTEX had positive target adjustment factors of 5.03 ppb and 3.44 µg/m^3 and $PM_{2.5}$ and PM_{10}had negative target adjustment factors of -0.72 µg/m^3 and -6.36 µg/m^3, respectively **(Table 4.9).**

These target adjustment factors were used to calculate bulk shift values for the three years of MASN data (**Equation 3.8, Table A3, and Table A4**). The 2008 bulk shift values were applied to adjust the 12-month series of spatially interpolated models for NO_2, total BTEX, $PM_{2.5}$, and PM_{10} for each month in 2008 (**Equation 3.9)**. In some cases, this resulted in an increase in values (positive bulk shift); in others, a decrease (negative bulk shift) (**Table 4.10**). In all cases, the adjustment enforced the target difference between the MASN and GeoDHOC averages uniformly across each time series of monthly estimates for each pollutant in Models 3 and 4 (**Figure 4.20**). Although only the 12 months of 2008 are applied in this thesis, the entire three-year monthly time series is shown in Figure 4.19 to include the GeoDHOC June 2009 Model 1 and Model 2 spatial averages, which were used to calculate the adjustment factor for

Models 3 and 4. The resulting spatially and temporally variable concentration models for each of the four pollutants during each of the 12 months incorporate the temporal trends present in the MASN data as well as the 300 by 300 m spatial resolution from the original GeoDHOC ordinary kriged maps (**Figure 4.21**).

Table 4.9. Comparison of Model 1 and Model 2 with MASN monthly averages.

Model 1

	September 2008			**June 2009**			
Analyte	GeoDHOC Spatial Model Average	MASN Average	Difference	GeoDHOC Spatial Model Average	MASN Average	Difference	Target Adjustment Factor (Average Difference)
NO_2 (ppb)	16.2	12.1	4.1	16.5	10.3	6.3	5.2
Total BTEX ($\mu g/m^3$)	10	7.7	2.3	6.7	1.9	4.8	3.5
$PM_{2.5}$ ($\mu g/m^3$)	7.9	9.9	-2.0	9.7	8.9	0.8	-0.6
PM_{10} ($\mu g/m^3$)	13.2	20.4	-7.2	16.2	22	-5.8	-6.5

Model 2

	September 2008			**June 2009**			
Analyte	GeoDHOC Spatial Model Average	MASN Average	Difference	GeoDHOC Spatial Model Average	MASN Average	Difference	Target Adjustment Factor (Average Difference)
NO^2 (ppb)	16.1	12.1	4.0	16.3	10.3	6.1	5.0
Total BTEX ($\mu g/m^3$)	9.8	7.7	2.1	6.7	1.9	4.8	3.4
$PM_{2.5}$ ($\mu g/m^3$)	7.9	9.9	-2.0	9.4	8.9	0.5	-0.7
PM_{10} ($\mu g/m3$)	13.5	20.4	-6.9	16.2	22	-5.8	-6.4

Table 4.10. Range of monthly bulk shift values for Model 1 and Model 2

	Analyte	Min	Max	Range
Model 1	NO_2 (ppb)	-3.8	8.3	12.1
	BTEX ($\mu g/m^3$)	-3.4	5. 6	9.0
	$PM_{2.5}$ ($\mu g/m^3$)	-4.1	8.0	12.1
	PM_{10} ($\mu g/m^3$)	-9.7	11.5	21.2
Model 2	NO_2 (ppb)	-3.7	8.3	12.1
	BTEX ($\mu g/m^3$)	-3.4	5.5	8.8
	$PM_{2.5}$ ($\mu g/m^3$)	-4.0	8.2	12.2
	PM_{10} ($\mu g/m^3$)	-9.7	9.9	19.6

4.3 Asthma Correlations

Four different models were created to assess the relationship between air pollution and asthma rates in Detroit and Windsor but two factors limited how the air pollutant models were used. First, because asthma data were available only for 2008 (Lemke et al., 2013), correlations were limited to the corresponding 2008 time period for this study. Second, temporal data were evaluated exclusively in Detroit for the perinatology study (O'Leary and Lemke, 2014). Therefore, Models 3 and 4 asthma correlations focused specifically on the Detroit airshed.

Air pollution Models 1 and 2 and 2008 asthma associations were evaluated for both Detroit (**Table 4.11**) and Windsor (**Table 4.12**). In Detroit, BTEX and VOCs were statistically significant for both Models 1 and 2, although the coefficients of determination (r^2) were weaker

than the original r^2 in Model 1 (**Table 4.11**). Although NO_2, $PM_{2.5}$, and PM_{10} were not statistically significant, r^2 values increased with the outliers removed. In Windsor, asthma associations with NO_2 for both Model 1 and 2, and VOCs and PM_{10} for Model 1 were statistically significant (**Table 4.12**). For all analytes in Windsor with the exception of $PM_{2.5}$, which was not statistically significant, r^2 values decreased indicating weaker correlations between asthma and air pollution.

Associations between asthma events in Detroit and each of the four air pollution models are presented in **Table 4.13**. NO_2 and BTEX associations were statistically significant for Model 3 and Model 4. The strength of NO_2 associations increased markedly in Model 3 and Model 4. Model 4 had the highest correlation of any model for BTEX. Although the Model 3 correlation was slightly lower than Model 1, it remained high. $PM_{2.5}$ associations were not statistically significant although incorporating the temporal trend in Model 3 and Model 4 increased the r^2 value. Conversely, PM_{10} was also not statistically significant but incorporating the temporal trend decreased the r^2 value.

Table 4.11. Model 1 and 2 asthma associations for Detroit. Statistically significant associations are highlighted in red.

GeoDHOC Detroit					
Year	Analyte	Model #	r^2	r	p-value
2008	NO_2	Model 1	0.03	0.17	0.40
2008	NO_2	Model 2	0.04	0.21	0.32
2008	BTEX	Model 1	0.28	0.53	0.01
2008	BTEX	Model 2	0.18	0.43	0.03
2008	VOC	Model 1	0.26	0.51	0.01
2008	VOC	Model 2	0.14	0.37	0.07
2008	$PM_{2.5}$	Model 1	0.00	0.04	0.84
2008	$PM_{2.5}$	Model 2	0.00	0.05	0.80
2008	PM_{10}	Model 1	0.00	0.00	1.00
2008	PM_{10}	Model 2	0.06	0.24	0.24

Table 4.12. Model 1 and 2 asthma associations for Windsor. Statistically significant associations are highlighted in red.

GeoDHOC Windsor					
Year	Analyte	Model #	r^2	r	p-value
2008	NO_2	Model 1	0.39	0.63	0.03
2008	NO_2	Model 2	0.35	0.59	0.04
2008	BTEX	Model 1	0.18	0.43	0.16
2008	BTEX	Model 2	0.13	0.36	0.24
2008	VOC	Model 1	0.34	0.58	0.05
2008	VOC	Model 2	0.25	0.50	0.10
2008	$PM_{2.5}$	Model 1	0.10	0.31	0.33
2008	$PM_{2.5}$	Model 2	0.16	0.40	0.21
2008	PM_{10}	Model 1	0.37	0.61	0.04
2008	PM_{10}	Model 2	0.23	0.48	0.11

Table 4.13. Comparison of Detroit asthma associations for Models 1-4. Statistically significant associations are highlighted in red.

GeoDHOC Detroit					
Year	Analyte	Model #	r^2	r	p
2008	NO_2	Model 1	0.03	0.17	0.40
2008	NO_2	Model 2	0.04	0.21	0.32
2008	NO_2	Model 3	0.19	0.44	0.03
2008	NO_2	Model 4	0.16	0.40	0.05
2008	BTEX	Model 1	0.28	0.53	0.01
2008	BTEX	Model 2	0.18	0.43	0.03
2008	BTEX	Model 3	0.26	0.51	0.01
2008	BTEX	Model 4	0.32	0.56	0.00
2008	$PM_{2.5}$	Model 1	0.00	0.04	0.84
2008	$PM_{2.5}$	Model 2	0.00	0.05	0.80
2008	$PM_{2.5}$	Model 3	0.02	0.13	0.52
2008	$PM_{2.5}$	Model 4	0.00	0.00	0.89
2008	PM_{10}	Model 1	0.00	0.00	1.00
2008	PM_{10}	Model 2	0.06	0.24	0.24
2008	PM_{10}	Model 3	0.04	0.19	0.36
2008	PM_{10}	Model 4	0.00	0.06	0.79

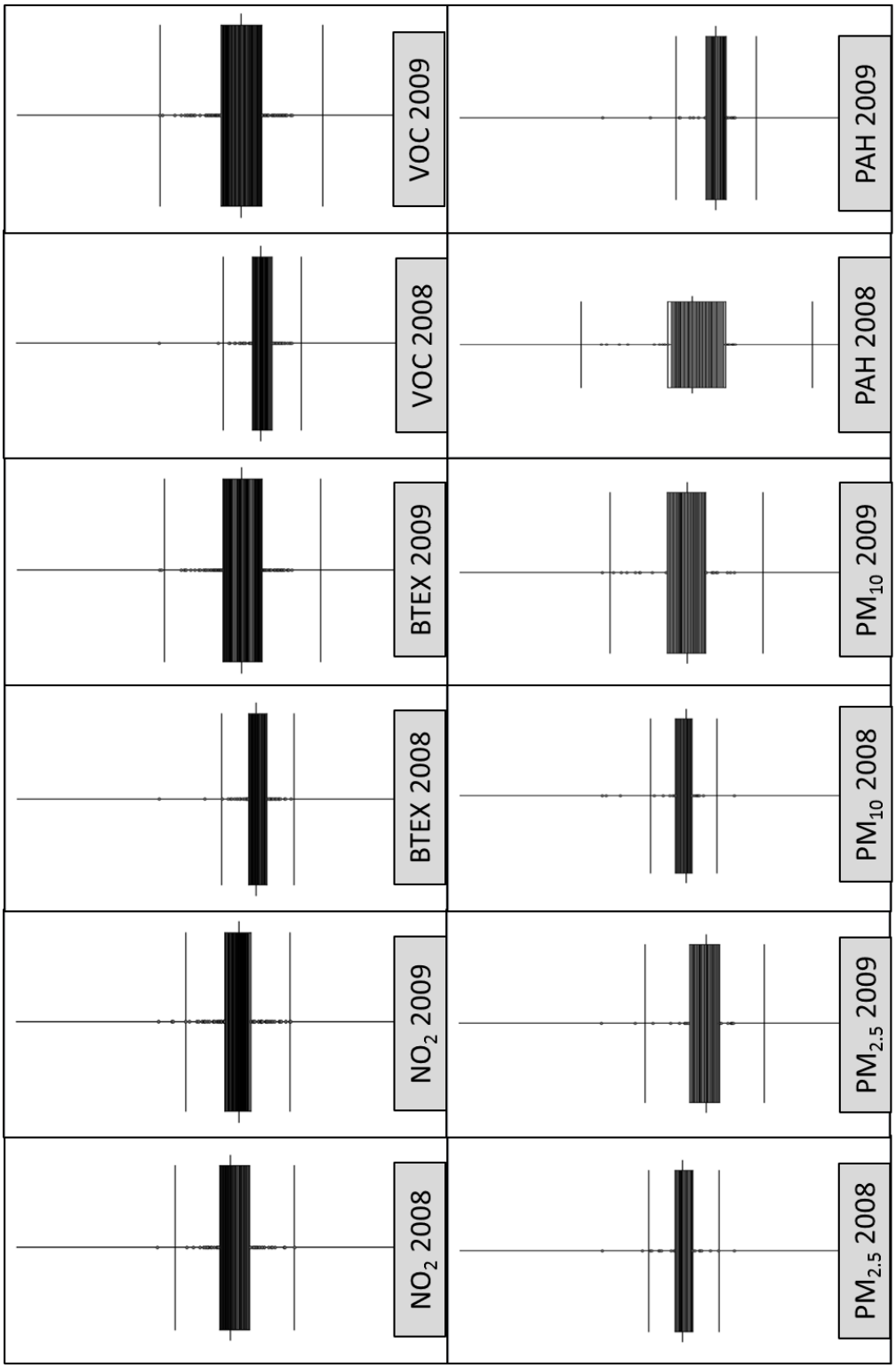


Figure 4.1. Box plots for each analyte in 2008 and 2009. Outliers are points above or below the whisker lines. See Table 4.1 for relevant plot values.

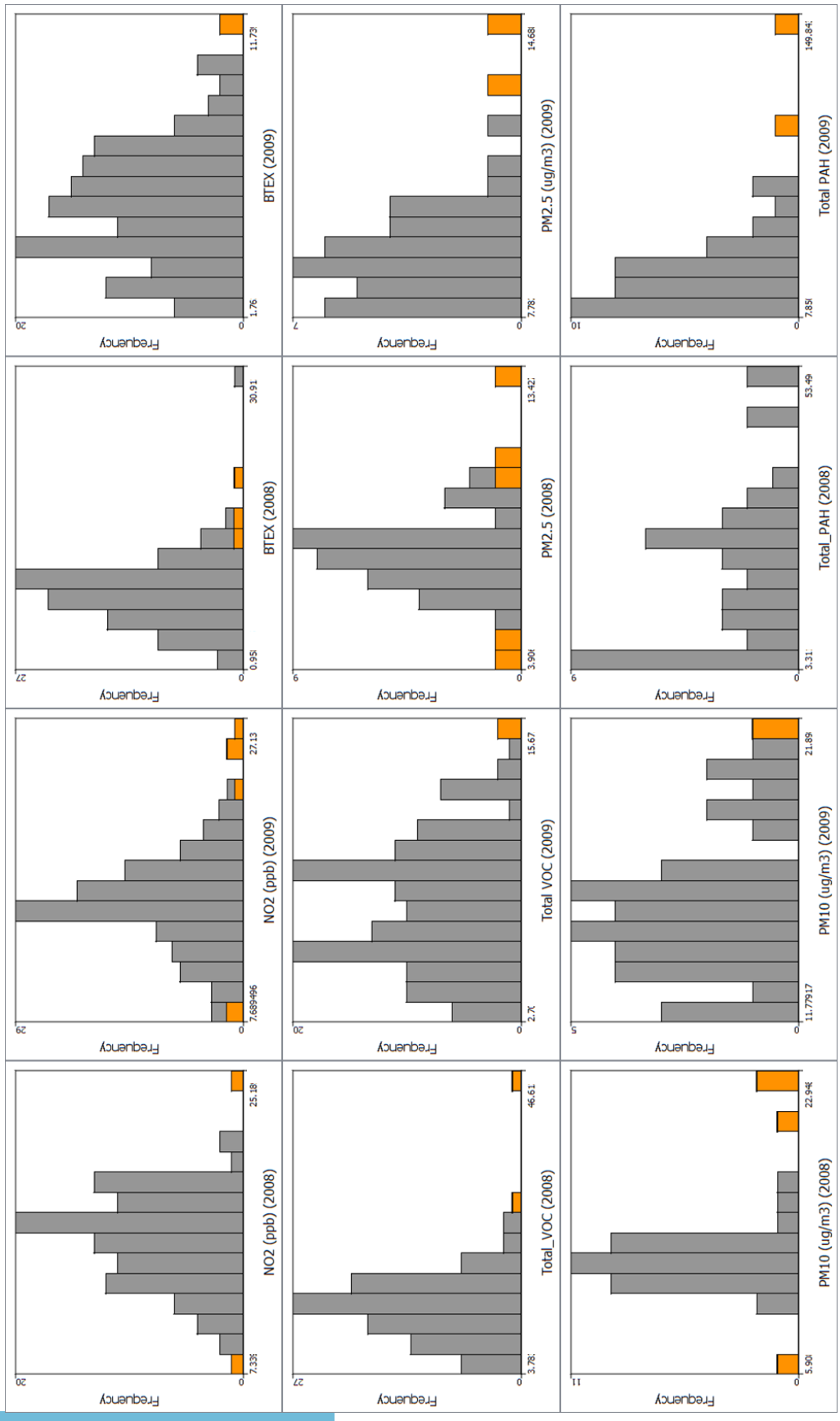


Figure 4.2. Histogram graphs for each analyte in 2008 and 2009. Yellow bars indicate points that were identified as potential outliers from the box plot method. See Table 4.1 for relevant plot values.

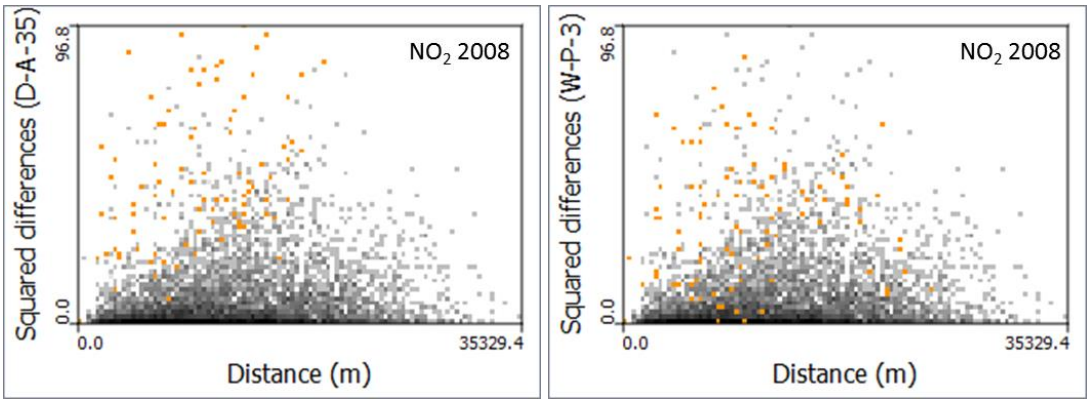


Figure 4.3. **Variogram cloud graph for NO_2 2008. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.**

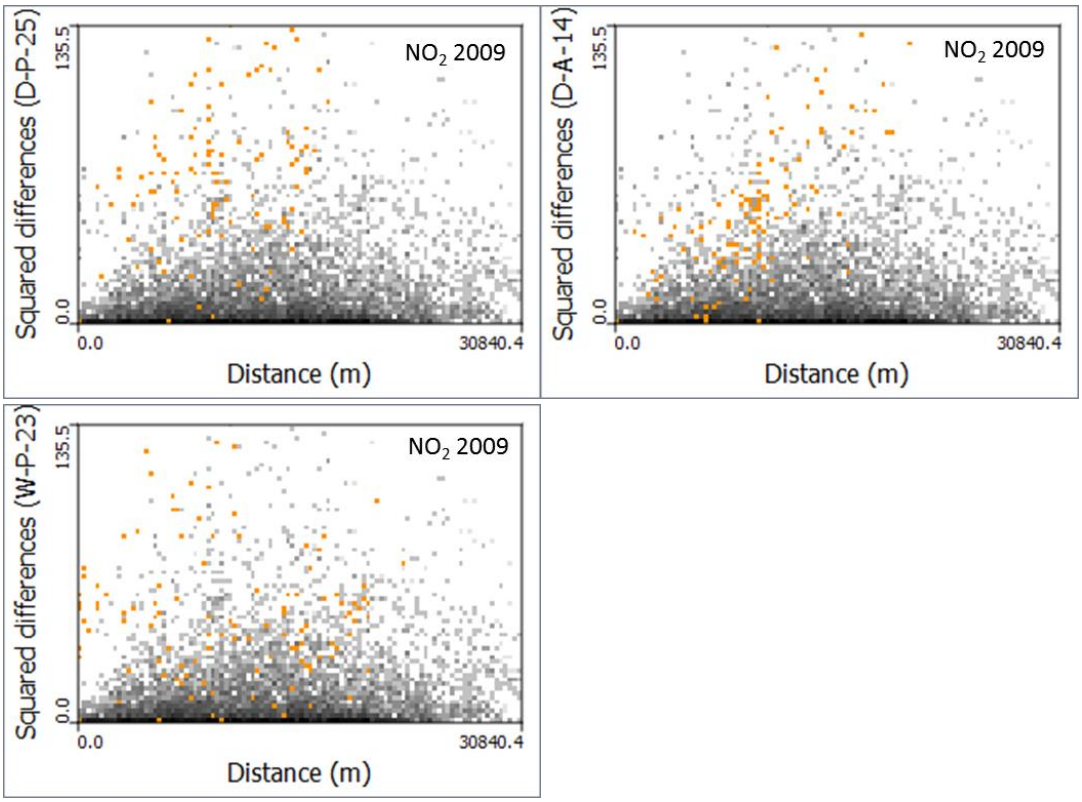


Figure 4.4. **Variogram cloud graph for NO_2 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.**

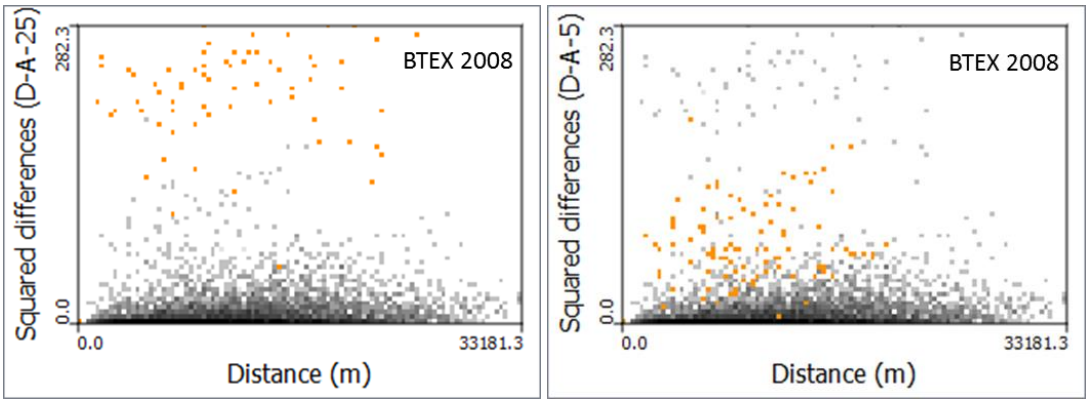


Figure 4.5. Variogram cloud graph for BTEX 2008. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.

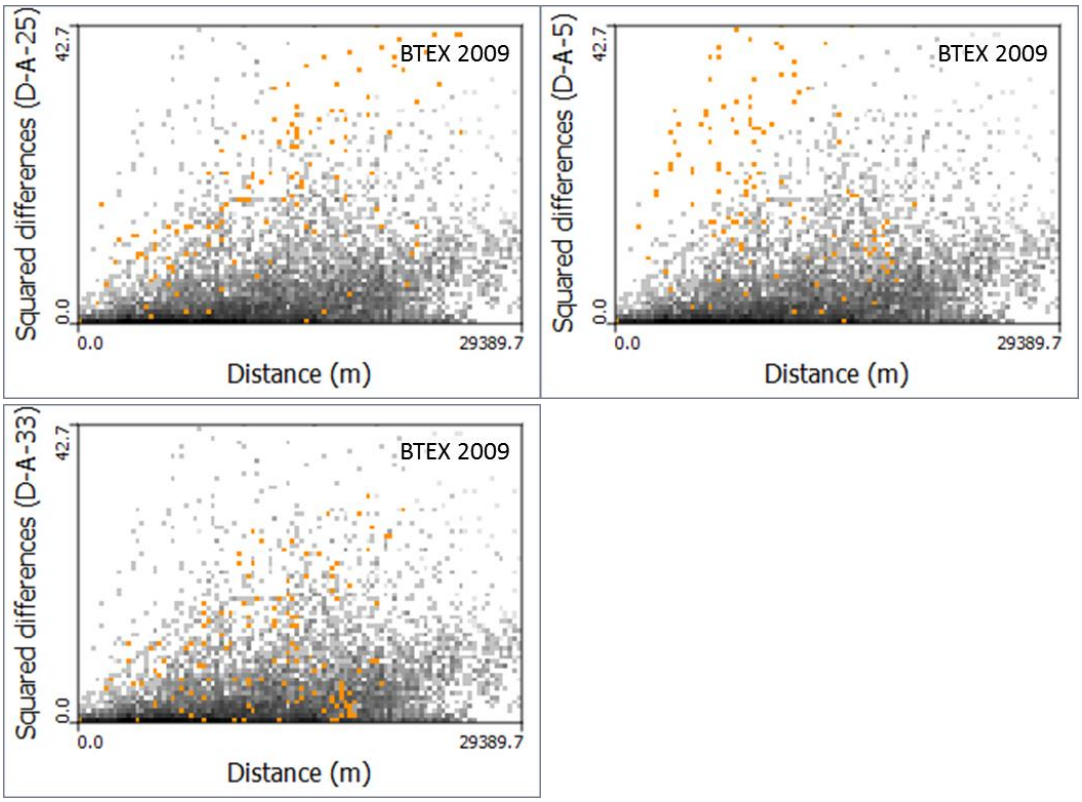


Figure 4.6. Variogram cloud graph for BTEX 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.

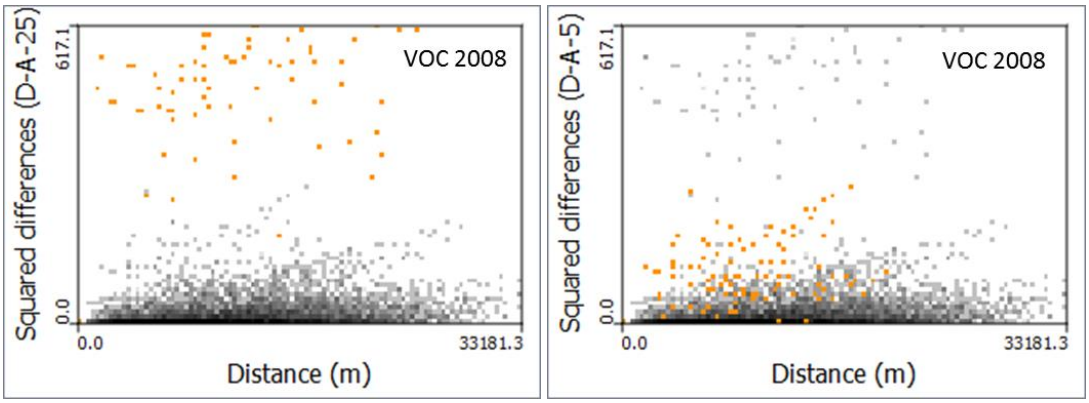


Figure 4.7. Variogram cloud graph for VOC 2008. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.

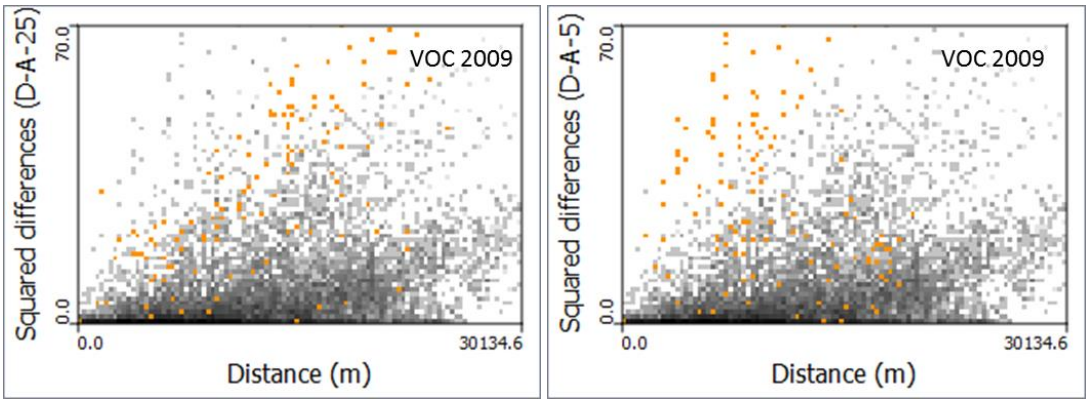


Figure 4.8. Variogram cloud graph for VOC 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.

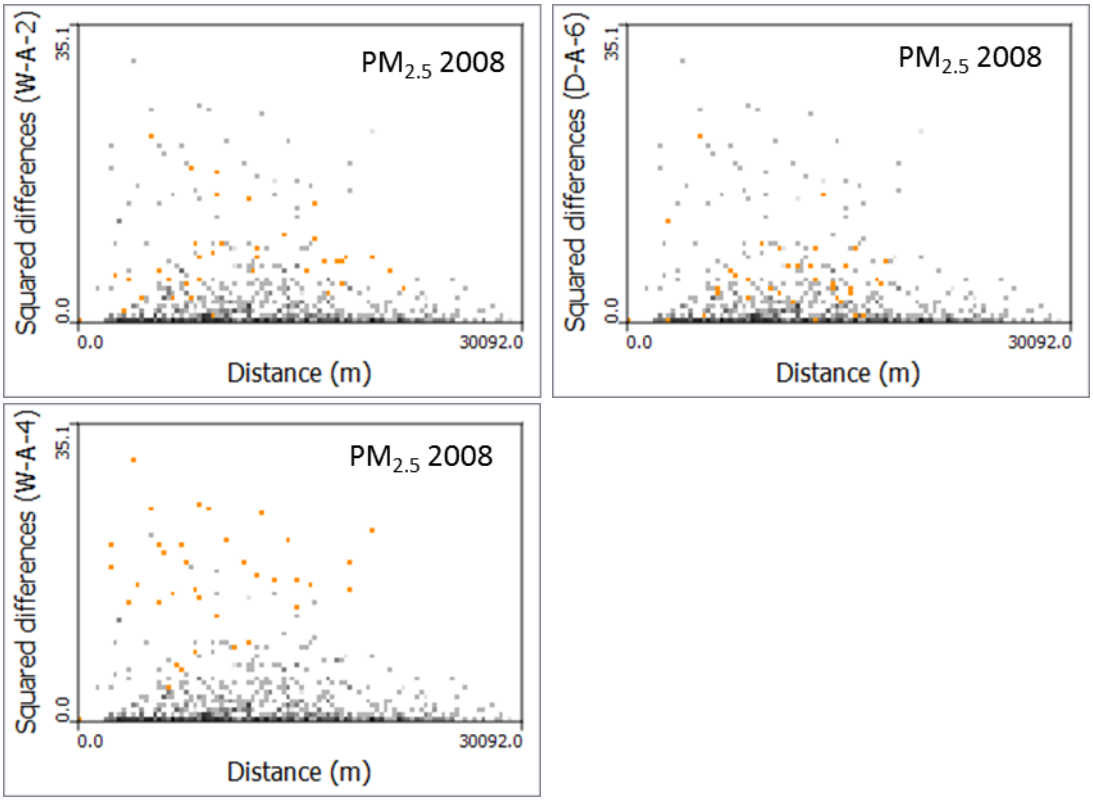


Figure 4.9. **Variogram cloud graph for $PM_{2.5}$ 2008. In each graph, the highlighted pairs are associated with a potential spatial utleir identified through the box plot method. Distance in meters.**

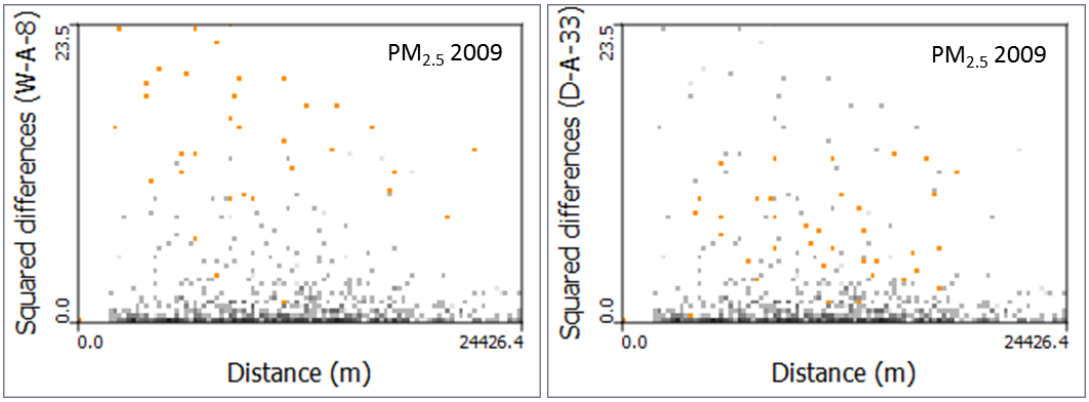


Figure 4.10. **Variogram cloud graph for $PM_{2.5}$ 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.**

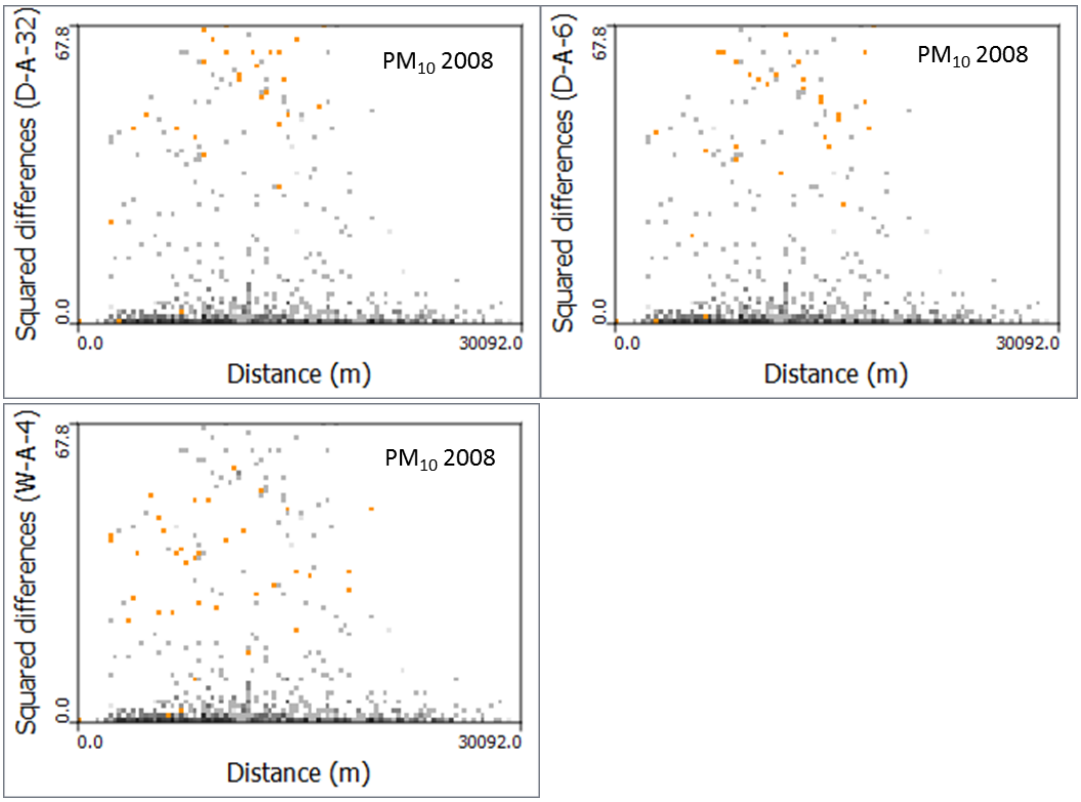


Figure 4.11. **Variogram cloud graph for PM_{10} 2008. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.**

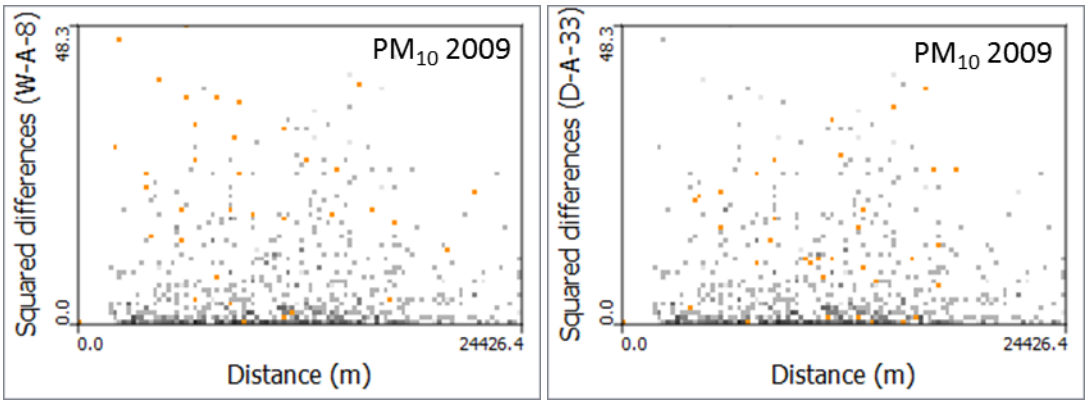


Figure 4.12. **Variogram cloud graph for PM_{10} 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.**

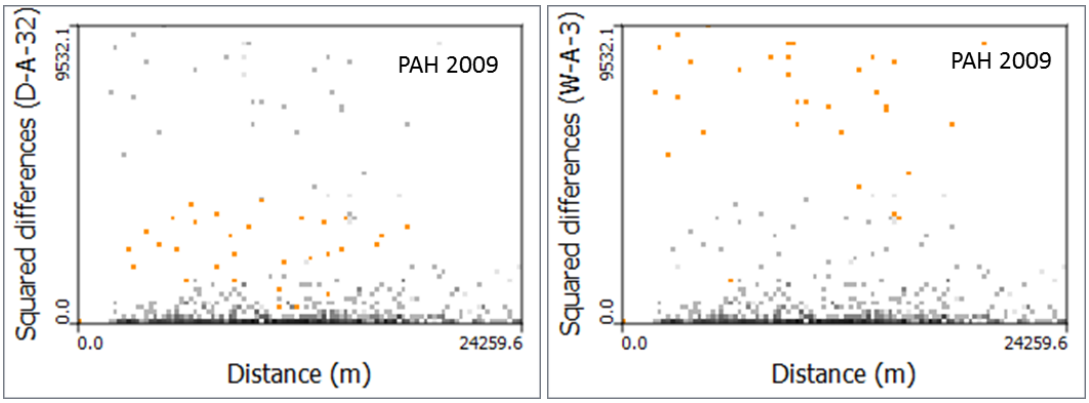


Figure 4.13. Variogram cloud graph for PAH 2009. In each graph, the highlighted pairs are associated with a potential spatial outleir identified through the box plot method. Distance in meters.

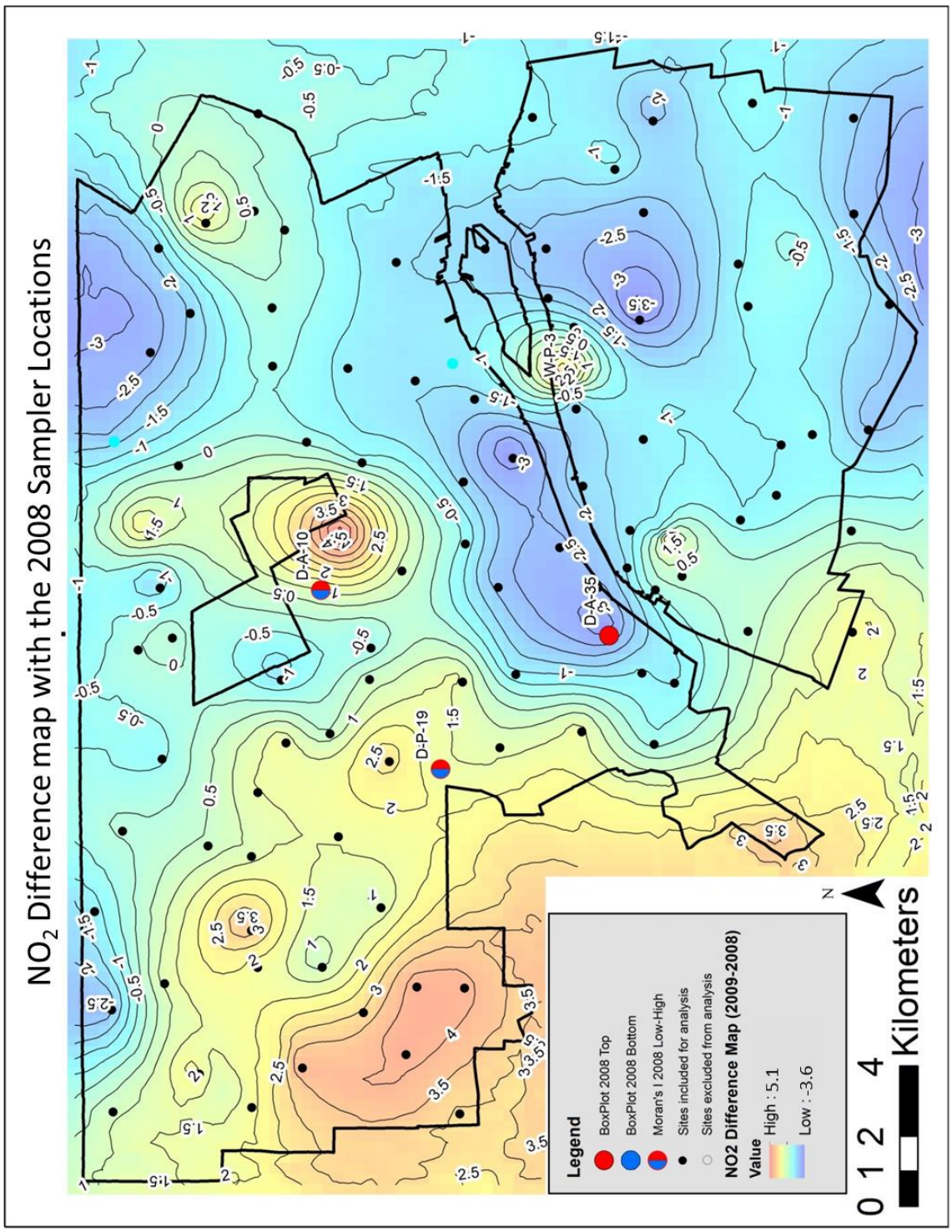


Figure 4.14a. NO_2 **difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in ppb.**

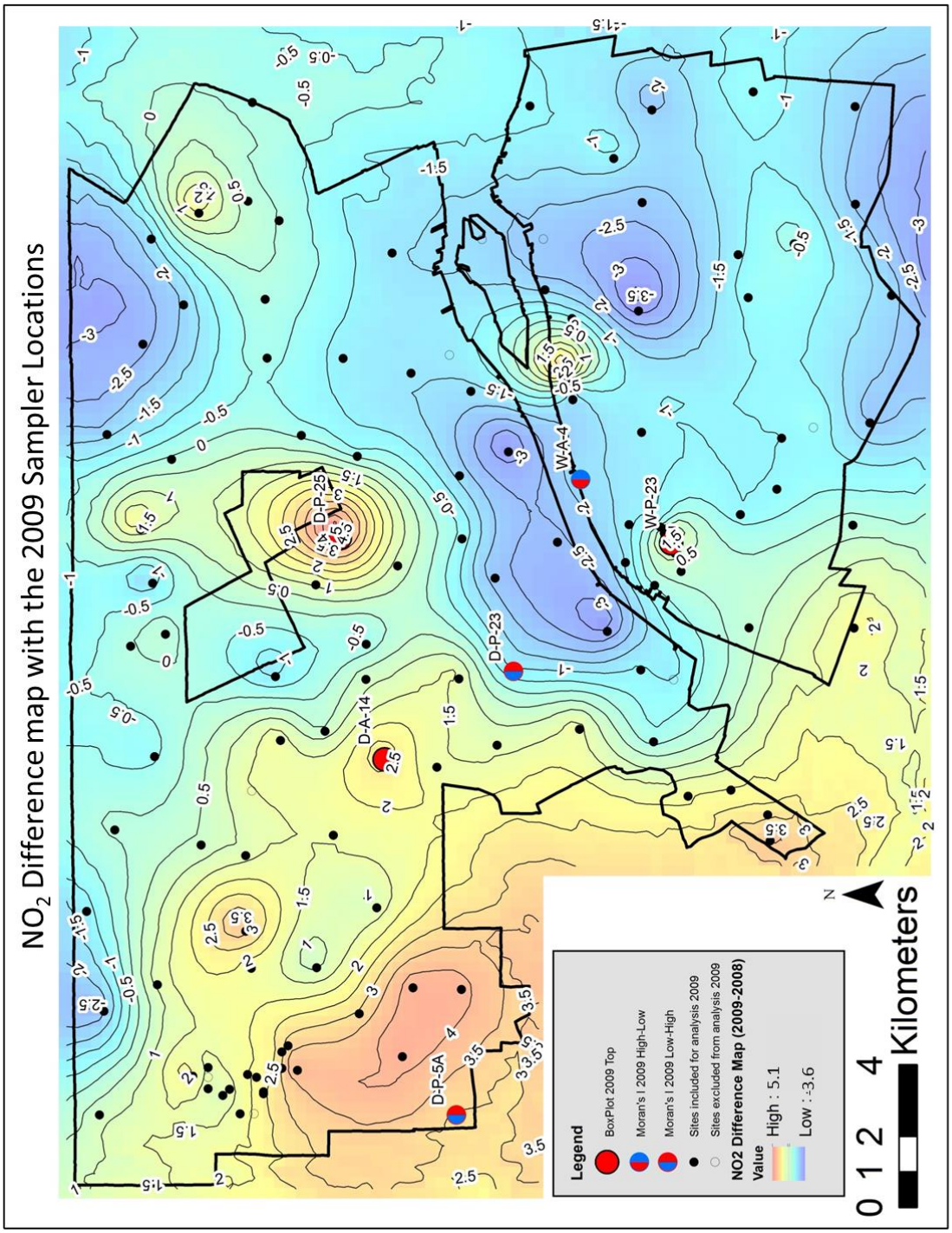


Figure 4.14b. NO_2 **difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in ppb.**

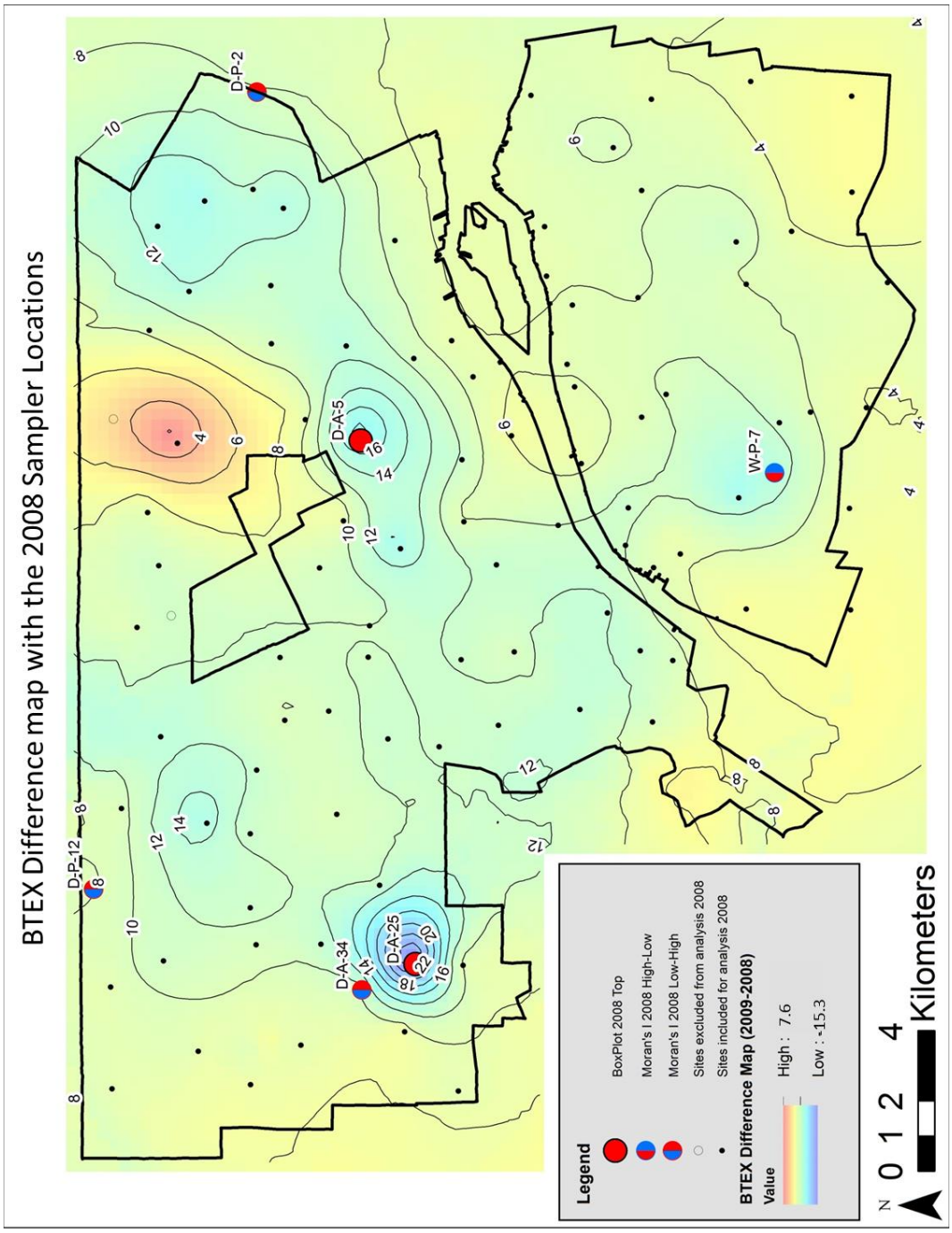


Figure 4.14c. BTEX difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.

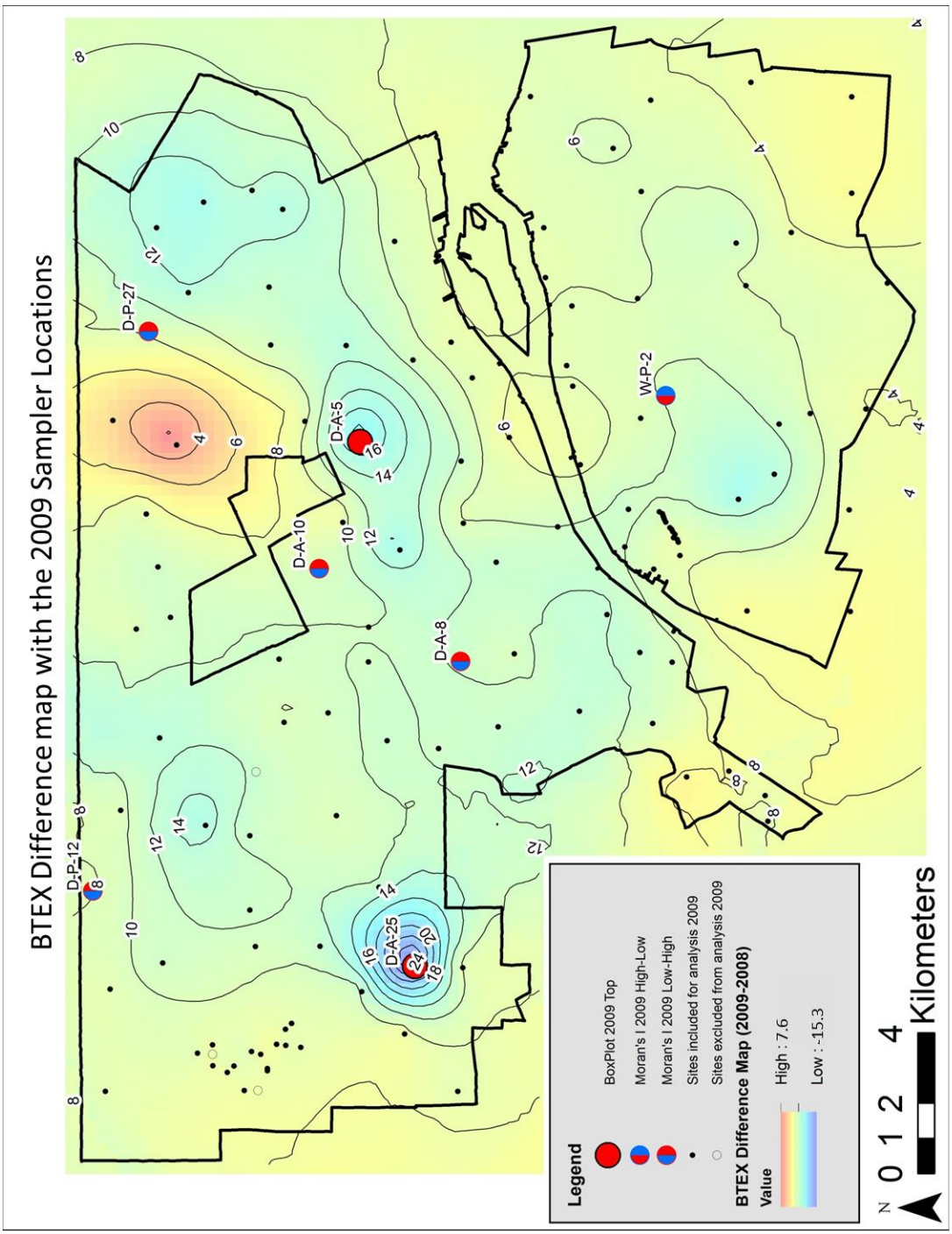


Figure 4.14d. BTEX difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.

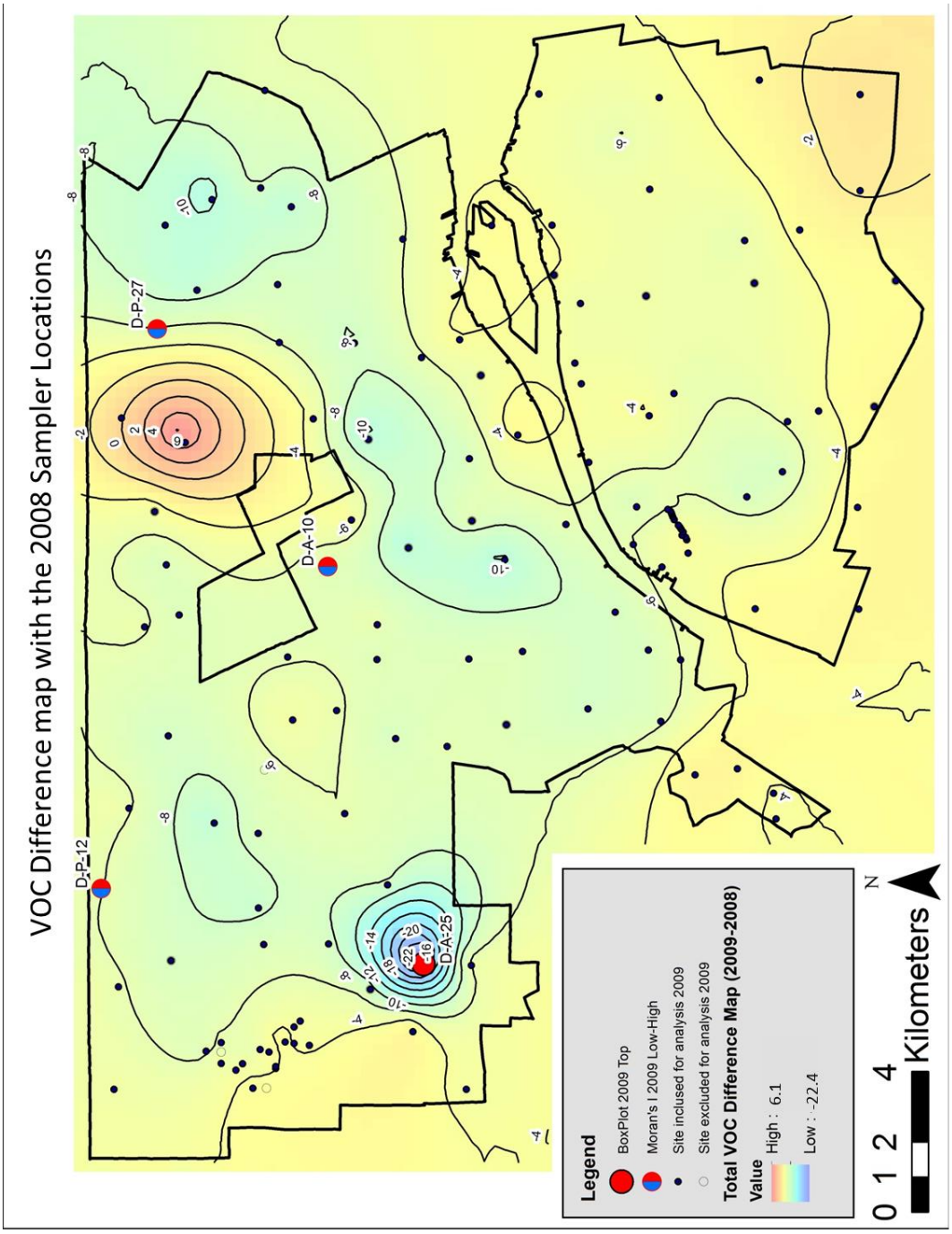


Figure 4.14e. VOC difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.

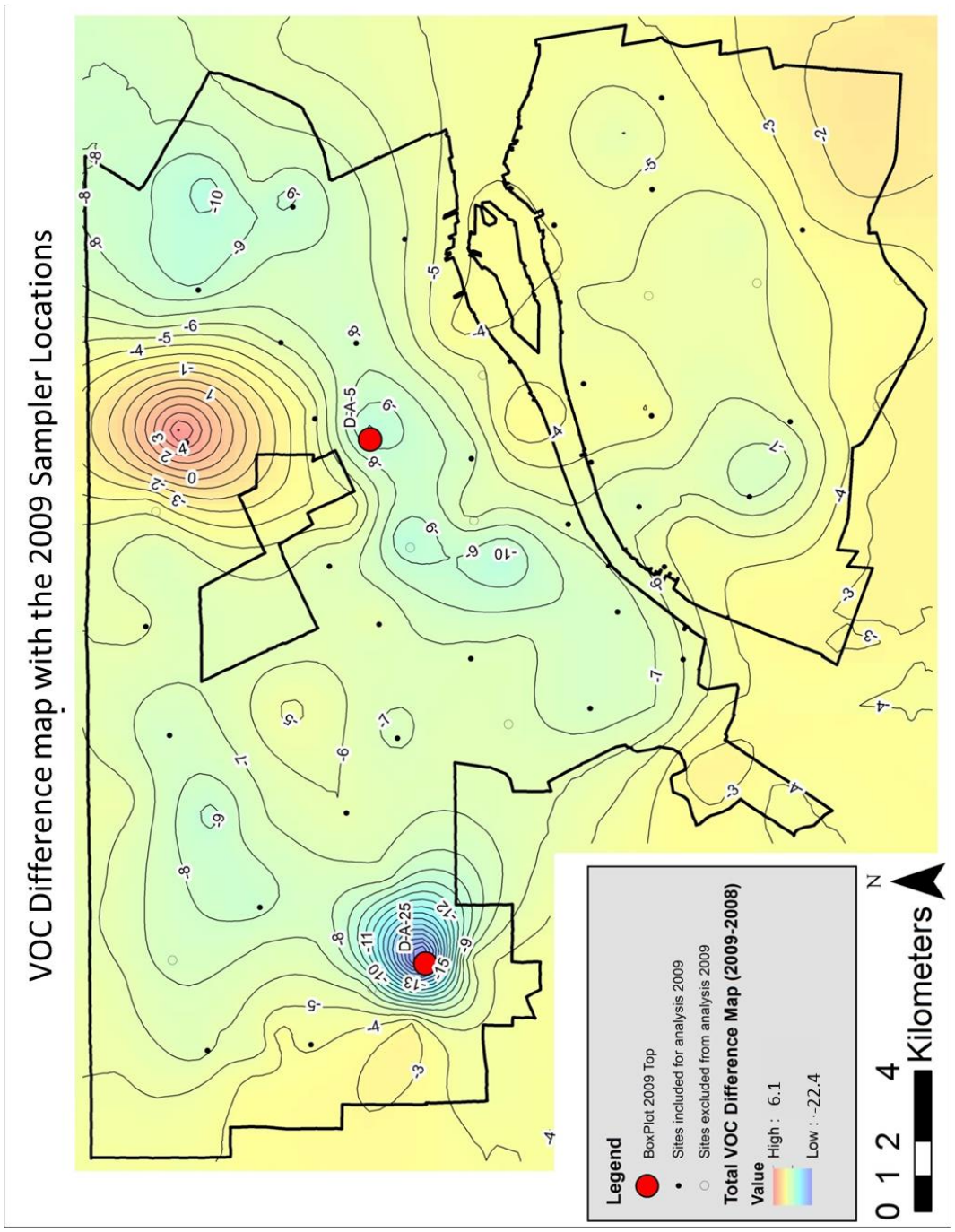


Figure 4.14f. VOC difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red. Concentration values are in µg/m³.

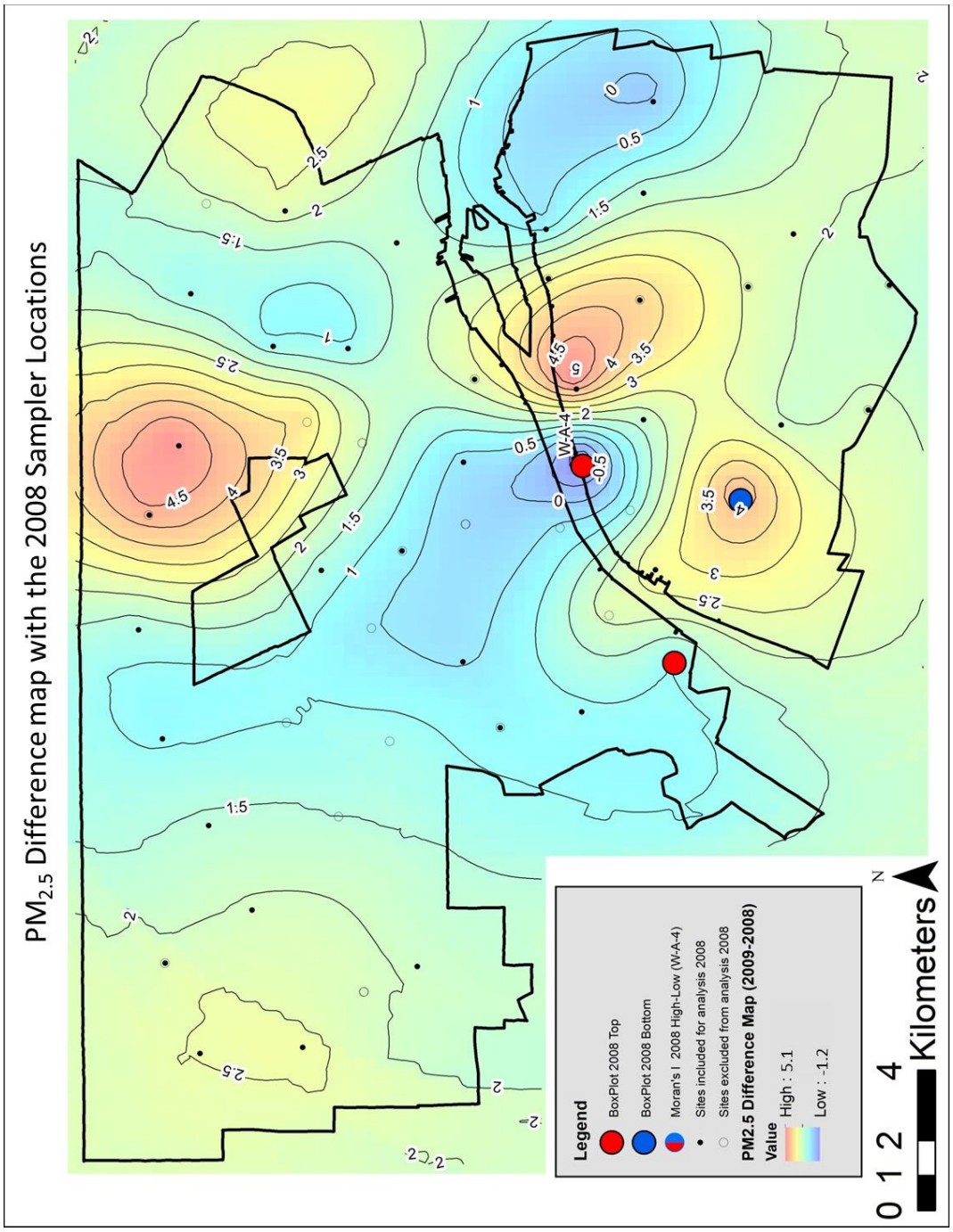


Figure 4.14g. $PM_{2.5}$ **difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in** $\mu g/m^3$.

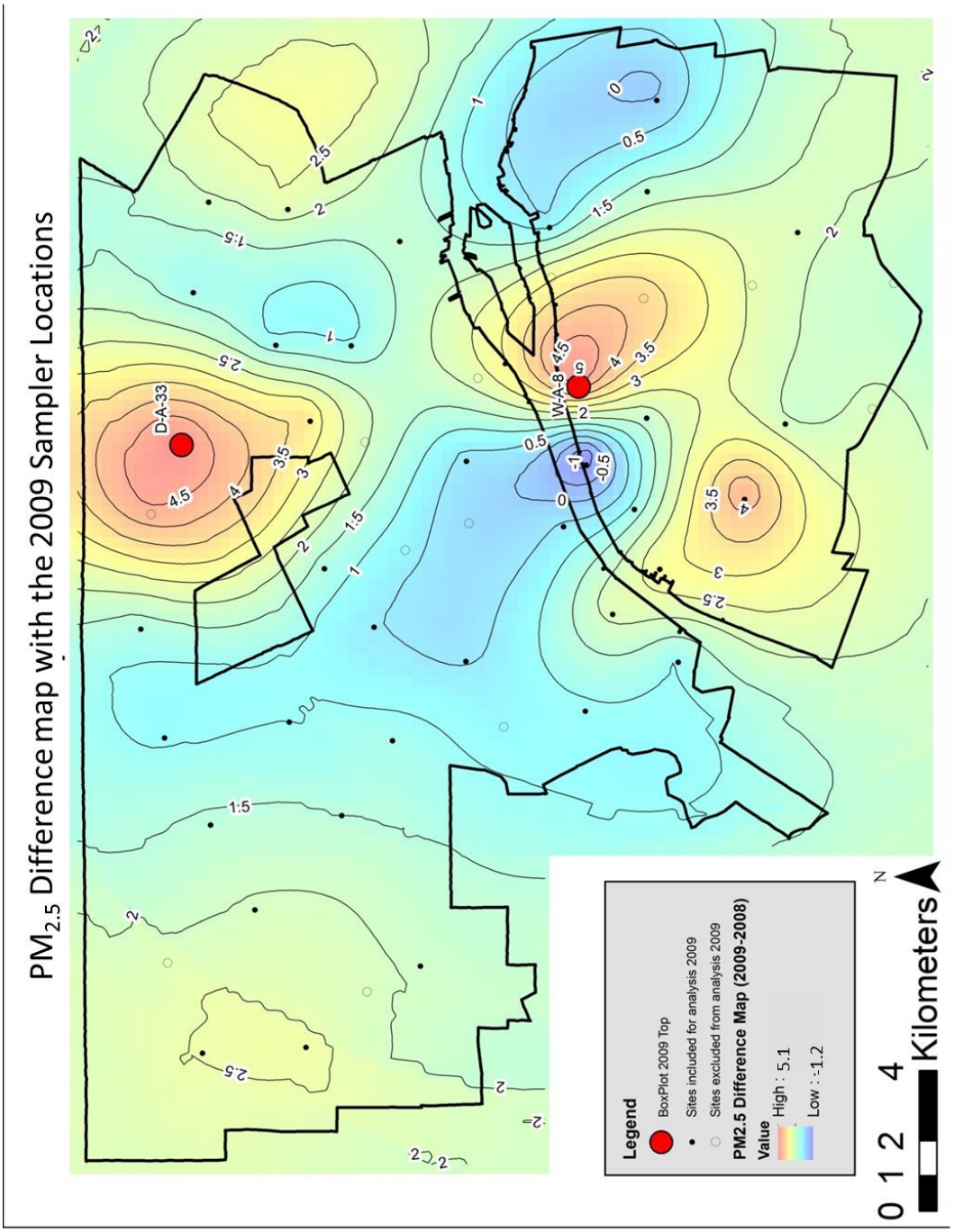


Figure 4.14h. **$PM_{2.5}$ difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.**

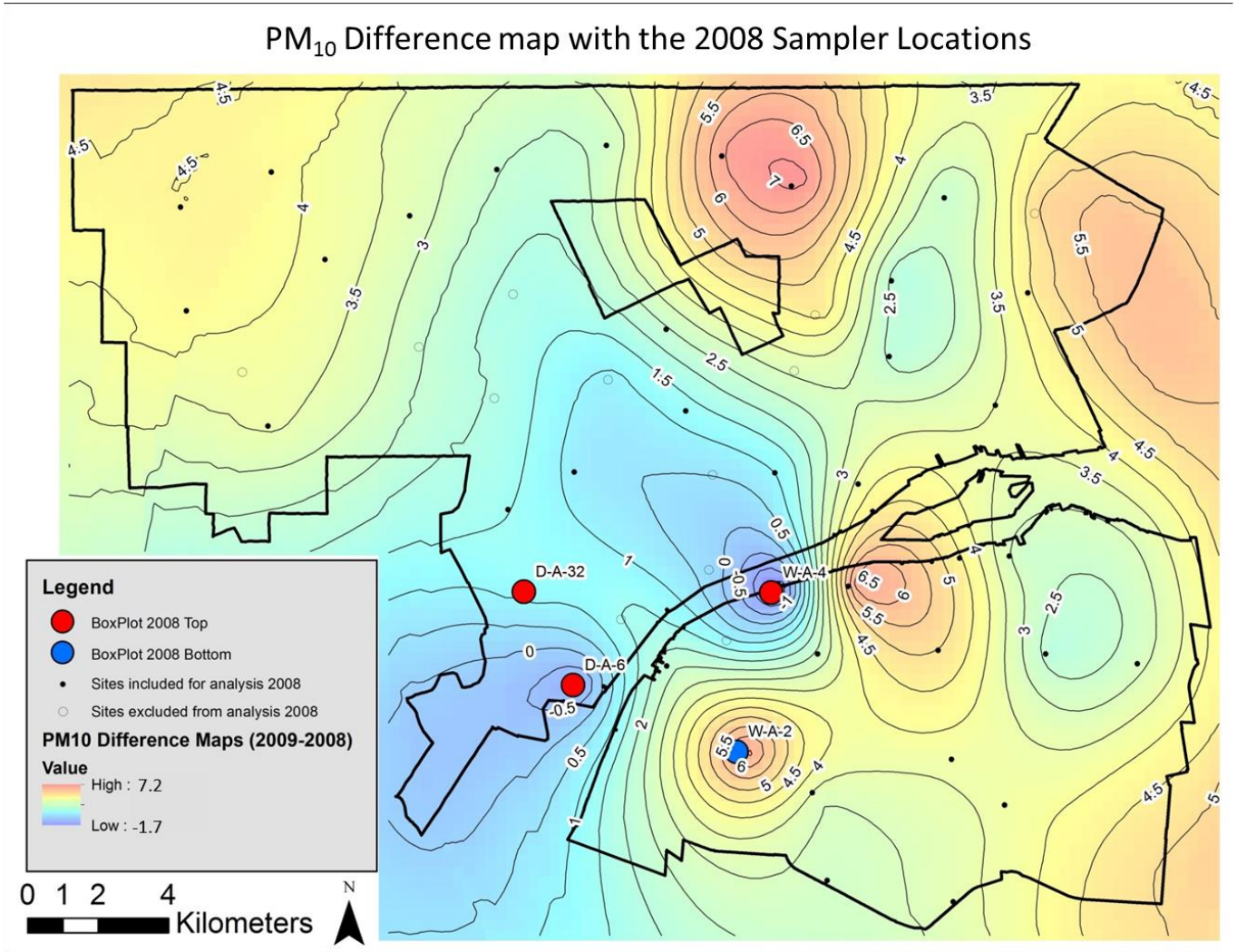


Figure 4.14i. **PM_{10} difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.**

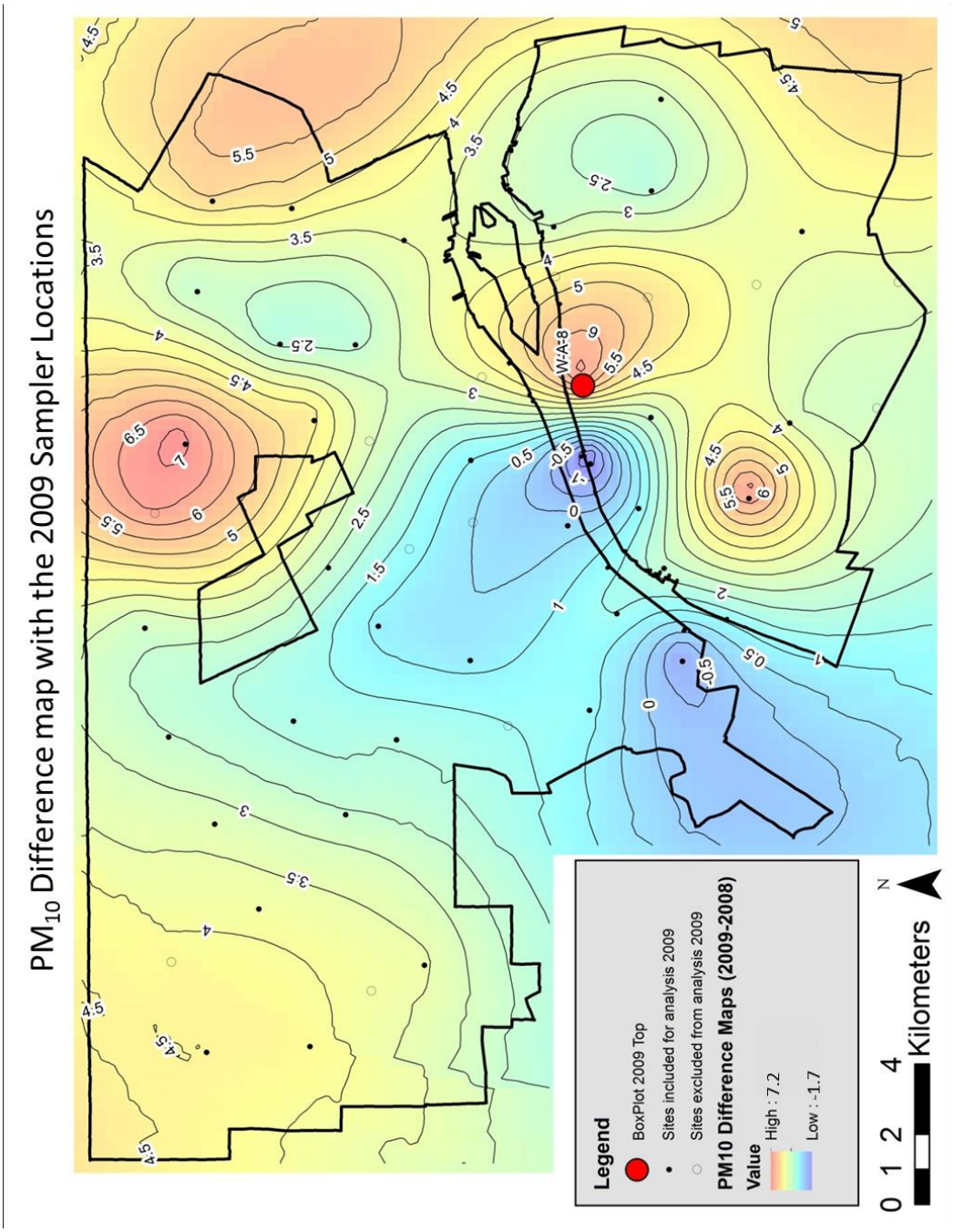


Figure 4.14j: PM_{10} difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.

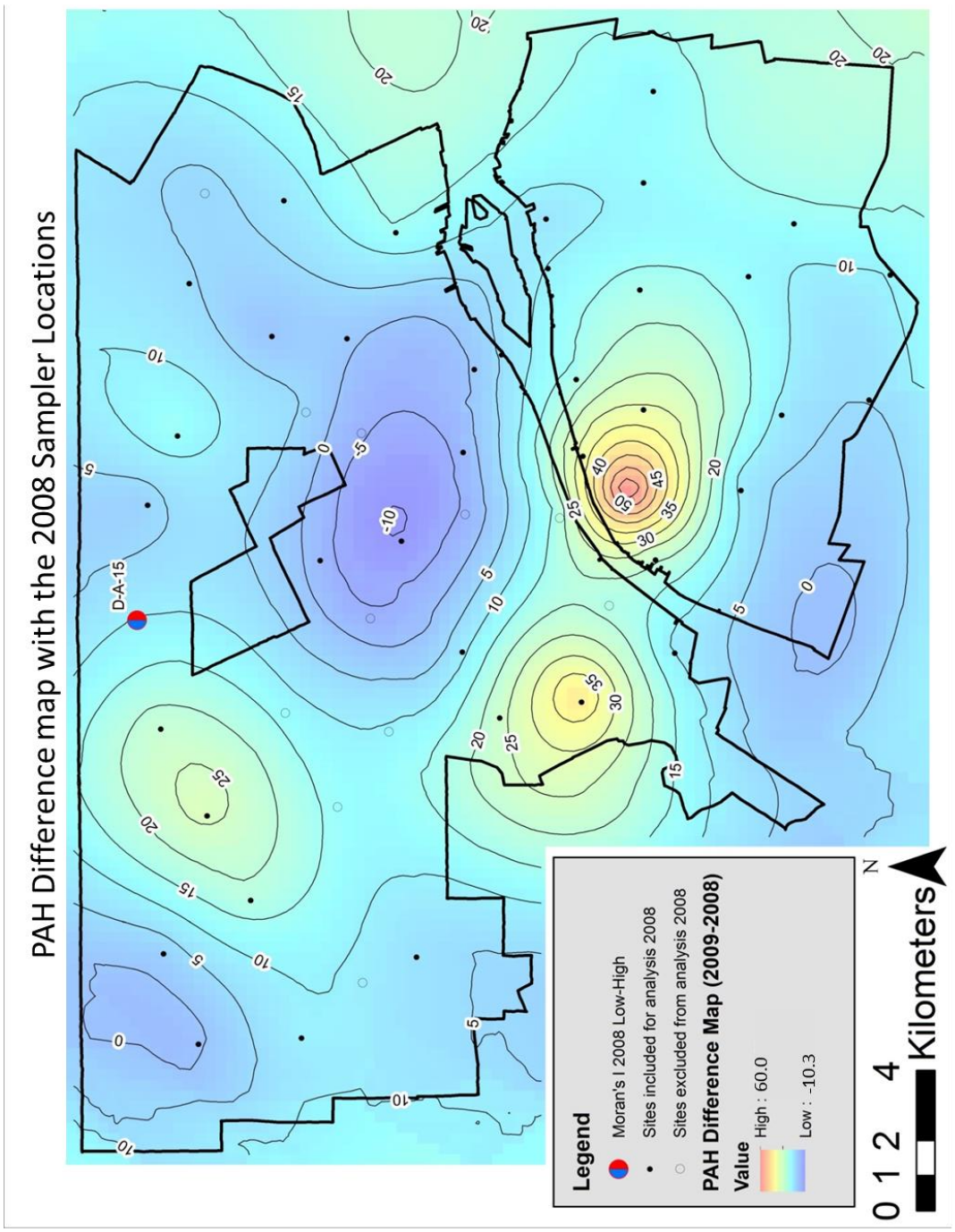


Figure 4.14k. PAH difference map with the September 2008 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m^3.

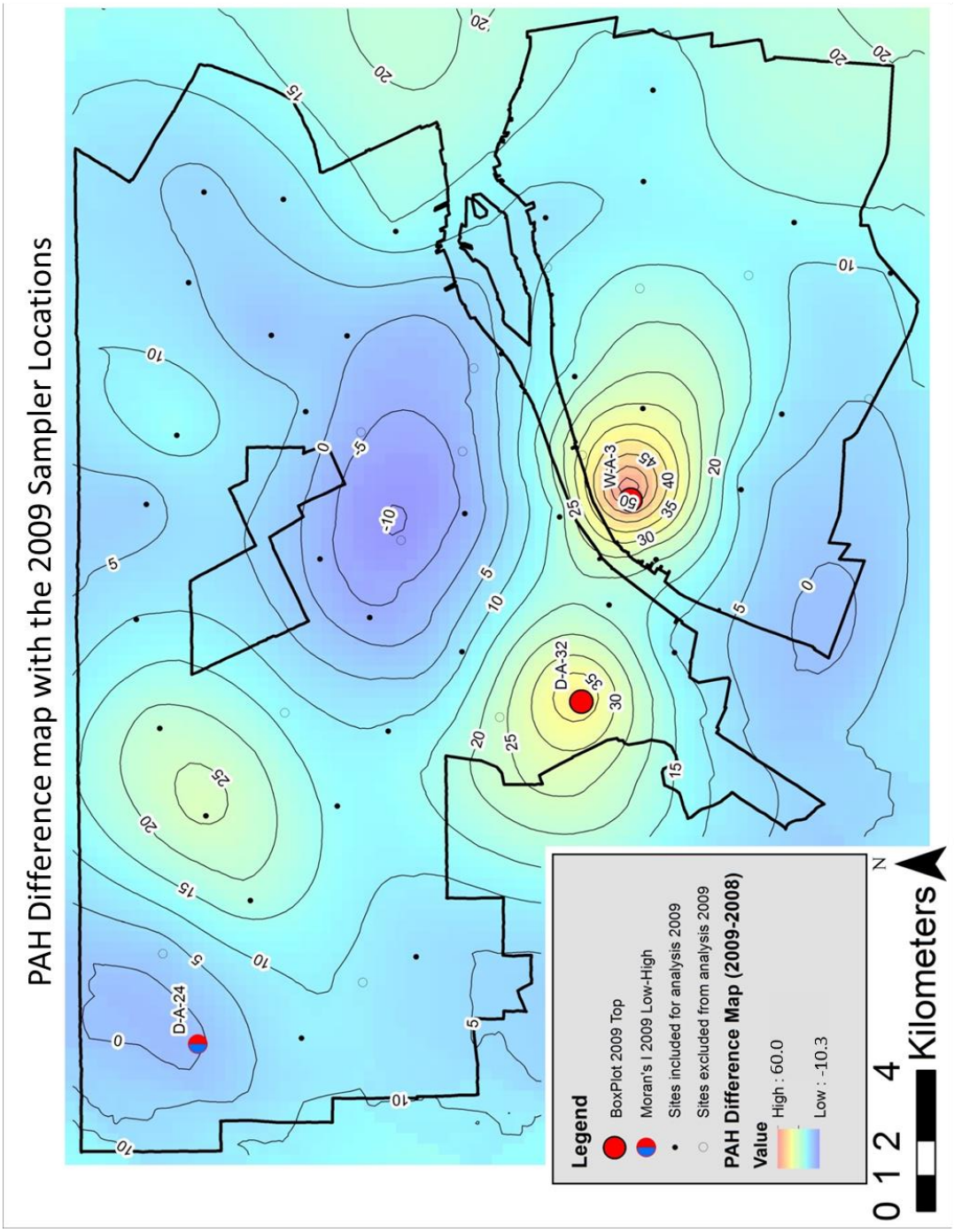


Figure 4.14l. PAH difference map with the September 2009 samples points plotted and potential spatial outliers highlighted in red or blue. Concentration values are in µg/m³.

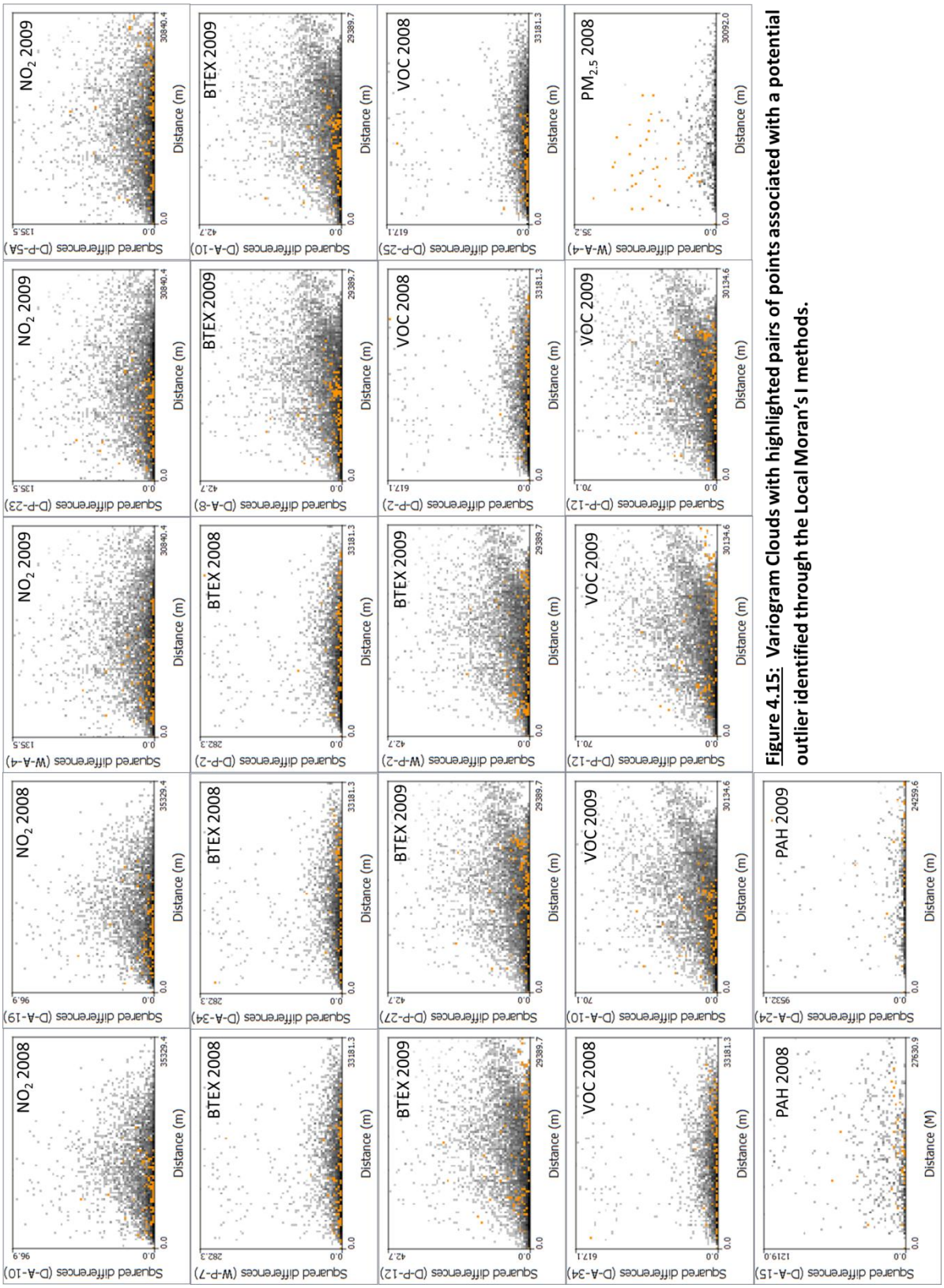


Figure 4.15: Variogram Clouds with highlighted pairs of points associated with a potential outlier identified through the Local Moran's I methods.

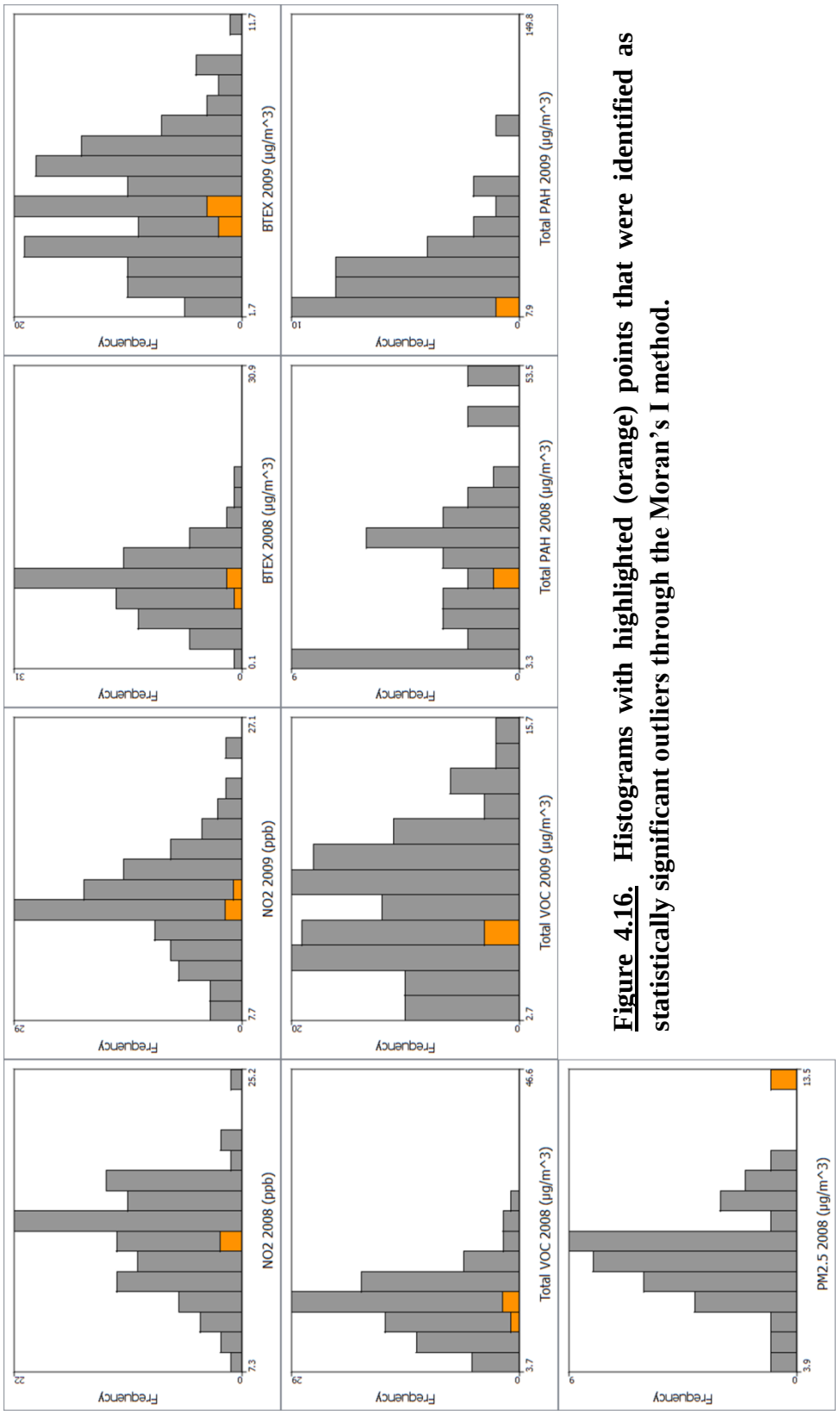


Figure 4.16. Histograms with highlighted (orange) points that were identified as statistically significant outliers through the Moran's I method.

Figure 4.17a.
Ordinary kriged maps for NO_2 2008. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in ppb.

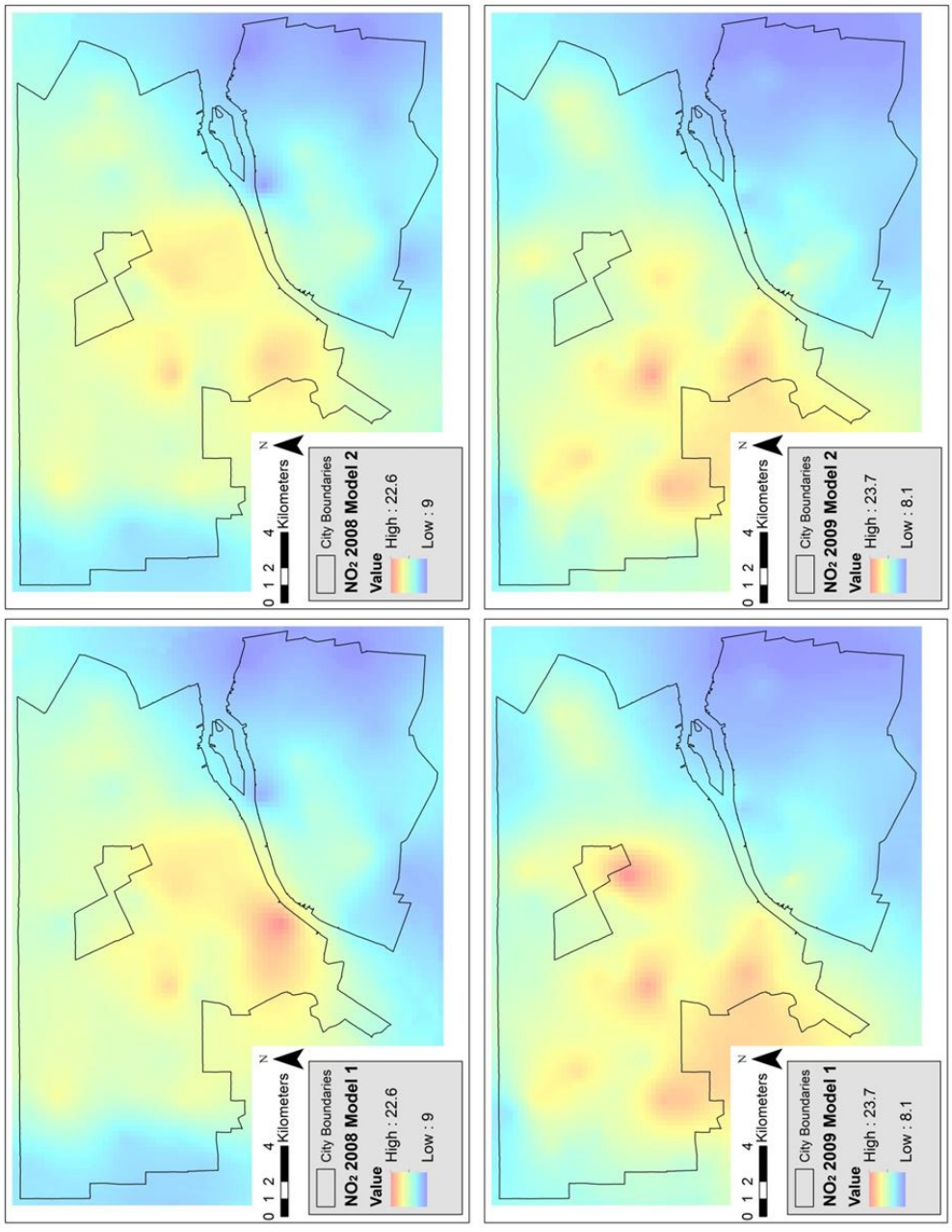


Figure 4.17b.
Ordinary kriged maps for NO_2 2009. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in ppb.

Figure 4.17c. Ordinary kriged maps for BTEX 2008. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in µg/m³.

Figure 4.17d. Ordinary kriged maps for VOC 2008. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in µg/m³.

Figure 4.17e. Ordinary kriged maps for $PM_{2.5}$ 2008. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in µg/m³.

Figure 4.17f. Ordinary kriged maps for $PM_{2.5}$ 2009. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in µg/m3.

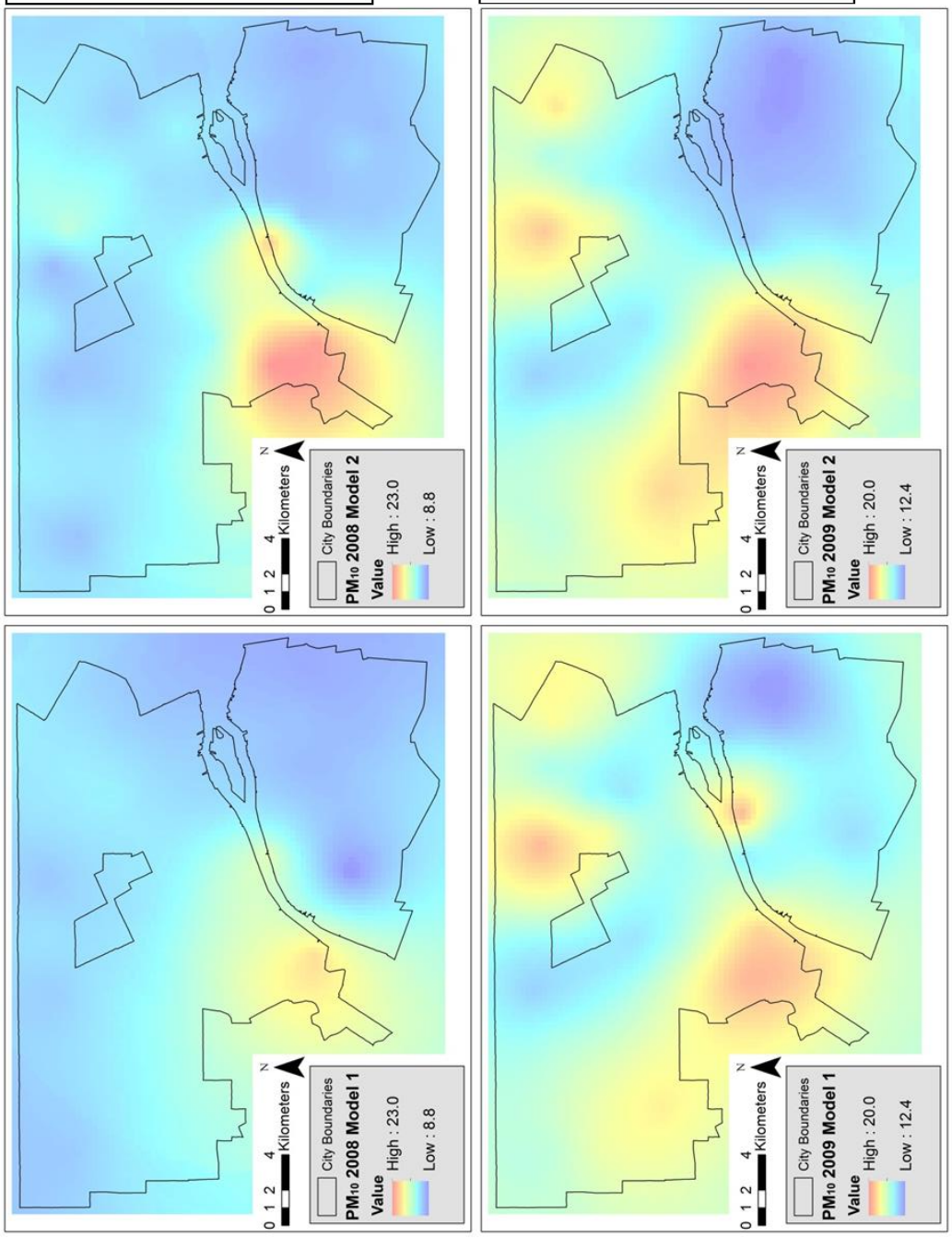


Figure 4.17g. Ordinary kriged maps for PM_{10} 2008. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in $\mu g/m^3$.

Figure 4.17h. Ordinary kriged maps for PM_{10} 2009. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in $\mu g/m^3$.

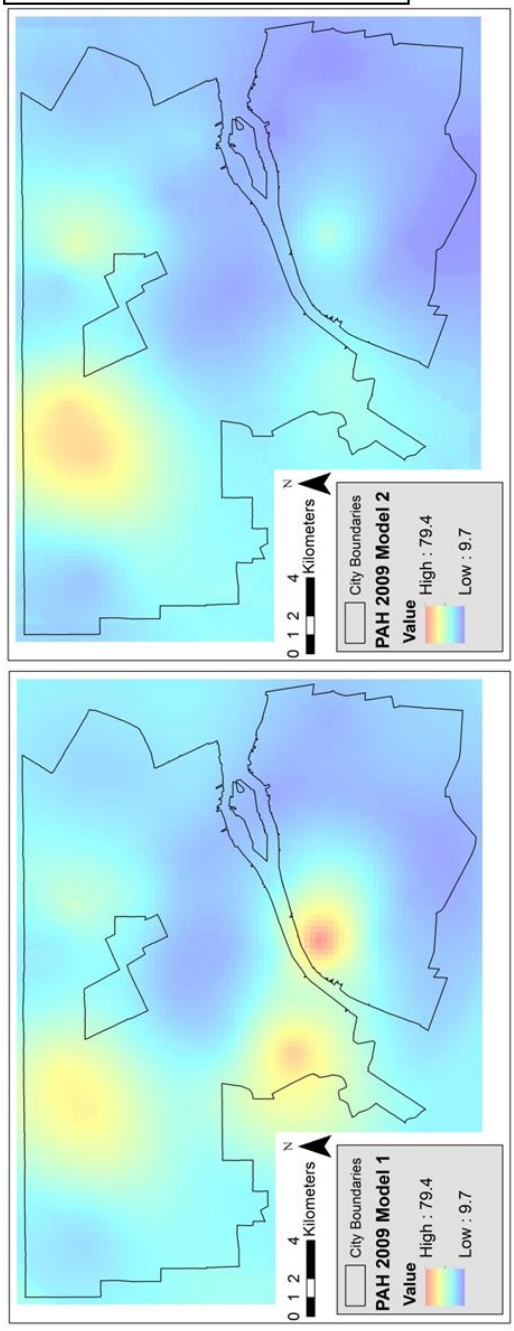


Figure 4.17i. Ordinary kriged maps for PAH 2009. Model 1 is on the left is Model 2 is on the right. The key is the same for each map and concentration values are in µg/m³.

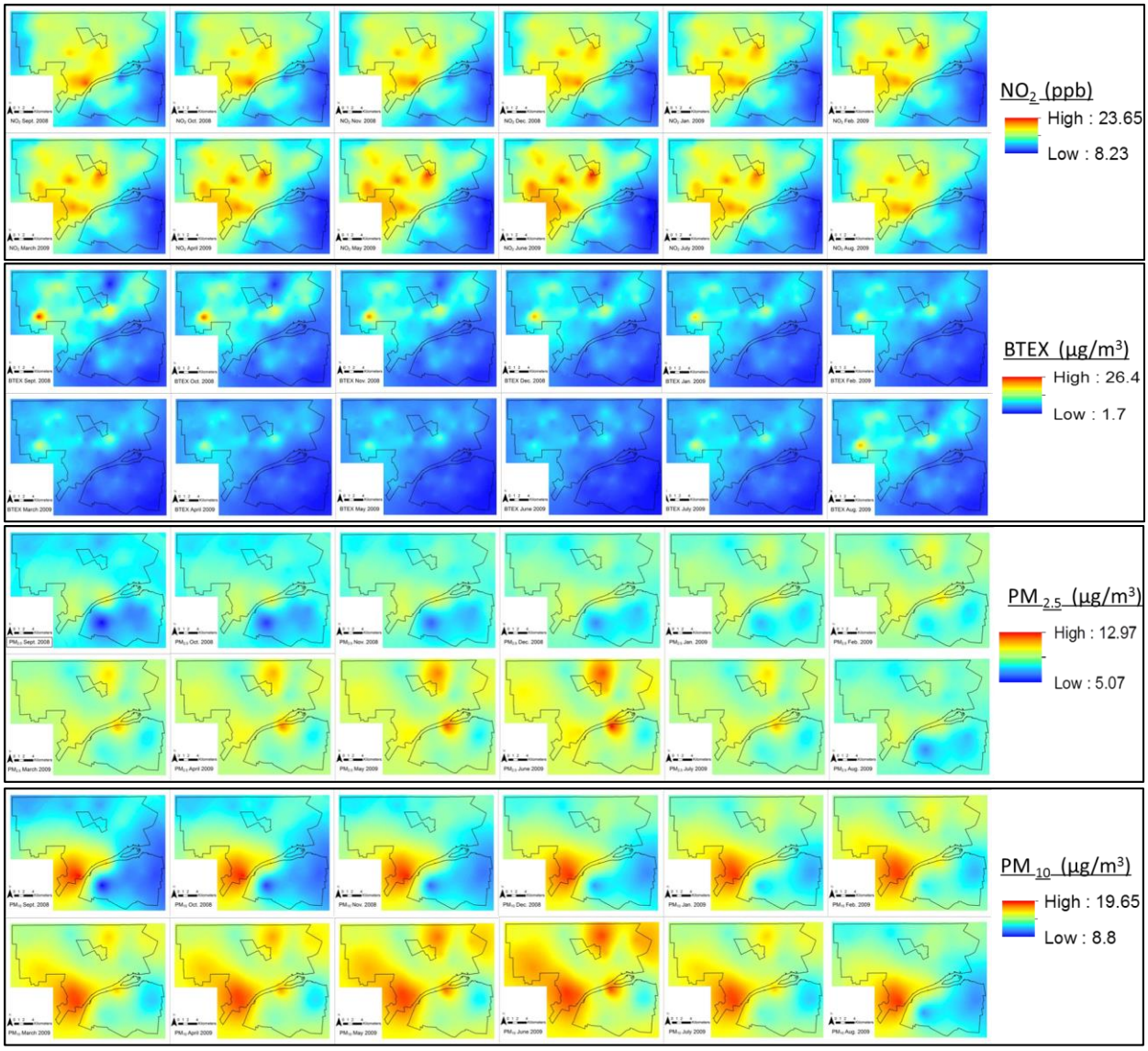


Figure 4.18a. Unadjusted 12-month Model 1 concentrations for NO_2, total BTEX, $PM_{2.5}$, and PM_{10} for September 2008 (upper left) through August 2009 (lower right) in each set.

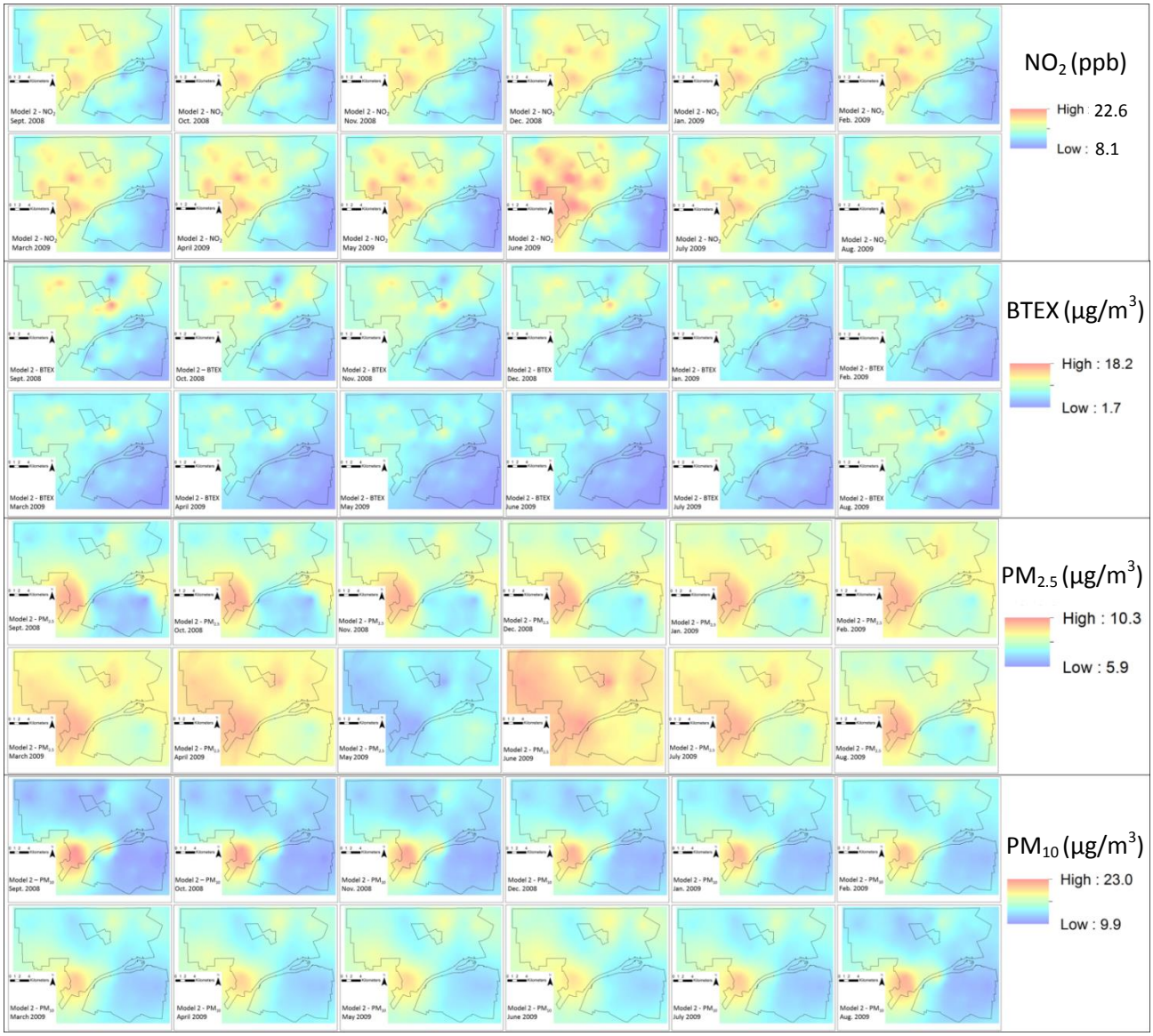


Figure 4.18b. Unadjusted 12-month Model 2 concentrations for NO_2, total BTEX, $PM_{2.5}$, and PM_{10} for September 2008 (upper left) through August 2009 (lower right) in each set.

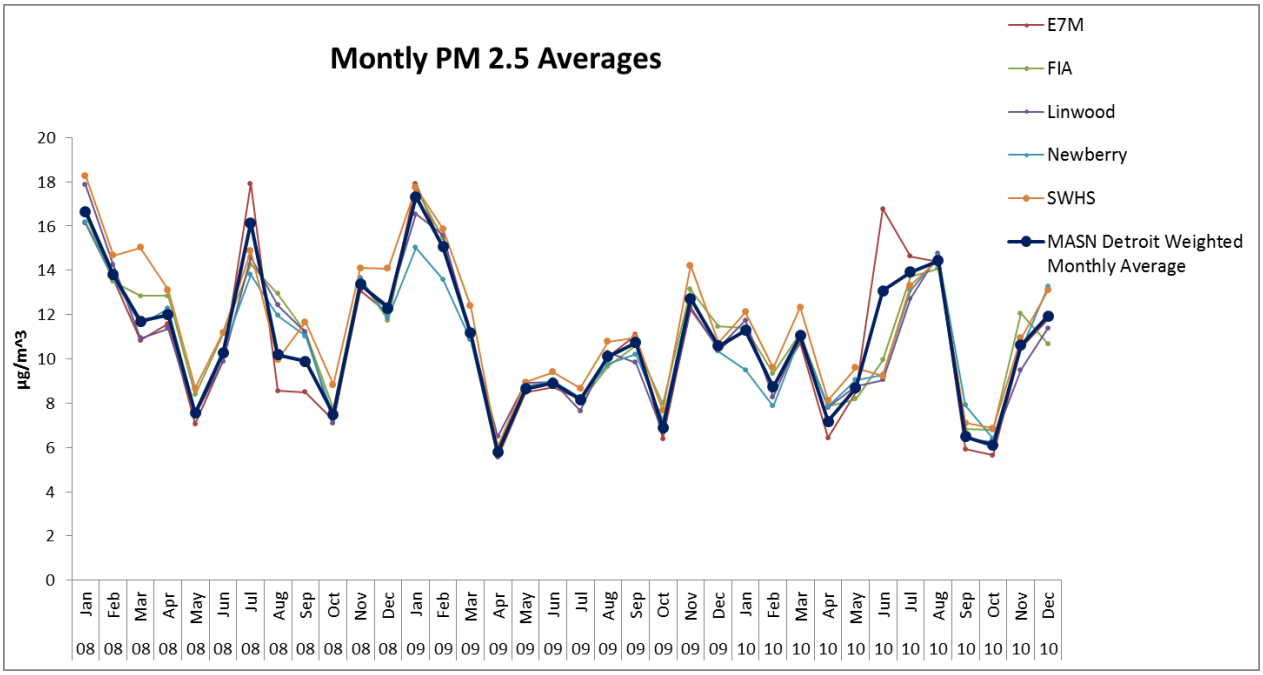


Figure 4.19. MASN Monthly average $PM_{2.5}$ values at five individual monitoring sites and declustered Detroit monthly average.

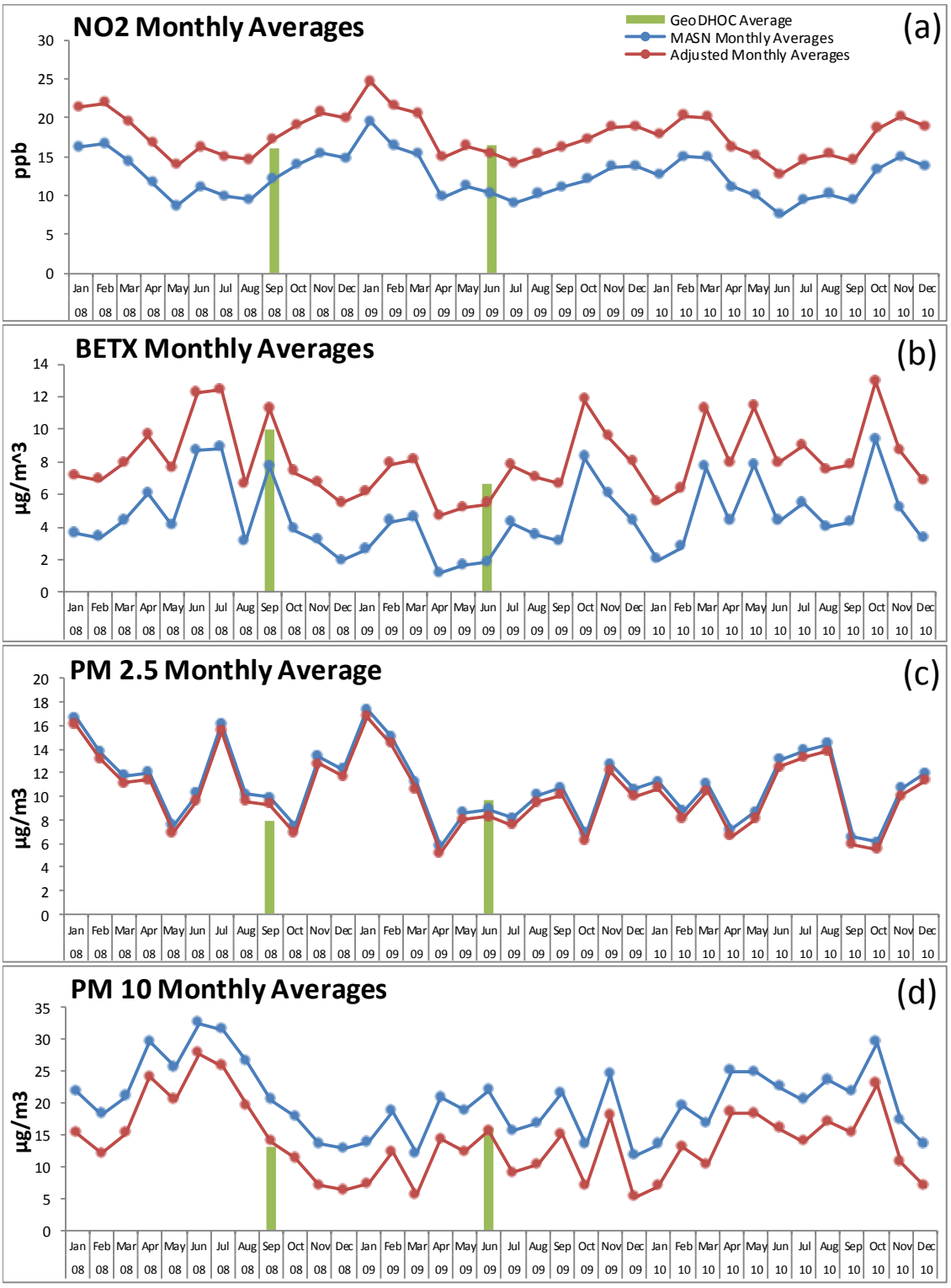


Figure 4.20a. **MASN averages and adjusted monthly averages for Model 1 for years 2008-2010: (a) NO_2, (b) total BTEX, (c) $PM_{2.5}$, and (d) PM_{10}. Unadjusted Model 1 GeoDHOC monthly averages for September 2008 and June 2009 shown as bars.**

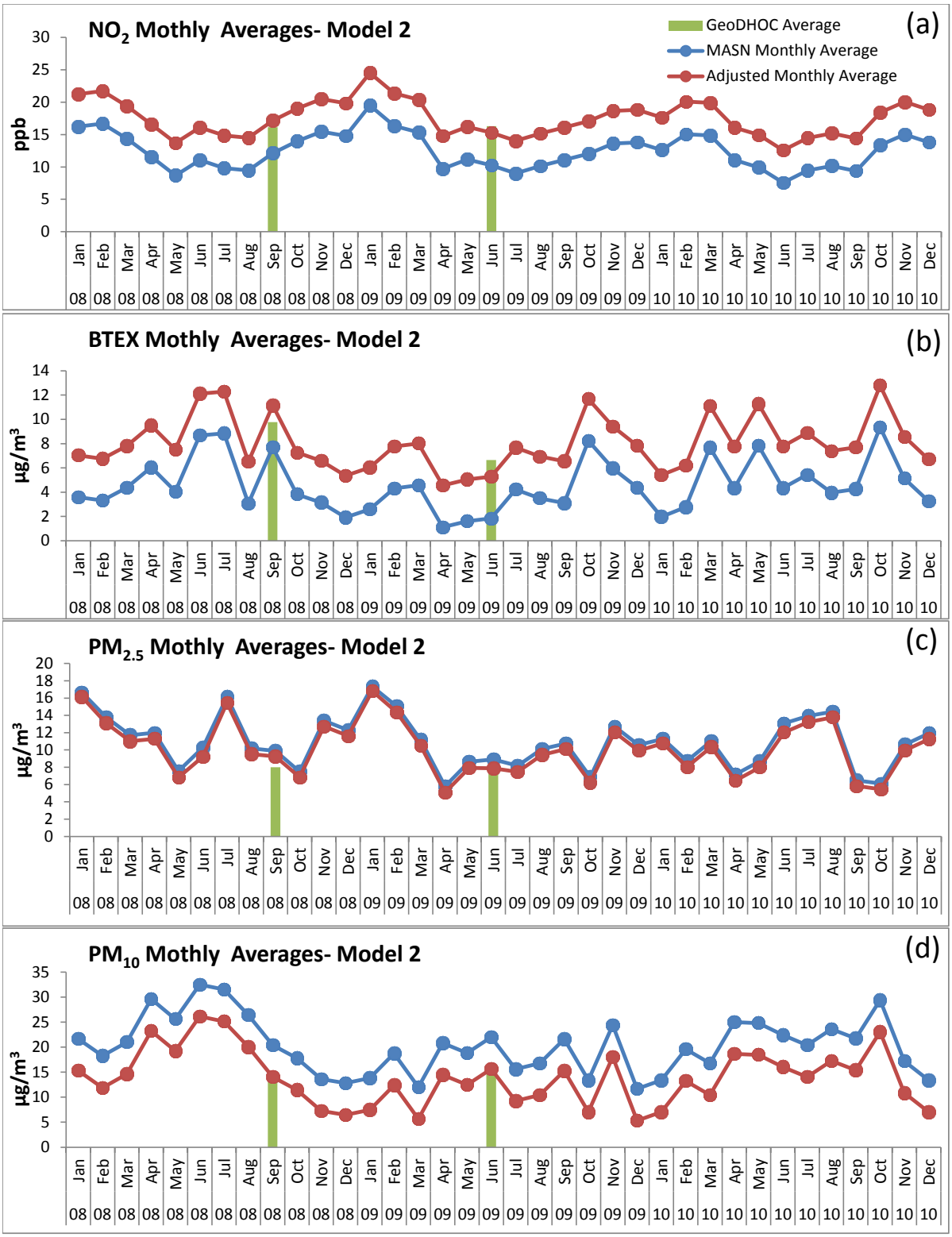


Figure 4.20b. **MASN averages and adjusted monthly averages for Model 2 for years 2008-2010: (a) NO_2, (b) total BTEX, (c) $PM_{2.5}$, and (d) PM_{10}. Unadjusted Model 2 GeoDHOC monthly averages for September 2008 and June 2009 shown as bars.**

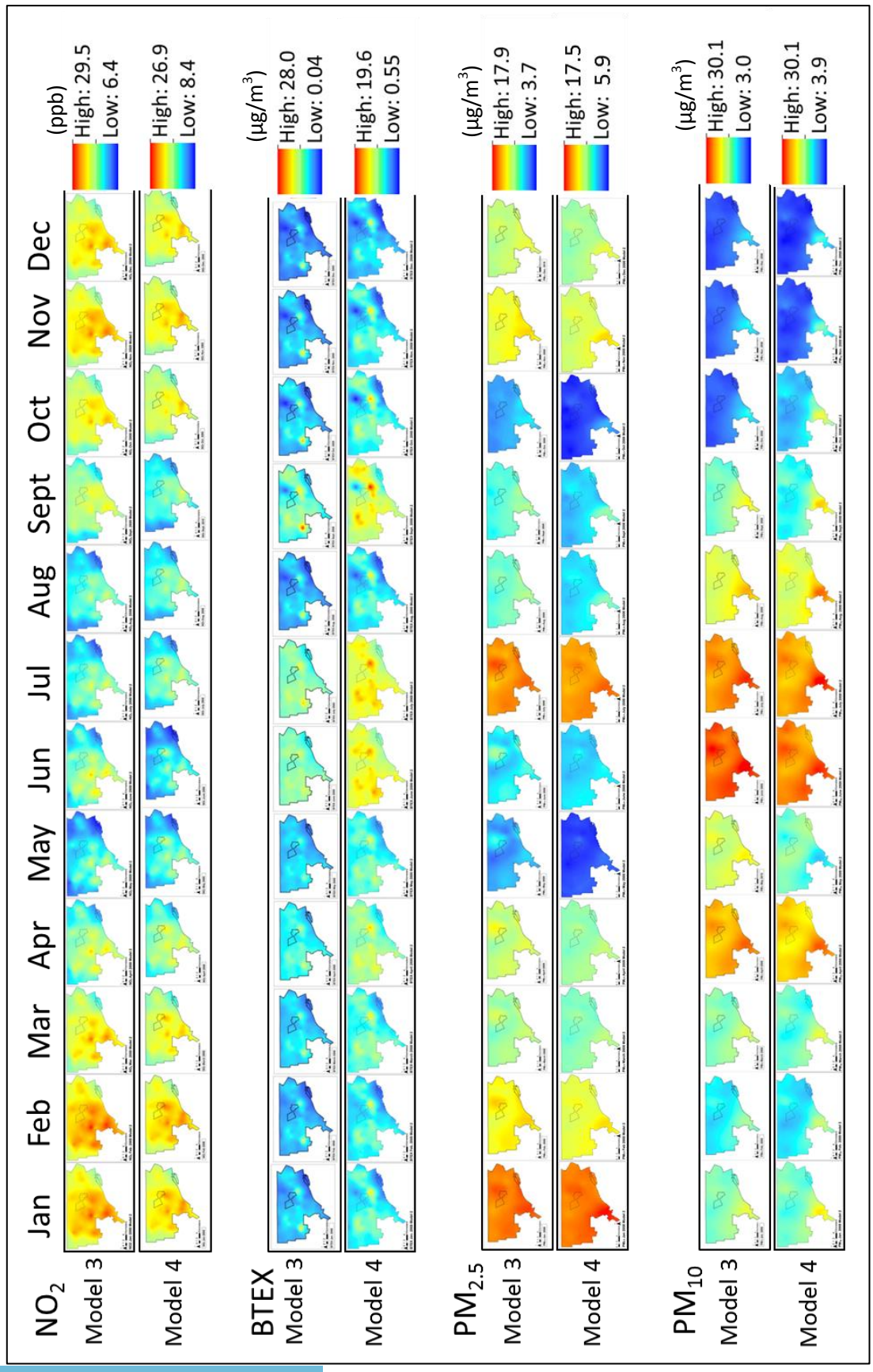


Figure 4.21. Monthly NO_2, total BTEX, $PM_{2.5}$, and PM_{10} concentrations across the Detroit airshed for 2008 through 2010.

CHAPTER 5. DISCUSSION

5.0 Introduction

Epidemiological studies of health outcomes related to chronic exposure to air pollutants increasingly rely upon models of urban air pollution to interpret complex urban airsheds. Practical methods are needed to assimilate short-term, spatially resolved air pollution measurements with widely spaced, long-term time series data to address chronic exposure. Incorporation of dense spatial information from numerous monitoring points into highly resolved models also increases the potential for erroneous measurement and intensifies the need for spatial outlier identification methods.

This thesis expands on the GeoDHOC investigation of spatial correlations between ambient air pollution concentration and asthma exacerbations in Detroit and Windsor. The discussion that follows examines the impact of outlier identification and temporal scaling to air pollutant models and subsequently, to air pollutant-asthma correlations. Conclusions and recommendations for future work are presented in the final section of this chapter.

5.1 Spatial outliers

A multi-step approach was employed to identify spatial outliers in the GeoDHOC air pollutant concentration measurement datasets for September 2008 and June 2009. In total, four alternative methods were evaluated for potential outlier identification. Unlike previous studies which relied upon a single method to identify outliers (e.g., Miller, 2012) or did not assess effects of outlier removal (e.g., Clougherty et al., 2013), this study employed multiple methods

and evaluated the consequences of outlier removal. Thus, the approach followed here arguably strengthens the overall interpretations of spatial outlier identification and influence.

The box plot method provided a simple, aspatial approach to the identification of extreme concentration values. The box plot is limited to the identification of global outliers. The majority of the sampling points identified by the box plot as outliers were extreme high values. This is not surprising because the bottom box plot whisker fell below zero in approximately half the measurement datasets (**Table 4.1**). This is a common occurrence and should be expected in environmental data that are bounded by zero. Because all observed values measured air pollution exceed zero, there were only two sampler locations measured as extremely low.

The variogram cloud is a more qualitative outlier identification method because, unlike the box plot, there are no specified boundaries within the graph to define what constitutes a spatial outlier. For this reason, the process of identifying spatial outliers with the variogram cloud is subjective. As a result, only sampling points identified using the other three methods were analyzed with the variogram cloud. For the datasets evaluated in this study, box plots and variogram clouds complemented each other well with the box plot providing definite boundaries and the variogram cloud contributing a spatial component (x-axis). In contrast, potential outliers identified with the local Moran's I did not correspond well to variogram cloud results.

Difference maps provide a comparison of modeled air pollutant concentrations over the same area during separate measurement events. In addition, the difference maps supply a visual summary tool facilitating additional evaluation of box plot and Moran's I results within a geospatial context. Each potential outlier identified by the box plot and the Moran's I was plotted on the difference maps to facilitate comparisons among these three methods. This

approach highlighted relationships between box plot and Moran's I outlier locations and significant difference gradients shown on the difference maps. In this study, box plot outliers were found to correspond well with areas of high concentration contrasts on the difference maps. Conversely, potential outliers identified using the Moran's I did not correspond to areas of high differences.

The Local Moran's I output identified 22 statistically significant points with high-low or low-high spatial structure. However, with the exception of one sampler location, these points did not correspond well with outliers found using any of the other methods. The sampler points identified with the Moran's I technique were not located in areas of large differences on the difference maps. These points also did not show patterns of local outliers on the variogram clouds (**Figure 4.15**). Furthermore, discounting the one exception (W-A-4 for $PM_{2.5}$ 2008), the Moran's I outliers fell in the middle of the histogram (**Figure 4.16**). This indicated that each identified outlier's concentration was within an expected range for the global dataset and were not global outliers.

Unlike the other three methods, the Local Moran's I did not take into account the entire dataset which is a potential reason for its lack of agreement with the other methods. Rather, it focused on a specific subset of the dataset, in this case 8 neighbors, to identify localized spatial differences. In addition, spatial weighting was 'standardized to neighbor count' which assigned uniform spatial weights to all neighbors in the Moran's I test. This option, as the default method in SpaceStat, provided a simple weighting method in the initial study of the outliers. Given a uniform spatial weight, sampler locations with extreme concentrations were expected to show up as statistically significant in the Moran's I test. However, this was not the case, which is potentially attributable to the narrow subset of data used with the Moran's I.

An alternative explanation for poor Moran's I performance involves the air pollution measurement datasets, whose number of samplers varied with each analyte. The minimum number of measurement points varied greatly for pollutants measured with active samplers (37) and passive samplers (98) (Table 2.1). Other studies applying the Local Moran's I used larger datasets (Li et al., 2013; Zhang et al., 2008; Zhang and McGrath, 2004; Zou et al., 2014). Of the sampler points identified as potential outliers with the Moran's I, 19 of 22 were analytes measured with passive samplers. $PM_{2.5}$ 2008 W-A-4, PAH 2008 D-A-15, and PAH 2009 D-A-24 points were the only active sampler location identified as a potential outlier with the Moran's I method. W-A-4 was the only point of agreement between the Moran's I and any of the other three methods.

The $PM_{2.5}$ W-A-4 point further highlights the difference in observed values versus modeled values because it was the only outlier that was not identified with the difference map. When W-A-4 was compared to the $PM_{2.5}$ 2008 dataset, it appeared as an outlier. The box plot, variogram cloud, and local Moran's I all use the 2008 observed dataset. The histogram plot of $PM_{2.5}$ 2008 also shows W-A-4 $PM_{2.5}$ concentration at the extreme end of the dataset. The difference map differed from the other approaches in that it used the $PM_{2.5}$ 2008 and 2009 modeled values and the W-A-4 location did not show an appreciable difference between each year with a difference of -1.0 $\mu g/m^3$. In this way, the difference map showed a consistency in the modeled concentration at this location. This consistency was not present in the measured difference of 5.5 $\mu g/m^3$ which is high for $PM_{2.5}$. Therefore, the $PM_{2.5}$ W-A-4 sampler location was included as an outlier based on the measured difference.

Given each method's strengths and limitations and a comparison of the results for each method, the box plots and difference maps were considered to be the most reliable outlier

indicators, followed by variogram clouds and the Local Moran's I. The box plots provided a quantitative means to identify outliers with discrete boundaries and the difference maps provided spatial as well as a temporal component when identifying outliers. Twelve of the 13 outliers in **Table 4.2** were identified using the box plot and difference map methods. The variogram cloud proved to be a weaker approach relative to the sensitivity of the box plots and difference maps. Nine of the 13 outliers were confirmed with the variogram cloud. Variogram clouds for the four outliers not identified by the variogram cloud method (W-A-2 $PM_{2.5}$ in 2008, D-A-33 $PM_{2.5}$ in 2009, W-A-2 PM_{10} in 2008, and D-A-32 PAH in 2009) exhibited some points that deviated from the majority of pairs, but did not show strong variance over a short geographic distance (Figure 4.9, 4.10, 4.13).

Four sampler locations, W-P-23 NO_2 2009; D-A-5 BTEX 2009; D-A-32 PM_{10} 2008; and D-A-6 PM_{10} 2008, were identified with the box plot and the variogram cloud methods. These points were not chosen as outliers because they did not show a larger difference on the difference maps. In two cases, D-A-25 for BTEX and VOC, there was a large measured concentration difference between values in 2008 and 2009. D-A-25 in 2008 was an outlier for BTEX and VOC but while this large difference remained in 2009, it was no longer anomalous for the D-A-25 2009 data sample.

The Moran's I was therefore considered the weakest outlier identification method, because of its almost uniformly poor agreement with the other identification methods. Given the relative congruence of results among the box plots, difference maps, and variogram cloud methods, coupled with their collective lack of agreement with the Moran's I results, the Moran's I approach was discounted as a method of outlier identification. In the end, the remaining three spatial data outlier identification methods determined the final list of outliers.

5.2 Kriging

Removing spatial outliers changed the overall concentration values and in some instances the spatial distribution of modeled concentration values in the revised ordinary kriged models. These changes are attributable to the removal of outliers and the subsequent adjustment of the variogram models from Model 1 to Model 2. These changes had a direct influence on the global statistics of the models (**Table 4.4**) and localized changes were observed immediately around the areas where outliers were removed (**Figure 4.16**).

There was relatively good agreement between Model 1 and 2 variogram characteristics for NO_2, BTEX and VOCs. This was expected given the large number of passive sampler points such that removal of one or two outliers had a smaller influence on global geostatistical relationships, including variogram characteristics. One important change from Model 1 to Model 2 for NO_2, BTEX, and VOC was lower sill contributions (**Table 3.2**) resulting from lower variance in the measurement distributions after the outliers were removed.

Conversely, there was poor consistency between the variogram characteristics for PAH and PM from Model 1 to Model 2. $PM_{2.5}$ behaved as expected with a lower standard deviation in Model 2 when compared to Model 1 (**Table 4.4**). PM_{10} and PAH did not share this outcome and the standard deviations increased in Model 2.

Direct comparisons between Model 1 and Model 2 were possible for PAH in 2009. Model 1 had a relative nugget of 30% compared to Model 2 nugget of 11%. This difference likely contributed to the unexpected increase in the standard deviation with Model 2 when compared to Model 1.

The particulate matter models ($PM_{2.5}$ and PM_{10}) are not directly comparable between Model 1 and Model 2 (Section 3.1.6). Nevertheless differences between the models were reviewed. $PM_{2.5}$ standard deviations for Model 2 were lower than Model 1, as expected; however, the percent difference of the standard deviations measured in 2008 and 2009 differed substantially (**Table 4.4**). in contrast, PM_{10} standard deviation increased with the removal of outliers, which was not expected. The large change in percent difference for $PM_{2.5}$ and the increase in standard deviation in Model 2 for PM_{10} may reflect differences in how each model was generated. The $PM_{2.5}$ Model 1 consists of a summation of independently generated PM_1 and $PM_{1\text{-}2.5}$ ordinary kriged models. Similarly, the PM_{10} Model 1 consists of a summation of independently generated PM_1, $PM_{1\text{-}2.5}$, and $PM_{2.5\text{-}10}$ ordinary kriged models. In contrast, $PM_{2.5}$ and PM_{10} for Model 2 were kriged directly from the summation of measured particulate mass concentration $\leq$2.5 microns for $PM_{2.5}$ and particulate mass concentration $\leq$10 microns for $PM_{10,}$ respectively. Additional investigation is needed to establish a direct comparison between the two models.

The removal of outliers from several models changed the grid interpolation values in areas relatively far away from the outlier locations (**Figure 4.16**). Although spatial changes were expected to occur in areas of close proximity to where outliers were removed, the degree of distance where grid nodes were affected was unanticipated. These changes resulted from a combination of effects stemming from the removal of the outlier including changes in the variogram model and the recalculation of kriging weights based on the revised distribution of control points. The variogram model has a global impact on the ordinary kriged model while the outliers have a local impact based on the search radius. In this way, changes to the models are most likely attributed to variogram revisions after the removal of the outliers.

5.3 Temporal Scaling

Air pollutant Models 1 and 2 were scaled to incorporate time varying concentrations observed in Detroit in the 2008 Michigan Air Sampling Network (MASN) dataset. Air pollution in Detroit is heterogeneous (Miller et al., 2010) and varies with time (**Figure 4.19**). Consequently, the goal of the temporal scaling was to generate a more complete exposure estimate for 2008 through the combination of a spatially resolved dataset (Model 1 and Model 2) with the temporally detailed MASN dataset. In the O'Leary and Lemke (2014) study, the GeoDHOC models were scaled for a three-year period. However, temporal scaling is evaluated only for 2008 in this thesis, corresponding to the period of available asthma data (Lemke et al., 2013).

This analysis incorporates air quality measurements for corresponding pollutants analyzed using different methods and varying time scales in the GeoDHOC and MASN data sets. It is therefore assumed that both of these datasets, the GeoDHOC and MASN, are compatible. Similar relationships between passive sampler measurements and continuous or periodic automated fixed-site measurements have been examined in other studies. For example, Vardoulakis et al. (2009) compared chemiluminescence and passive NO_2 measurements from collocated samplers and found satisfactory agreement (relative bias and coefficient of variation < 5%) during four or five week measurement periods over thirteen months. Mukerjee et al. (2004) found BTEX measurements made with 3M organic vapor monitors over three to seven day sampling periods agreed within 10% of automated gas chromatograph measurements.

In the present study, collocated GeoDHOC and MASN measurements for NO_2 and total BTEX agreed well, with the exception of June 2009 BTEX (**Table 4.5**). GeoDHOC $PM_{2.5}$ and PM_{10} concentrations measured within 1000 m of MASN stations also compared well to MASN

measurements (**Table 4.6**). However, GeoDHOC concentrations represent integrated measurements for two-week periods and therefore might not be expected to agree well, particularly when MASN measurements are discontinuous and infrequent during the comparative time window. For example, the June 2009 MASN monitor sampled total BTEX twice in June 2009, with 24-hour reported concentrations of 2.0 $\mu g/m^3$ on June 6, and 1.7 $\mu g/m^3$ on June 18. These values agree poorly with the collocated GeoDHOC sample measurement of 4.4 $\mu g/m^3$ (**Table 4.5**). On the other hand, integrated measurements of longer duration are arguably more useful than infrequent central monitoring measurements for long term exposure estimation if the detection of peak concentrations of short duration is not essential. Given the well documented intra-urban variability of NO_2 (e.g., Hewitt, 1991; Jerrett et al., 2007; Ross et al., 2013), BTEX (e.g., Miller et al., 2012b; Vardoulakis et al., 2011), and PM (e.g., Brook et al., 1999; Rodes et al., 2010; Wilson et al., 2005) measurements, this thesis chose not to treat the MASN values were not considered to be representative of the entire city. Consequently, no attempt to incorporate a systematic bias between collocated GeoDHOC and MASN measurements was undertaken in this study. Rather, computation and subsequent modification focused on comparisons between GeoDHOC model values (spatially averaged over the entire Detroit study area) and corresponding MASN measurements for which time series measurements were available. The rationale for this approach requires that the GeoDHOC datasets and interpolated monthly models adequately capture spatial variability of the air pollutants across the city of Detroit. This is supported through the semi-variogram models (**Figure A2**) and sampler spacing (**Figure 2.1**). The GeoDHOC samplers, which ranged in spatial density from 5 km^2 to 10 km^2, were spaced at intervals well below the variogram ranges (**Table 3.2**).

The approach followed here also assumes that the GeoDHOC ordinary kriging models (Models 1 and 2), based on two-week measurements, are representative of the distribution of the mapped pollutants for the entire months of September 2008 and June 2009. It was further assumed that these models can be interpolated to represent spatial distributions in the ten remaining unsampled months of the year (Section 3.2.2) and that these inferred spatial distributions can be extrapolated for a year-long study period (Section 3.2.3). The first assumption is supported by the constancy of MASN pollutant concentration running averages calculated for each month (**Figures A3, A4, A5, and A6**). Rapid, large magnitude changes in mean concentrations were not observed and therefore not expected over the course of a few weeks for any of the pollutants considered. The second assumption implies that the location and relative magnitude of stationary and mobile sources is consistent throughout the year and that meteorological conditions were similar enough to allow spatial distributions modeled in September 2008 to serve as a proxy for September 2009. Alternatively, extrapolating between the months from June 2008 to September 2008 was considered. However, the meteorological conditions between June 2008 and June 2009 showed distinct differences in dominant wind directions, therefore making the September comparison more favorable.

Continuity of source distribution is supported by prior studies in Windsor, a segment of the Detroit-Windsor airshed that established significant correlations across winter, spring, summer, and fall seasons for NO_2 and BTEX (Miller et al., 2012b; Wheeler et al., 2008). Seasonal correlations in the Windsor observations suggested consistency in source distributions, and hence spatial variability, for NO_2 and VOCs throughout the year. This supposition was more likely for pollutants derived from local sources such as NO_2, VOCs, and PM_{10}, than for

pollutants of secondary origin like $PM_{2.5}$, which may also contain a regionally-sourced component that can depend on wind direction and other meteorological conditions.

The approach outlined here rests upon the additional assumption that the available MASN time series measurements reflect temporal trends across the city of Detroit throughout the three-year study period. Although differing measurement frequencies were employed for each analyte (**Table 2.1**), records were 95% or more complete at each MASN site during 2008-2010. In the case of $PM_{2.5}$, concentrations measured at five MASN locations across Detroit (**Figure 2.1**) tracked consistently with each other during the three-year period of interest (**Figure 4.18**). Sajani et al. (2004) reported similar contemporaneous temporal trends for NO_2 measured at four different stations located throughout the urban area of Bologna, Italy. This implied that although air pollutant measurements made at widely-spaced regulatory monitoring sites may fail to capture significant spatial variability in the surrounding area (Baxter et al., 2013; Ozkaynak et al., 2013), relative changes in these measurements over time reflected temporal trends affecting the larger surrounding urban area.

Unquestionably, such temporal trends are influenced by daily and seasonal meteorological conditions. Weather conditions were not explicitly factored into the modeling procedures employed here and the inability of kriged maps to incorporate changes in monthly average wind direction is a limitation of this study's approach. However, the influence of major weather changes is assumed to directly influence the temporal trends recorded at MASN sites. Seasonal variations in concentrations occur in the 12 monthly NO_2 concentration models (**Figure 4.19**), with higher concentrations in winter months. Consistent seasonal variations are not evident in BTEX, $PM_{2.5}$ or PM_{10} models, however (**Figure 4.19**). This study attributed the lack of seasonality in BTEX to greater variability in local VOC source distribution (as reflected in the

high coefficient of variation for BTEX) (**Table 4.8**). Elsewhere, other researchers have observed a lack of seasonality in PM measurements. Brook et al. (1999), for example, found significant overlap between summer and winter seasons in 24-hour $PM_{2.5}$ mass concentration distributions measured in fourteen Canadian cities and Johnson et al. (2013) found that seasonal $PM_{2.5}$ models did not predict daily concentrations better than annual models in Windsor, Ontario.

Finally, in the absence of other information, it was assumed that the modeled spatial distribution of pollutant concentrations remained constant while the magnitude of the concentrations fluctuated uniformly across Detroit throughout the yearlong study period. As a consequence of this assumption, a bulk shift was employed to translate temporal trends from the MASN time series measurements to the monthly GeoDHOC estimates (Section 3.3). Alternative shifting techniques for incorporating the temporal trend were considered, including a ratio technique employed by Ross et al. (2013) to adjust two-week spatially interpolated air pollutant concentrations to temporal trends from continuous stationary monitors in New York City. The ratio technique was rejected because it resulted in localized concentration estimates far outside (in some cases two to three times higher than) the range of measured values for several pollutant models. In contrast, the bulk shift generated concentration values within the range of observed values for each pollutant (**Table 4.10**, **Table A3, and Table A4**), except in the case of negative concentration values present in two months in Model 1 and Model 2. BTEX values in these cells were replaced with 1/2 the method detection limit ($0.1 \mu g/m^3$) to maintain physically realistic concentrations. The underlying assumption that the magnitude of pollutant distributions fluctuate uniformly throughout Detroit becomes more tenuous when extended over longer periods of time during which major changes in infrastructure or economic conditions may occur.

5.4 Asthma correlations

The thesis examined the relationship between asthma exacerbations and air pollution in Detroit and Windsor with a specific focus on spatial outlier analysis (hypothesis #1) and temporal scaling (hypothesis #2). Initially it was hypothesized that with the removal of the spatial outliers, health correlations for asthma exacerbations would improve (hypothesis #1). The second hypothesis stated that with increasing temporal resolution, correlations between model air pollutant concentrations and for asthma exacerbations would also improve. Evaluation of each hypothesis yielded differing results.

The diminished association between air pollution and asthma exacerbations in Model 2 compared to Model 1 does not support the first hypothesis. Although the strength of the correlations improved for some analytes, the removal of spatial outliers in Model 2 reduced the number of statically significant correlations. Specifically, the removal of outliers improved correlations in Detroit for NO_2, $PM_{2.5}$, and PM_{10} (**Table 4.11**). However, this improvement was not enough to reach the level of statistical significance at $p \leq 0.05$. Conversely, in Windsor, removing outliers in Windsor decreased the strength of the correlation for all analytes except $PM_{2.5}$ (**Table 4.12**). All of the outliers removed were extreme data values and by removing the outliers the range of modeled values decreased in Model 2. This decreased range may have had an adverse effect on the sensitivity of the statistical relationship between air pollution and asthma exacerbations.

The improved correlations between air pollution and asthma exacerbations in Models 3 and 4 support the second hypothesis that correlations between air pollutant models and asthma exacerbations increase with temporal resolution. The incorporation of the MASN dataset, in Models 3 and 4, consistently increased the strength of the correlation in Detroit except for BTEX

for Model 3 and $PM_{2.5}$ and PM_{10} for Model 4. With the inclusion of a time trend in the air pollution data, both the air pollution models and asthma data represent aggregated values for the entire 2008 year. The improved correlations potentially are a result of the incorporation of additional information in the form of MASN time-series measurements throughout 2008 and the revision of the postal code pollution estimates for each analyte to represent an equivalent period of time to the asthma dataset.

The analytes measured with passive samplers tended to correlate better with asthma than analytes measured with active samplers in all models and, as a result, the statistically significant points identified in Models 3 and 4 (**Table 4.13**) were analytes measure with passive samplers. This is potentially attributable to the number of passive versus active samples rather than the type of sampling method. The number of passive samplers ranged from 97 to 98 locations across the entire study area while the number of active samplers ranged from 37 to 38 samplers in the same area. In Detroit the number of passive samplers ranged from 65 to 66 samplers and the number of active samplers was 23 samplers. The asthma correlation statistical results suggest higher sensitivity with higher spatial sample density and may indicate that a minimum sample spacing of approximately 1 per 5 km^2 is needed to accurately model neighborhood spatial variability of the pollutants analyzed.

5.5 Model Limitations

Overall, the asthma correlations in this study were limited by the assignment of air pollution exposure estimates based on modeled ambient air concentrations rather than personal exposure which can lead to potential error (Kearney et al., 2011). Moreover, the asthma events

were reported by postal code which decreased the effective spatial resolution of the air pollutant models from a 300 by 300 meter grid spacing to aggregate zip code or forward sortation area scale averages. In addition, the asthma data were aggregated on a yearly basis and did not account for temporal variation throughout the year. Finally, the results of the correlation do not account for differences in socioeconomic demographics or medical management of asthma.

The air pollution models had a number of specific limitations which differ between Models 1 through 4. Air pollution Models 1 and 2 were limited in their temporal resolution. These models relied exclusively on the September 2008 air pollutant data for the asthma correlation. These two-week air pollutant concentrations did not account for daily air pollution fluxes or air pollutant concentrations during the rest of 2008. Air pollution Models 3 and 4 incorporated temporal trends for all of 2008, but the resulting monthly air pollution models were subsequently aggregated to annual zip code concentration values in order to study the relationship to the asthma data. This process reduced the temporal resolution of the models.

Future studies should reassess the asthma associations using better resolved spatial and temporal asthma data. Refining the spatial resolution of asthma events to the neighborhood level using residential addresses and increasing the temporal resolution of the asthma data to monthly counts, as found in the air pollution models, may improve asthma-pollutant associations. Applying the temporal scaling technique used in the study to the Windsor air pollution models may further refine the air pollution dataset for the Detroit-Windsor international airshed. Given that Detroit and Windsor differed in their response to the spatial data outlier analysis, this type of future study could support or refute the results of temporal scaling in Detroit (hypothesis #2).

5.6 Conclusions and Recommendations

This thesis developed three new sets of air pollutant models that can be utilized for future epidemiological studies. Overall, the correlations between air pollution and asthma hospitalizations were weaker with the spatial outliers removed but improved with the addition of temporal data. When outliers were removed, statistically significant correlations between air pollution and asthma decreased. Incorporating a temporal time trend increased the number of statistically significant correlations. The resulting models generated by this study provided a more detailed analysis of the air pollution in Detroit and Windsor. The study improves model integrity by increasing the spatial integrity and temporal resolution of air pollution estimates. Removing outliers from the datasets reduced the variance of the model for a number of analytes in Detroit and Windsor. The results of temporal scaling preserve spatial variability captured by the two detailed GeoDHOC air sampling campaigns and incorporate temporal variability present in MASN data in Detroit. Limitations of the asthma data necessitated less detailed regression analysis using spatially and temporally aggregated air pollution models. Nevertheless, the aggregated air pollution models still contain the spatial and temporal content embedded in the newly created models. This allowed for new correlations between air pollution and asthma exacerbations.

In conclusion, this thesis expands on previous GeoDHOC studies (Lemke et al., 2013; Miller et al., 2010; O'Leary and Lemke, 2014) and provides additional insights into the association between asthma exacerbations in both Detroit and Windsor and the international airshed spanning both cities. The spatial data outlier identification and temporal scaling approaches outlined here can be applied to other datasets or cities where long-term time series measurements are available to supplement spatially variable air pollutant datasets. Additional

spatially or temporally resolved asthma exacerbation data can lead to future studies that incorporate more detailed aspects of the air pollutant models. In particular, incorporation of Windsor NAPS data through temporal scaling can help confirm the findings of the Detroit temporal scaling.

APPENDICES

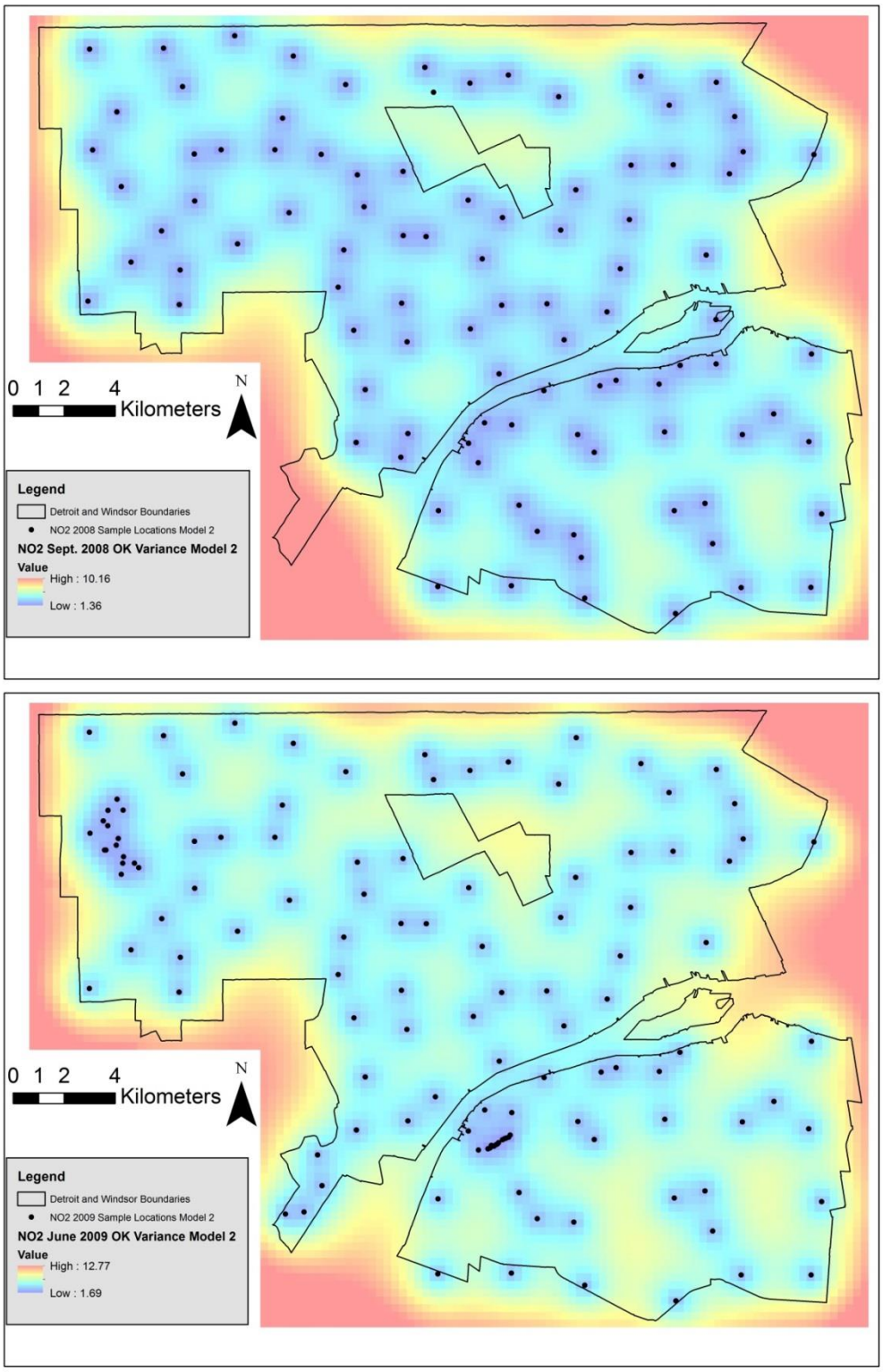


Figure A1f. Model 2 NO_2 variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

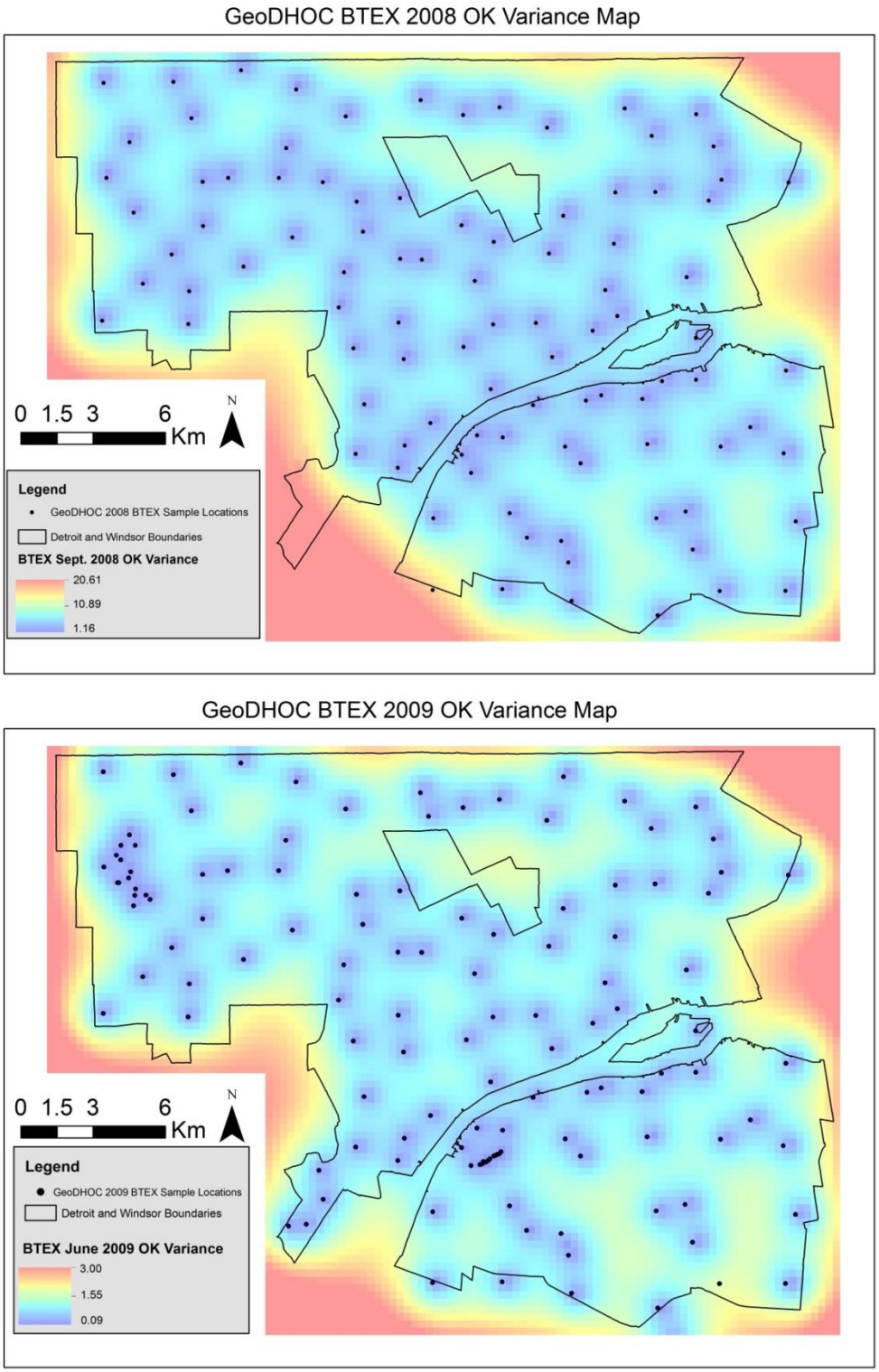


Figure A1b. Model 1 BTEX variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

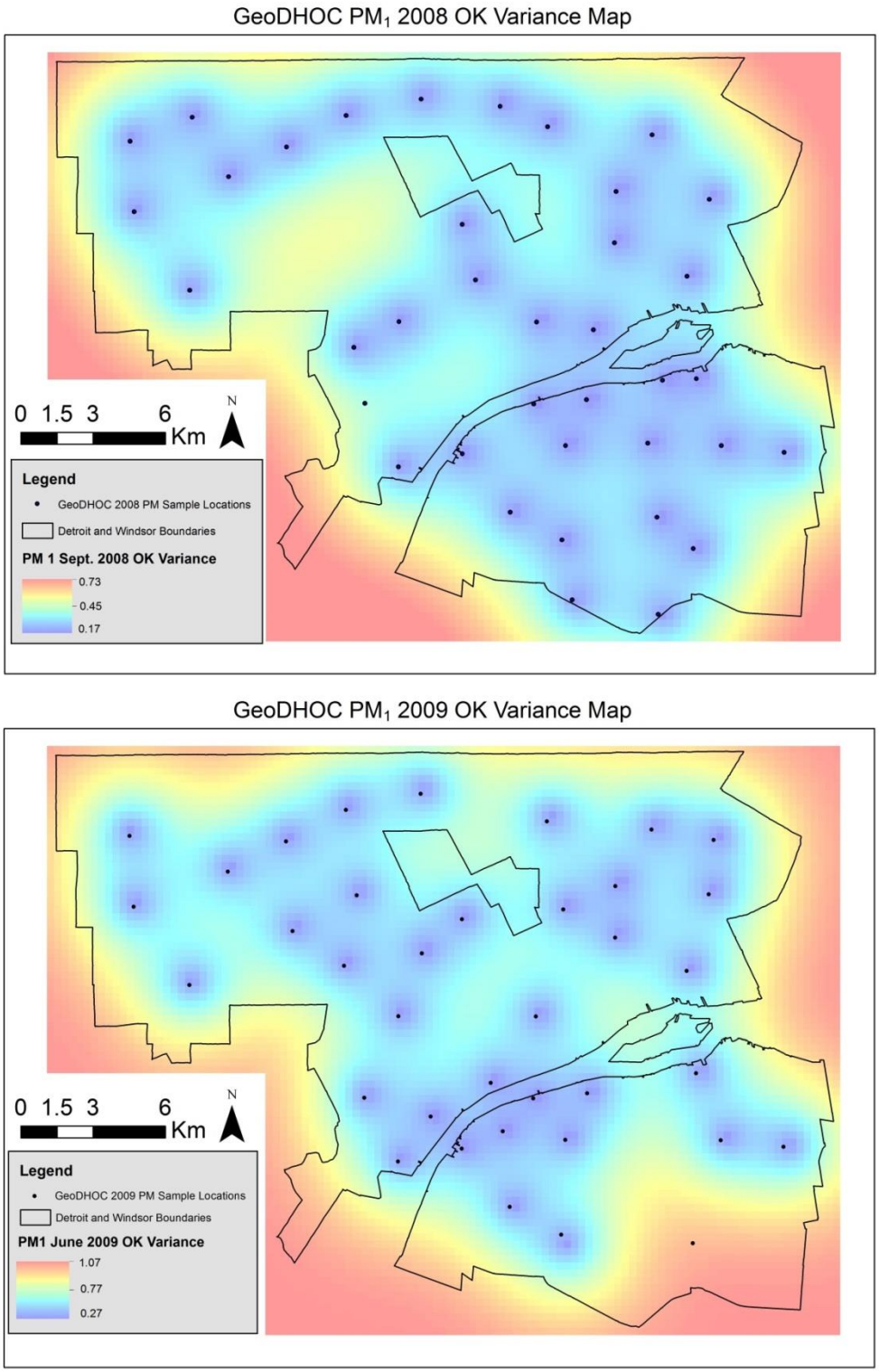
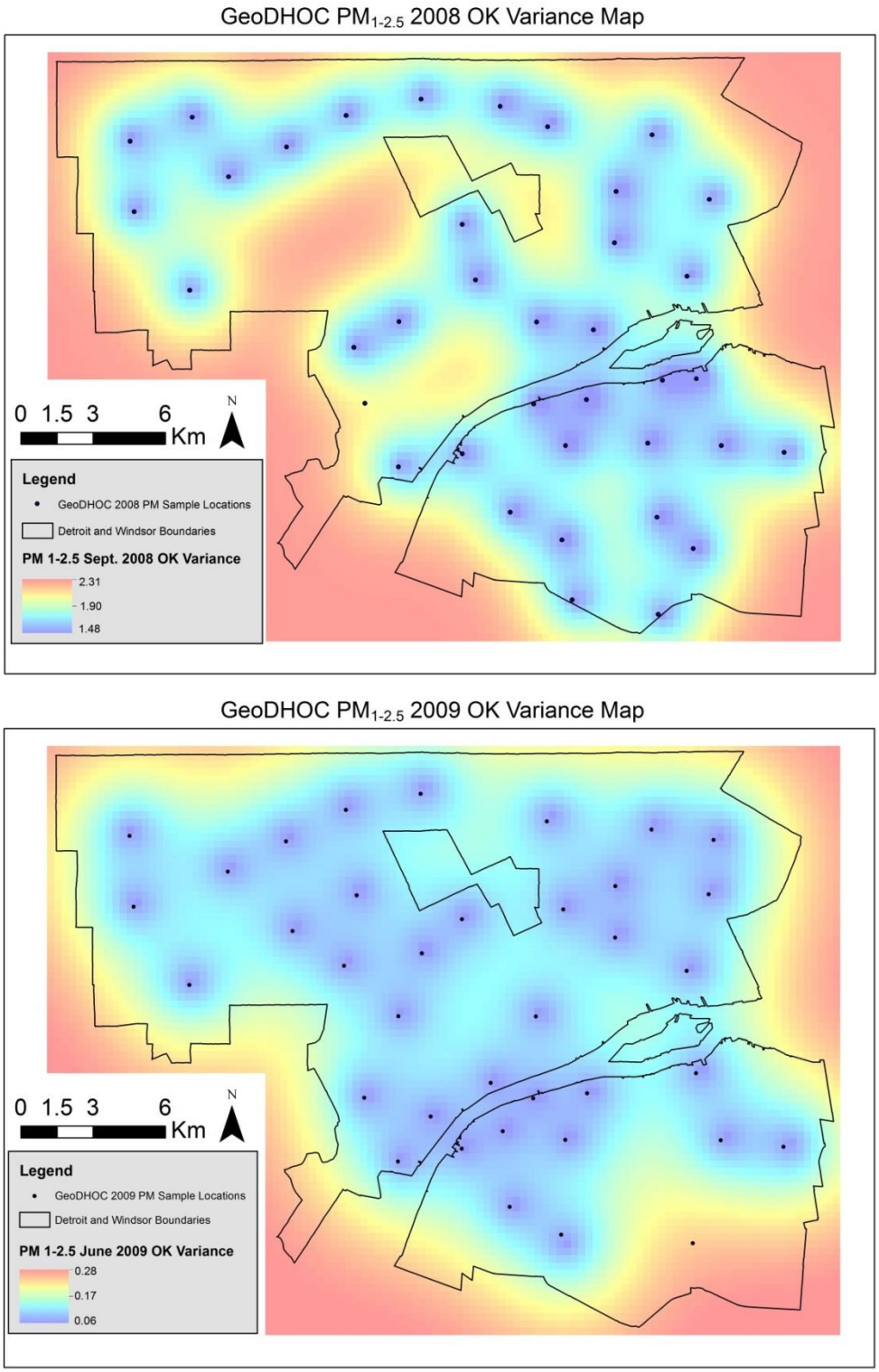


Figure A1c. Model 1 PM_1 variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).



<u>Figure A1d.</u> Model 1 $PM_{1\text{-}2.5}$ variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

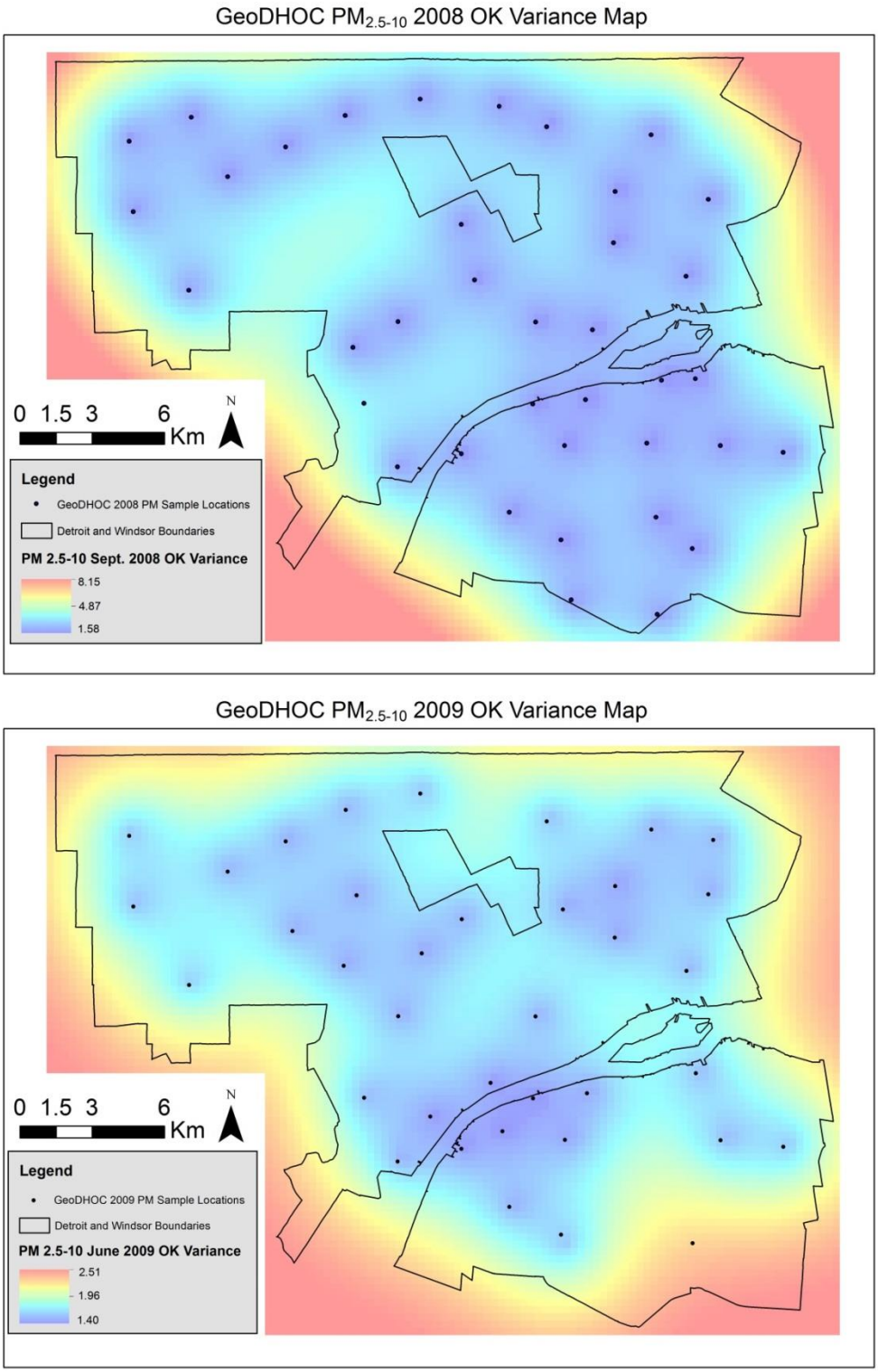


Figure A1e. Model 1 $PM_{2.5-10}$ variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

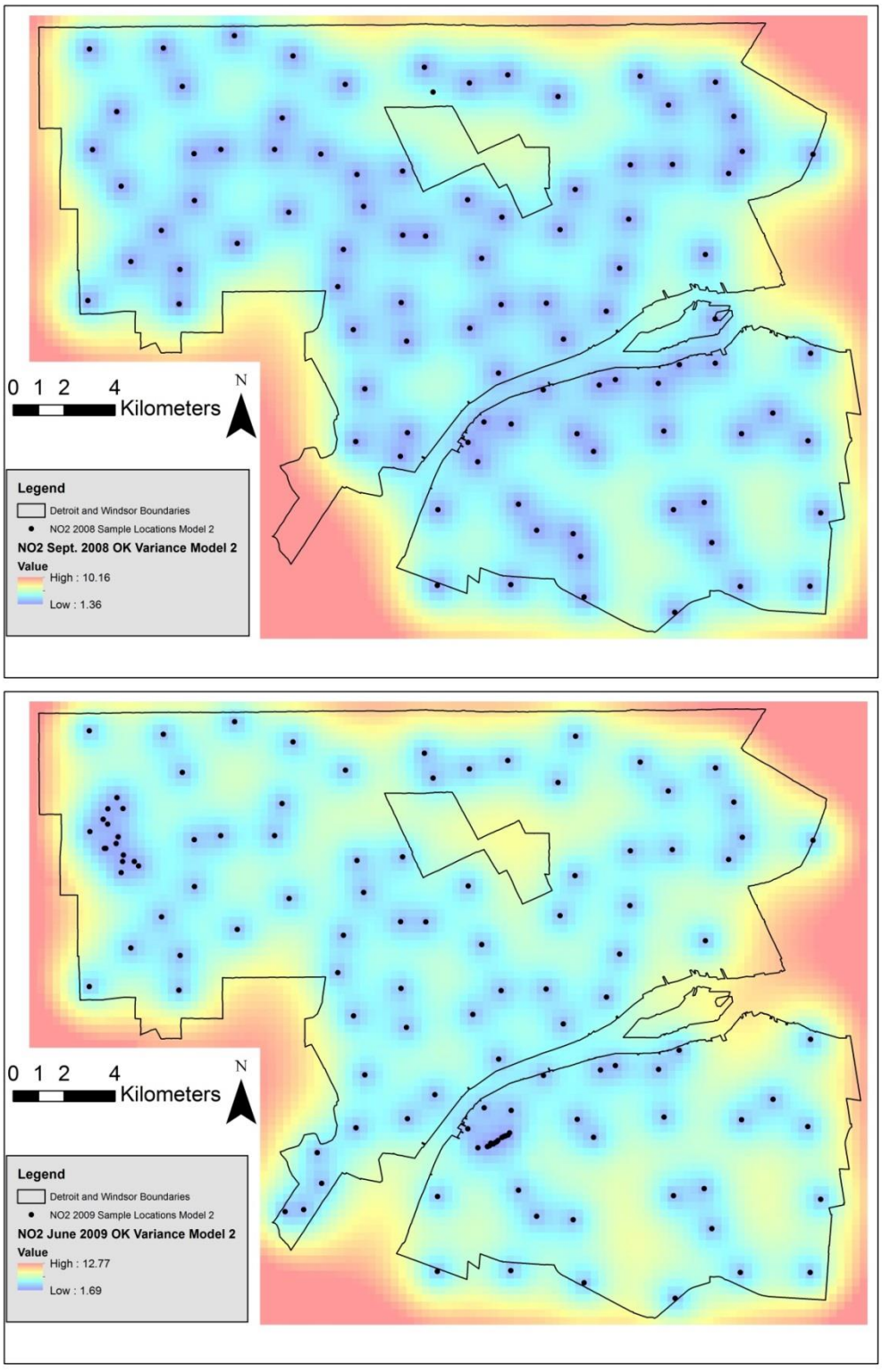


Figure A1f. Model 2 NO_2 variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

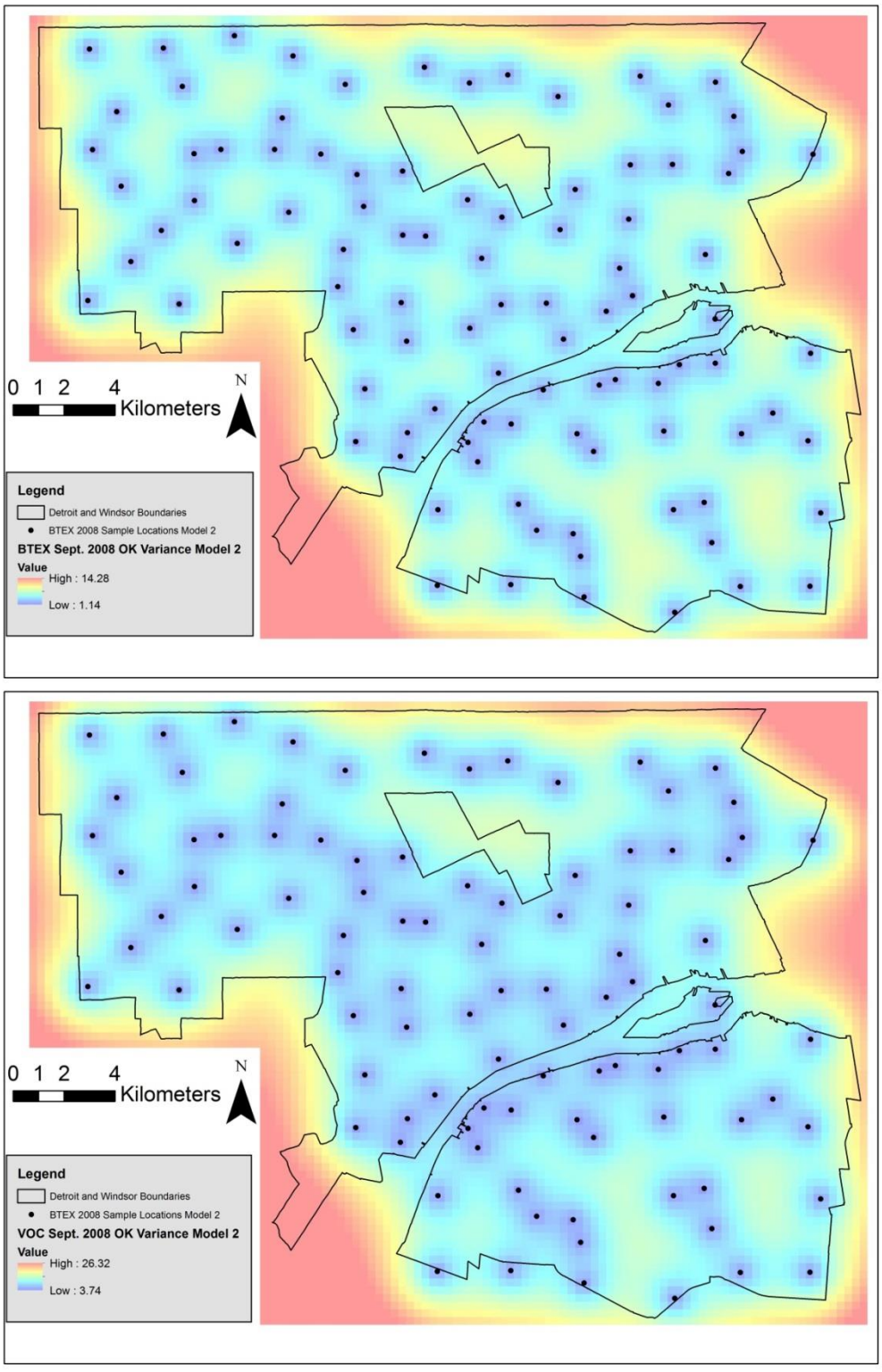


Figure A1g. Ordinary kriged variance maps for Model 2 in September 2008 with BTEX variance concentrations on top and VOC variance concentrations below.

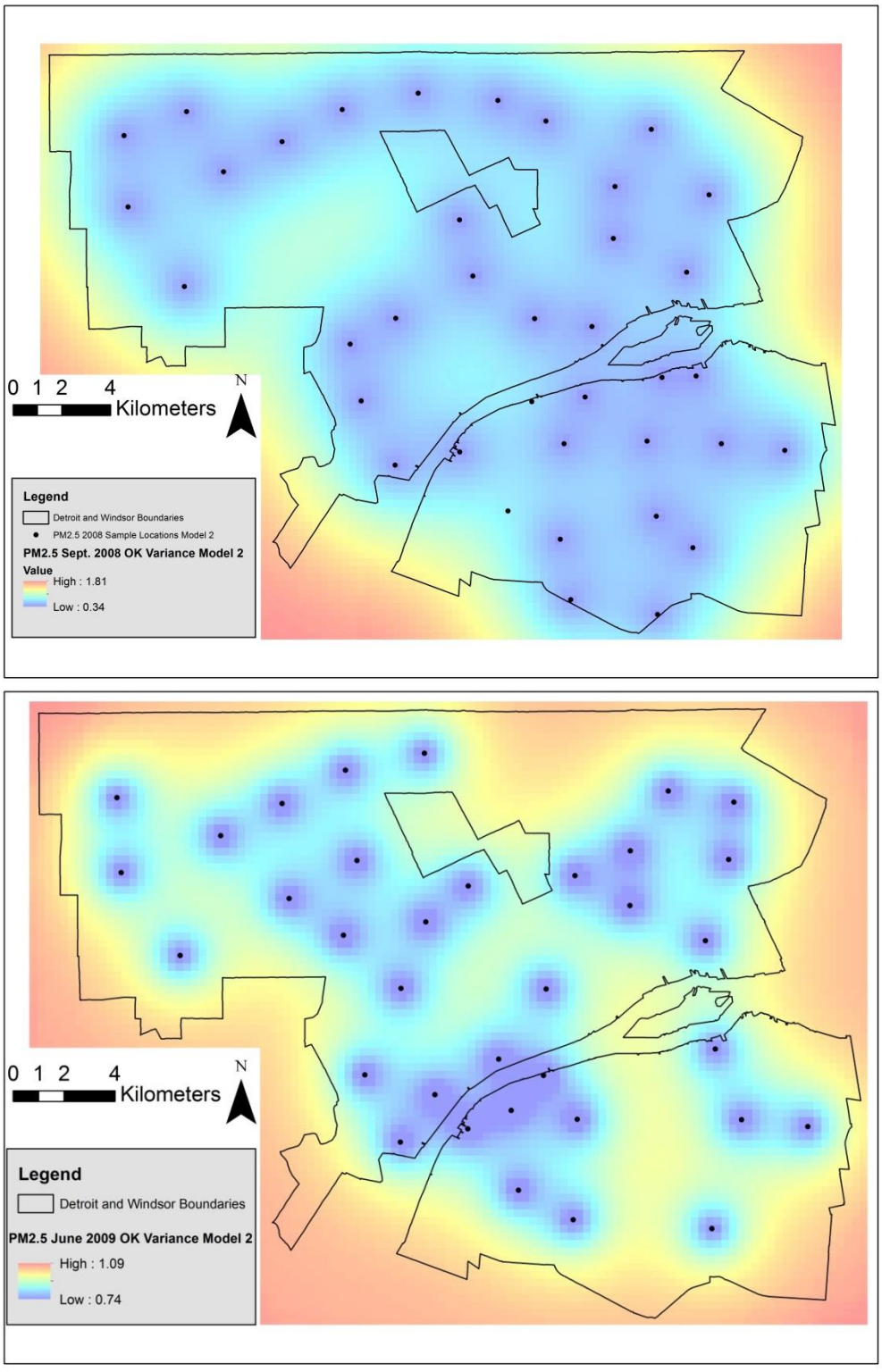


Figure A1h. Model 2 $PM_{2.5}$ variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

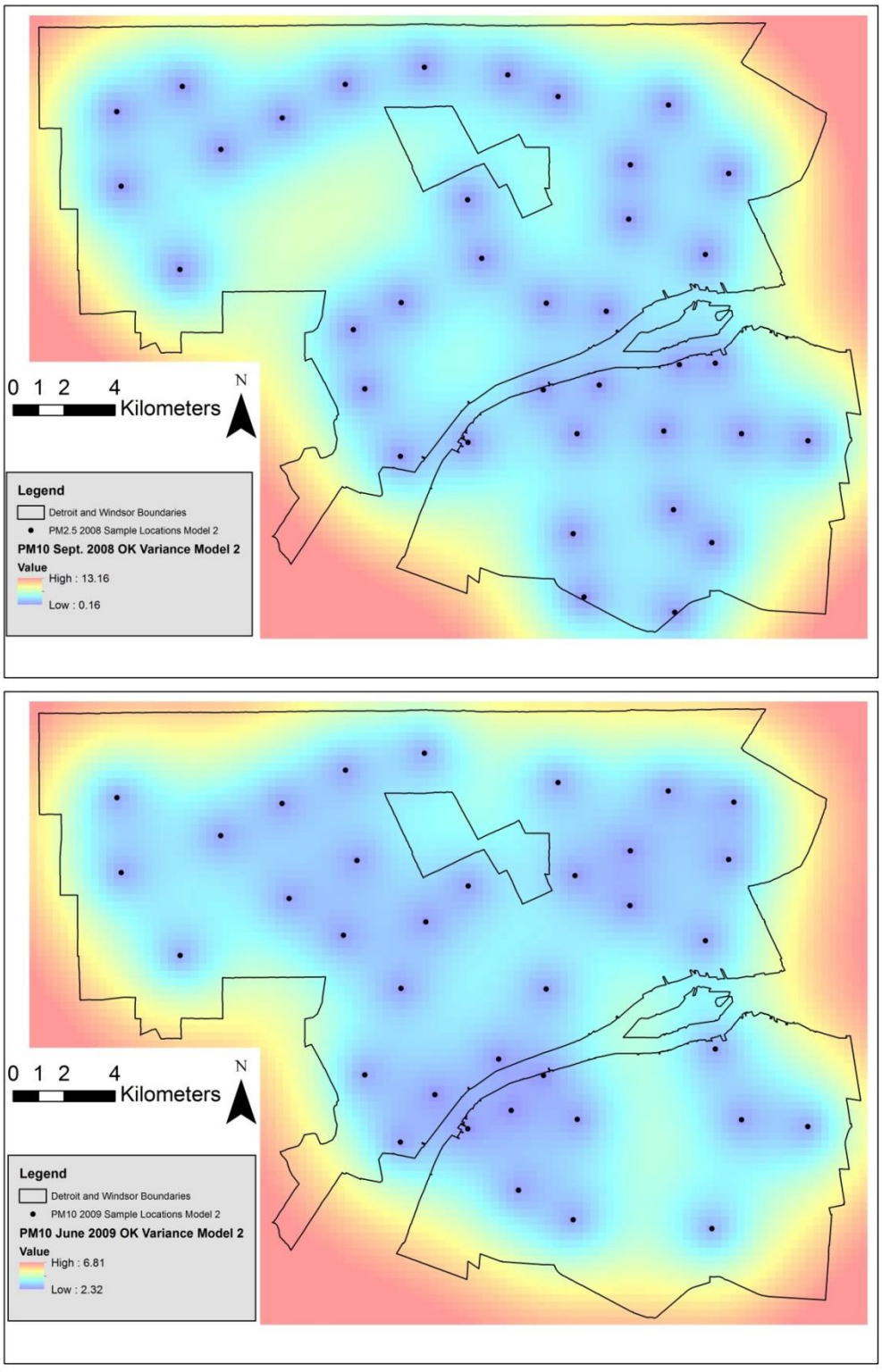


Figure A1i. Model 2 PM_{10} variance maps for the ordinary kriged models for September 2008 (top) and June 2009 (bottom).

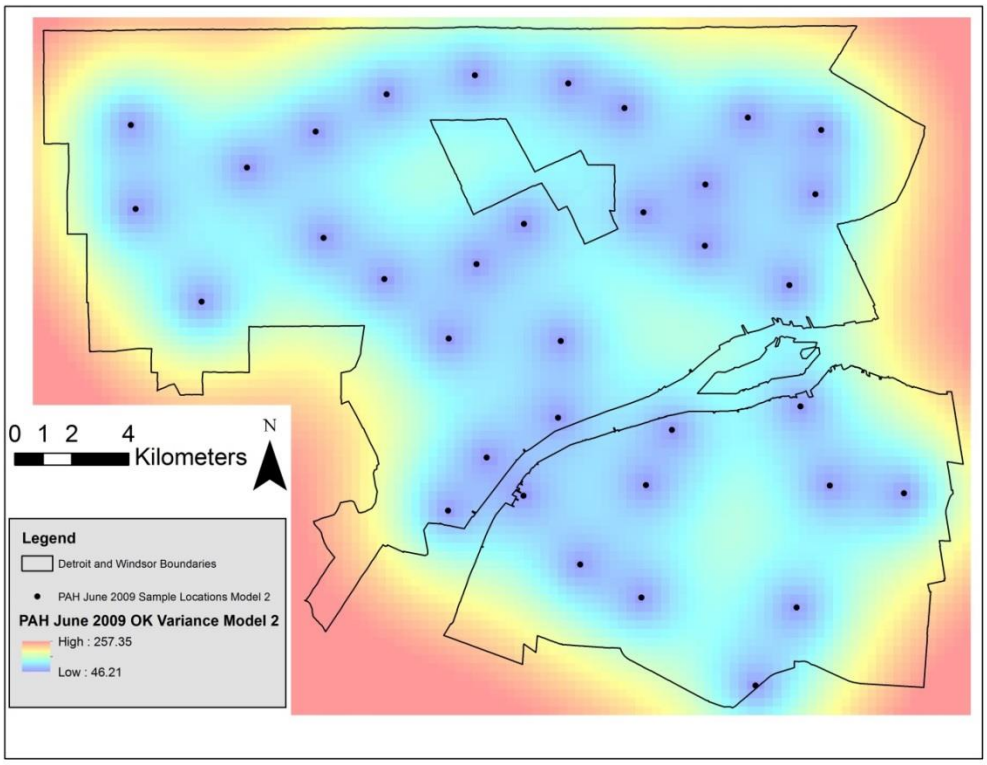


Figure A1j. Model 2 PAH variance maps for the ordinary kriged models for June 2009.

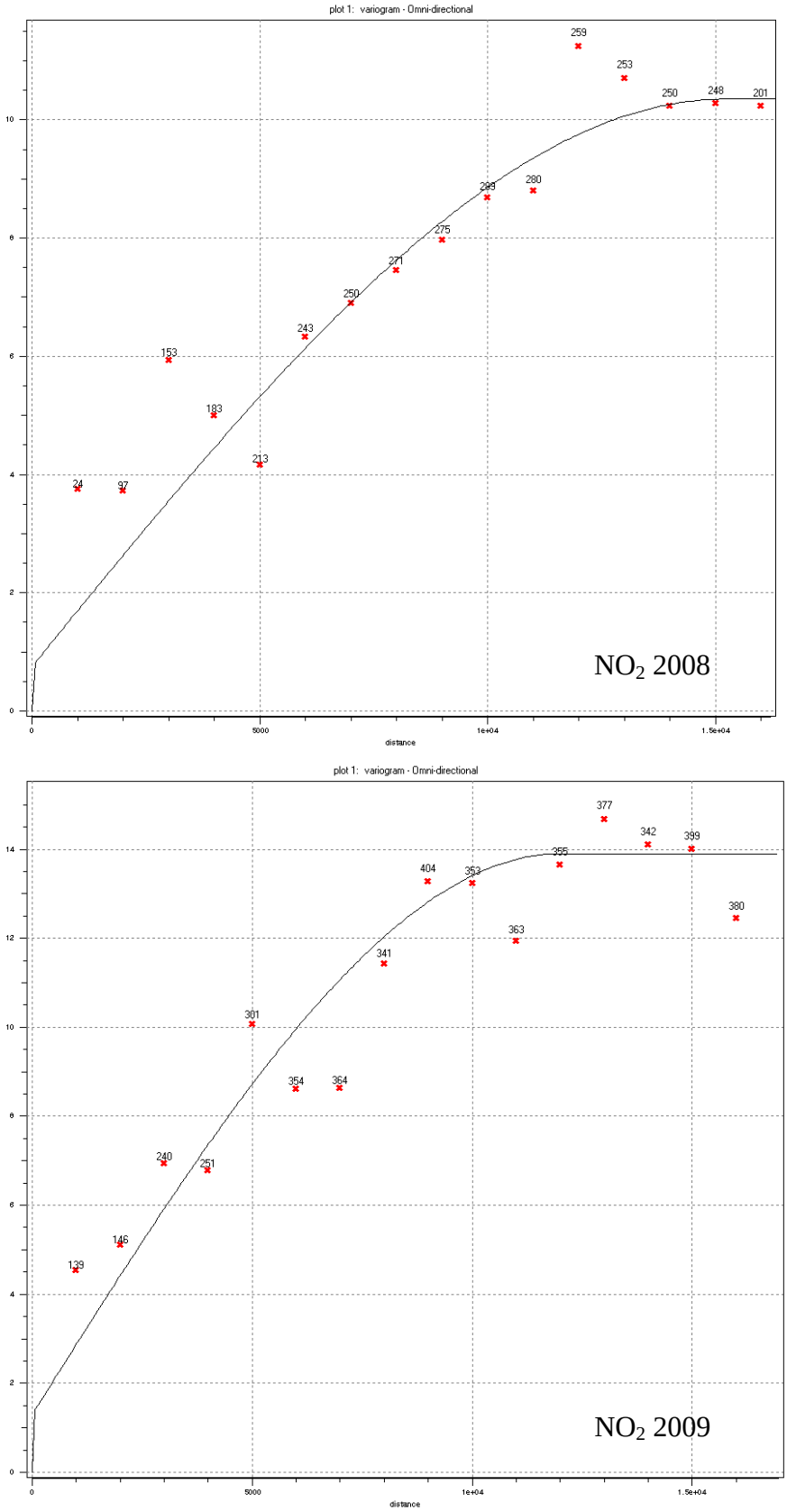
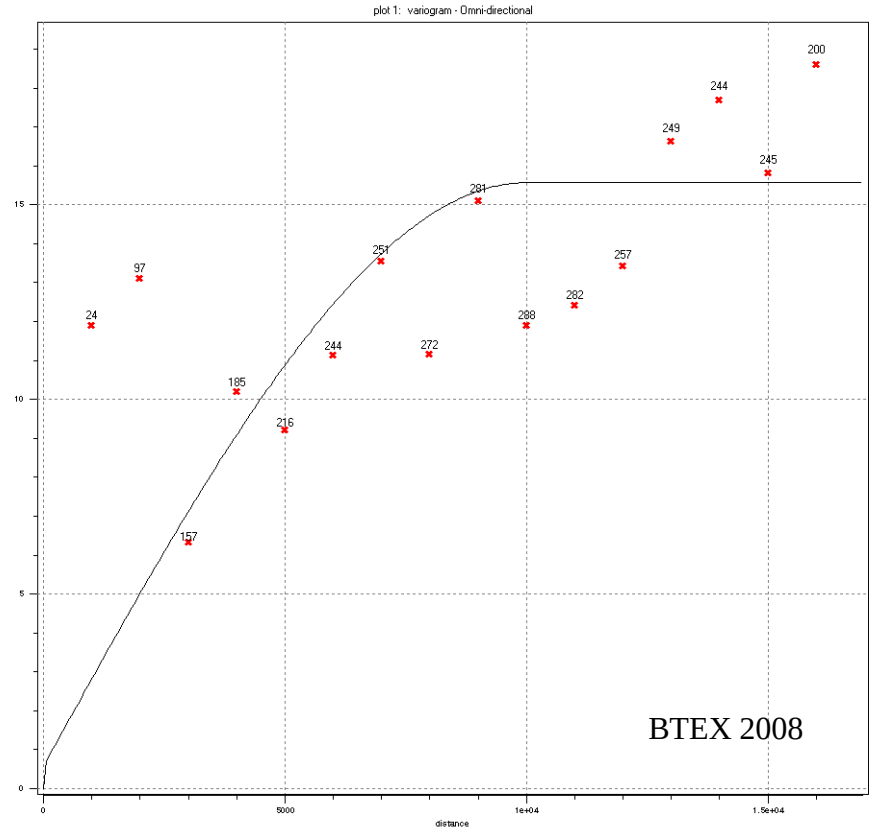


Figure A2a. Model 1 NO_2 sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).



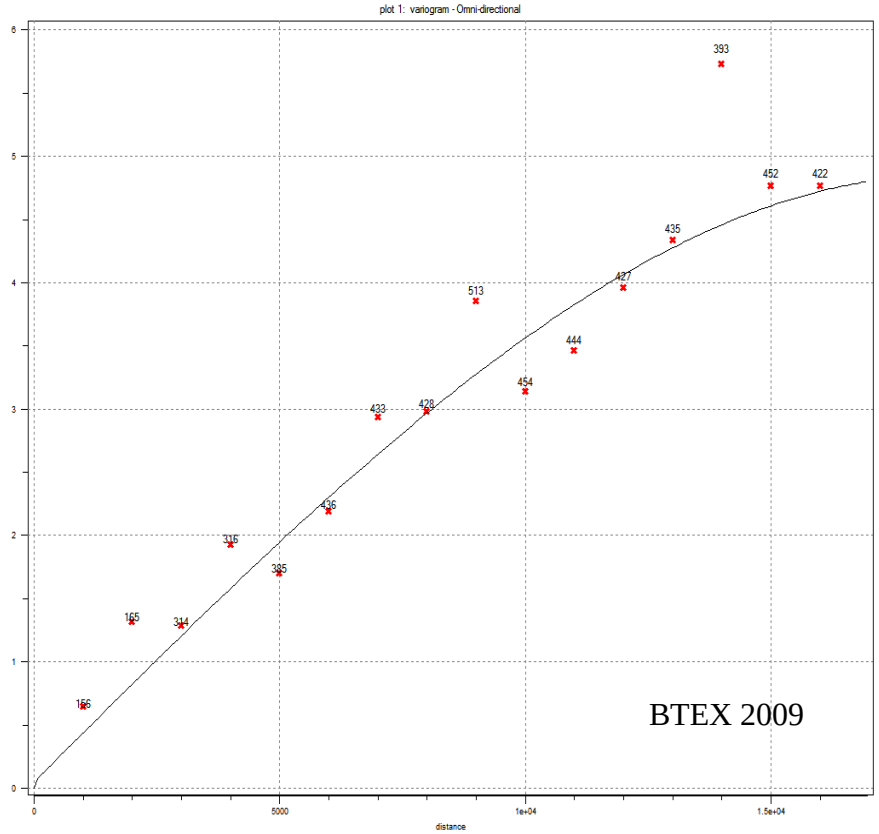
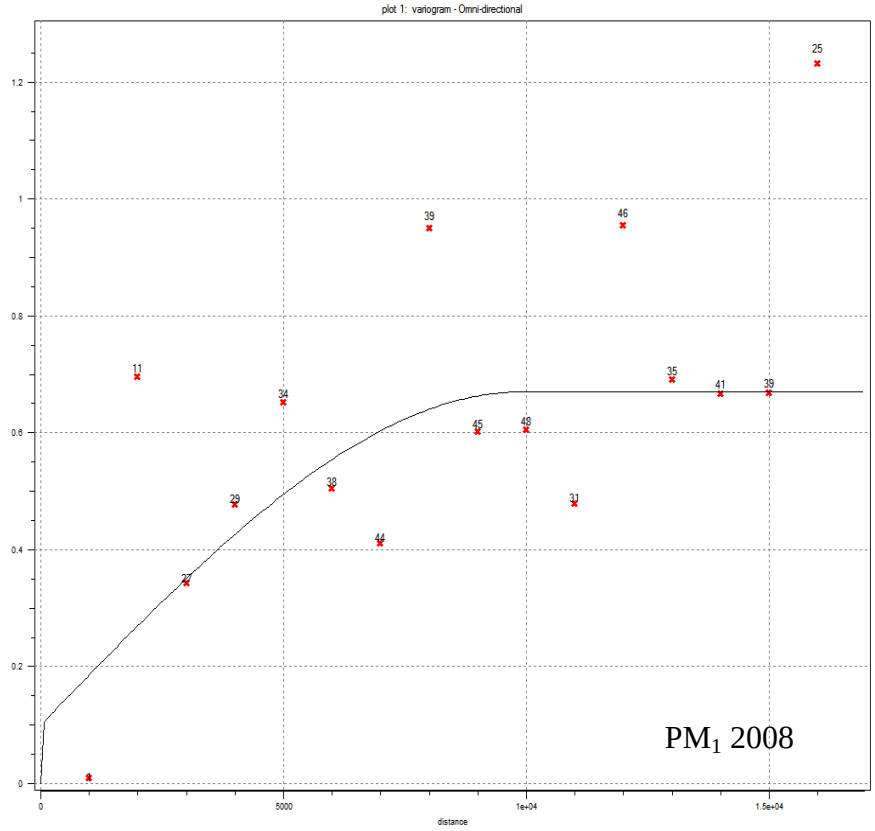


Figure A2b. Model 1 BTEX sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).



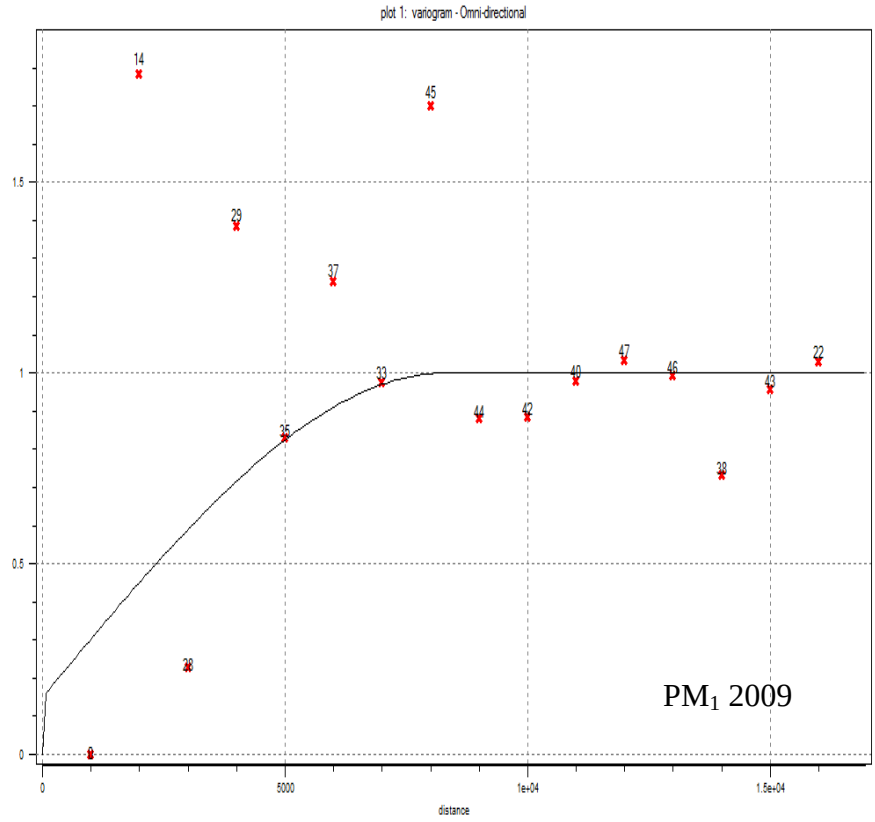
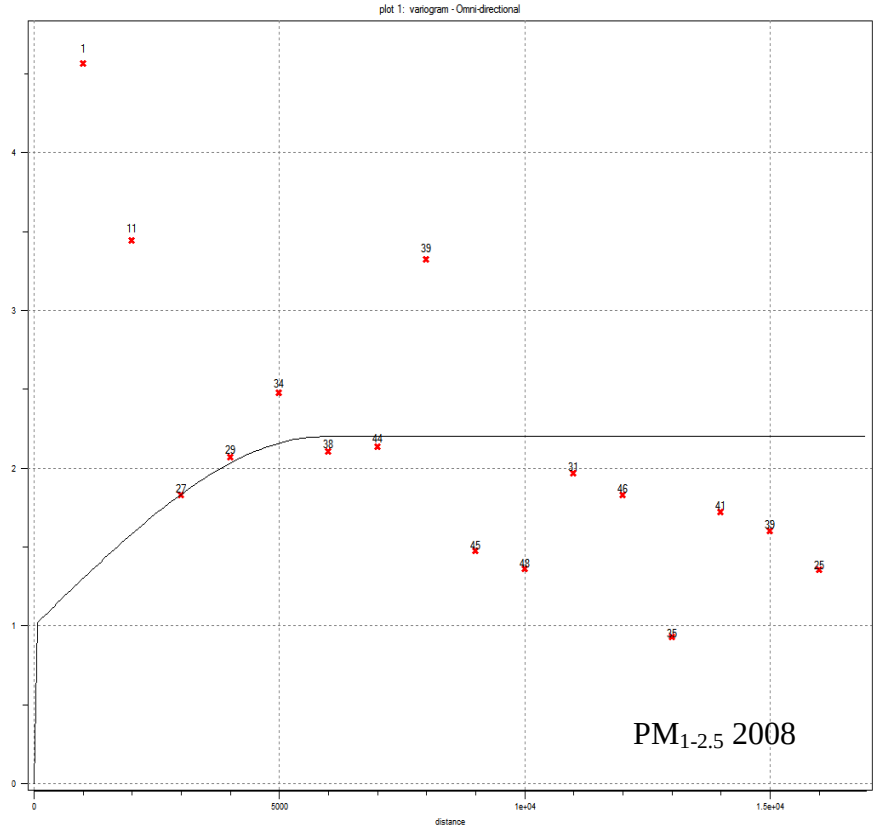


Figure A2c. Model 1 PM_1 sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).



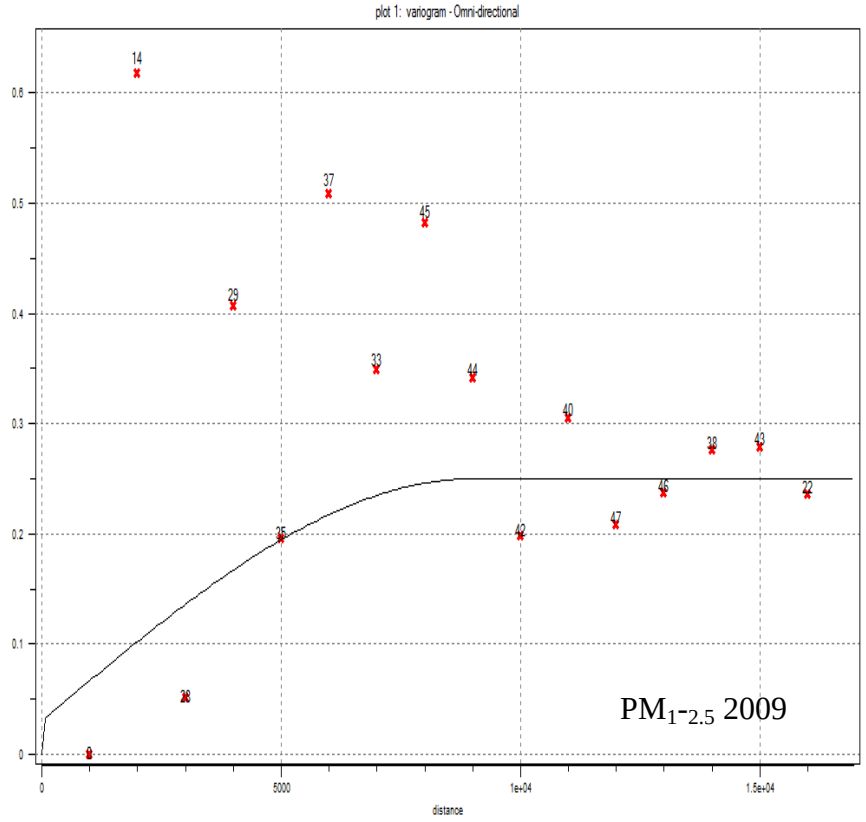
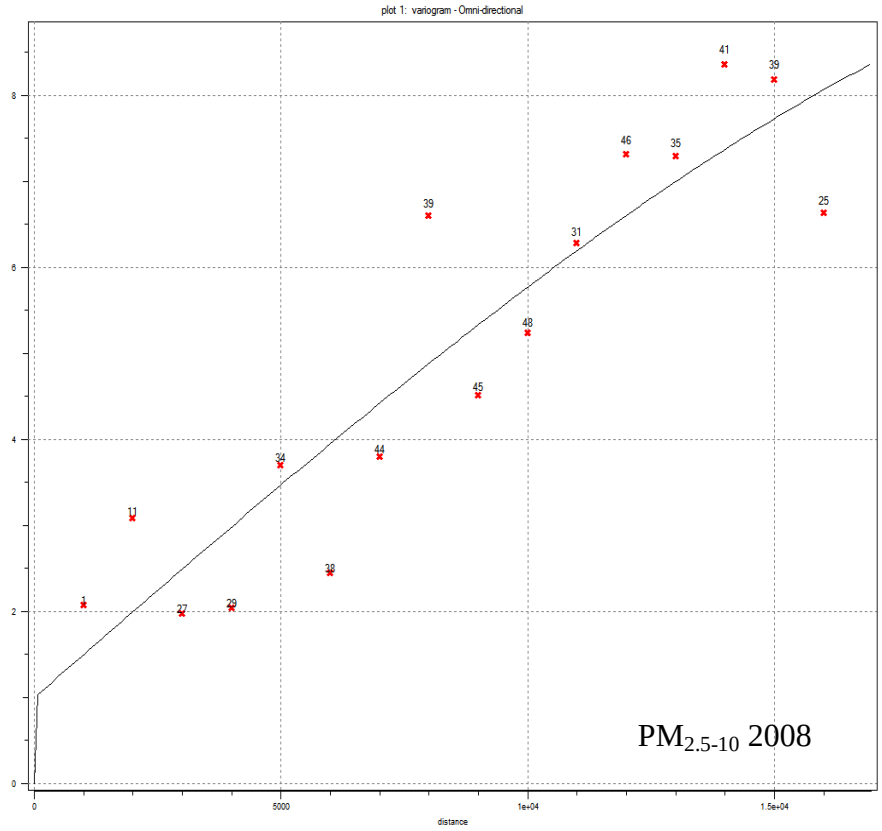


Figure A2d. Model 1 $PM_{1\text{-}2.5}$ sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).



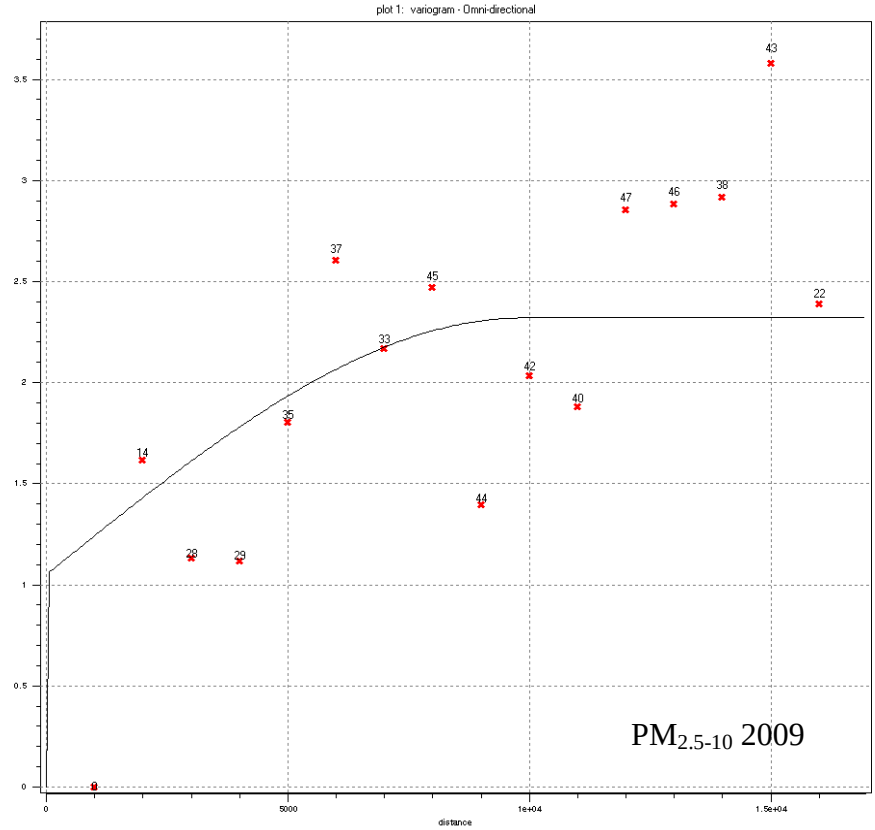


Figure A2e. Model 1 $PM_{2.5-10}$ sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).

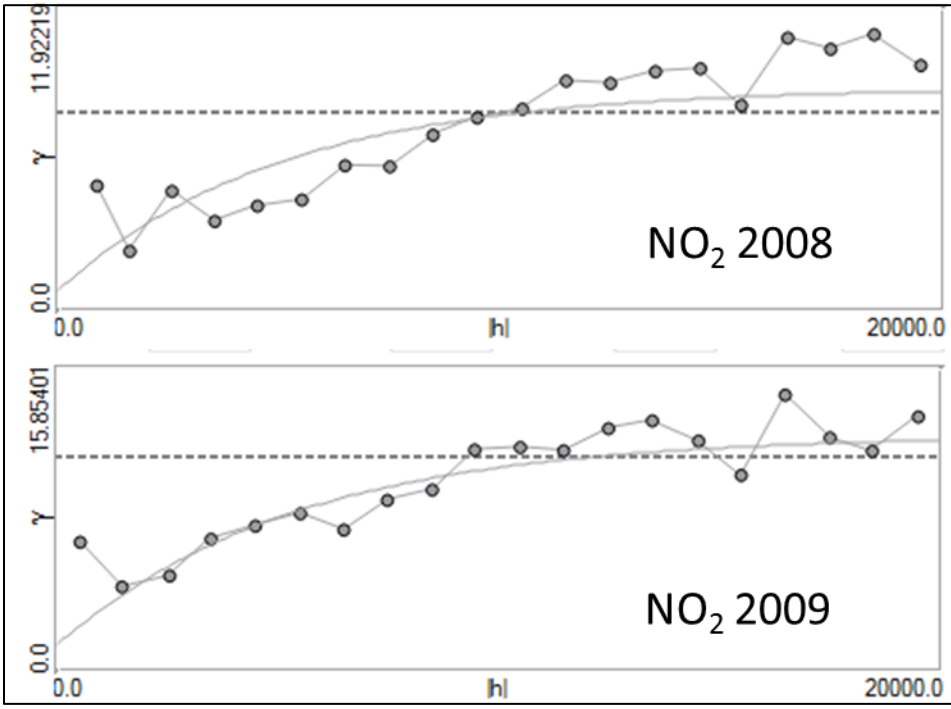


Figure A3a. **Model 2 NO_2 sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).**

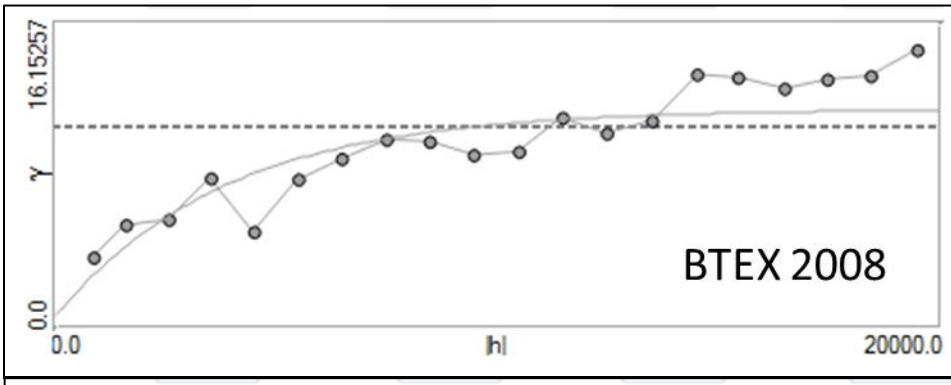


Figure A3b. **Model 2 BTEX sample semi-variogram and variogram models for September 2008.**

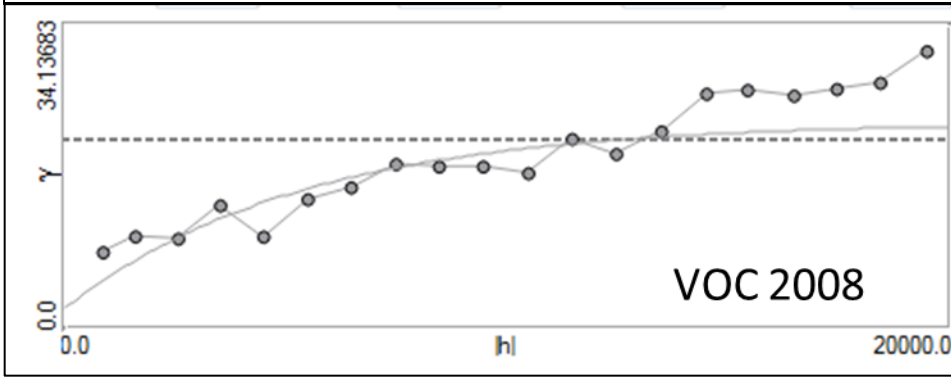


Figure A3c. **Model 2 VOC sample semi-variogram and variogram models for September 2008.**

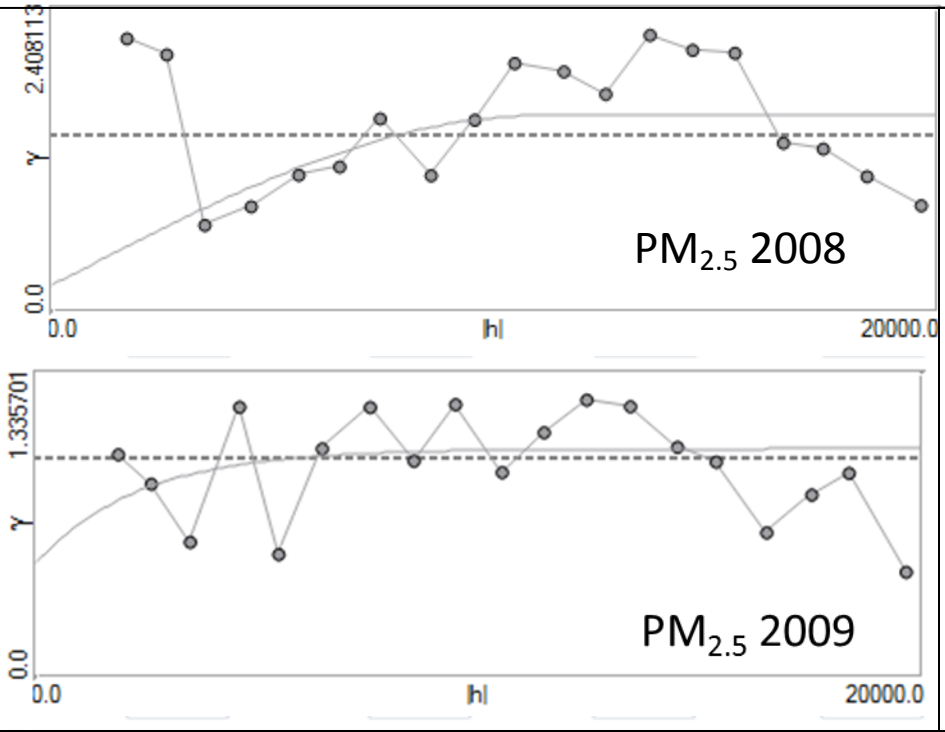


Figure A3c. **Model 2 $PM_{2.5}$ sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).**

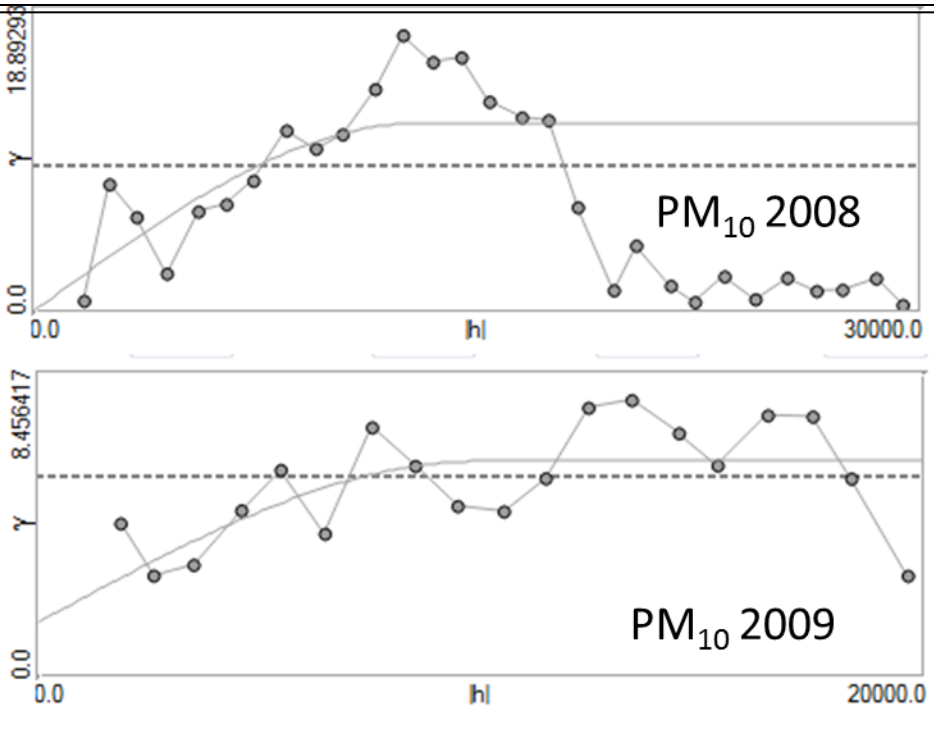


Figure A3d. **Model 2 PM10 sample semi-variogram and variogram models for September 2008 (above) and June 2009 (below).**

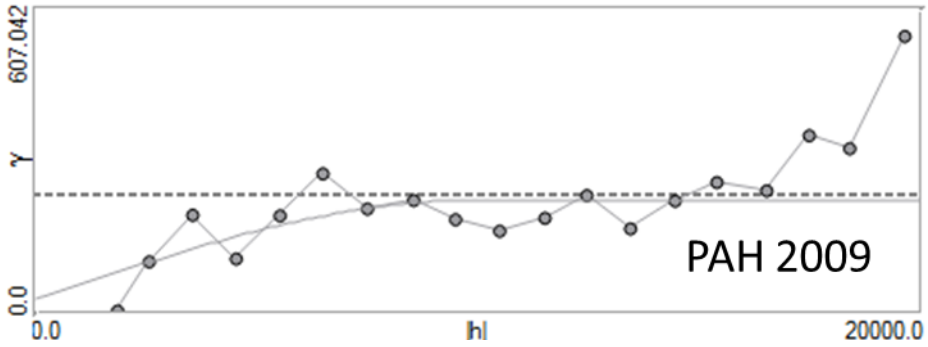


Figure A3e. **Model 2 VOC sample semi-variogram and variogram models for June 2009.**

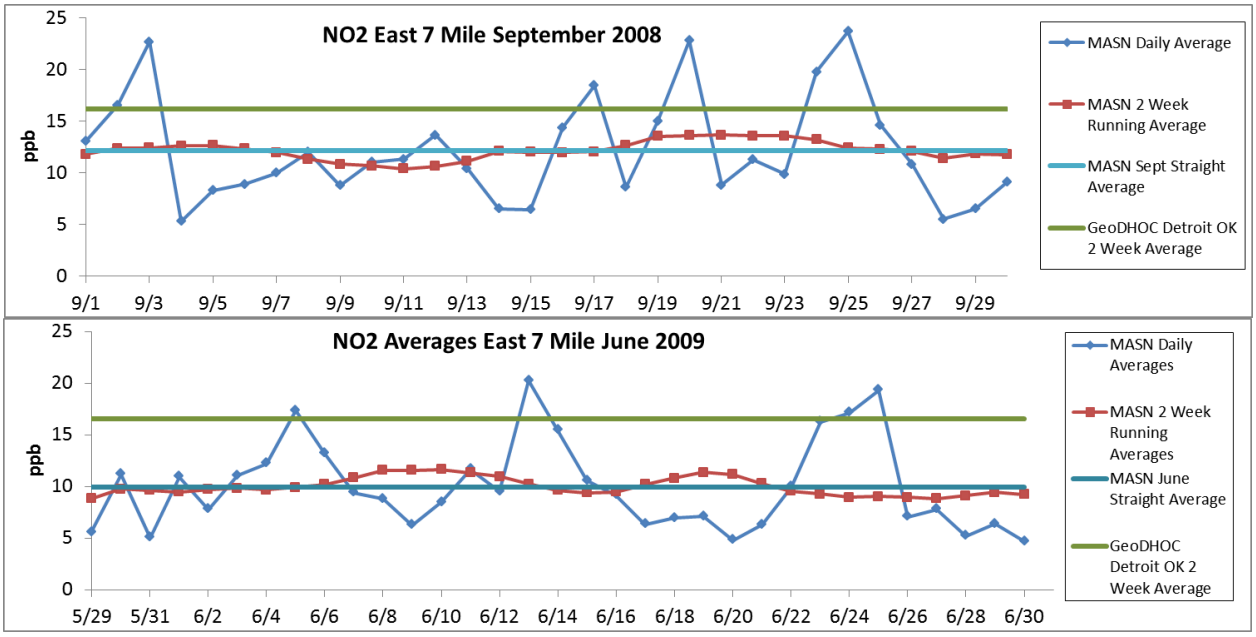


Figure A4. **NO_2 concentrations for September 2008 and June 2009.**

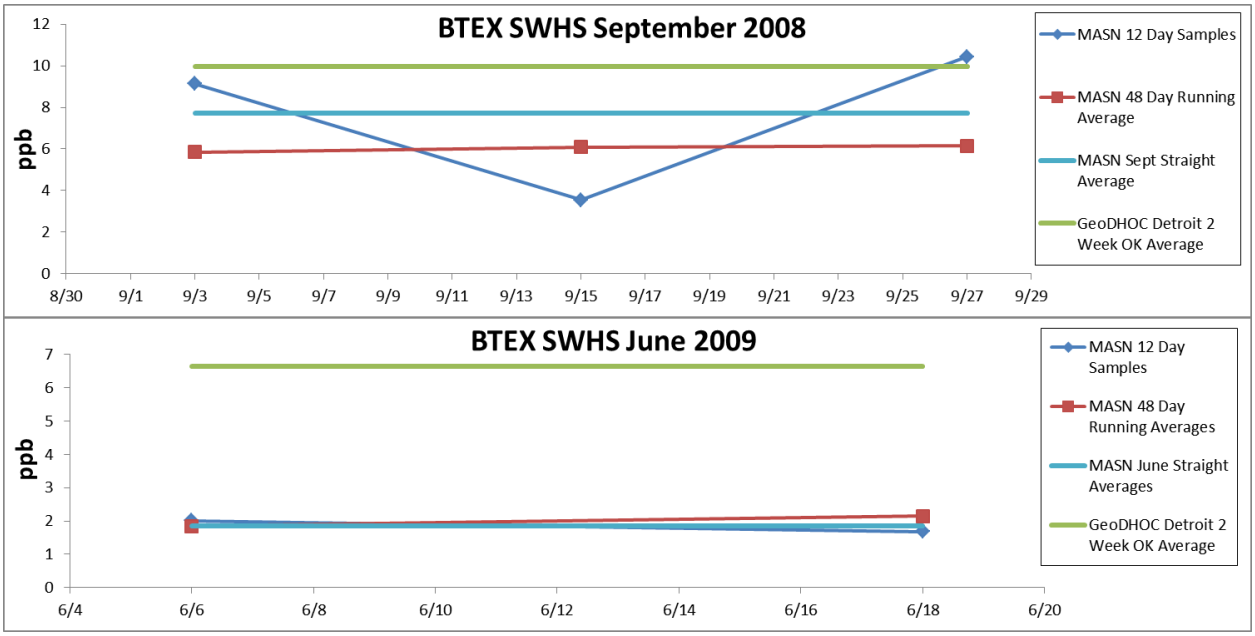


Figure A5. **Total BTEX concentrations for September 2008 and June 2009.**

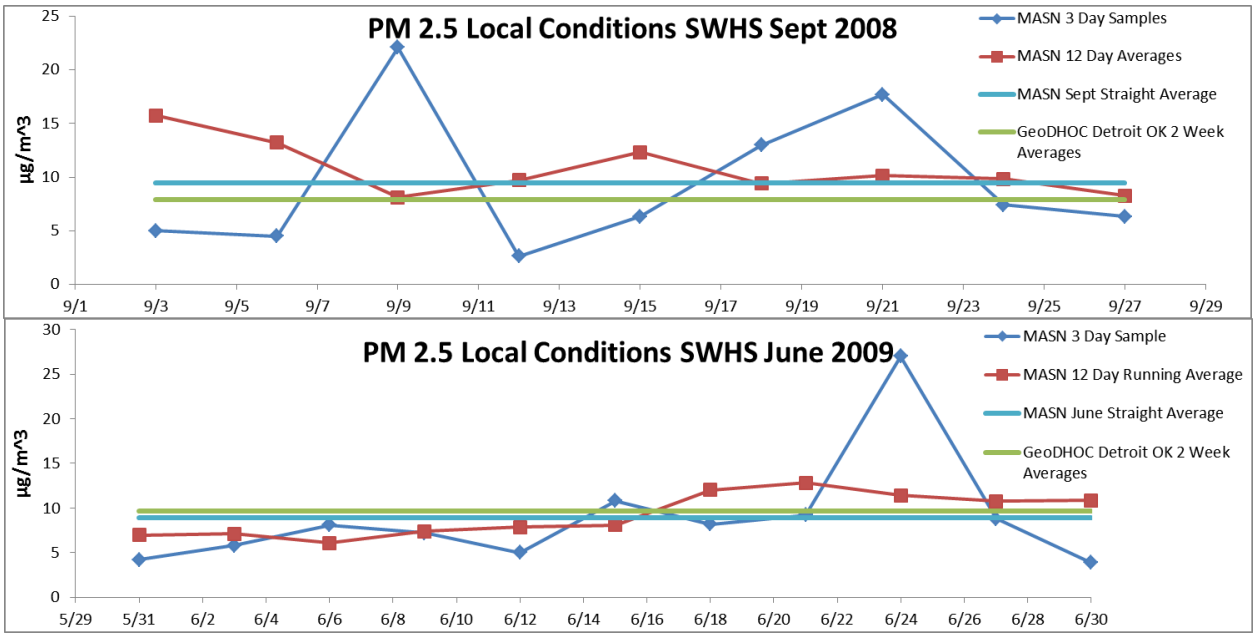


Figure A6. **$PM_{2.5}$ concentrations for September 2008 and June 2009.**

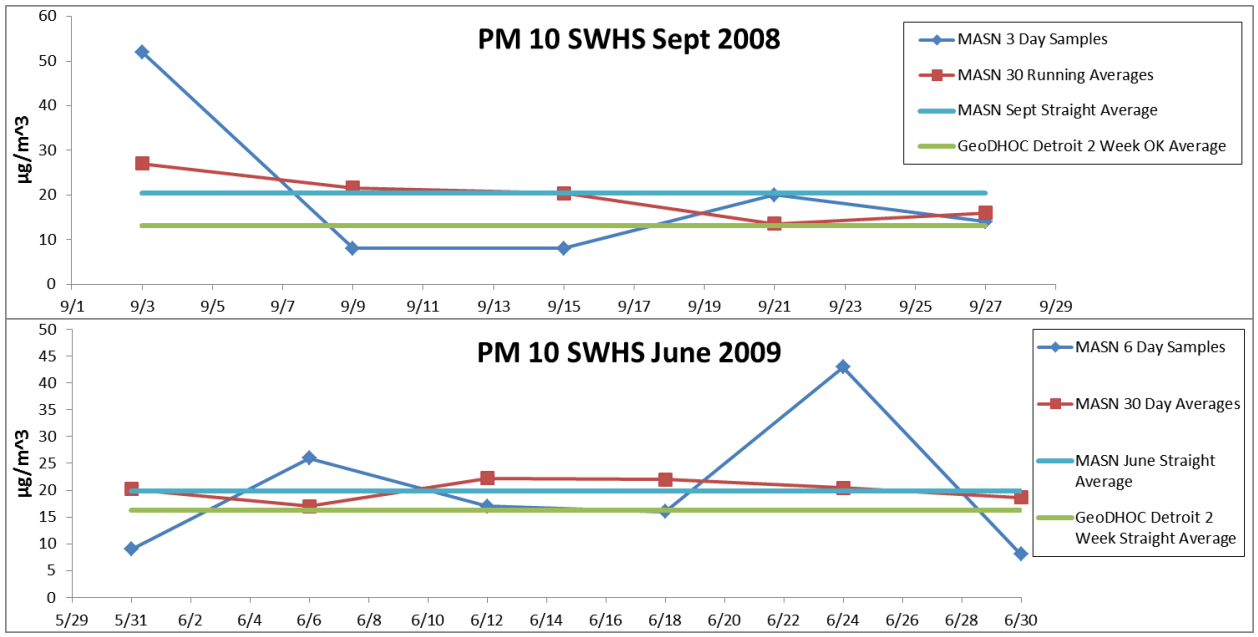


Figure A7. **PM_{10} concentration for September 2008 and June 2009.**

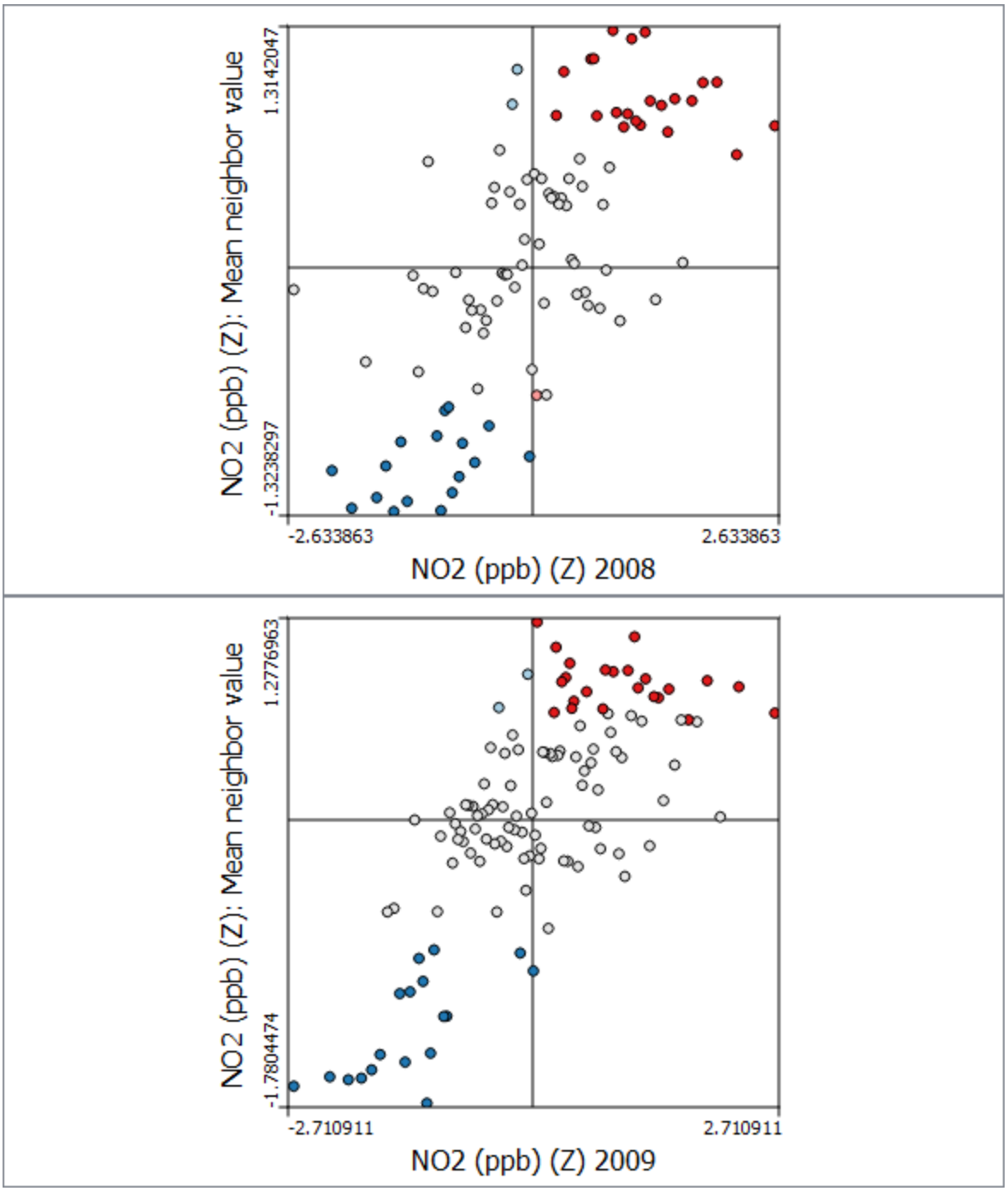


Figure A8a. Local Moran's I scatter plot for NO_2 with September 2008 on top and June 2009 below.

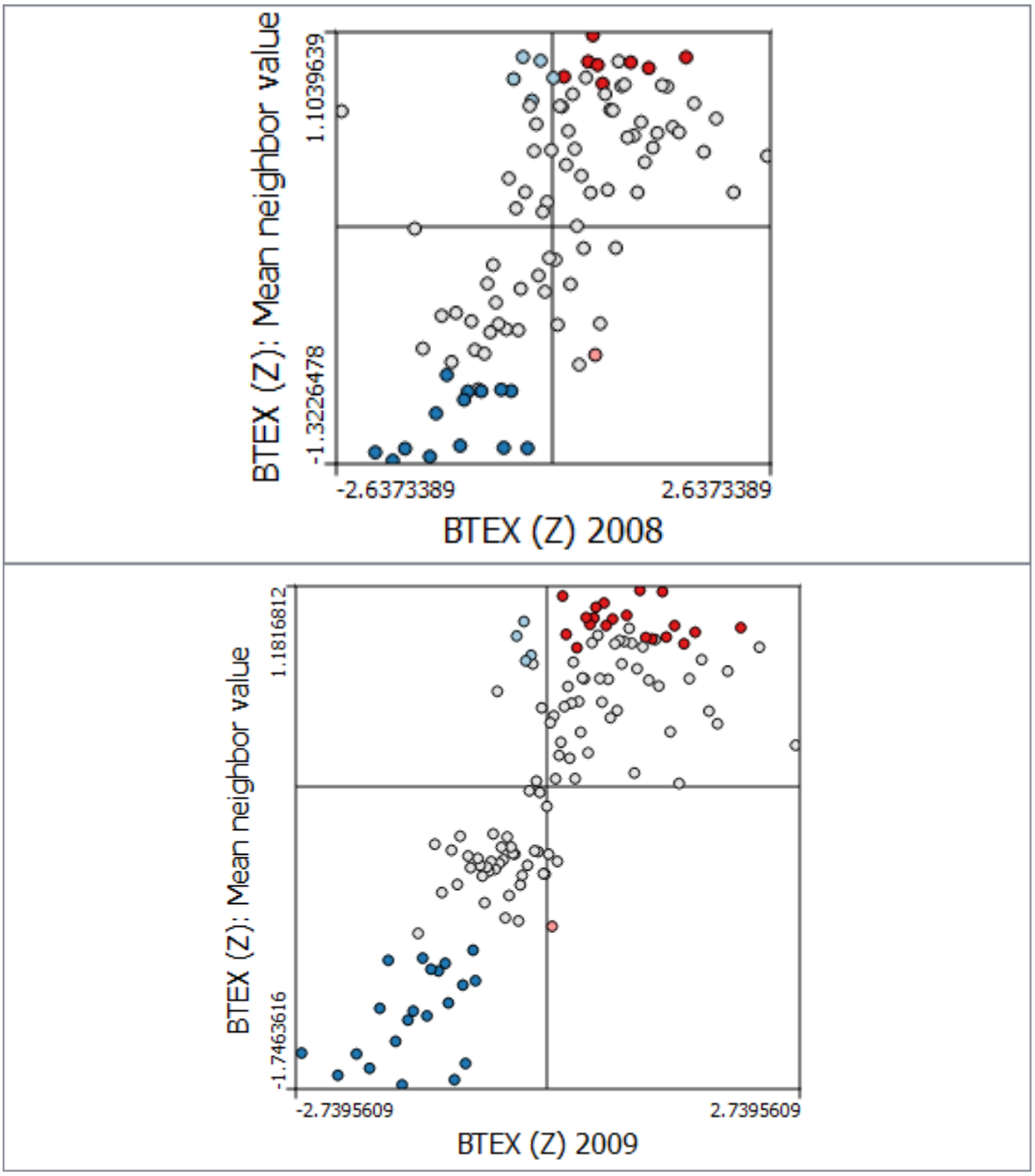


Figure A8b. Local Moran's I scatter plot for BTEX with September 2008 on top and June 2009 below.

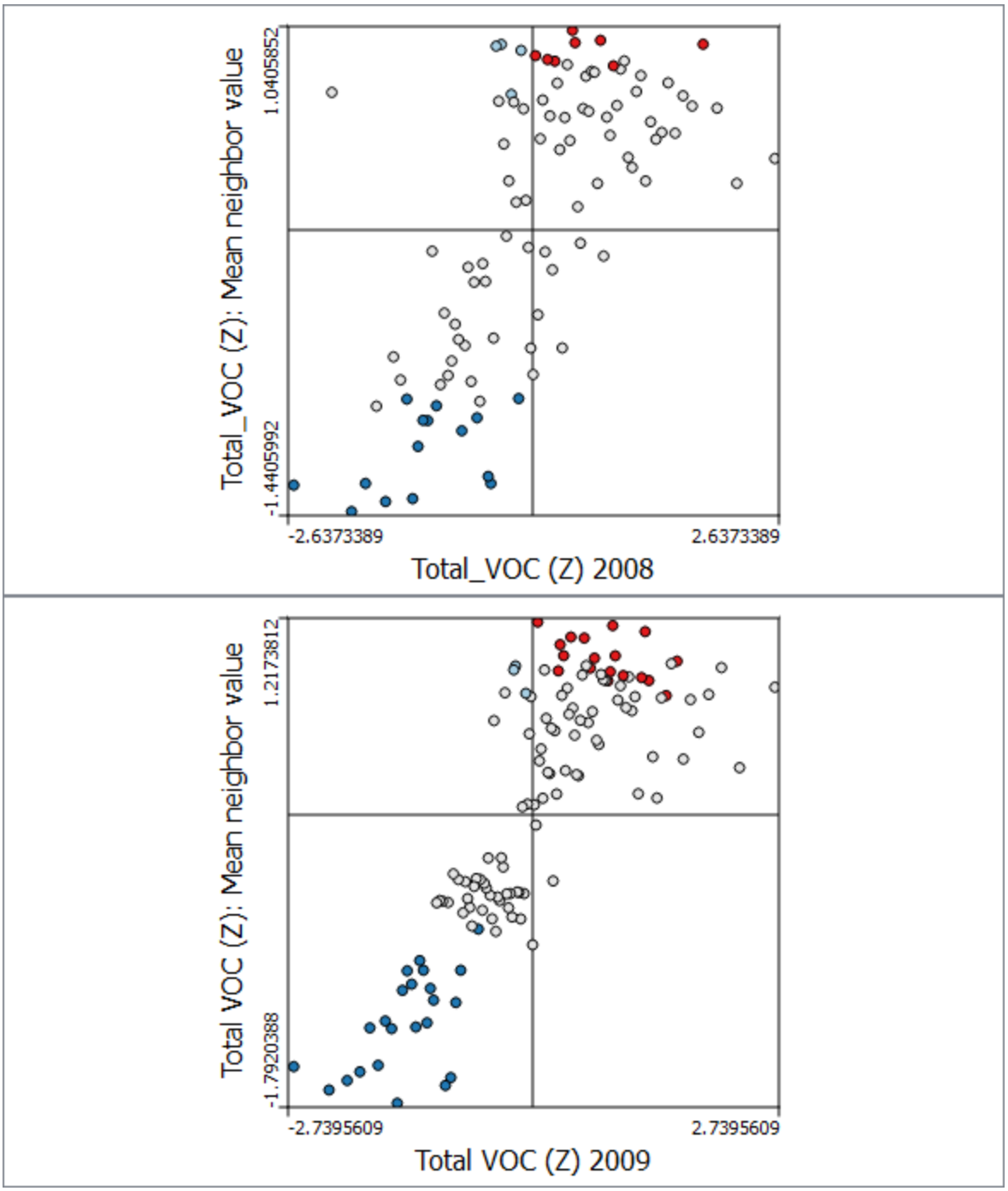


Figure A8c. Local Moran's I scatter plot for VOC with September 2008 on top and June 2009 below.

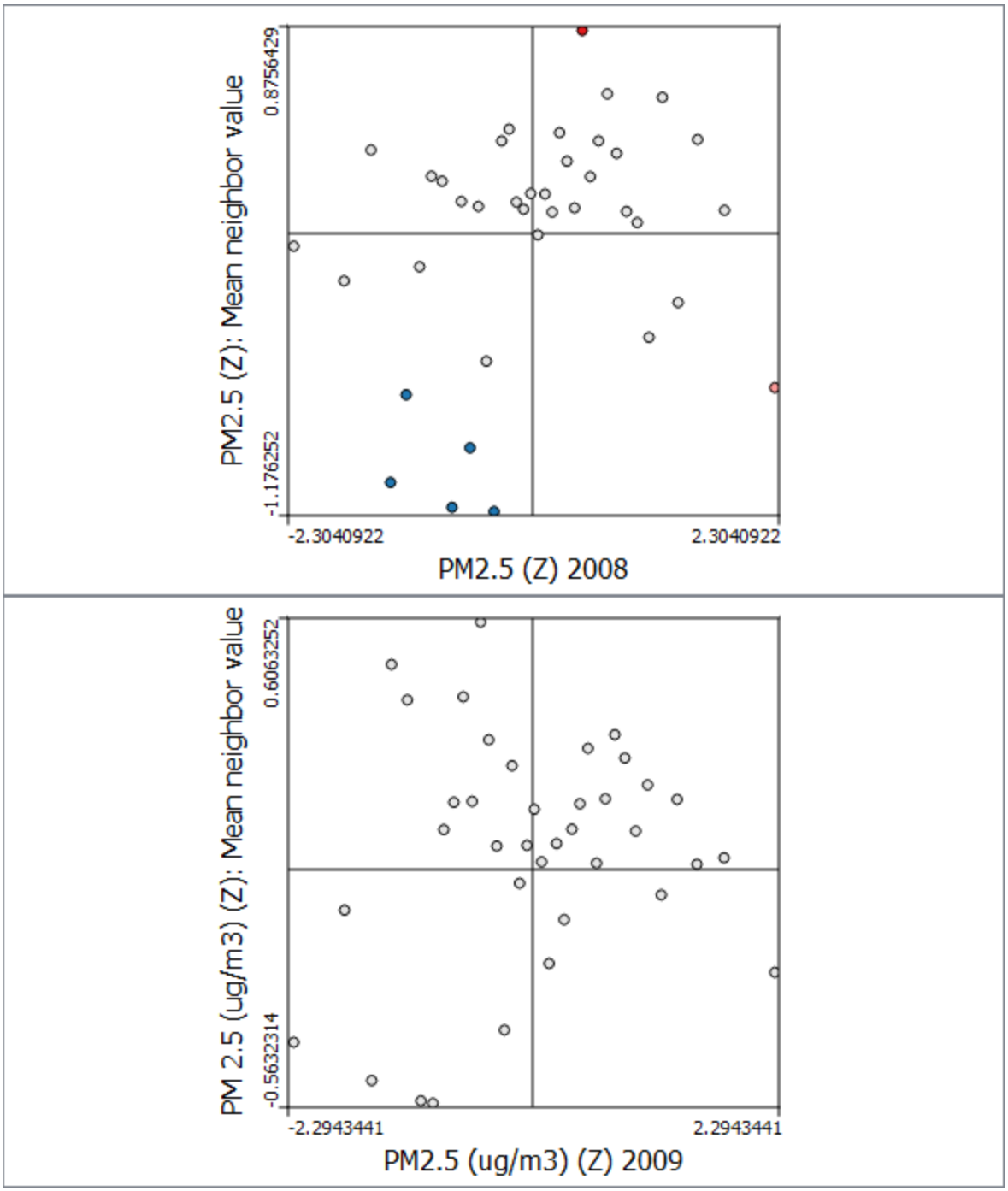


Figure A8d. Local Moran's I scatter plot for $PM_{2.5}$ with September 2008 on top and June 2009 below.

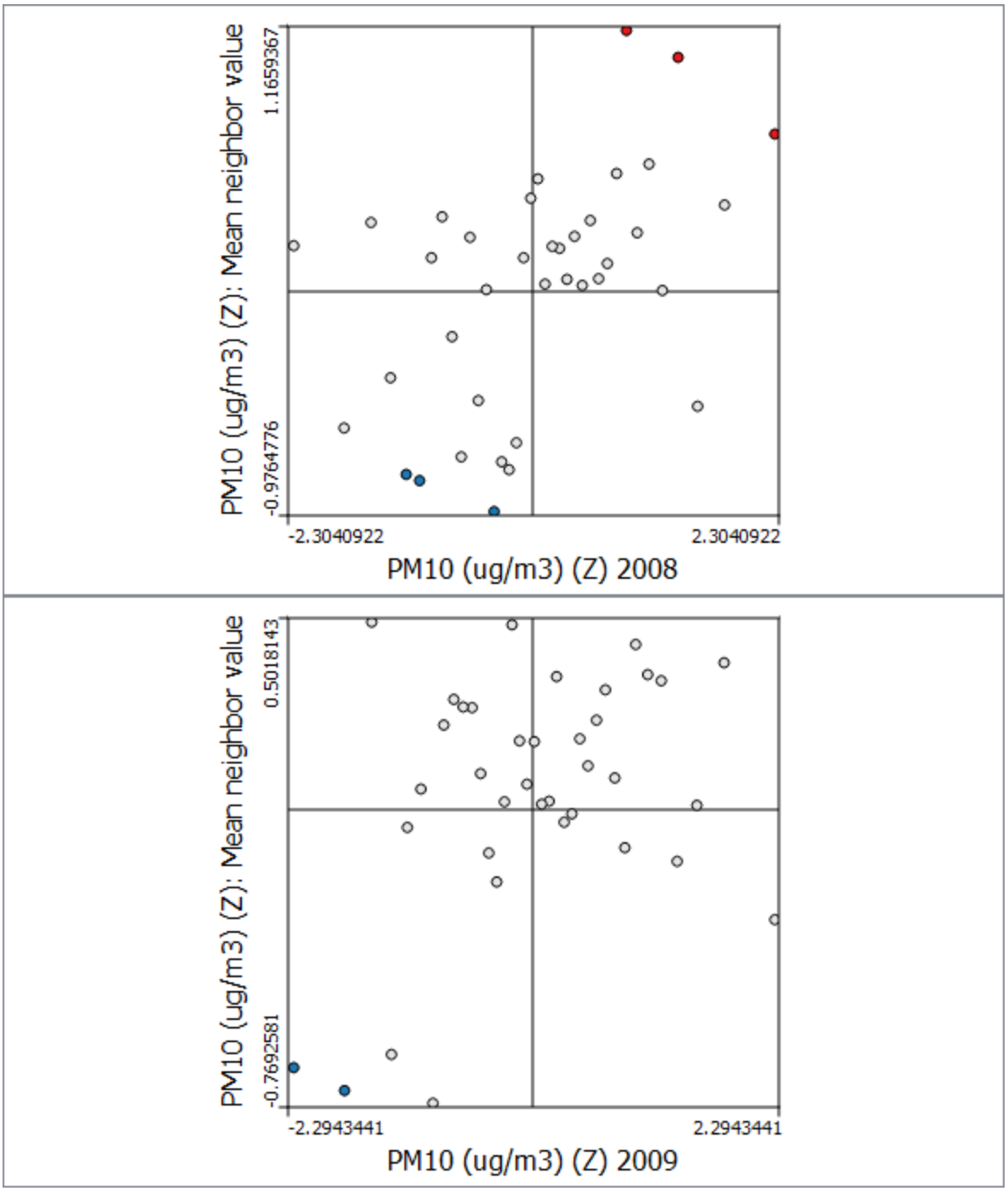
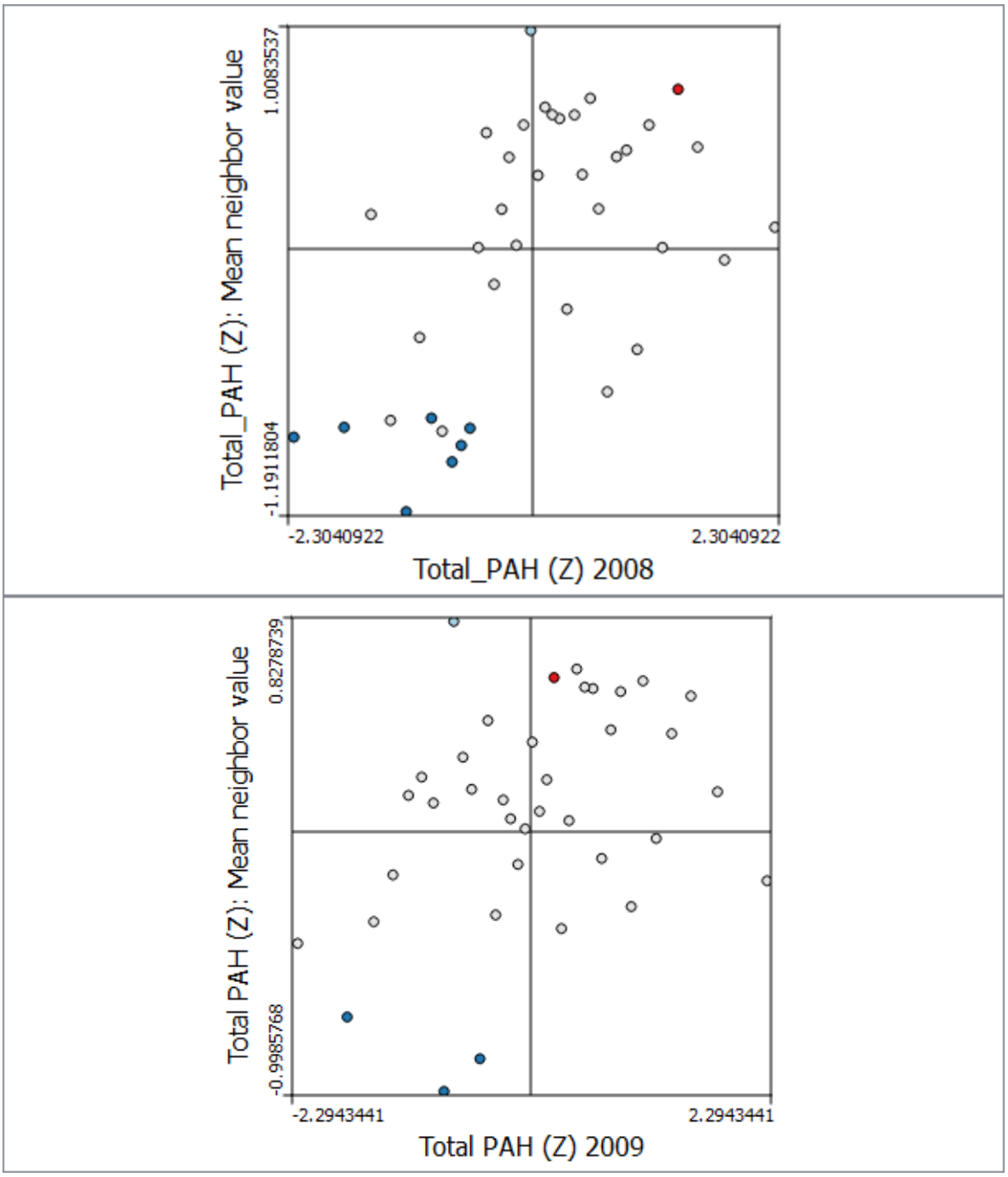


Figure A8e. Local Moran's I scatter plot for PM_{10} with September 2008 on top and June 2009 below.



<u>Figure A8f.</u> Local Moran's I scatter plot for PAH with September 2008 on top and June 2009 below.

Table A1. Cross validation results for Model 1 and Model 2 in September 2008 and June 2009 ordinary kriged models.

Model 1				
Year	Analyte	Mean error for interval	Mean absolute error for interval	Std. Dev.
2008	NO2	7.50E-02	1.5	2
2009	NO2	1.10E-01	1.9	2.6
2008	BTEX	1.40E-01	2.4	3.7
2009	BTEX	2.50E-02	0.29	1.6
2008	PM1	9.40E-04	0.55	0.82
2009	PM1	-1.70E-03	0.72	1.1
2008	PM1-2.5	-2.70E-03	0.84	1.5
2009	PM 1-2.5	1.60E-03	0.35	0.57
2008	Pm 2.5-10	2.20E-02	1.1	1.7
2009	Pm 2.5-10	-4.80E-02	0.9	1.3
Model 2				
Year	Analyte	Mean error for interval	Mean absolute error for interval	Std. Dev.
2008	NO2	4.63E-02	1.42	1.8
2009	NO2	7.53E-02	1.78	2.4
2008	BTEX	8.05E-02	1.84	2.6
2008	VOC	8.80E-02	2.58	3.5
2008	PM2.5	-2.54E-03	0.96	0.34
2009	PM2.5	1.96E-02	0.74	0.94
2008	PM10	3.61E-02	1.65	2.6
2009	PM10	-2.02E-02	1.44	2.0
2009	PAH	6.29E-01	9.46	12.0

Table A2. Detroit September 2008 through June 2009 monthly spatial model statistics.

Spatially interpolated 12 month series from Model 1													
		NO2			BTEX			PM2.5			PM10		
Year		Mean (ppb)	Std (ppb)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)
2008	Sep.	16.2	1.8	11.3	10	2.47	24.8	7.9	0.53	6.75	13.2	1.81	13.8
2008	Oct.	16.2	1.8	11.1	9.6	2.29	23.8	8.1	0.48	5.92	13.5	1.68	12.5
2008	Nov.	16.3	1.8	11.1	9.2	2.11	22.8	8.3	0.44	5.34	13.8	1.56	11.3
2008	Dec.	16.3	1.81	11.1	8.9	1.93	21.8	8.5	0.43	5.02	14.2	1.45	10.2
2009	Jan.	16.3	1.85	11.3	8.5	1.78	20.9	8.7	0.43	4.98	14.5	1.35	9.3
2009	Feb.	16.4	1.9	11.6	8.1	1.63	20.1	8.9	0.46	5.19	14.9	1.27	8.51
2009	Mar.	16.4	1.98	12.0	7.8	1.51	19.4	9.1	0.51	5.60	15.2	1.2	7.90
2009	Apr.	16.5	2.06	12.5	7.4	1.41	19.0	9.3	0.57	6.14	15.6	1.16	7.47
2009	May	16.5	2.16	13.1	7	1.34	19.0	9.5	0.64	6.75	15.9	1.15	7.21
2009	Jun.	16.5	2.27	13.8	6.7	1.3	19.6	9.7	0.72	7.42	16.2	1.16	7.12
2009	Jul.	16.4	1.98	12.0	7.8	1.51	19.4	9.1	0.51	5.6	15.2	1.20	7.9
2009	Aug.	16.3	1.81	11.1	8.9	1.93	21.8	8.5	0.43	5.02	14.2	1.45	10.2
Spatially interpolated 12 month series from Model 2													
		NO2			BTEX			Pm2.5			PM10		
Year		Mean (ppb)	Std (ppb)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)	Mean (µg/m^3)	Std (µg/m^3)	CV (%)
2008	Sep.	16.1	1.67	10.3	9.8	1.91	19.57	7.9	0.58	7.30	13.5	2.40	17.8
2008	Oct.	16.1	1.64	10.2	9.4	1.78	18.85	8.1	0.53	6.53	13.8	2.24	16.2
2008	Nov.	16.2	1.64	10.1	9.1	1.65	18.19	8.2	0.48	5.83	14.1	2.08	14.7
2008	Dec.	16.2	1.65	10.2	8.7	1.54	17.62	8.4	0.43	5.18	14.4	1.92	13.4
2009	Jan.	16.2	1.68	10.4	8.4	1.44	17.19	8.6	0.39	4.60	14.7	1.78	12.1
2009	Feb.	16.2	1.73	10.7	8.0	1.36	16.97	8.7	0.36	4.10	15.0	1.65	11.0
2009	Mar.	16.2	1.80	11.1	7.7	1.31	17.01	8.9	0.33	3.69	15.3	1.54	10.1
2009	Apr.	16.3	1.87	11.5	7.3	1.28	17.40	9.0	0.31	3.39	15.6	1.45	9.28
2009	May	16.3	1.97	12.1	7.0	1.28	18.22	9.2	0.30	3.22	15.9	1.38	8.67
2009	Jun.	16.3	2.07	12.7	6.7	1.3	19.6	9.4	0.30	3.19	16.2	1.34	8.25
2009	Jul.	16.2	1.80	11.1	7.7	1.31	17.01	8.9	0.33	3.69	15.3	1.54	10.1
2009	Aug.	16.2	1.65	10.2	8.7	1.54	17.62	8.4	0.43	5.18	14.4	1.92	13.4

Table A3. Model 1 monthly bulk shift values for NO_2, total BTEX, $PM_{2.5}$, and PM_{10}

Year	Month	NO2 (ppb)	BTEX ($\mu g/m^3$)	PM2.5 ($\mu g/m^3$)	PM10 ($\mu g/m^3$)
2008	Jan	5.1	-1.4	7.4	0.6
2008	Feb	5.5	-1.3	4.3	-2.8
2008	March	3.1	0.2	2.0	0.0
2008	April	0.3	2.2	2.1	8.6
2008	May	-2.6	0.6	-2.6	4.6
2008	June	-0.3	5.6	0.0	11.5
2008	July	-1.4	4.6	6.4	10.5
2008	Aug	-1.7	-2.3	1.1	5.4
2008	Sept	1.1	1.3	1.4	0.7
2008	Oct	2.9	-2.2	-1.2	-2.2
2008	Nov	4.4	-2.6	4.5	-6.7
2008	Dec	3.7	-3.4	3.2	-7.9
2009	Jan	8.3	-2.4	8.0	-7.2
2009	Feb	5.1	-0.3	5.6	-2.6
2009	March	4.1	0.4	1.5	-9.7
2009	April	-1.6	-2.7	-4.1	-1.3
2009	May	-0.2	-1.9	-1.5	-3.6
2009	June	-1.1	-1.3	-1.4	-0.7
2009	July	-2.2	0.0	-1.5	-6.1
2009	Aug	-1.0	-1.8	1.0	-3.9
2009	Sept	0.1	-3.3	2.2	1.9
2009	Oct	1.0	2.2	-1.8	-6.6
2009	Nov	2.5	0.3	3.8	4.1
2009	Dec	2.7	-0.9	1.5	-9.0
2010	Jan	1.4	-3.0	2.0	-7.6
2010	Feb	3.8	-1.8	-0.8	-1.8
2010	Mar	3.6	3.5	1.4	-4.9
2010	Apr	-0.2	0.5	-2.7	3.0
2010	May	-1.4	4.3	-1.4	2.4
2010	June	-3.8	1.2	2.8	-0.3
2010	July	-1.8	1.2	4.2	-1.3
2010	Aug	-1.0	-1.4	5.3	2.9
2010	Sept	-1.6	-2.2	-2.0	2.1
2010	Oct	2.3	3.3	-2.6	9.4
2010	Nov	3.9	-0.6	1.7	-3.1
2010	Dec	2.7	-2.1	2.8	-7.3

Table A4: Model 2 monthly bulk shift values for NO_2, total BTEX, $PM_{2.5}$, and PM_{10}.

		NO_2	BTEX	$PM_{2.5}$	PM_{10}
Year	Month	(ppb)	($\mu g/m^3$)	($\mu g/m^3$)	($\mu g/m^3$)
2008	Jan	5.1	-1.4	7.6	0.6
2008	Feb	5.5	-1.3	4.4	-3.1
2008	Mar	3.2	0.1	2.1	-0.7
2008	Apr	0.3	2.2	2.2	7.6
2008	May	-2.6	0.5	-2.4	3.3
2008	Jun	-0.2	5.5	-0.1	9.9
2008	Jul	-1.4	4.6	6.5	9.9
2008	Aug	-1.7	-2.2	1.1	5.7
2008	Sep	1	1.4	1.3	0.6
2008	Oct	2.9	-2.2	-1.3	-2.3
2008	Nov	4.3	-2.5	4.4	-6.8
2008	Dec	3.7	-3.4	3.2	-7.9
2009	Jan	8.3	-2.4	8.2	-7.2
2009	Feb	5.2	-0.3	5.6	-2.6
2009	Mar	4.1	0.3	1.6	-9.7
2009	Apr	-1.5	-2.8	-4.0	-1.2
2009	May	-0.1	-2	-1.3	-3.4
2009	Jun	-1	-1.4	-1.5	-0.6
2009	Jul	-2.2	0	-1.4	-6.1
2009	Aug	-1	-1.8	1.0	-3.9
2009	Sep	0	-3.2	2.1	1.8
2009	Oct	0.9	2.2	-1.9	-6.7
2009	Nov	2.5	0.3	3.8	4
2009	Dec	2.6	-0.9	1.5	-9.1
2010	Jan	1.4	-3	2.2	-7.7
2010	Feb	3.9	-1.8	-0.7	-1.8
2010	Mar	3.7	3.4	1.4	-4.9
2010	Apr	-0.2	0.4	-2.6	3
2010	May	-1.3	4.3	-1.2	2.5
2010	Jun	-3.7	1.1	2.7	-0.2
2010	Jul	-1.8	1.2	4.3	-1.3
2010	Aug	-1	-1.3	5.3	2.9
2010	Sep	-1.7	-2.1	-2.2	1.9
2010	Oct	2.3	3.4	-2.7	9.3
2010	Nov	3.9	-0.5	1.7	-3.2
2010	Dec	2.6	-2	2.8	-7.3

REFERENCES

Anselin, L., 1995, Local Indicators of Spatial Association—LISA: Geographical Analysis, v. 27, no. 2, p. 93-115.

Barnett, V., and Lewis, T., 1994, Outliers in Statistical Data, New York, New York, John Wiley, 584 p.

Batterman, S. A., Peng, C.-Y., and Braun, J., 2002, Levels and composition of volatile organic compounds on commuting routes in Detroit, Michigan: Atmospheric Environment, v. 36, no. 39–40, p. 6015-6030.

Baxter, L. K., Dionisio, K. L., Burke, J., Ebelt Sarnat, S., Sarnat, J. A., Hodas, N., Rich, D. Q., Turpin, B. J., Jones, R. R., Mannshardt, E., Kumar, N., Beevers, S. D., and Ozkaynak, H., 2013, Exposure prediction approaches used in air pollution epidemiology studies: key findings and future recommendations: Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 6, p. 654-659.

Beevers, S. D., Kitwiroon, N., Williams, M. L., Kelly, F. J., Ross Anderson, H., and Carslaw, D. C., 2013, Air pollution dispersion models for human exposure predictions in London: Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 6, p. 647-653.

Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P., Gehring, U., Heinrich, J., Cyrys, J., Bellander, T., Lewne, M., and Brunekreef, B., 2003, Estimating Long-Term Average Particulate Air Pollution Concentrations: Application of Traffic Indicators and Geographic Information Systems: Epidemiology, v. 14, no. 2, p. 228-239.

Brook, J. R., Dann, T. F., and Bonvalot, Y., 1999, Observations and interpretations from the Canadian fine particle monitoring program: Journal of the Air & Waste Management Association, v. 49, no. 9, p. 35-44.

Chang-Tien, L., Dechang, C., and Yufeng, K., 2003, Detecting spatial outliers with multiple attributes *in* Proceedings 15th IEEE International Conference 3-5 Nov. 2003 of Conference, p. 122-128.

Chow, J., Bachmann, J., Hsu, Y.-C., Chen, S.-K., Tsai, J.-H., Chiang, H.-L., Fujita, E., Zielinska, B., Campbell, D., Arnott, W., Sagebiel, L. J., Gabele, P., Crews, W., Snow, R., Clark, N., Wayne, W., Lawson, D., Arnott, W., Brown, S., Frankel, A., Raffuse, S., Roberts, P., Hafner, H., Anderson, D., Eklund, B., Simon, M., Regmi, S., Ongwandee, M., Morrison, G., Fitch, M., and Surampalli, R., 2007, The A&WMA 2007 Critical Review -- Will the Circle Be Unbroken: A History of the U.S. National Ambient Air Quality Standards: Journal of the Air & Waste Management Association, v. 57, no. 6, p. 652-697.

Clougherty, J. E., Kheirbek, I., Eisl, H. M., Ross, Z., Pezeshki, G., Gorczynski, J. E., Johnson, S., Markowitz, S., Kass, D., and Matte, T., 2013, Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS): Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 3, p. 232-240.

Cocheo, C., Sacco, P., Ballesta, P. P., Donato, E., Garcia, S., Gerboles, M., Gombert, D., McManus, B., Patier, R. F., Roth, C., de Saeger, E., and Wright, E., 2008, Evaluation of the best compromise between the urban air quality monitoring resolution by diffusive sampling and resource requirements: Journal of Environmental Monitoring, v. 10, no. 8, p. 941-950.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., and Speizer, F. E., 1993, An association between air pollution and mortality in six U.S. cities: The New England Journal of Medicine, v. 329, no. 24, p. 1753-1759.

Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation, New York, New York, Oxford University Press, Applied Geostatistics Series, 496 p.

Health Canada, 2000, Detroit River Ara of Concern: Health Data and Satistics for the Population of Windsor and Region (1986-1992). A Technical Report for the RAP Community, <http://web.archive.org/web/20030416041841/www.hc-sc.gc.ca/ehp/ehd/bch/bioregional/detroit_river.pdf>.

Hewitt, N. C., 1991, Spatial variations in nitrogen dioxide concentrations in an urban area: Atmospheric Environment. Part B. Urban Atmosphere, v. 25, no. 3, p. 429-434.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D., 2008, A review of land-use regression models to assess spatial variation of outdoor air pollution: Atmospheric Environment, v. 42, no. 33, p. 7561-7578.

Isaaks, E. H., and Srivastava, R. M., 1989, An Introduction to Applied Geostatistics, New York , New York, Oxford University Press, 561 p.

Isakov, V., Touma, J., Johnson, M. M., and Ozkaynak., H. A., 2011, Development and evaluation of land-use regression models using modeled air quality concentrations, NATO - Air Pollution Modeling and its Application, Volume XXI: Springer Netherlands, Netherlands, p. 717-722.

Jerrett, M., Arain, M. A., Kanaroglou, P., Beckerman, B., Crouse, D., Gilbert, N. L., Brook, J. R., Finkelstein, N., and Finkelstein, M. M., 2007, Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada: Journal of Toxicology and Environmental Health, Part A, v. 70, no. 3-4, p. 200-212.

Jerrett, M., Burnett, R. T., Ma, R., Pope, C. A. I., Krewski, D., Newbold, K. B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E. E., and Thun, M. J., 2005, Spatial analysis of air pollution and mortality in Los Angeles: Epidemiology, v. 16, no. 6, p. 727-736.

Johnson, M., Macneill, M., Grgicak-Mannion, A., Nethery, E., Xu, X., Dales, R., Rasmussen, P., and Wheeler, A., 2013, Development of temporally refined land-use regression models predicting daily household-level air pollution in a panel study of lung function among asthmatic children: Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 3, p. 259-267.

Kearney, J., Wallace, L., MacNeill, M., Xu, X., VanRyswyk, K., You, H., Kulka, R., and Wheeler, A. J., 2011, Residential indoor and outdoor ultrafine particles in Windsor, Ontario: Atmospheric Environment, v. 45, no. 40, p. 7583-7593.

Kiefer, N. a., and Salmon, M., 1983, Testing normality in econometric models: Economic Letters, v. 11, p. 123-127.

Kim, E., Hopke, P. K., Pinto, J. P., and Wilson, W. E., 2005, Spatial variability of fine particle mass, components, and source contributions during the Regional Air Pollution Study in St. Louis: Environmental Science & Technology, v. 39, no. 11, p. 4172-4179.

Kracht, O., Reuter, H. I., and Gerboles, M., 2013, A spatio-temporal screening tool for outlier detection in long term / large scale air quality observation time series and monitoring networks [abs.], EGU General Assembly 2013, Vienna, Austria, 7-12 April of Conference.

Künzli, N., Jerrett, M., Mack, W. J., Beckerman, B., LaBree, L., Gilliland, F., Thomas, D., Peters, J., and Hodis, H. N., 2004, Ambient Air Pollution and Atherosclerosis in Los Angeles: Environmental Health Perspectives, v. 113, no. 2, p. 201-206.

Lemke, L. D., Lamerato, L. E., Xu, X., Booza, J. C., Reiners, J. J., Jr., Raymond Iii, D. M., Villeneuve, P. J., Lavigne, E., Larkin, D., and Krouse, H. J., 2013, Geospatial relationships of air pollution and acute asthma events across the Detroit-Windsor international border: Study design and preliminary results: Journal of Exposure Science and Environmental Epidemiology.

Li, L., Wu, J., Ghosh, J. K., and Ritz, B., 2013, Estimating spatiotemporal variability of ambient air pollutant concentrations with a hierarchical model: Atmospheric Environment, v. 71, p. 54-63.

Luginaah, I., Xu, X., Fung, K. Y., Grgicak-Mannion, A., Wintermute, J., Wheeler, A., and Brook, J., 2006, Establishing the spatial variability of ambient nitrogen dioxide in Windsor, Ontario: International Journal of Environmental Studies, v. 63, no. 4, p. 487-500.

Mayer, H., 1999, Air pollution in cities, *in* Proceedings Proceedings of the 1996 International Conference on Urban Climatology, Essen, Germany, June 10, 1996 - June 14, 1996 of Conference, Volume 33, Elsevier Ltd, p. 4029-4037.

McGrath, D., and Zhang, C., 2003, Spatial distribution of soil organic carbon concentrations in grassland of Ireland: Applied Geochemistry, v. 18, no. 10, p. 1629-1639.

MDEQ, 2008, 2008 Annual Air Quality Report, http://www.michigan.gov/documents/deq/deq-aqd-air-2008-Air-Quality-Report_296426_7.pdf

MDEQ, 2013, Michigan's 2014 Ambient Air Monitoring Network Review, http://www.michigan.gov/documents/deq/deq-aqd-aqe-monitoring-network-review-2014_426389_7.pdf

Mercer, L. D., Szpiro, A. A., Sheppard, L., Lindstrom, J., Adar, S. D., Allen, R. W., Avol, E. L., Oron, A. P., Larson, T., Liu, L. J., and Kaufman, J. D., 2011, Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NOx) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air): Atmospheric Environment, v. 45, no. 26, p. 4412-4420.

Miller, L., 2012, A long term geospatial investigation of air quality in Windsor, Ontario, Canada - Comparison to nearby cities and applications of inter-species ratios [NR78853 Ph.D.]: University of Windsor (Canada), 239 p.

Miller, L., Lemke, L. D., Xu, X., Molaroni, S. M., You, H., Wheeler, A. J., Booza, J., Grgicak-Mannion, A., Krajenta, R., and Graniero, P., 2010, Intra-urban correlation and spatial variability of air toxics across an international airshed in Detroit, Michigan (USA) and Windsor, Ontario (Canada): Atmospheric Environment, v. 44, no. 9, p. 1162-1174.

Miller, L., Lemke, L. D., Xu, X., Molaroni, S. M., You, H., Wheeler, A. J., Booza, J., Grgicak-Mannion, A., Krajenta, R., Graniero, P., Krouse, H., Lamerato, L., Raymond, D., Reiners, J., and Weglicki, L., 2012a, Corrigendum to "Intra-urban correlation and spatial variability of air toxics across an international airshed in Detroit, Michigan (USA) and Windsor, Ontario (Canada)" [Atmos. Environ. 44 (2010) 1162–1174]: Atmospheric Environment, v. 46, p. 683-686.

Miller, L., Xu, X., Grgicak-Mannion, A., Brook, J., and Wheeler, A., 2012b, Multi-season, multi-year concentrations and correlations amongst the BTEX group of VOCs in an urbanized industrial city: Atmospheric Environment, v. 61, p. 305-315.

Molaroni, S. M., 2010, Modeling ambient air quality in the Detroit-Windsor airshed [Masters Thesis]: Wayne State University Thesis, 148 p.

Mölter, A., Lindley, S., de Vocht, F., Simpson, A., and Agius, R., 2010, Modelling air pollution for epidemiologic research – Part II: Predicting temporal variation through land use regression: Science of The Total Environment, v. 409, no. 1, p. 211-217.

Mukerjee, S., Smith, L. A., Norris, G. A., Morandi, M. T., Gonzales, M., Noble, C. A., Neas, L. M., and Özkaynak, A. H., 2004, Field method comparison between passive air samplers and continuous monitors for VOCs and NO2 in El Paso, Texas: Journal of the Air & Waste Management Association, v. 54, no. 3, p. 307-319.

National Climatic Data Center, 2014, NOAA ISD-Lite 2008 FTP Directory, ftp://ftp.ncdc.noaa.gov./pub/data/noaa (March 12)

O'Leary, B. F., and Lemke, L. D., 2014, Modeling spatiotemporal variability of intra-urban air pollutants in Detroit: A pragmatic approach: Atmospheric Environment, v. 94, p. 417-427.

Ozkaynak, H., Baxter, L. K., Dionisio, K. L., and Burke, J., 2013, Air pollution exposure prediction approaches used in air pollution epidemiology studies: Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 6, p. 566-572.

Pinto, J. P., Lefohn, A. S., and Shadwick, D. S., 2004, Spatial sariability of PM2.5 in urban areas in the United States: Journal of the Air & Waste Management Association, v. 54, no. 4, p. 440-449.

Pope, C. A. I., Ezzati, M., and Dockery, D. W., 2009, Fine-particulate air pollution and life expectancy in the United State: New England Journal of Medicine, v. 360, p. 376-386.

Reiners, J., Cassidy-Bushrow, A., Goovaerts, P., Lemke, L., Lamerato, L., O'Leary, B., Mathieu, M., and Sperone, G., 2014, Ambient air quality in Detroit: Relationship to adverse birth

outcomes [abs.], *in* Proceedings C.S. Mott Center Scientific Retreat, Atheneum Hotel, Detroit, MI, of Conference, Volume O/09, p. 9.

Rodes, C. E., Lawless, P. A., Thornburg, J. W., Williams, R. W., and Croghan, C. W., 2010, DEARS particulate matter relationships for personal, indoor, outdoor, and central site settings for a general population: Atmospheric Environment, v. 44, no. 11, p. 1386-1399.

Romanowicz, R., Young, P., Brown, P., and Diggle, P., 2006, A recursive estimation approach to the spatio-temporal analysis and modelling of air quality data: Environmental Modelling & Software, v. 21, no. 6, p. 759-769.

Ross, Z., Ito, K., Johnson, S., Yee, M., Pezeshki, G., Clougherty, J. E., Savitz, D., and Matte, T., 2013, Spatial and temporal estimation of air pollutants in New York City: Exposure assignment for use in a birth outcomes study: Environ Health, v. 12, p. 51.

Sajani, S. Z., Scotto, F., Lauriola, P., Galassi, F., and Montanari, A., 2004, Urban air pollution monitoring and correlation properties between fixed-site stations: Journal of the Air & Waste Management Association, v. 54, no. 10, p. 1236-1241.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., and Zeger, S. L., 2000, Fine particulate air pollution and mortaility in 20 U.S. cities, 1987-1994: The New England Journal of Medicine, v. 343, no. 24, p. 1742-1749.

Sampson, P. D., Szpiro, A. A., Sheppard, L., Lindström, J., and Kaufman, J. D., 2011, Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data: Atmospheric Environment, v. 45, no. 36, p. 6593-6606.

Sarnat, S. E., Sarnat, J. A., Mulholland, J., Isakov, V., Ozkaynak, H., Chang, H. H., Klein, M., and Tolbert, P. E., 2013, Application of alternative spatiotemporal metrics of ambient air

pollution exposure in a time-series epidemiological study in Atlanta: Journal of Exposure Science and Environmental Epidemiology, v. 23, no. 6, p. 593-605.

Simon, C., Sills, R., Depa, M., Sadoff, M., Kim, A., and Heindorf, M., 2005, Detroit air toxics initiative risk assessment report, 228 p., http://www.michigan.gov/documents/DATI_-_COMPLETE_FINAL_REPORT_11-9-05_142053_7.pdf

Stocco, C., MacNeill, M., Wang, D., Xu, X., Guay, M., Brook, J., and Wheeler, A. J., 2008, Predicting personal exposure of Windsor, Ontario residents to volatile organic compounds using indoor measurements and survey data: Atmospheric Environment, v. 42, no. 23, p. 5905-5912.

Torres, J. M., Garcia Nieto, P. J., Alejano, L., and Reyes, A. N., 2011, Detection of outliers in gas emissions from urban areas using functional data analysis: Journal of Hazardous Materials, v. 186, no. 1, p. 144-149.

U.S. EPA, 1999, Compendium Method TO-15: Determination Of Volatile Organic Compounds (VOCs) In Air Collected In Specially-Prepared Canisters And Analyzed By Gas Chromatography/ Mass Spectrometry GC/MS): U.S. Environmental Protection Agency, EPA/625/R-96/010b, 67 p., http://www.epa.gov/ttnamti1/files/ambient/airtox/to-15r.pdf

U.S. EPA, 2009, Detroit River-Western Lake Erie Basin Indicator Project, http://www.epa.gov/med/grosseile_site/indicators/air-pollution.html (March 2014)

U.S. EPA, 2013, List of designated reference and equivalent methods: U.S. Environmental Protection Agency, 60 p., http://epa.gov/ttn/amtic/files/ambient/criteria/reference-equivalent-methods-list.pdf

Vardoulakis, S., Lumbreras, J., and Solazzo, E., 2009, Comparative evaluation of nitrogen oxides and ozone passive diffusion tubes for exposure studies: Atmospheric Environment, v. 43, no. 16, p. 2509-2517.

Vardoulakis, S., Solazzo, E., and Lumbreras, J., 2011, Intra-urban and street scale variability of BTEX, NO_2 and O_3 in Birmingham, UK: Implications for exposure assessment: Atmospheric Environment, v. 45, no. 29, p. 5069-5078.

Vicedo-Cabrera, A. M., A., B., Grisotto, L., Barbone, F., and Catelan, D., 2013, A Bayesian kriging model for estimating residential exposure to air pollution of children living in a high-risk area in Italy.: Geospatial Health, v. 8, no. 1, p. 87-95.

Walker, I. J., Eamer, J. B. R., and Darke, I. B., 2013, Assessing significant geomorphic changes and effectiveness of dynamic restoration in a coastal dune ecosystem: Geomorphology, v. 199, p. 192-204.

Webster, R., and Oliver, M. A., 2007, Geostatistics for Environmental Scientists, West Sussex, England, John Wiley and Sons Ltd., Statistics in Practice, 300 p.

Wheeler, A., Smithdoiron, M., Xu, X., Gilbert, N., and Brook, J., 2008, Intra-urban variability of air pollution in Windsor, Ontario—Measurement and modeling for human exposure assessment: Environmental Research, v. 106, no. 1, p. 7-16.

Williams, R., Rea, A., Vette, A., Croghan, C., Whitaker, D., Stevens, C., McDow, S., Fortmann, R., Sheldon, L., Wilson, H., Thornburg, J., Phillips, M., Lawless, P., Rodes, C., and Daughtrey, H., 2009, The design and field implementation of the Detroit Exposure and Aerosol Research Study: Journal of Exposure Science and Environmental Epidemiology, v. 19, no. 7, p. 643-659.

Wilson, J. G., Kingham, S., Pearce, J., and Sturman, A. P., 2005, A review of intraurban variations in particulate air pollution: Implications for epidemiological research: Atmospheric Environment, v. 39, no. 34, p. 6444-6462.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Peacock, J., Anderson, R. H., Le Tertre, A., Bobros, J., Celko, M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Hoyos, S. P., Wichmann, H. E., and Katsouyanni, K., 2003, The temporal pattern of respiratory and heart disease mortality in response to air pollution: Environmental Health Perspectives, v. 111, no. 9, p. 1188-1193.

Zhang, C., Luo, L., Xu, W., and Ledwith, V., 2008, Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland: Science of the Total Environment, v. 398, no. 1-3, p. 212-221.

Zhang, C., and McGrath, D., 2004, Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods: Geoderma, v. 119, no. 3-4, p. 261-275.

Zou, B., Peng, F., Wan, N., Mamady, K., and Wilson, G. J., 2014, Spatial cluster detection of air pollution exposure inequities across the United States: PLoS ONE, v. 9, no. 3, p. e91917.

ABSTRACT

EVALUATING SPATIAL OUTLIERS AND INTEGRATING TEMPORAL DATA IN AIR POLLUTION MODELS FOR THE DETROIT-WINDSOR AIRSHED

by

Brendan Francis O'Leary

August 2014

Advisor: Dr. Lawrence Lemke

Major: Geology

Degree: Master of Science

The heterogeneous nature of urban air complicates human exposure estimates and creates a need for accurate, highly detailed spatiotemporal air contaminant models. The study expands on previous investigations by the Geospatial Determinants of Health Outcomes Consortium that examined relationships between air pollutant distributions and asthma exacerbations. Two approaches, the removal of spatial data outliers and the integration of spatial and temporal data, were used to refine air quality models in the Detroit and Windsor international airshed. The evaluation of associations between the resulting air quality models and asthma exacerbations in Detroit and Windsor revealed weaker correlations with spatial outliers removed but improved correlations with the addition of temporal data. Recommendations for future work include increasing the spatial and temporal resolution of the asthma datasets and incorporating Windsor NAPS data through temporal scaling to help confirm the findings of the Detroit temporal scaling.

AUTOBIOGRAPHICAL STATEMENT

I began my education at Wayne State University in the preschool program at the Childhood Development Lab, now known as the Early Childhood Center of the Merrill-Palmer Skillman Institute. I completed my high school education at the University of Detroit Jesuit High School and my Bachelors of Science Degree in Environmental Geology and Environmental Science at Allegheny College in Meadville, Pennsylvania. During my undergraduate program I complete the Semester in Environmental Science research program at Marine Biology Laboratory in Woods Hole, Massachusetts. My research project focused on how permeable reactive barriers mitigate eutrophication of estuaries around Cape Cod. My undergraduate comprehensive research project at Allegheny College evaluated the influence of geomorphic controls on pre-settlement forest distribution. After Allegheny College, I was fortunate to return to Wayne State University to complete my Master of Science Degree in Geology. I hope to remain in Detroit for the foreseeable future and look to use my knowledge in geospatial modeling in my new position as a geologist with ARCADIS.