11-10-2014

Statistical Analysis and Modeling of PM$_{2.5}$ Speciation Metals and Their Mixtures

Boubakari Ibrahimou
University of South Florida, bibrahim@fiu.edu

Statistical Analysis and Modeling of $PM_{2.5}$ Speciation Metals and Their Mixtures

by

Boubakari Ibrahimou

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Epidemiology & Biostatistics
College of Public Health
University of South Florida

Major Professor: Yiliang Zhu, Ph.D.
Getachew Dagne, Ph.D.
Foday Jaward, Ph.D.
Hamisu Salihu, MD, Ph.D.
Wei Wang, Ph.D.

Date of Approval:
November 10, 2014

Keywords: Air pollution, $PM_{2.5}$, metals, mixtures, Bayesian inference,
Spatio-temporal, Markov chain Monte Carlo.

To the memories of my Parents

Madji Nassourou and Yerima Boubakari Ahmadou

TABLE OF CONTENTS

ABSTRACT

Exposure to fine particulate matter ($PM_{2.5}$) in the ambient air is associated with various health effects. There is increasing evidence which implicates the central role played by specific chemical components such as heavy metals of $PM_{2.5}$. Given the fact that humans are exposed to complex mixtures of environmental pollutants such as $PM_{2.5}$, research efforts are intensifying to study the mixtures composition and the emission sources of ambient PM, and the exposure-related health effects. Factor analysis as well source apportionment models are statistical tools potentially useful for characterizing mixtures in $PM_{2.5}$. However, classic factor analysis is designed to analyze samples of independent data. To handle (spatio-)temporally correlated $PM_{2.5}$ data, a Bayesian approach is developed and using source apportionment, a latent factor is converted to a mixture by utilizing loadings to compute mixture coefficients. Additionally there have been intensified efforts in studying the metal composition and variation in ambient PM as well as its association with health outcomes. We use non parametric smoothing methods to study the spatio-temporal patterns and variation of common PM metals and their mixtures. Lastly the risk of low birth weight following exposure to metal mixtures during pregnancy is being investigated.

1 INTRODUCTION

1.1 Air Pollution and $PM_{2.5}$

Air pollution is a mixture of solid particles, gases, biological molecules, or other harmful materials into the air. It has the potential of causing diseases, death to humans, damage to other living organisms such as food crops, vegetations, or the natural or built environment. To protect people and the environment, the Clean Air Act, which was last amended by the congress in 1990, requires the Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS) for pollutants considered harmful to public health and the environment. EPA has set NAAQS for six principal pollutants, which are called "criteria" pollutants: Carbon Monoxide (CO), Lead (pb), Nitogen Dioxide (NO_2), Ozone (O_3), Sulfur Dioxide (SO_2), and Particulate Matter (PM_{10}, $PM_{2.5}$). Air quality data such as CO, pb, NO_2, O_3, SO_2, PM_{10} and $PM_{2.5}$, as well as $PM_{2.5}$ speciation chemicals data are housed in the US EPA Air Quality System (AQS) data base. AQS is divided into several groups around the continental US called Metropolitan Statistical Area (MSA) with each MSA containing several monitoring stations.

Particulate matter, also known as particle pollution or PM, is a complex mixture of extremely small particles and liquid droplets. Particulate matter comes from many different sources such as factories, power plants, dry cleaners, cars, buses, trucks and even windblown dust and wildfires. Particle pollution is made up of a number of components, including metals, acids (e.g. nitrates and sulfates), organic chemicals,

and soil or dust particles. Particles are either directly emitted into the air or formed in the atmosphere by transformation of emitted gases such as SO_2. They come in many different size ranges such as coarse, fine and ultrafine. They also vary in composition and origin. The size of particles is directly linked to their potential for causing health problems. EPA is concerned about particles that are 10 micrometers in diameter or smaller (PM_{10}) because those are the particles that generally pass through the throat and nose and enter the lungs. Once inhaled, these particles can affect the heart and lungs and cause serious health effects. Because of different chemical compositions, some airborne particles are more toxic than others (e.g., U.S. EPA, 2004). In this analysis we are interested in fine particles, those less than $2.5\mu m$ aerodynamic diameter ($PM_{2.5}$) and especially their metal components.

1.2 Research Motivation

Recent epidemiologic studies showed statistically significant associations of various ambient PM indicators with a variety of respiratory health endpoints, from physiologic changes in pulmonary function, respiratory illness and symptoms, emergency department visits, hospital admissions, to mortality. Children, elderly, and individuals with asthma are among the most susceptible to PM effects (U.S. EPA, 2004).

Although ambient air particulate matter (PM) has been clearly associated with adverse human health outcomes (NRC, 2004; U.S. EPA, 2004), the relationship between specific physicochemical properties of PM and these health effects remains largely unresolved (Schlesinger et.al 2006). In a study linking $PM_{2.5}$ for motor vehicles and coal combustion to mortality, Laden et. al. (2000) reported that a $10\mu g/m^3$ increase in $PM_{2.5}$ from mobile sources accounted for a 3.4% increase in daily mortality, while the equivalent increase in fine particles from coal combustion sources accounted for a 1.1% increase. Fine particles from crustal sources were not associated with mortality. They conclude that "the results indicate that combustion particles in the fine fraction from

mobile and coal combustion sources, but not fine crustal particles, are associated with increased mortality". This suggests that maybe not all $PM_{2.5}$ chemical components are associated with health effects. PM comprises a mixture of several compounds, including carbon-centered combustion particles, secondary inorganics, and crustal derived particles. These compounds may contribute, with different potential, to the PM-induced health effects (Schwarze et al 2006).

The multi-centre time series studies, Air Pollution and Health - A European Approach (APHEA) and National Mortality Morbidity Air Pollution Study (NMMAPS) (Samet et Al. 2000A, 200B, ALA, 2001), indicate the occurrence of some heterogeneity with respect to risk estimates between locations. This observation is further corroborated by seasonal analyses, which showed stronger effects on mortality in the summer in the Northeast of the US, whereas there were smaller or no differences between seasons in other areas (Peng et al. 2005). Schwarze et al 2006 suggested that these variations in risk estimates between different cities could be due to differences in particle composition, although other factors may also be involved. In their final report, the NRC Committee on Research Priorities for Airborne Particulate Matter (NRC, 2004) provides a summary table of PM characteristics that may be important to health responses including PM metals.

The importance of combined exposure to different PM-components is far from elucidated and needs to be determined. Indeed, there seems to be increasing support for the idea that the chemical characteristics are important for the adverse health effects of ambient PM. Thus, there has been a call for more specific exposure metrics which account for the composition or origin of ambient PM (Grahame et al 2005, Forsberg et al 2005). To comply with this requires the identification of the hazardous and non-hazardous PM components. Several epidemiological studies have identified PM metals as one of these components. A good understanding of the PM metals and the interplay between them become therefore critical. We would like to understand why concentration of PM metals speciation differs from one city to another or from

one season to the next. A good knowledge and understanding of the pattern will be critical in quantifying personal exposure. Additionally we might be able to relate the exposure to potential health outcomes, such as birth outcomes, respiratory, and cardiovascular diseases.

Given the fact that humans are exposed to complex mixtures of environmental pollutants such as $PM_{2.5}$, research efforts are intensifying to study the mixtures composition and the emission sources of ambient PM, and the exposure-related health effects. Developing statistical tools that are able to identify chemicals mixtures are therefore critical and important. Factor analysis as well source apportionment models are statistical tools potentially useful for characterizing mixtures in $PM_{2.5}$. However, classic factor analysis is designed to analyze samples of independent data and Air pollution data in general and $PM_{2.5}$ speciation metals in particular are generally temporally correlated within each monitoring station and spatio-temporally correlated when an MSA (several monitoring stations) is considered. Methods to address this type of data correlation need to be considered.

1.3   Specific Aims

As mentioned earlier, $PM_{2.5}$ is a complex mixtures of extremely small particles and liquid droplets. It is made up of a number of different components, including metals, acids and organic chemicals which form the $PM_{2.5}$ chemicals speciation. Several studies have identified $PM_{2.5}$ metals as some of the chemicals of great concern for human health. There is a huge variation of these chemicals in the air around the country and nothing is known about them,

- First, a spatio temporal analysis in order to understand their distribution pattern is therefore critical.

- Second, our exposure is believed to be in form of mixtures and we are unlikely to be exposed to only one chemical from the air environment at any given time. It

is therefore important to understand and be able to quantify mixture exposure. Unfortunately there is no known theory to do so especially for correlated data. Chemicals speciation data, a large dataset containing more than 100 different chemicals including 35 metals are measured every three to six days from the environment with different source origins in some select stations. Within each station we have a time series data (temporal correlation) and when several stations are considered we have multiple time series data with possible correlation between stations (spatio-temporal correlation).

- Third, Factor analysis in conjunction with source apportionment model will be used to determine empirical mixture observed.

In this dissertation we therefore sought to understand the temporal and spatial variation of PM metals in three geographic areas in the US: Tampa Bay Area in Florida, Houston area in Texas and Pittsburgh area in Pennsylvania. Statistical modeling, especially those based on nonparametric approach will be used and health impact will follow subsequently. Additionally theories on metals mixtures will be developed and applied to one station in Tampa and discussion on extension to several stations will be added. The followings specific aims will then be addressed.

1. **Specific aim 1**: In order to improve our understanding of $PM_{2.5}$ special metals, we will in this aim carry an analysis of the spatio-temporal patterns and variation of individual metals and some mixtures in $PM_{2.5}$

2. **Specific aim 2**: In this aim we look at the risk of low birth weight (LBW) from exposure to particulate matter ($PM_{2.5}$) speciation metals mixtures during Pregnancy

3. **Specific aim 3**: In this aim we develop method of identifying mixtures via a combination of factor analysis and source apportionment methods. As $PM_{2.5}$ speciation data are (spatio-) temporally correlated, an extension of the regular

factor analysis is needed. We will develop a Bayesian factor analysis for temporally correlated $PM_{2.5}$ speciation data. And discuss the extension to spatio-temporally correlated data. Using a combination of factor analysis and source apportionment models, we transform latent factors into mixtures.

2 Spatio-Temporal Patterns and Variation of Common Particulate Matter Speciation Metals

2.1 Introduction

Particulate matter, also known as particle pollution or PM, is a complex mixture of extremely small particles and liquid droplets. Particulate matter comes from many different sources such as factories, power plants, dry cleaners, cars, buses, trucks and even windblown dust and wildfires. Particle pollution is made up of a number of components, including metals, acids (e.g. nitrates and sulfates), organic chemicals, and soil or dust particles. Particles are either directly emitted into the air or formed in the atmosphere by transformation of emitted gases such as SO_2. They come in many different size ranges such as coarse, fine and ultrafine. They also vary in composition and origin.

The size of particles is directly linked to their potential for causing health problems. EPA is concerned about particles that are 10 micrometers in diameter or smaller because those are the particles that generally pass through the throat and nose and enter the lungs. Once inhaled, these particles can affect the heart and lungs and cause serious health effects. Because of different chemical compositions, some airborne particles are more toxic than others (e.g., U.S. EPA, 2004). In this analysis we are interested in fine particles, those less than $2.5\mu m$ aerodynamic diameter ($PM_{2.5}$).

2.1.1 Health effects of Total Particulate Matter

Recent epidemiologic studies show statistically significant associations of various ambient PM indicators with a variety of respiratory health endpoints, from physiologic changes in pulmonary function, respiratory illness and symptoms, emergency department visits, hospital admissions, to mortality. Children, elderly, and individuals with asthma are among the most susceptible to PM effects (U.S. EPA, 2004).

Two landmark prospective cohort studies, the 1993 Six Cities Study and the 1995 American Cancer Society study reported that chronic exposure to particulate pollution increases the risk of premature mortality. In the 1993 Six Cities Study, Harvard University researchers followed the health of more than 8,000 people in six small cities that fell along a gradient of air pollution concentrations for a period of 14 to 16 years. As particle concentrations increased, there was an almost directly proportional increase in the death rate in the residents studied. Residents of the most polluted city in the study, Steubenville, Ohio, had a 26 percent increased risk of premature mortality, compared to the residents of the cleanest city studied, Portage, Wisconsin. The increased risks were associated with a difference in ambient fine particle concentrations of 18.6 micrograms per cubic meter (Dockery et. al.1993). The 1995 American Cancer Society study reported an association between fine particle air pollution and premature death by cardio-pulmonary and other causes in a study group of over half a million people in 151 U.S. cities. All cause mortality increased by 17 percent with a 24.5 microgram per cubic meter difference in fine particle pollution between the cleanest and dirtiest city studied (Pope et. al. 1995).

Although ambient air particulate matter (PM) has been clearly associated with adverse human health outcomes (NRC, 2004; U.S. EPA, 2004), the relationship between specific physicochemical properties of PM and these health effects remains largely unresolved (Schlesinger et.al 2006). In a study linking $PM_{2.5}$ for motor vehicles and coal combustion to mortality, Laden et. al. (2000) reported that a $10 \mu g/m^3$ increase in

$PM_{2.5}$ from mobile sources accounted for a 3.4% increase in daily mortality, while the equivalent increase in fine particles from coal combustion sources accounted for a 1.1% increase. Fine particles from crustal sources were not associated with mortality. They conclude that "the results indicate that combustion particles in the fine fraction from mobile and coal combustion sources, but not fine crustal particles, are associated with increased mortality". This suggests that maybe not all $PM_{2.5}$ chemical components are associated with health effects. PM comprises a mixture of several compounds, including carbon-centered combustion particles, secondary inorganics, and crustal derived particles. These compounds may contribute, with different potential, to the PM-induced health effects (Schwarze et al 2006).

The multi-centre time series studies, Air Pollution and Health - A European Approach (APHEA) and National Mortality Morbidity Air Pollution Study (NMMAPS), indicate the occurrence of some heterogeneity with respect to risk estimates between locations. This observation is further corroborated by seasonal analyses, which showed stronger effects on mortality in the summer in the Northeast of the US, whereas there were smaller or no differences between seasons in other areas (Peng et al. 2005). Schwarze et al 2006 suggested that these variations in risk estimates between different cities could be due to differences in particle composition, although other factors may also be involved. In their final report, the NRC Committee on Research Priorities for Airborne Particulate Matter (NRC, 2004) provides a summary table of PM characteristics that may be important to health responses including PM metals.

2.1.2   Health effects of PM Metals

The importance of combined exposure to different PM-components is far from elucidated and needs to be determined. Indeed, there seems to be increasing support for the idea that the chemical characteristics are important for the adverse health effects of ambient PM. Thus, there has been a call for more specific exposure metrics which

account for the composition or origin of ambient PM (Grahame et al 2005, Forsberg et al 2005). To comply with this requires the identification of the hazardous and non-hazardous PM components. Several epidemiological studies have identified PM metals as one of these components.

A role for transition metals such as iron (Fe), manganese (Mn), Nickel (Ni), copper (Co), Silver (Ag) in producing adverse health effects is based on their potential for oxidative activity and the production of reactive oxygen species. Similar to the quinones, soluble forms of these metals can be involved in a Fenton-type reaction (Schlesinger et al 2006). Elevated oxidative stress in the lungs and hearts of rats exposed to concentrated ambient particles (CAPs) in Boston and residual oil fly ash (ROFA) was most strongly associated with metal fractions of these particles (Gurgueira et al., 2002). Molinelli et al. (2002) exposed a human airway epithelial cell line to aqueous extracts of PM collected in the Utah Valley. In this study, part of the extract was treated to remove cations, including transition metals. Cells exposed to the untreated extract showed a concentration-dependent increase in the inflammatory mediator interleukin (IL)-8 compared to controls; cells incubated with the treated extract showed no such change. This suggests that the removal of metal cations attenuated cellular response to the aqueous extract, and supports a role for transition metal involvement in PM toxicity. In this regard, cultured human T cells exposed to $1\mu m$ carbon particles or particles containing both carbon and iron showed increased production of reactive oxygen species with the latter, but not with the former (Long et al., 2005). Finally, Sorensen et al. (2005) found a relationship between the vanadium (V) and chromium (Cr) components of fine particulate CAPs and oxidative damage to DNA. It is unclear whether all PM-associated transition metals are equally toxic, or whether there can be a ranking of toxicity related to specific metal content or metal valence state (Schlesinger et al 2006).

Furthermore, relative water solubility may also be a factor in modulating biological response. When ambient PM from St. Louis, MO, Washington, DC, Dusseldorf,

Germany, and Ottawa, Canada, were tested for toxicity, the observed greater response to Ottawa PM was postulated to be due to its higher content of water soluble metals (Costa & Dreher, 1997). Other studies have indicated that Zn in PM may be responsible for various pulmonary effects, such as inflammation, necrosis, and airway hyperreactivity (Adamson et al., 2000; Dye et al., 2001; Kodavanti et al., 2002a; Gavett et al., 1997; Schlesinger et al., 2006). Human bronchial epithelial cells exposed to extracts of PM collected in Taiwan showed a correlation between cytokine production and metal content, with effects on some cytokines correlating with Cr and Mn, and others with Fe and Cr (Huang et al., 2003). In a study using ROFA, particles with higher Zn content resulted in greater pulmonary inflammation and airway responsiveness than did particles with higher Ni or V content (Gavett et al., 1997). Schlesinger et al. (2006) suggested while the apparent differences in response to various metals may seem to add to inconsistencies between toxicological studies, they do, in fact, support the idea that the endpoint examined is critical in reaching any conclusion as to the efficacy of specific metals and/or that effects may be linked to specific valence state.

In most epidemiological studies, the metal content of the PM has not been analysed. However, in some special locations, the PM-sources are known to be rich in metals, for instance in areas with metallurgic industries. In the Utah Valley area, air pollution changes during a transient closure of a steel mill were associated with changes in mortality and morbidity. PM levels and mortality and morbidity declined during the closure of the mill, but increased again when the mill was reopened (Gheo et al. 2004).The changes in health effects were not fully accounted for by PM mass. Though metals were not modeled in the epidemiological study, it seems likely that the number of metal particles and amount of metals in particles was considerably reduced during the closure period. The mortality risk estimates in the Utah Valley studies appeared to be in the upper range compared to other investigations, in which exposure to traffic dominates.

Several years ago, WHO published air quality guidelines on metals as air pollutants (WHO 1987). More recent information about the different metals has resulted in metals being scrutinized as important constituents of PM. In brief, transition metals, such as iron and copper, are believed to contribute to particle induced formation of reactive oxygen species (ROS) through the Fenton reaction, and have, therefore, been considered important for particle toxicity (Donaldson et. al. 2003, Fubini et. al. 2003, Barchowsky et. al. 2003). Other metals, such as zinc, may trigger effects more directly by interacting with cellular proteins (Haase et. al. 2005, Tal et. al. 2006).

In a study of Canadian cities, Burnett et al., found that iron, nickel and zinc, in addition to sulphates, were associated with increased mortality. The associations of these constituents were even better than for total mass, indicating that they were better predictors for mortality than mass. However, the larger variation in the metal-associated effects than those found for mass also indicated that there were other important contributing components in the mixture (Burnett et al 2005).

A cross-sectional study in Eastern Germany showed higher lifetime prevalence of respiratory disorders and allergic sensitization in children living near industrial sites compared to children without such exposure (Heinrich et al 1999). A decline in pollution reduced respiratory symptoms in children (Heinrich et al 2005). Later analyses revealed higher levels of particles and higher metal content near the industrial site compared to the more rural area (Schaumann et al 2004). Though metal composition in these studies was not modeled, the subsequent experimental results suggested that a reduction in effects might be achieved by the reduction in particle-associated metal exposure.

An analysis of the Six Cities data, including elemental composition, revealed the importance of nickel, lead and sulphur on the increased mortality of long-term increased air pollution (Laden et. al 2000). Several experimental studies suggest a role of metals in PM-induced cardiovascular effects. Longterm inhalation studies have shown that zinc-containing PM may cause myocardial injury in rats (Kodavantiet al

2003). Copper, zinc and vanadium have been shown to induce a range of different cardiovascular effects, including increased expression of different cytokines and stress proteins, reduction in spontaneous beat rate, vasoconstriction and vasodilation (Graff et al. 2004, Li et al 2005, Bagate et al. 2005). Notably, the effects may be triggered through a complex interplay between different metals. Campen et al., 2001 reported that nickel and vanadium may interact synergistically to cause immediate and delayed cardiovascular effects. Nickel-exposure was found to cause delayed bradycardia, hypothermia and arrhythmogenesis, whereas vanadium did not cause any significant delayed effects alone, but enhanced the effect of nickel (Campen et al., 2001). In contrast, vanadium, but not nickel or iron exposure, resulted in immediate responses on the same cardiovascular parameters. Moreover, nickel was also found to exacerbate the immediate effects of vanadium, whereas iron attenuated the vanadium-induced effects(Campen et al 2002). Metal-rich particles have also been found to enhance allergic responses to ovalbumin and house dust mite (Gavett et al. 2003, Lambert et al 1999, Lambert et al 2000), and to induce the increased release of allergy-related cytokines, eosinophil recruitment and airway hyper-responsiveness in mice (Gavett et al 1999). Moreover, metal ions, such as aluminium, cadmium, nickel and strontium ions, have been found to enhance IL-4 release and degranulation of mast cells (Walczak-Drzewiecka et al. 2003). Thus, there seems to be some support for the idea that soluble metals from PM may be involved in allergic responses.

Based on the aforementioned epidemiological studies and findings, a good understanding of the PM metals and the interplay between them become critical. We would like to understand why concentration of PM metals speciation differs from one city to another or from one season to the next. A good knowledge and understanding of the pattern will be critical in quantifying personal exposure. Additionally we might be able to relate the exposure to potential health outcome, such as respiratory and cardiovascular diseases. In this study we address the temporal and spatial variation of PM metals in three geographic areas in the US: Tampa Bay Area in Florida, Houston

area in Texas and Pittsburgh area in Pennsylvania. Statistical modeling, especially those based on nonparametric approach will be used and health impact will follow in subsequent studies.

As mentioned above, we have very limited knowledge on health impact of PM metals and more importantly very little is known on even the distribution and variation of PM metals in the country. Among the very few studies that attempted work in the area of particulate matter metals we cite the work done by Bell et. al. 2007, Mubiana et. al. 2005, Mugica et al. 2002. To our knowledge, no study has focussed on spatial and temporal variation of particulate matter speciation despite their proven epidemiologic importance in public health.

In this paper we use nonparametric statistical methods to describe the temporal and spatial variations of selected metal constituents in $PM_{2.5}$ using the US EPA's National Air Quality Monitoring Network data collected from three metropolitan areas. Based on the spatiotemporal models we may be able to propose in subsequent studies ways to estimate potential inhalation dose, both for lower dose chronic exposure and possible acute exposure due to short term surge in air pollutants concentration.

2.2   Particulate Matter Metals Speciation Data

Based on the geographic location and more importantly the existing type of industries, three metropolitan statistical areas (MSA) in the US have been selected for our study: Pittsburgh MSA belongs to the Ohio valley region dominated by steel industries. Houston area in addition to being a metropolitan area where traffic is an important source of particulate matter, is in addition a region dominated by oil industries. Tampa Bay MSA, the third region is comparable to Houston as for its traffic and geographic location but yet very different in term of industry types. Tampa Bay is extremely poor in industries with power plants for electricity generation as major industries in the region.

Table 2.1: Study sites and total number of observation per site together with the starting date and ending date of data collection in Tampa, Houston and Pittsburgh MSAs

Site ID	length	Start Date	End date	Location
120571075	489	02/09/2000	12/29/2003	Tampa, FL
120573002	472	01/01/2004	12/29/2007	Plant City, FL
121030026	328	09/06/2004	12/29/2007	Pinellas Park, FL
482010024	798	08/17/2000	12/26/2007	Not a City, TX
482010026	620	08/17/2000	08/29/2005	Channelview, TX
482010055	640	08/18/2000	08/29/2005	Houston, TX
482010803	173	08/17/2000	11/06/2001	Houston, TX
482011034	226	01/02/2002	08/26/2005	Houston, TX
482011039	1582	02/15/2000	12/29/2007	Deer Park, TX
483390078	464	10/25/2001	08/29/2005	Conroe, TX
483390089	124	08/15/2000	09/19/2001	Conroe, TX
420030008	842	06/30/2001	12/29/2007	Pittsburgh, PA
420030021	146	06/30/2001	09/30/2003	Pittsburgh, PA
420030064	226	10/06/2003	12/26/2007	Liberty, PA
421255001	414	06/30/2001	12/26/2007	Not a City, PA
421290008	423	06/30/2001	12/26/2007	Greensburg, PA

Data have been obtained from the US EPA's National Air Quality Monitoring Network covering a period of eight years from 2000 to 2007. Table 2.1 presents the study sites together with the total number of data points, the starting and ending dates of data collections for each sites. It shows that some sites were not operational during the whole study period. For the analysis, we use three separate datasets representing the aforementioned three metropolitan statistical areas (MSA): Tampa-Saint Pittersburg-Clearwater MSA in Florida, Houston MSA in Texas and Pittsburgh MSA in Pennsylvania.

The dataset in Tampa includes 1289 observations days across all sites (i.e., monitors-days of data). Except for few days, data were collected every three days during the collection periods. Houston MSA data and Pittsburgh MSA data include 4627 and 2051 observations days across all sites respectively. Each dataset contains several co-variates including: dates of data collection, site numbers, county, state and address indicating site locations, daily average temperature, daily average barometric pressure,

15
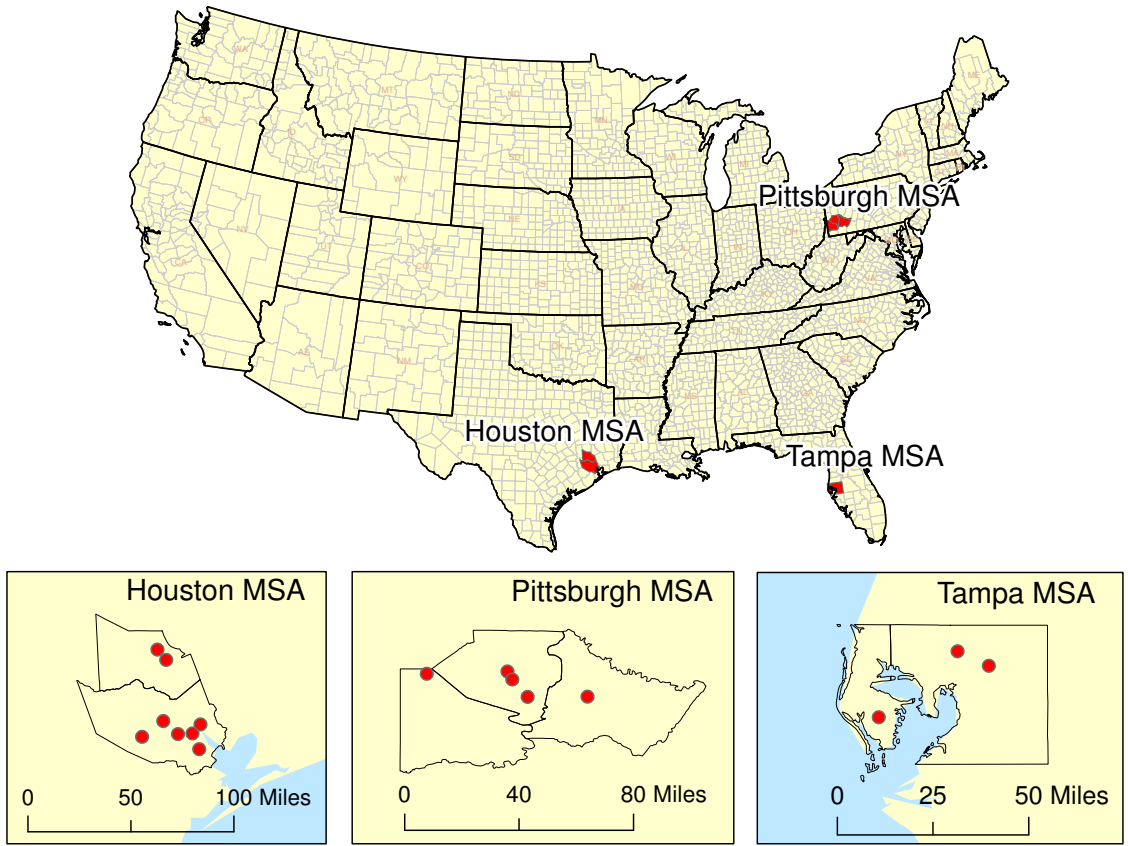
Figure 2.1: Study Selected Stations

latitude, longitude and $PM_{2.5}$ total mass and speciation chemical concentrations.

There are three speciation sites in the Tampa Bay Area with an average of 415 observations per site, covering a population of approximately 2,723,949 as of July 1st 2006. The site located in Tampa have been relocated in 2003 and has data only from February 09, 2000 to December 29, 2003. The other two currently operational sites are both of residential land use and have started operating in January 01, 2004 and September 06, 2004 respectively (Table 2.1). They are both of population exposure monitoring objective with rural and suburban location. Data of all three stations have been used in the analysis. Houston MSA has eight sites with an average of 489 observations per site, while Pittsburgh MSA has five operational sites covering a population of roughly two million (Wittig et. al. 2004) with an average of 386 observations per site. While a comprehensive analysis and discussion of all PM metals is sought, we will concentrate our effort on highly concentrated metals or metals where a clear pattern and noticeable differences between regions are observed.

There is an increasing belief in environmental research studies that, in the natural environment, we are not exposed to one chemical at the time but rather to a mixture of chemicals which is defined as a combination of several chemicals. In addition to individual metals, we will analyze and discuss some mixture of metals, whose selection will be solely based on known emission sources and factor analysis.

Table 2.2 gives the average daily concentration of each metal and total PM mass together with their standard errors in Tampa, Houston and Pittsburgh MSAs. Except for the total PM mass which is in microgram per cubic meter, all other concentrations are in nanogram per cubic meter. Aluminum, calcium, iron, potassium and sodium are the most abundant metals in all three areas. Aluminum and calcium daily average concentration are lowest in Pittsburgh and highest in Houston. Iron is lowest in Tampa and highest in Pittsburgh probably because of steel industries, while potassium and sodium are lowest in Pittsburgh and highest in Tampa. Among other heavy metals with huge variation in concentration between regions are chromium, lead,
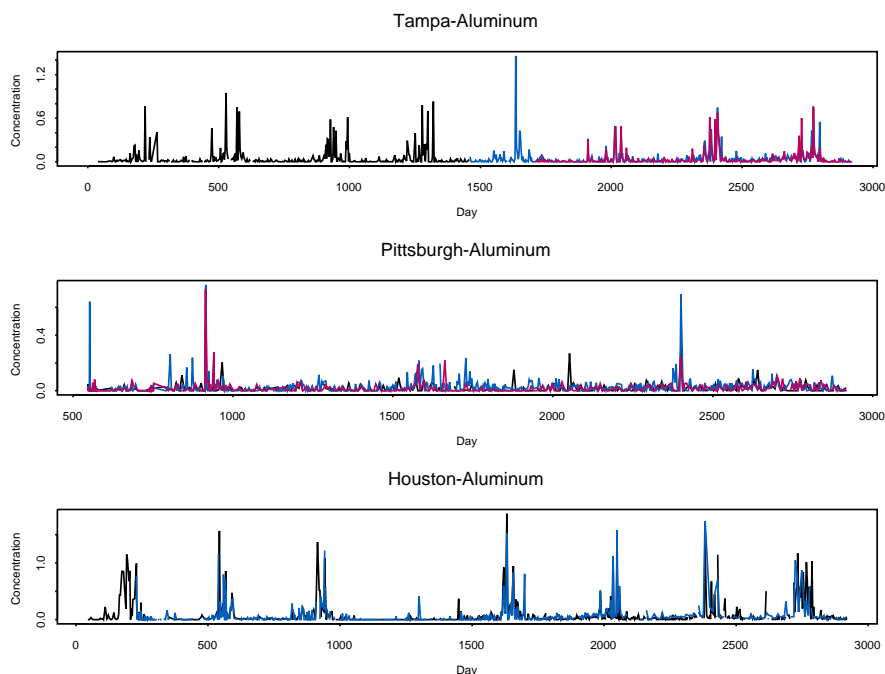
Figure 2.2: Plot of Aluminum in Tampa, Pittsburgh and Houston

manganese, magnesium and zinc. These metals are heavily concentrated in Pittsburgh
with concentration in general two to three times higher than that of the other two
regions. Magnesium in the contrary is very low in Pittsburgh. It is half the amount
found in Tampa and one third of that of Houston.

Time series plots of aluminum (Figure 2.2) indicates that the metal is periodic
with low concentration in the winter and high in the summer in Tampa and Houston,
while it is variable in the Pittsburgh area. Calcium although variable in all three
regions, is generally high in the summer (Figure 2.3). Sodium does not show any
noticeable pattern in all three regions (Figure 2.4) and is generally high compare to
aluminum and calcium. Aluminum has no noticeable trend in Tampa. The aluminum
average concentration in Tampa was 33.26 in 2000, its increase to the highest average
concentration of 53.09 in 2001 followed by a decreasing trend leading to the average
lowest concentration of 30.64 in 2005. There seems to be a huge increase of aluminum
concentration in 2006 and 2007. On average the lowest concentrations are recorded in
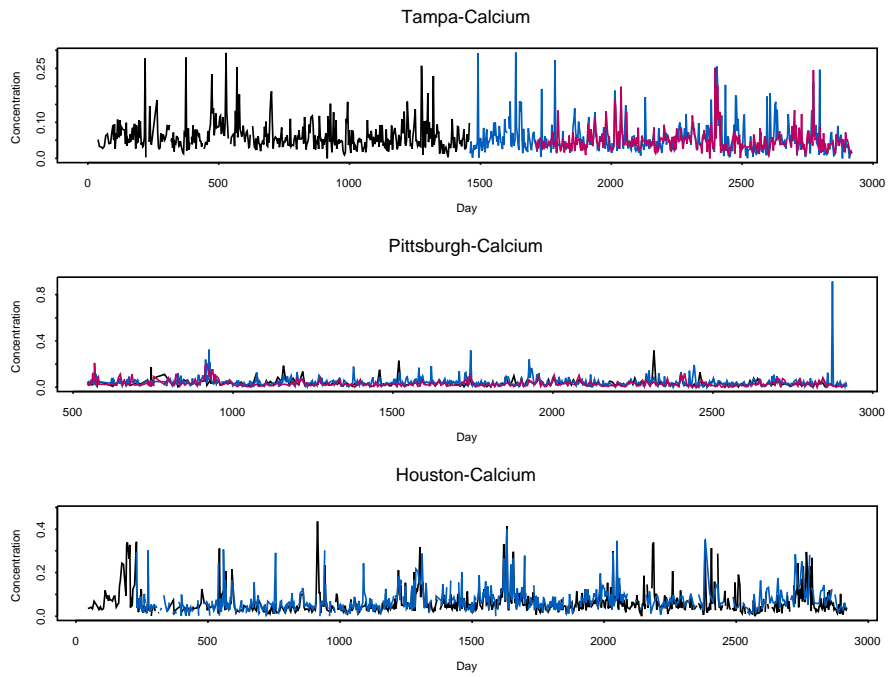
Figure 2.3: Plot of Calcium in Tampa, Pittsburgh and Houston
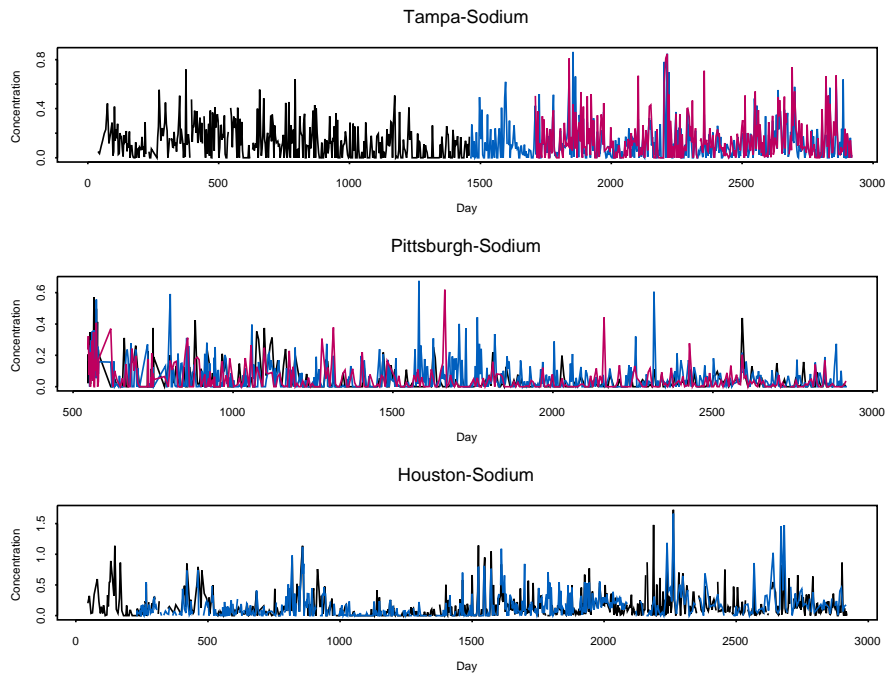
Figure 2.4: Plot of Sodium in Tampa, Pittsburgh and Houston

19

site 3 located in Pinellas Park. Concentrations per site are comparable, except in 2004 where from Table 2.3 we see that the average daily concentration in Site 2 in Tampa (43.49) is approximately five times that of site 3 in Tampa (8.10). Aluminum trend in Houston is in general increasing, with a huge drop in concentration to a minimum of 5.75 in 2003. The highest concentration of 131.13 is observed in 2007. There is a noticeable site to site variation, with year 2003 seing the lowest observation across all sites. Pittsburgh aluminum is moderately increasing. Although moderate as it is the case with the general trend, there is a site to site variation. Here the concentrations are low compared to the other two MSAs. Among the three regions, Houston sees the highest concentration of aluminum and Pittsburgh having the lowest. This may suggest that the steel industries are not the main source of PM aluminum.

Table 2.4 presents the average daily concentration of calcium per year and per site in Tampa, Houston and Pittsburgh. Calcium is in general decreasing in Tampa-St. Petersburg-Clearwater MSA with a minimum of 48.91 in 2007 and a maximum of 74.41 in 2001. Site to site variations are comparable. Calcium concentrations in Houston are decreasing till 2003 then increased afterward. Here site to site concentrations are variable. The concentrations seems to generally decreased in Pittsburgh. The maximum concentration of 49.14 is recorded in 2001 while the minimum of 35.11 is in 2006. Sites concentrations are moderately variable.

From Table 2.5, we see that sodium concentrations decreased from 2001 to a low level of 74.59 in 2003 and increased to more than double in 2007 in Tampa. Site concentrations are variable. In Houston the concentrations see a yearly variation with a huge increase starting from 2005. Here also there is a site to site variation. Pittsburgh sees in general a decrease in it sodium concentrations. This decrease in also noticeable from site to site except in site 3 with data available from 2003 where we see yearly variation.

Table 2.2: Average daily concentration and standard error in ng/m^3 of PM2.5 speciation metals in Tampa, Houston and Pittsburgh

Metal	Tampa	Houston	Pittsburgh
Aluminum	43.70(0.086)	68.54(0.040)	24.06(0.028)
Barium	17.49(0.028)	11.62(0.004)	13.36(0.010)
Cadmium	1.81(0.003)	1.14(0.000)	2.12 (0.002)
Calcium	56.03(0.031)	62.98(0.012)	41.45(0.020)
Chromium	1.64(0.003)	0.97(0.000)	3.01(0.003)
Cobalt	0.18(0.000)	0.14(0.000)	0.23(0.000))
Copper	4.49(0.010)	3.79(0.002)	4.82(0.002)
Cesium	3.15(0.008)	2.31(0.001)	3.08(0.004)
Gallium	0.62(0.001)	0.32(0.000)	0.48(0.000)
Iron	62.30(0.055)	93.11(0.032)	107.88(0.048)
Hafnium	1.89(0.004)	1.98(0.001)	1.96(0.003)
Lead	2.93(0.003)	2.74(0.001)	9.55(0.006)
Indium	2.15(0.003)	1.34(0.000)	2.18(0.002)
Manganese	1.31(0.001)	2.30(0.001)	5.06(0.003)
Iridium	0.96(0.001)	0.56(0.000)	1.19(0.001)
Molybdenum	0.71(0.001)	1.02(0.001)	1.30(0.001)
Nickel	2.69(0.017)	1.57(0.001)	1.64(0.001)
Magnesium	11.51(0.030)	18.77(0.010)	6.70(0.011)
Mercury	1.29(0.002)	0.60(0.000)	1.43(0.002)
Gold	1.05(0.001)	0.69(0.000)	1.11(0.001)
Lanthanum	6.51(0.012)	3.37(0.002)	5.36(0.006)
Niobium	0.47(0.001)	0.28(0.000)	0.48(0.001)
Tin	6.41(0.007)	3.98(0.001)	6.71(0.005)
Titanium	5.52(0.007)	7.49(0.003)	4.73(0.003)
Scandium	0.17(0.000)	0.22(0.000)	0.18(0.000)
Vanadium	2.73(0.002)	3.72(0.001)	1.28(0.001)
Silver	2.30(0.002)	1.25(0.000)	2.21(0.002)
Zinc	5.21(0.005)	12.33(0.003)	25.71(0.018)
Strontium	1.18(0.004)	1.07(0.000)	0.98(0.001)
Tantalum	5.10(0.007)	2.95(0.001)	5.65(0.006)
Rubidium	0.37(0.001)	0.23(0.000)	0.36(0.000)
Potassium	92.77(0.282)	82.31(0.024)	67.36(0.041)
Yttrium	0.41(0.001)	0.28(0.000)	0.42(0.000)
Sodium	130.11(0.113)	127.66(0.040)	57.45(0.046)
Zirconium	0.89(0.003)	0.53(0.000)	0.68(0.001)
PM Mass	12.65(0.006)	11.60(0.001)	17.13(0.005)

Table 2.3: Average daily concentration of aluminum in ng/m^3 per Year and site in Tampa, Houston and Pittsburgh

Site	2000	2001	2002	2003	2004	2005	2006	2007
Tampa	33.26	53.09	44.45	47.13	35.65	30.64	52.63	52.18
Tampa 1	33.26	53.09	44.45	47.13	NA	NA	NA	NA
Tampa 2	NA	NA	NA	NA	43.49	29.88	54.86	53.23
Tampa 3	NA	NA	NA	NA	8.10	31.55	49.91	50.92
Houston	65.78	72.61	71.25	5.58	78.26	95.19	98.30	127.66
Houston 1	37.79	73.02	66.05	8.41	75.42	114.56	144.18	128.2
Houston 2	31.13	74.08	30.43	2.59	79.21	131.01	NA	NA
Houston 3	24.66	68.12	77.22	2.76	38.77	114.98	NA	NA
Houston 4	49.91	93.96	NA	NA	NA	NA	NA	NA
Houston 5	NA	NA	82.55	16.16	90.39	125.14	NA	NA
Houston 6	131.36	64.79	66.71	5.76	98.06	62.22	84.57	127.45
Houston 7	NA	32.54	108.46	2.55	68.38	105.75	NA	NA
Houston 8	49.87	70.79	NA	NA	NA	NA	NA	NA
Pittsburgh	NA	11.77	28.94	15.87	26.92	19.07	30.24	31.08
Pittsburgh 1	NA	18.28	29.30	17.85	37.04	22.62	32.57	31.67
Pittsburgh 2	NA	7.17	31.57	22.09	NA	NA	NA	NA
Pittsburgh 3	NA	NA	NA	5.90	16.36	21.50	44.42	45.96
Pittsburgh 4	NA	12.70	35.76	12.25	15.47	9.83	22.17	25.56
Pittsburgh 5	NA	5.83	19.45	13.25	14.73	15.80	21.44	21.46

Table 2.4: Average daily concentration of calcium in ng/m^3 per Year and site in Tampa, Houston and Pittsburgh

Site	2000	2001	2002	2003	2004	2005	2006	2007
Tampa	63.68	74.41	53.90	58.37	53.88	51.35	55.10	48.91
Tampa 1	63.68	74.41	53.90	58.37	NA	NA	NA	NA
Tampa 2	NA	NA	NA	NA	57.11	52.43	59.27	50.41
Tampa 3	NA	NA	NA	NA	42.55	50.04	49.97	47.11
Houston	61.21	56.28	51.09	62.57	71.01	70.24	66.64	71.11
Houston 1	57.33	55.97	57.13	79.66	83.16	78.92	74.74	93.68
Houston 2	51.48	61.30	31.69	42.71	77.57	89.88	NA	NA
Houston 3	44.82	48.24	49.43	53.41	40.93	57.21	NA	NA
Houston 4	77.09	82.78	NA	NA	NA	NA	NA	NA
Houston 5	NA	NA	79.61	99.45	115.12	104.28	NA	NA
Houston 6	78.02	48.13	44.49	64.08	69.35	64.08	64.21	62.11
Houston 7	NA	29.41	56.41	51.29	65.04	52.13	NA	NA
Houston 8	41.41	42.00	NA	NA	NA	NA	NA	NA
Pittsburgh	NA	44.41	49.14	38.79	39.49	42.06	35.11	40.30
Pittsburgh 1	NA	46.22	53.16	43.27	47.10	48.60	38.94	44.99
Pittsburgh 2	NA	44.19	46.92	40.87	NA	NA	NA	NA
Pittsburgh 3	NA	NA	NA	30.34	34.99	39.23	35.50	62.14
Pittsburgh 4	NA	51.96	50.27	28.21	25.29	30.62	26.51	26.48
Pittsburgh 5	NA	34.49	43.60	41.13	33.05	37.66	36.75	25.44

Table 2.5: Average daily concentration of sodium in ng/m^3 per Year and site in Tampa, Houston and Pittsburgh

Site	2000	2001	2002	2003	2004	2005	2006	2007
Tampa	139.38	173.15	109.16	74.59	116.61	131.64	131.63	150.03
Tampa 1	139.38	173.15	109.16	74.59	NA	NA	NA	NA
Tampa 2	NA	NA	NA	NA	111.88	121.97	113.56	136.02
Tampa 3	NA	NA	NA	NA	133.24	143.25	153.79	166.94
Houston	116.13	101.81	135.57	36.20	145.16	186.05	212.41	202.58
Houston 1	85.00	89.15	147.74	35.93	171.55	196.55	226.31	267.90
Houston 2	100.54	105.65	73.25	21.33	136.10	206.97	NA	NA
Houston 3	88.92	70.46	125.43	30.11	101.70	198.90	NA	NA
Houston 4	98.40	102.95	NA	NA	NA	NA	NA	NA
Houston 5	NA	NA	171.47	37.08	207.30	266.94	NA	NA
Houston 6	174.64	143.72	165.55	44.99	155.06	162.09	208.24	176.54
Houston 7	NA	49.11	137.72	38.81	116.39	175.13	NA	NA
Houston 8	85.63	96.15	NA	NA	NA	NA	NA	NA
Pittsburgh	NA	165.48	59.93	46.23	56.64	35.63	41.12	33.59
Pittsburgh 1	NA	146.10	59.14	43.78	73.26	40.62	47.69	38.47
Pittsburgh 2	NA	209.60	65.58	44.75	NA	NA	NA	NA
Pittsburgh 3	NA	NA	NA	22.90	36.52	50.24	56.60	38.09
Pittsburgh 4	NA	136.00	56.40	51.12	43.67	23.07	30.95	23.53
Pittsburgh 5	NA	191.09	59.33	50.81	33.03	19.81	26.19	31.08

2.3 Statistical Methods

The traditional time series models are useful in describing the temporal variations, including seasonality and trend, in ambient air $PM_{2.5}$ metals. However, preliminary examination of the $PM_{2.5}$ metals (Figures 2.2-2.4) suggested variations that are less regular and less appropriate for time series modeling. As a result, we will use additive models, a type of non-parametric methods to describe the temporal variations.

Being time series, $PM_{2.5}$ metals speciation data are likely to be correlated. Several data values are also recorded from each sites giving it a structure of clustered data with site as a cluster. A model that can take into account this nature of the data is needed. In general, when dealing with correlated data such as clustered, hierarchical and spatial designs data, model that could account for the correlation need to be used and random effect models are good candidate. For these reasons we will use the Generalized Additive Mixed Models(GAMM) described below for the modeling.

Suppose that y_i is the ith observations of the random variable y and p covariates $x_i = (1, x_{i1}, \cdots, x_{ip})^T$ associated with fixed effects and a $q \times 1$ vector of covariates z_i associated with random effects. Given a $q \times 1$ vector b of random effects, the observation y_i are assumed to be independent with means $E(y_i|b) = \mu_i^b$ and variances $var(y_i|b) = \phi m_i^{-1} v(\mu_i^b)$, where $v(.)$ is a specified variance function, m_i is a prior weight (e.g. a binomial denominator) and ϕ is a scale parameter. A generalized additive model is given by

$$g(\mu_i^b) = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) + z_i^T b, \qquad (2.3.1)$$

where $g(.)$ is a monotonic differentiable link function, $f_j(.)$ is a centered twice-differentiable smooth function, the random effects b are assumed to be distributed as $N(0, D(\theta))$ and θ is a $c \times 1$ vector of variance components.

A key feature of the GAMM(2.3.1) is that additive nonparametric functions are used to model covariates and random effects are used to model correlation between

observations. If $f_j(.)$ is a linear function, the GAMM(2.3.1) reduce to Generalized Linear Mixed Model (GLMM). In addition if the link function is taking to be identity, then they will be reduced to additive mixed models.

Model (2.3.1) encompasses various study designs, including clustered, hierarchical and spatial designs. This is because a flexible covariance structure of the random effects b can be specified. For longitudinal data, the random effects b can be decomposed into a random intercept and a stochastic process (Zeger and Diggle, 1994; Zhang et al., 1998). For hierarchical (multilevel) data, they can be partitioned to represent different levels of a hierarchy, e.g. a centre, physician and patient in a multicentre clinical trial (lin and Breslaw, 1996). For spatial data, which is common in disease mapping and ecological studies, they can be used to model spatial correlation (Cressie, 1993; Breslaw and Cleton, 1993).

The multiple smoothing parameter estimation by generalized cross validation (mgcv) package is part of the recommended suite that comes with the default installation of R and is used by GAMM for the model fitting. It is based on methods described in Wood (2000). Different packages are available in R for fitting additive model in general. The gam package allows more choice in the smoothers used while the mgcv package has an automatic choice in the amount of smoothing as well as wider functionality. The gss package of Gu (2002) takes a spline-based approach. The fitting algorithm depends on the package used. The penalized smoothing spline approach is used in the mgcv package and works as follow:

Suppose we represent $f_j(x_j) = \sum_i \beta_{ij}\phi_i(x_j)$ for a family of spline basis functions, ϕ_i. We impose a penalty $\int [f_j''(x)]^2 dx$ which can be shown to be of the form $\beta_j^T S_j \beta_j$, for a suitable matrix S_j that depends on the choice of basis. The model is fitted by minimizing,

$$||y - X\beta||^2 + \sum_j \lambda_j \beta_j^T S_j \beta_j$$

with respect to β and the $\lambda_j s$ control the amount of smoothing for each variable. The

problem of estimating the degree of smoothness for the model is now the problem of estimating the smoothing parameter $\lambda_j s$. Generalized cross-validation method (GCV) is used to select the $\lambda_j s$.

In GCV, λ is chosen to minimize:

$$V_g = \frac{n \sum_{i=1}^{n} (y_i - \hat{f}_i)^2}{[tr(I - A)]^2}.$$

where $\hat{f}$ is the estimate of f from fitting all the data, and A is the corresponding influence matrix.

For this special case of $PM_{2.5}$ speciation data, a possible model can be given by

$$y_{is} = \beta_0 + f_{1s}(year_{is}) + f_{2s}(month_{is}) + f_{3s}(temp_{is}) + f_{4s}(press_{is}) + Site_i + \epsilon_{is}$$

The problem with this model is as the number of site augment, we will pay the price of having to estimate a lot of parameters for the site effect. Another difficulty in using this model is that, we can only make statements about stations given in the data and won't be able to generalize to a neighboring station in the region. The best approach is to use a random intercept for the site and the following model is proposed.

$$y_{is} = \beta_0 + f_{1s}(year_{is}) + f_{2s}(month_{is}) + f_{3s}(temp_{is}) + f_{4s}(press_{is}) + b_s + \epsilon_{is}$$

where $y_i s$ is the ith observations of the random variable y and s represent the MSA. f_{js} refer to a smooth function for the covariate x_s^j of a given metropolitan statistical area s with ith observation x_{is}^j. b_s is the random effect representing sites in a given MSA s. Within the random effect different correlation structure could be specified. In our present situation an autoregressive of order one (AR1) is found to be suitable. ϵ_{is} is the noise and is assumed to be independently distributed with mean 0 and variance σ^2.

Metal mixtures is defined as a combination of several metals. A metal will be con-

sidered part of a mixture based on several criteria including factor analysis, knowledge of association with health effect and source origin. All metals coming from the same leading factors will be grouped together to form a mixture. Metals from significant factors with high loading coefficients (> 0.05) could also be grouped to constitute a mixture. Additionally some mixtures are formed based on knowledge of their significant association with health outcome such as low birth weight and preterm birth either during the first trimester of pregnancy or the entire pregnancy period. Summary of mixtures analyzed in this report are giving in Tables 2.8 and 2.9.

2.4 Results

2.4.1 Individual metals

Aluminum has no trend in Tampa (p=0.13), while the trend is significant in both Houston ($p < 0.0001$) and Pittsburgh ($p = 0.0009$). It is a seasonal metal in Tampa ($p < 0.0001$) and Houston ($p < 0.0001$) but not in Pittsburgh ($p = 0.042$). Temperature is highly associated with aluminum variation in Houston ($p < 0.0001$) and Pittsburgh ($p < 0.0001$) but that association is not clear in Tampa ($p = 0.025$). Barometric pressure has no statistically significant effect on aluminum in Houston and Pittsburgh ($p = 0.06$) and ($p = 0.29$) respectively, see Table 2.6. Table 2.6 shows that calcium is highly dependent upon temperature and barometric pressure in all three regions. Calcium also has a clear seasonal pattern in all three regions but the trend in Pittsburgh is somehow not so clear (p=0.037). Excepts for barometric pressure in Pittsburgh ($p = 0.08$), sodium has a clear trend, seasonality and is dependent upon temperature and barometric pressure in all three regions. Iron is highly seasonal, it has a clear trend except in Tampa ($p = 0.82$) and depend upon temperature and pressure in all three regions. Chromium has a statistically significant trend in Tampa, questionable in Houston and no trend in Pittsburgh. It is also not a seasonal metal in both Tampa ($p = 0.92$) and Pittsburgh ($p = 0.70$). Temperature has no significant

effect on chromium in both Tampa ($p = 0.44$) and Houston ($p = 0.46$) and is highly dependent upon pressure only in Pittsburgh ($p < 0.0001$). In Houston lead has a clear trend, seasonality and dependent upon temperature and pressure. The metal is not seasonal in both Tampa ($p = 0.45$) and Pittsburgh ($p = 0.07$). Pressure is not a significant predictor of lead in Tampa ($p = 0.21$). Manganese has highly significant trend and seasonality in all tree regions. Temperature has no clear significant effect in Tampa ($p = 0.04$), while pressure is not statistically significant in Tampa ($p = 0.05$) and Houston ($p = 0.07$). Trend, seasonality and pressure effect are not present on magnesium in Tampa and Pittsburgh, while temperature is highly significant in all three regions. Zinc has a highly significant trend in all regions. Seasonality is not present in Tampa ($p = 0.52$) and Houston ($p = 0.31$), while temperature has no effect in Pittsburgh ($p = 0.47$). These results are summarized in Table 2.6.

Standard deviations of the fitted random effects gamm models in Tampa, Houston and Pittsburgh are summarized in Table 2.7. It also gives the order one autoregressive correlation coefficients fitted to the pollution data. Aluminum, calcium, sodium and zinc are variable among sites in Tampa. All metals show variation between sites in Houston with calcium, sodium and iron showing the highest variations. In Pittsburgh meanwhile, sodium and magnesium show little variation between sites. Table 2.7 also suggests the presence of spatial variation among sites within a region and between regions.

Aluminum

Figures 2.5-2.7 show the results of fitting the above gamm models to aluminum in Tampa, Houston and Pittsburgh respectively. Trend are similar in Tampa and Pittsburgh while seasonality are similar in Tampa and Houston with pick in the summer. Although increasing, trend in Houston is not linear as in the other two regions. Trend similarity in Tampa and Pittsburgh may suggest similar aluminum sources in these

Table 2.6: Significance of Year, Month, Temperature and Barometric Pressure on some selected heavy metals in Tampa, Houston and Pittsburgh.

Aluminum				
	Year	Month	temp	Pressure
Tampa	0.1310	3.66×10^{-11}	0.0254	0.0106
Houston	2.80×10^{-10}	$< 2 \times 10^{-16}$	7.55×10^{-10}	0.0633
Pittsburgh	0.000886	0.041634	5.67×10^{-12}	0.289741
Calcium				
Tampa	0.000481	9.80×10^{-06}	1.07×10^{-06}	8.66×10^{-05}
Houston	4.55×10^{-10}	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	1.09×10^{-06}
Pittsburgh	0.037	4.90×10^{-13}	$< 2 \times 10^{-16}$	4.89×10^{-07}
Sodium				
Tampa	0.001394	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.0000459
Houston	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.000966
Pittsburgh	$< 2 \times 10^{-16}$	0.0104	1.33×2^{-11}	0.0785
Iron				
Tampa	0.820395	3.06×10^{-11}	0.000187	0.013248
Houston	2.06×10^{-07}	$< 2 \times 10^{-16}$	1.40×10^{-08}	3.00×10^{-05}
Pittsburgh	8.98×10^{-05}	6.86×10^{-07}	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
Chromium				
Tampa	0.0108	0.9214	0.4414	0.6051
Houston	0.0214	0.0123	0.4554	0.0331
Pittsburgh	0.81747	0.69980	0.00211	2.02×10^{-06}
Lead				
Tampa	5.58×10^{-12}	0.44955	0.00654	0.20542
Houston	0.002709	0.000127	1.72×10^{-06}	1.34×10^{-10}
Pittsburgh	0.00253	0.06698	3.63×2^{-06}	2.62×2^{-14}
Manganese				
Tampa	5.48×10^{-14}	0.00694	0.04116	0.04987
Houston	$< 2 \times 10^{-16}$	0.000808	0.000197	0.072555
Pittsburgh	7.85×10^{-08}	1.41×10^{-07}	1.19×10^{-11}	9.86×10^{-12}
Magnesium				
Tampa	0.229674	0.079368	0.000503	0.333945
Houston	$< 2 \times 10^{-16}$	1.71×10^{-11}	$< 2 \times 10^{-16}$	0.184
Pittsburgh	0.0708	0.6234	4.35×10^{-07}	0.3729
Zinc				
Tampa	0.010822	0.524571	0.000213	0.005207
Houston	2.72×10^{-11}	2.46×10^{-09}	1.85×10^{-05}	0.000312
Pittsburgh	0.000194	0.311186	0.468151	1.44×10^{-06}

Table 2.7: Standard deviation of sites random effects and ρ values for the fitted autoregressive of order one correlation.

Metal	Tampa	Houston	Pittsburgh
Aluminum	0.02(0.091)	0.06 (0.224)	0.02 (0.166)
Calcium	1.2 (0.154)	14.2(0.237)	3.2 (0.160)
Sodium	0.7 (0.056)	19.3 (0.225)	2.57×10^{-5} (0.095)
Iron	9.59×10^{-4} (0.162)	31.5 (0.237)	9.2 (0.0702)
Chromium	6.29×10^{-7} (-0.016)	0.3 (0.295)	0.3 (0.0134)
Lead	6.76×10^{-6} (-0.028)	0.8 (0.063)	4.2 (0.055)
Manganese	5.13×10^{-5} (0.073)	1.0(NA)	0.8 (NA)
Magnesium	1.56×10^{-4} (0.021)	3.3 (0.233)	8.33×10^{-8} (0.065)
Zinc	0.5 (0.008)	5.0 (0.071)	5.5 (0.016)

Figure 2.5: Non parametric estimates of trend (p=0.131), seasonality ($p < 0.001$), temperature (p=0.025) and barometric pressure (p=0.0106) effect of aluminum in Tampa

31

Figure 2.6: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure (p=0.063) effect of aluminum in Houston

Figure 2.7: Non parametric estimates of trend ($p < 0.001$), seasonality ($p = 0.041$), temperature ($p < 0.001$) and barometric pressure (p=0.290) effect of aluminum in Pittsburgh

Figure 2.8: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of calcium in Tampa

two geographical areas. Mobile sources might be the most important contributors of aluminum in these two regions. Both Tampa and Houston see a high elevation of aluminum in the summer with the reverse situation in Pittsburgh. The reason might be the high demand of electricity and more importantly an increased traffic in the summer due to high tourism activity during that period. In all three regions high temperatures are associated with higher aluminum levels, while pressure has effect only in Houston. It is not clear, what role the steel and oil industries play in aluminum concentration as they are probably not the major sources.

Calcium

Calcium level is decreasing in Tampa, increasing in Houston and nearly constant in Pittsburgh (Figures 2.8-2.10). The concentration level may be driven by the type of industries and not other sources. Except for the industry types, Tampa is very much

Figure 2.9: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of calcium in Houston

Figure 2.10: Non parametric estimates of trend (p=0.037), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of calcium in Pittsburgh

Figure 2.11: Non parametric estimates of trend ($p = 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of sodium in Tampa

comparable to Houston but yet we see completely different trend of calcium. As for seasonality, we see similar trend in Tampa and Houston, probably due to similar summer activities in these two regions. In both regions, calcium concentrations are high in the summer and low in the winter and have tendency to increase with high level of temperature. To the contrary, in Pittsburgh summer see the lowest concentrations. Similar temperature effect as in the previous two regions is observed. Pressure plays the same role in all these regions having an increasing level of calcium associated with high level of barometric pressure.

Sodium

Sodium in moderately decreasing in Tampa and Pittsburgh but increasing in Houston (Figures 2.11-2.13). Summer months seem to see the lowest level of sodium in both Tampa and Houston while the level is closely constant in Pittsburgh. Can the type
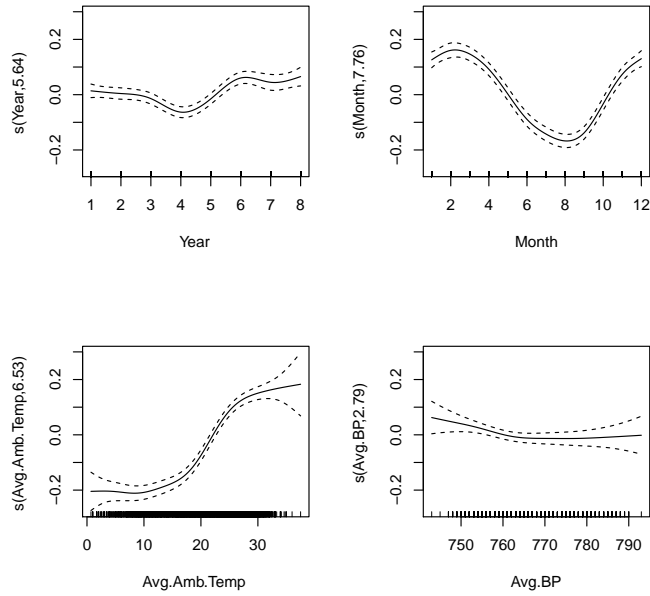
Figure 2.12: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of sodium in Houston
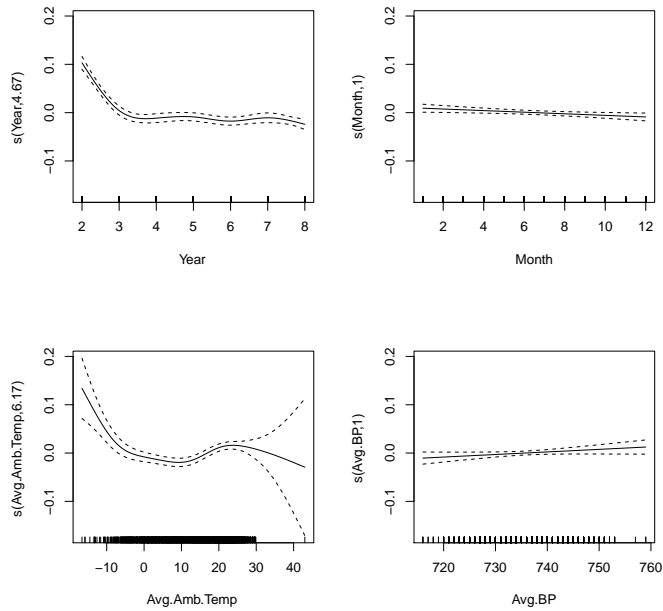
Figure 2.13: Non parametric estimates of trend ($p < 0.001$), seasonality ($p = 0.010$), temperature ($p < 0.001$) and barometric pressure (p=0.079) effect of sodium in Pittsburgh
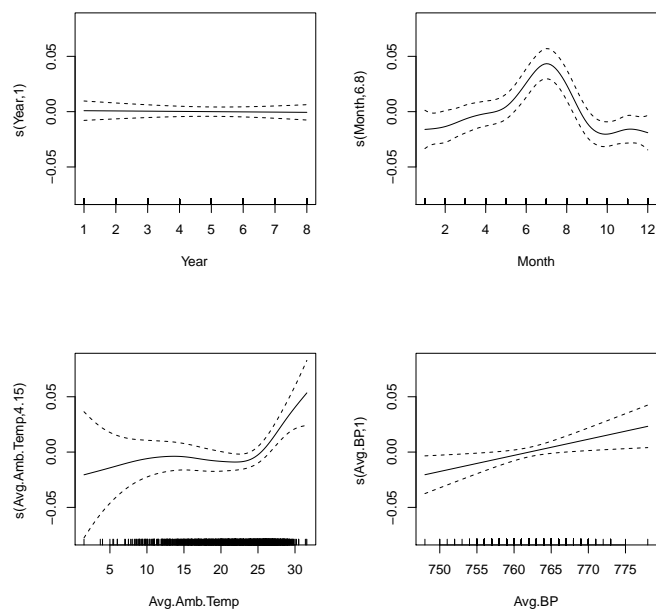
Figure 2.14: Non parametric estimates of trend (p=0.820), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure (p=0.013) effect of iron in Tampa

of industries be a determinant factor for this metal too? Temperature has opposite effect by regions. In Tampa and Houston, high level of temperatures are associated with high levels of sodium while Pittsburgh sees the opposite: with high level of temperatures associated with low levels of sodium. Probably not only temperature play a role but rather climate including rain and wind may play a significant role in these differences. Barometric pressure has also a slightly different contribution depending on regions. While it has a decreasing effect in Pittsburgh, the pattern is more complex in Tampa and Houston.

Iron

Trend, seasonality, temperature and pressure effects on iron are similar to aluminum in all three regions, Figures 2.14-2.16. Iron has no trend in Tampa. We see a nearly constant iron level during the eight years study period. It varies by season with high
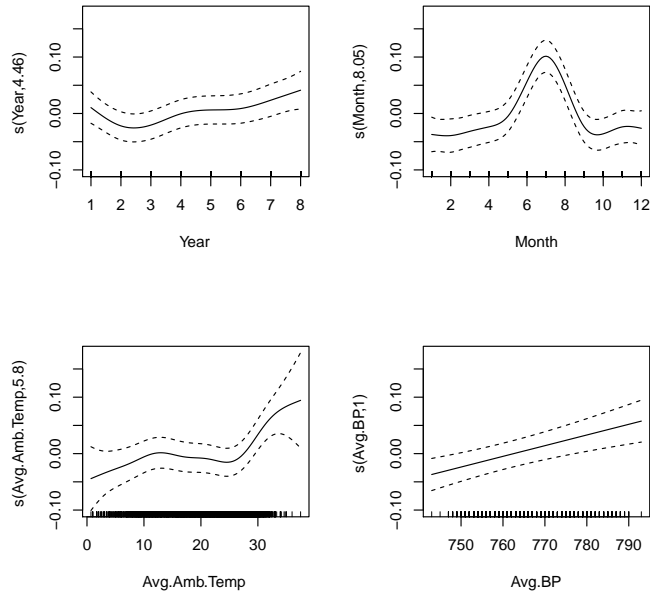
Figure 2.15: Non parametric estimates of trend ($p < 0.001$) , seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure($p < 0.001$) effect of iron in Houston
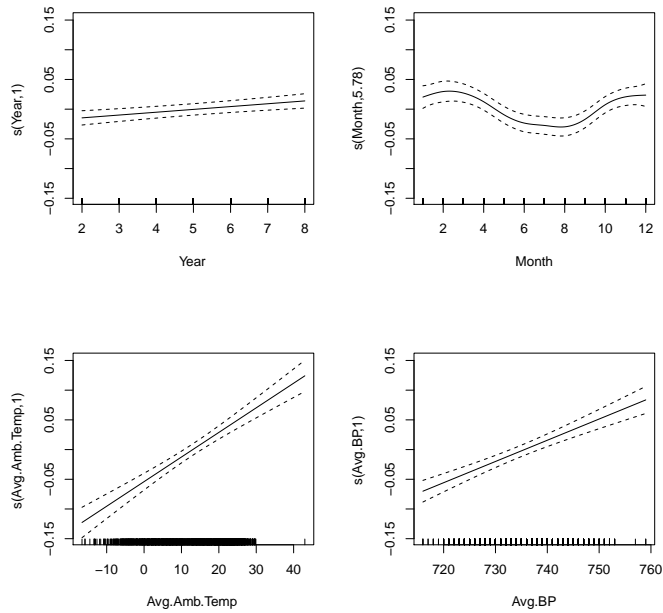
Figure 2.16: Non parametric estimates of trend ($p < 0.001$) , seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure($p < 0.001$) effect of iron in Pittsburgh

concentrations in the summer yielding to a pick in July. Temperatures above 25^0C are associated with high elevations of iron in the Tampa bay area. In Houston iron has an increasing trend with maximum concentrations in 2007. The seasonality is closely similar to the Tampa bay seasonality which may suggest again a non industries type pollution source. On road sources rather that non road pollution sources may have played an important role with heavy summer vacation traffic. Temperature has a similar effect as in Tampa with high temperatures associated with high iron level. In all three regions Tampa, Houston and Pittsburgh, pressure has a linear increasing effect. Pittsburgh meanwhile has a moderate increasing trend of iron. Industries might be playing a big role in iron level as both regions, Houston and Pittsburgh with more presence of industries show some trend level of iron while with almost no industry, Tampa shows a near constant trend of the metal. These findings also suggest iron origin might not be from road and non road sources alone. Seasonality in Pittsburgh is more variable. March and November see the highest concentration levels.

Chromium

Chromium has a convex down shape in Tampa and Houston and is constant in Pittsburgh. In Houston it reaches it maximum in 2003 and 2004. The seasonality is constant in Tampa and Pittsburgh. It is periodic in Houston with maximum in the summer. June sees the highest maximum concentrations in Houston. The concentration increases linearly with temperature in all three regions. Barometric pressure has a similar effect in Houston and Pittsburgh while concentrations decrease with increasing pressures in Tampa (Figures 2.17-2.19).

Lead

Lead, Figures 2.20-2.22 has a highly decreasing trend in Tampa, moderately decreasing
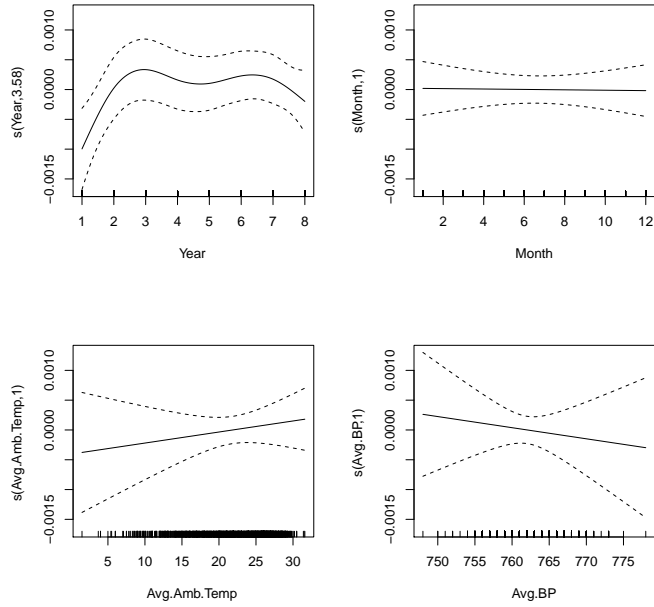
Figure 2.17: Non parametric estimates of trend (p=0.011), seasonality (p=0.921), temperature (p=0.441) and barometric pressure (p=0.605) effect of chromium in Tampa
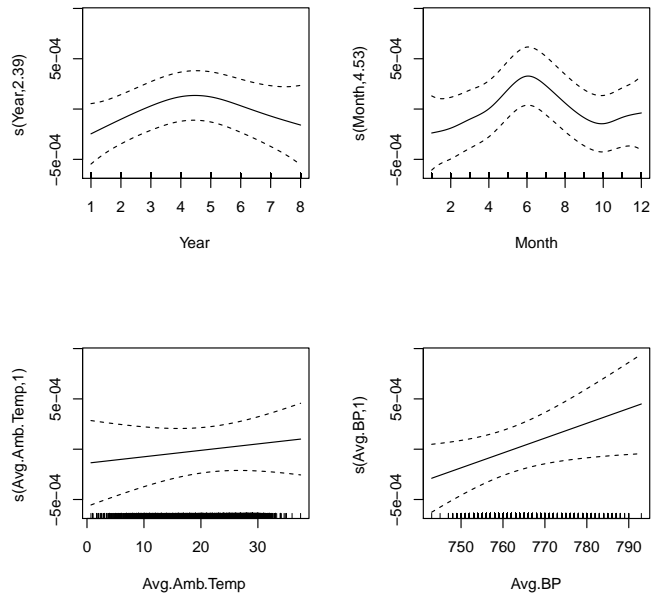
Figure 2.18: Non parametric estimates of trend (p=0.021), seasonality (p=0.012), temperature (p=0.455) and barometric pressure (p=0.033) effect of chromium in Houston
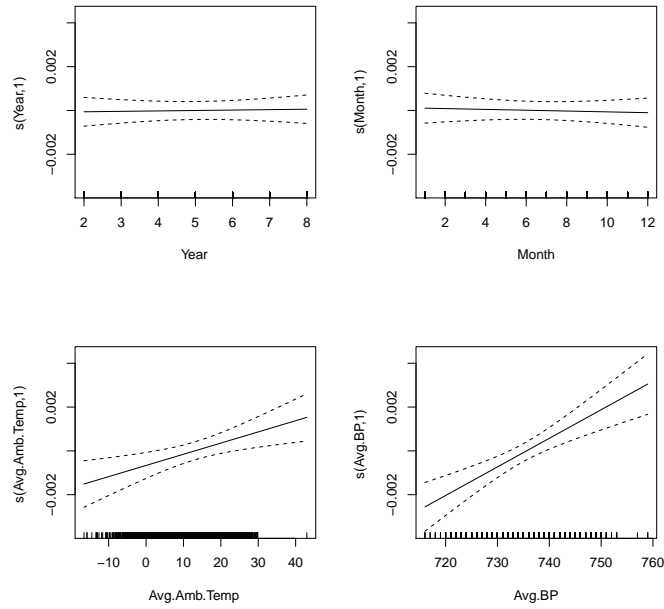
Figure 2.19: Non parametric estimates of trend (p=0.817), seasonality (p=0.700), temperature (p=0.002) and barometric pressure ($p < 0.001$) effect of chromium in Pittsburgh
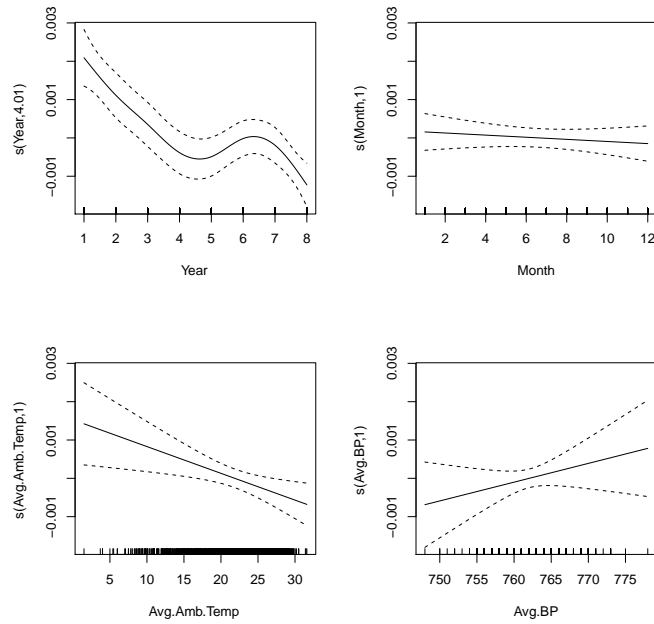
Figure 2.20: Non parametric estimates of trend ($p < 0.001$), seasonality (P=0.450), temperature (P=0.007) and barometric pressure (P=0.205) effect of lead in Tampa
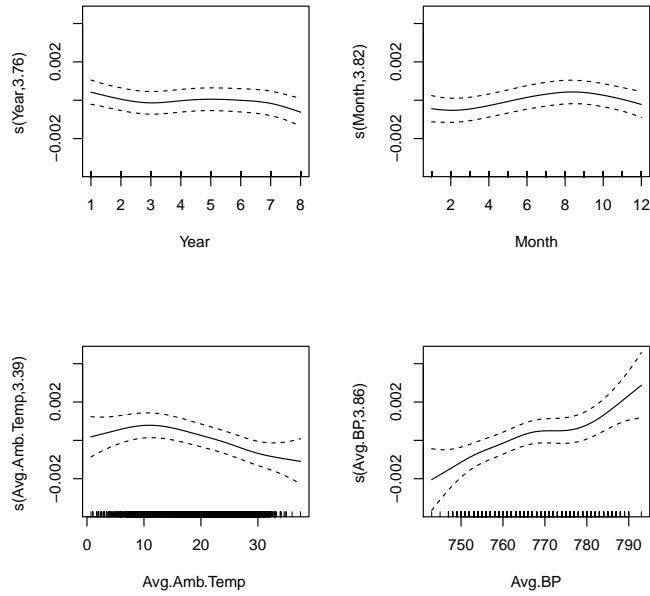
Figure 2.21: Non parametric estimates of trend ($p = 0.003$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of lead in Houston
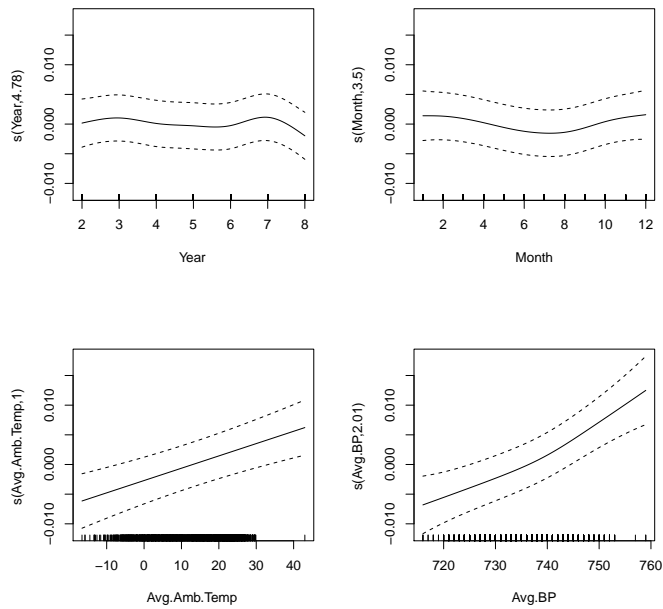
Figure 2.22: Non parametric estimates of trend (P=0.003), seasonality (P=0.067), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of lead in Pittsburgh
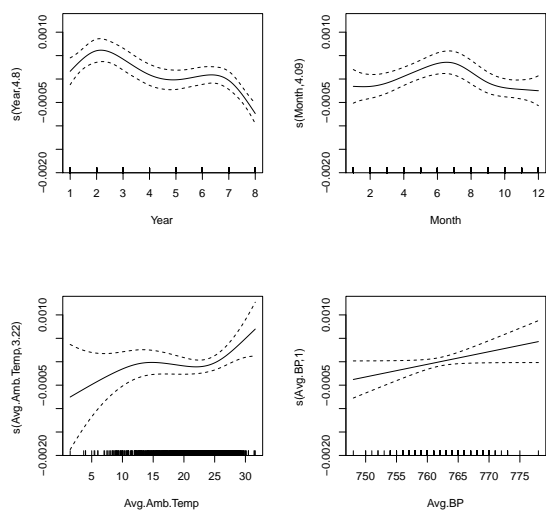
Figure 2.23: Non parametric estimates of trend ($p < 0.001$), seasonality (p=0.007), temperature (p=0.041) and barometric pressure (p=0.050) effect of manganese in Tampa

in Houston and almost constant in Pittsburgh. Seasonality is quite different in all three regions. Linear decreasing in Tampa, convex down in Houston with maximum in August and convex down in Pittsburgh with minimum in July and August. Lead concentrations decrease with increasing temperatures in Tampa and have opposite effect in Pittsburgh. All three regions see increasing concentrations with increasing barometric pressures.

Manganese

Manganese, Figures 2.23-2.25 has a decreasing trend in Tampa and V-shape form in Houston with low concentrations in 2002. It is linearly decreasing in Pittsburgh. We see similar convex up periodicity in Tampa and Houston with maximum in June-July and opposite figures in Pittsburgh with minimum in June-July. While temperatures have a mixed effects on manganese concentrations in Houston, we see increasing concentrations with increasing temperatures in the other two regions. Pressure shows a linear increasing effect on manganese in Tampa and Houston with mixed results in
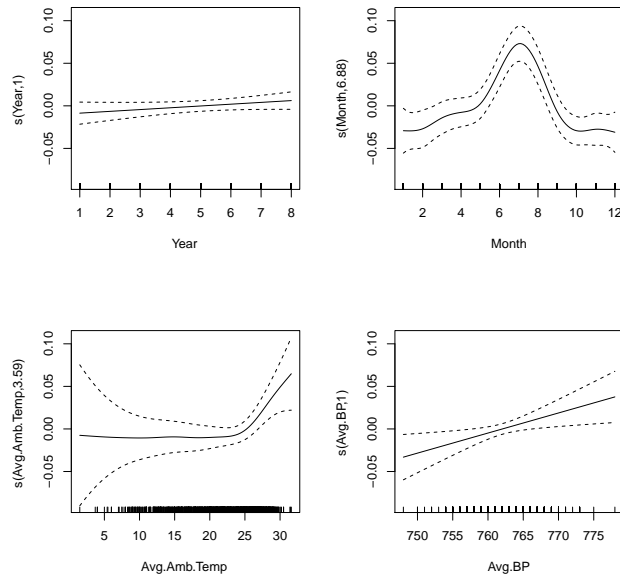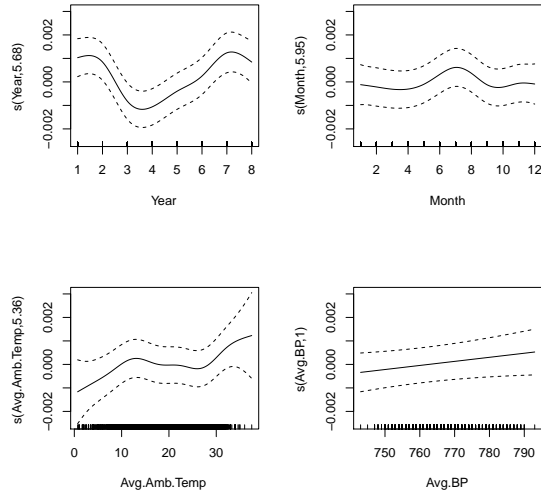
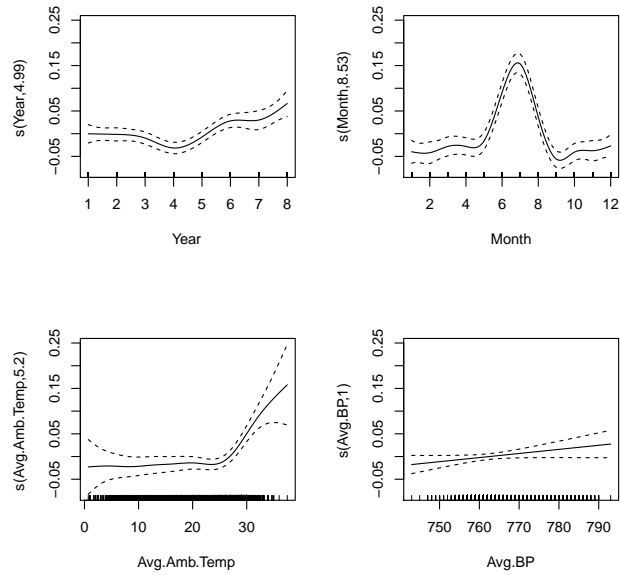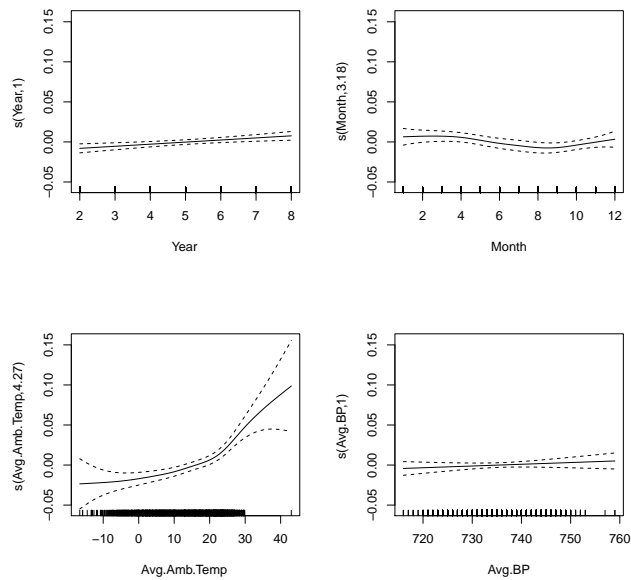Figure 2.24: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure (p=0.073) effect of manganese in Houston
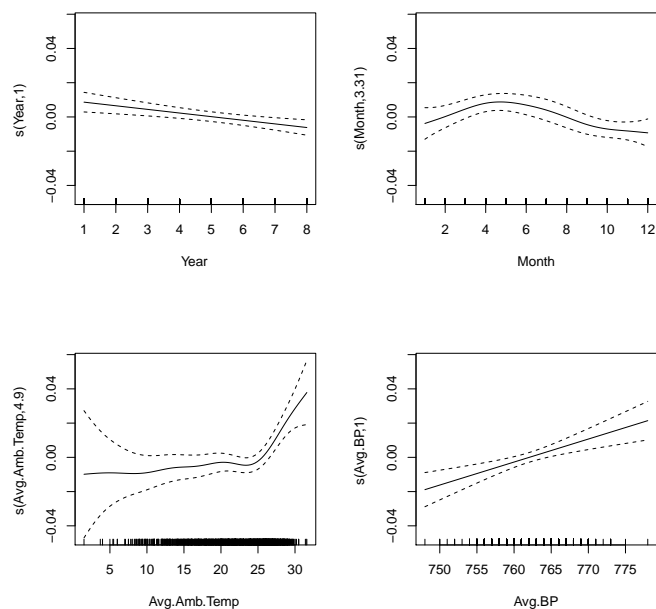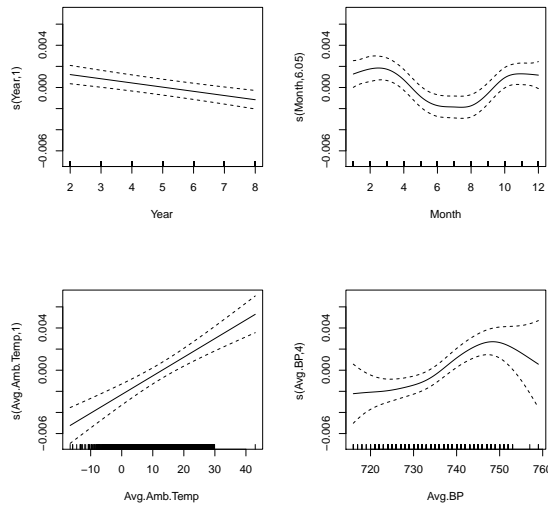

Figure 2.25: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of manganese in Pittsburgh

Figure 2.26: Non parametric estimates of trend (p=0.230), seasonality (p=0.079), temperature ($p < 0.001$) and barometric pressure ($p = 0.334$) effect of magnesium in Tampa

Pittsburgh.

Magnesium

Magnesium, Figures 2.26-2.28 has a linear increasing trend in Tampa, sinusoidal in Houston and convex up in Pittsburgh. The high contrast of this metal within the three regions make it difficult to predict it main sources. Seasonality is similar in Tampa and Pittsburgh and irregular in Houston with maximum in July and minimum in September. Increasing temperatures are associated with increasing magnesium levels in all three regions. Pressure shows a similar effect everywhere.

Zinc

Increasing barometric pressure see increasing zinc concentrations with near constant seasonality in all three regions. Concentrations decrease with increasing temperatures in Tampa, increase with increasing temperatures in Pittsburgh and variable in Houston. The trend is mostly increasing almost everywhere Figures 2.29-2.31.

Figure 2.27: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.184$) effect of magnesium in Houston

Figure 2.28: Non parametric estimates of trend (p=0.071), seasonality (p=0.623), temperature ($p < 0.001$) and barometric pressure ($p = 0.373$)effect of magnesium in Pittsburgh

Figure 2.29: Non parametric estimates of trend (p=0.011), seasonality (p=0.525), temperature ($p < 0.001$) and barometric pressure (p=0.005) effect of zinc in Tampa
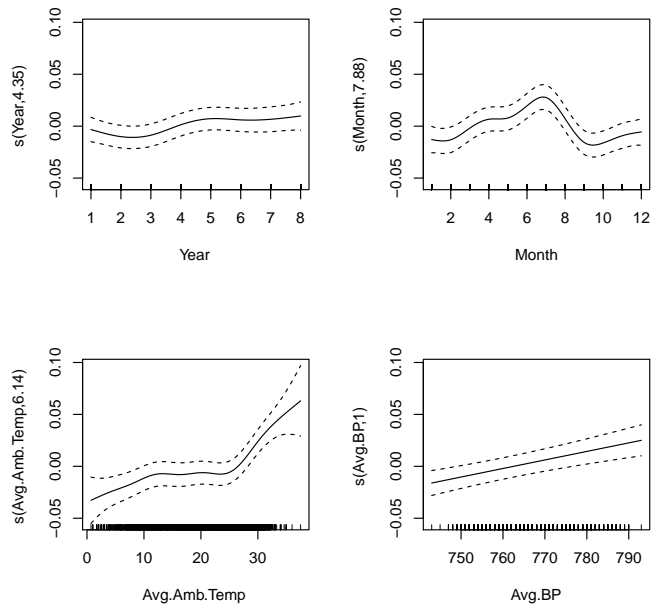

Figure 2.30: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of zinc in Houston
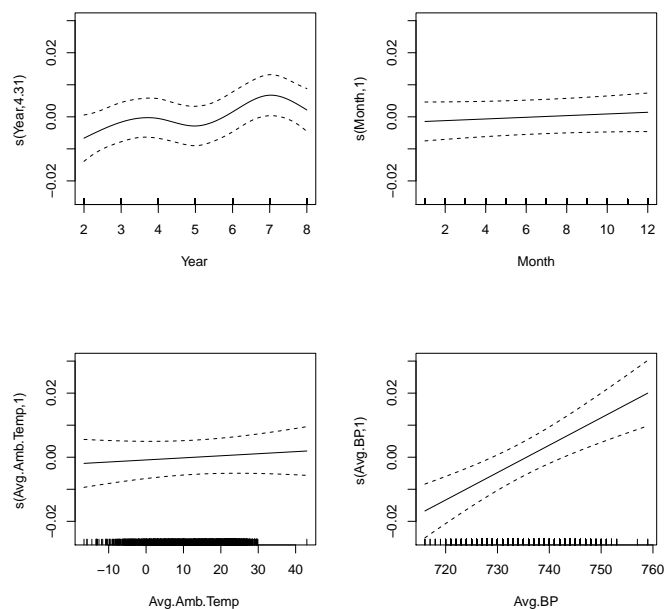
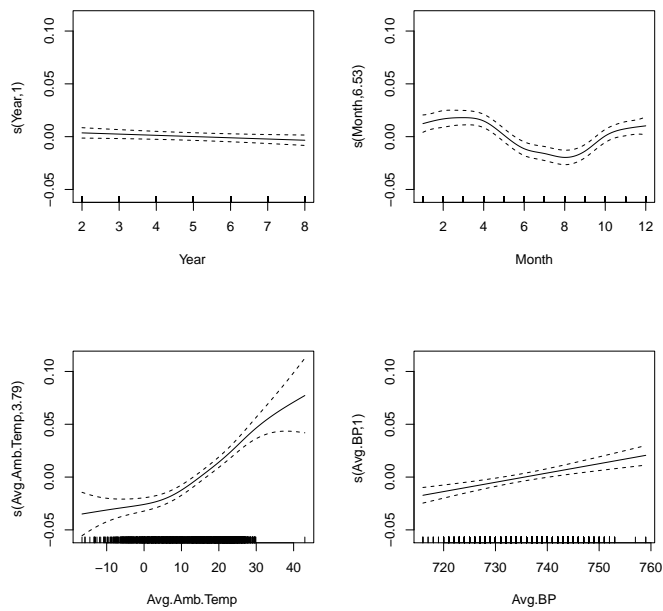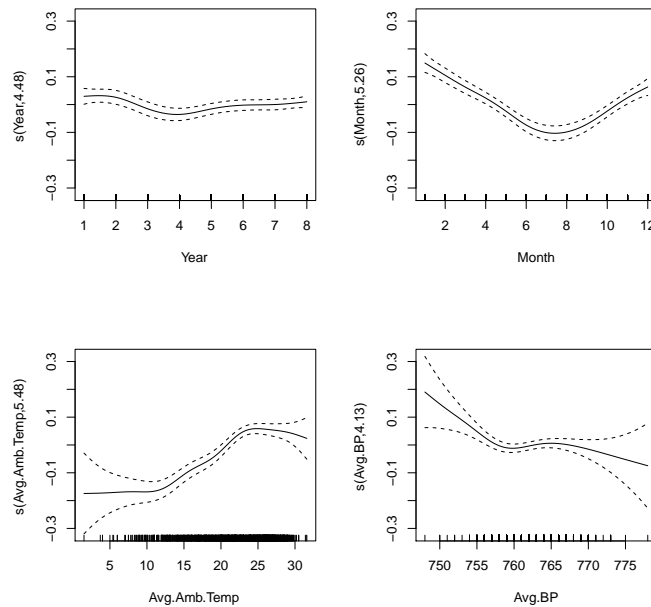Figure 2.31: Non parametric estimates of trend ($p < 0.001$), seasonality (p=0.311), temperature (p=0.468) and barometric pressure ($p < 0.001$) effect of zinc in Pittsburgh

2.4.2 Metal mixtures

Table 2.8 presents the first 9 mixtures of metals based on the first three leading factors from factor analysis. Mixtures 1-3 are from the three leading factors in Tampa, mixtures 4-6 are from the three leading factors in Houston and mixtures 7-9 are from the three leading factors in Pittsburgh. As with individual metals, differences in mixtures from the factor analysis is an evidence of huge spatial variations between regions. Mixtures 10-18 are presented in Table 2.9. Mixtures 10-12 from Tampa, 13 from Houston and 14-16 from Pittsburgh are based on factor loadings greater or equal to 0.05. Mixtures 17 and 18 are based on significance with low birth and preterm birth. Summaries of fitting the above generalized additive mixed models to the mixtures are presented in tables 2.10 and 2.11[1]. They show the significance of trend, seasonality, temperature and barometric pressure effects in Tampa, Houston and Pittsburgh.

---

[1]**Notation**: (I) increasing, (D) decreasing, (N) undecided, (LS) low in September, (HS) high in the summer, (C) constant, (ID) increasing followed by decreasing

Table 2.8: Mixture of Metals, Mixtures 1-3 in Tampa, 4-6 in Houston and 7-9 in Pittsburgh are based on on the first three leading factors from factor analysis

Metal	Mix1	Mix2	Mix3	Mix4	Mix5	Mix6	Mix7	Mix8	Mix9
Aluminum	+	+		+			+	+	
Barium	+		+		+	+		+	+
Cadmium									
Calcium		+	+	+	+		+		+
Chromium				+	+		+		
Cobalt		+	+	+	+	+	+		+
Copper	+				+		+	+	
Cesium			+			+			+
Gallium			+			+	+		
Iron		+		+	+		+		
Hafnium			+	+					+
Lead	+		+		+	+	+	+	
Indium									
Manganese	+	+	+	+	+	+	+		+
Iridium			+			+			+
Molybdenum			+		+		+		+
Nickel			+		+		+		
Magnesium	+				+	+		+	+
Mercury			+						
Gold			+						+
Lanthanum	+		+			+			+
Niobium			+			+			+
Tin			+		+	+			+
Titanium	+	+		+	+	+	+		+
Scandium			+	+					
Vanadium	+	+	+	+	+		+	+	
Silver			+						+
Zinc	+				+		+		
Strontium	+			+	+	+		+	
Tantalum			+			+			+
Rubidium			+	+		+			
Potassium	+			+	+	+	+	+	
Yttrium			+	+					+
Sodium				+	+		+		+
Zirconium				+					+

Table 2.9: Mixture of Metals, Mixtures 10-12 in Tampa, 13 in Houston and 14-16 in Pittsburgh are based on loading greater or equal to 0.05. Mixture 17 and 18 are based on significance with low birth.

Metal	Mix10	Mix11	Mix12	Mix13	Mix14	Mix15	Mix16	Mix17	Mix18
Alum		+		+				+	
Barium	+						+		
Cadmium								+	
Calcium		+		+	+				
Chrom									
Cobalt									
Copper	+					+		+	
Cesium									
Gallium								+	+
Iron		+		+	+			+	
Hafnium								+	+
Lead									
Indium									
Manganes		+			+				
Iridium			+					+	+
Molyb								+	
Nickel					+			+	+
Magnes	+			+					
Mercury								+	+
Gold								+	+
Lanthan									
Niobium									
Tin								+	+
Titanium		+		+					
Scandium									
Vanadium									
Silver									
Zinc					+				
Strontium	+					+			
Tantalum			+				+	+	+
Rubidium									
Potassium	+					+			
Yttrium									
Sodium								+	+
Zirconium									

Table 2.10: Significance of Trend, Seasonality, Temperature and Pressure of some metal mixtures in Tampa, Houston and Pittsburgh.

Mixture 1				
	Year	Month	temp	Pressure
Tampa	0.617 (N)	< 0.001(HS)	0.701 (N)	0.226(N)
Houston	< 0.001(I)	< 0.001(HS)	< 0.001 (I)	0.003(I)
Pittsburgh	0.004(N)	< 0.001(LS)	< 0.001 (I)	< 0.001 (I)
Mixture 2				
Tampa	0.797 (N)	< 0.001(HS)	< 0.001 (I)	0.004 (I)
Houston	< 0.001 (I)	< 0.001(HS)	< 0.001(I)	0.001(I)
Pittsburgh	0.003(N)	< 0.001 (LS)	< 0.001 (I)	< 0.001(I)
Mixture 3				
Tampa	< 0.001 (D)	< 0.001(N)	< 0.001(I)	0.201(N)
Houston	< 0.001(D)	< 0.001(HS)	< 0.001(I)	< 0.001(I)
Pittsburgh	< 0.001(D)	< 0.001(LS)	< 0.001(I)	< 0.001(I)
Mixture 4				
Tampa	0.983 (N)	< 0.001(HS)	< 0.001(I)	0.197
Houston	< 0.001(I)	< 0.001(LS)	< 0.001(I)	0.013 (I)
Pittsburgh	0.002(D)	< 0.001(LS)	< 0.001(I)	< 0.001(I)
Mixture 5				
Tampa	0.124	< 0.001(HS)	< 0.001(I)	0.516
Houston	< 0.001(I)	< 0.001(LS)	< 0.001(I)	0.013(I)
Pittsburgh	< 0.001(D)	< 0.001(LS)	< 0.001(I)	< 0.001(I)
Mixture 6				
Tampa	0.043(D)	0.007(HS)	0.994	0.513
Houston	< 0.001(D)	< 0.001(HS)	< 0.001(I)	0.009(I)
Pittsburgh	< 0.001(D)	< 0.001(HS)	< 0.001(I)	< 0.001(I)
Mixture 7				
Tampa	0.965(N)	< 0.001 (HS)	< 0.001(I)	0.197
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	0.005(I)
Pittsburgh	0.002(N)	< 0.001(LS)	< 0.001(I)	< 0.001
Mixture 8				
Tampa	0.880 (N)	< 0.001(HS)	0.719	0.238
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	0.009(I)
Pittsburgh	0.165 (N)	< 0.001(LS)	< 0.001(I)	< 0.001(I)
Mixture 9				
Tampa	< 0.001(D)	< 0.001(LS)	< 0.001(I)	0.006(D)
Houston	< 0.001(I)	< 0.001(LS)	< 0.001(I)	0.056
Pittsburgh	< 0.001 (D)	< 0.001(LS)	< 0.001(N)	< 0.001 (I)

Table 2.11: Significance of Trend, Seasonality, Temperature and Pressure of some metal mixtures in Tampa, Houston and Pittsburgh.

Mixture 10				
	Year	Month	temp	Pressure
Tampa	0.611	0.009 (HS)	0.996	0.469
Houston	< 0.001(ID)	< 0.001(HS)	< 0.001 (I)	0.021(I)
Pittsburgh	0.999	< 0.001(HS)	< 0.001(I)	< 0.001(I)
Mixture 11				
Tampa	0.814	< 0.001(HS)	< 0.001 (I)	0.003(I)
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	< 0.001(I)
Pittsburgh	0.003(N)	< 0.001(LS)	< 0.001(I)	< 0.001
Mixture 12				
Tampa	< 0.001 (D)	0.031(N)	0.183	0.067
Houston	< 0.001(D)	< 0.001(HS)	0.759	0.001(N)
Pittsburgh	< 0.001(D)	< 0.001(HS)	0.815	0.987
Mixture 13				
Tampa	0.989(N)	< 0.001(HS)	< 0.001 (I)	0.004(I)
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	< 0.001 (I)
Pittsburgh	< 0.001(I)	< 0.001(LS)	< 0.001(I)	< 0.001 (I)
Mixture 14				
Tampa	0.037 (D)	< 0.001(HS)	< 0.001(I)	0.002(I)
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	< 0.001 (I)
Pittsburgh	0.020(D)	< 0.001 (LS)	< 0.001 (I)	< 0.001(I)
Mixture 15				
Tampa	0.567(D)	0.007 (HS)	0.846	0.394
Houston	0.029(I)	< 0.001(HS)	0.001 (I)	0.004(I)
Pittsburgh	< 0.001(I)	< 0.001(HS)	< 0.001(I)	< 0.001(I)
Mixture 16				
Tampa	< 0.001 (D)	0.068	0.392	0.112
Houston	< 0.001(D)	< 0.001(HS)	0.026(N)	0.099
Pittsburgh	0.001(D)	< 0.001(N)	0.027(N)	0.875
Mixture 17				
Tampa	0.137	< 0.001(N)	< 0.001(I)	0.032(D)
Houston	< 0.001(I)	< 0.001(HS)	< 0.001(I)	0.105
Pittsburgh	< 0.001(D)	< 0.001(LS)	< 0.001(I)	< 0.001(I)
Mixture 18				
Tampa	< 0.001 (D)	< 0.001 (LS)	< 0.001 (I)	< 0.001 (D)
Houston	< 0.001(I)	< 0.001(LS)	< 0.001 (I)	0.001 (D)
Pittsburgh	< 0.001(D)	0.004 (N)	< 0.001(D)	0.110

Table 2.12: Standard deviation of sites random effects and ρ values for the coefficient of the autoregressive of order one correlation.

|	Tampa	Houston	Pittsburgh
Mix1	$3.19e^{-5}(0.011)$	0.0248(0.173)	0.0151(0.170)
Mix2	$4.59e^{-6}(0.138)$	0.0395(0.188)	0.010(0.128)
Mix3	$1.89e^{-5}(0.130)$	0.0223(0.187)	0.009(0.153)
Mix4	$1.18e^{-5}(0.047)$	0.0509(0.203)	0.0180(0.124)
Mix5	$1.69e^{-5}(0.027)$	0.0623(0.198)	0.0254(0.068)
Mix6	$1.67e^{-5}(0.002)$	0.0152(0.117)	0.004(0.146)
Mix7	$1.99e^{-5}(0.045)$	0.0572(0.203)	0.0293(0.098)
Mix8	NA(0.007)	0.0191(0.175)	0.0075(0.212)
Mix9	0.006(0.103)	0.0349(0.244)	$2.84e^{-5}(0.074)$
Mix10	$4.23e^{-7}(-0.005)$	0.0131(0.114)	$6.15e^{-6}(0.171)$
Mix11	$4.27e^{-6}(0.135)$	0.0365(0.190)	0.010(0.137)
Mix12	$4.41e^{-8}(0.130)$	0.0001(0.174)	$1.80e^{-8}(0.189)$
Mix13	$4.79e^{-6}(0.136)$	0.0364(0.188)	0.009(0.130)
Mix14	$1.18e^{-6}(0.146)$	0.0477(0.212)	0.0135(0.084)
Mix15	$2.34e^{-6}(-0.009)$	0.0106(0.099)	$7.14e^{-6}(0.151)$
Mix16	$4.93e^{-7}(0.075)$	NA(0.183)	$9.35e^{-11}(0.085)$
Mix17	$6.07e^{-7}(0.130)$	0.0311(0.223)	0.0155(0.062)
Mix18	0.007(0.053)	0.0199(0.235)	0.003(0.116)
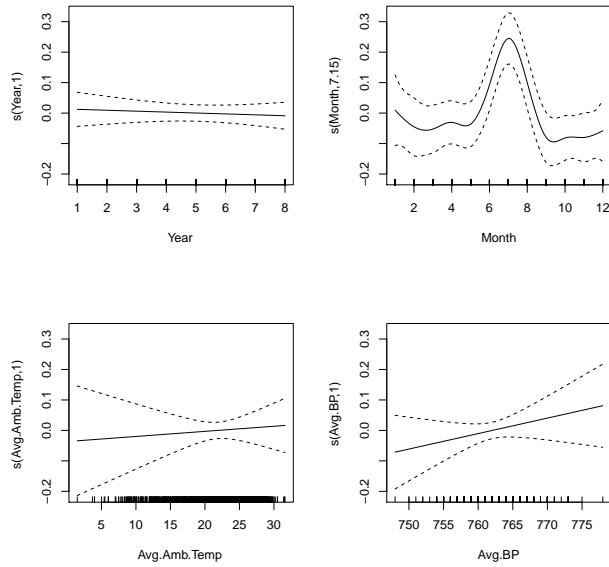
Figure 2.32: Non parametric estimates of trend ($p = 0.617$), seasonality ($p < 0.001$), temperature ($p = 0.701$) and barometric pressure (p=226) effect of mixture 1 in Tampa

Trend

Mixtures can be grouped in term of their trend similarity in Tampa, Houston and Pittsburgh. Mixtures 1, 2, 7, 8 and 11 show no apparent trend in both Tampa and Pittsburgh but an increasing trend in Houston. A decreasing trend in Tampa and Pittsburgh with an increasing trend in Houston is observed with mixtures 9, 14 and 18 . Mixture 4, 5 and 17 have no apparent trend in Tampa, increasing in Houston and decreasing in Pittsburgh. A decreasing trend in all three regions is observed with mixtures 3, 6, 12 and 16. Mixtures 13 and 15 present no trend in Tampa and an decreasing trend in both Houston and Pittsburgh. Mixture 10 shows an initial increase followed by a decrease in trend in Houston and no trend everywhere else.

Seasonality
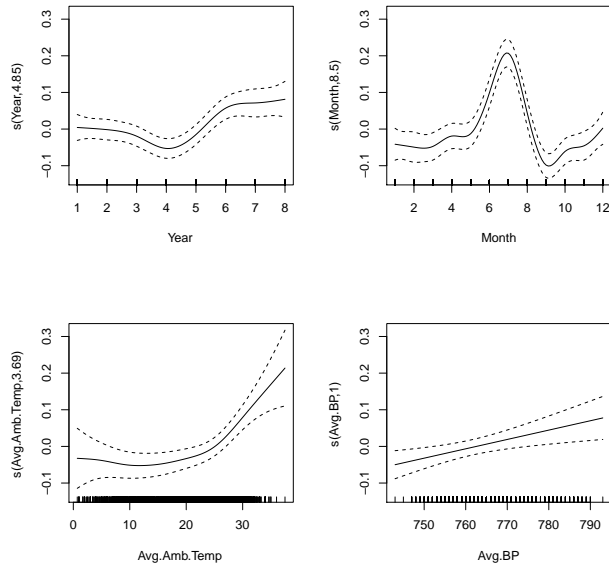
Figure 2.33: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure (p=0.003) effect of mixture 1 in Houston
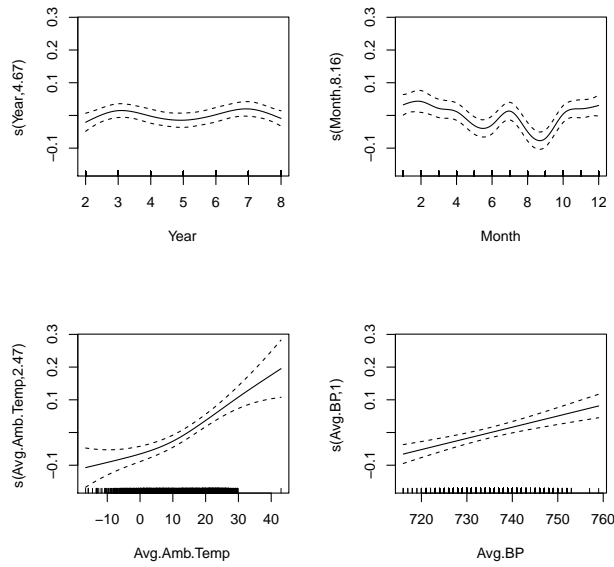
Figure 2.34: Non parametric estimates of trend ($p = 0.004$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of mixture 1 in Pittsburgh
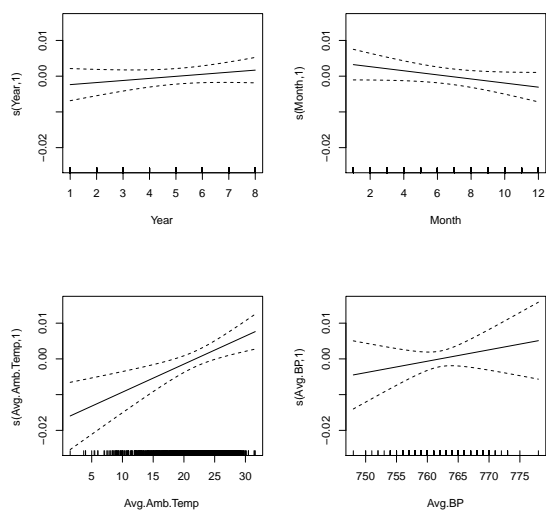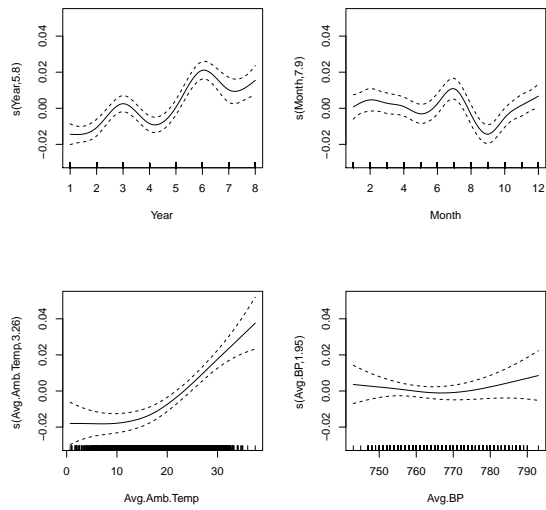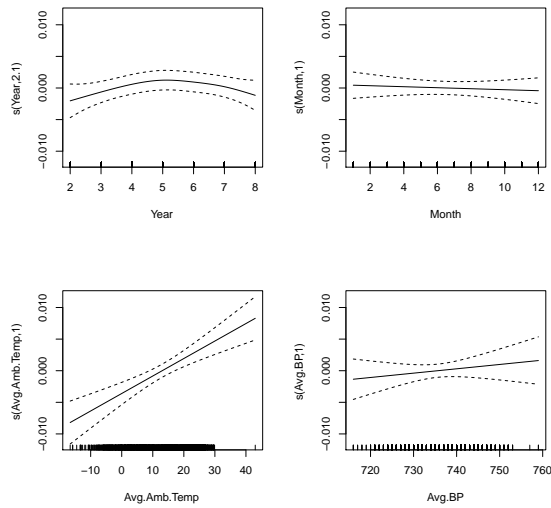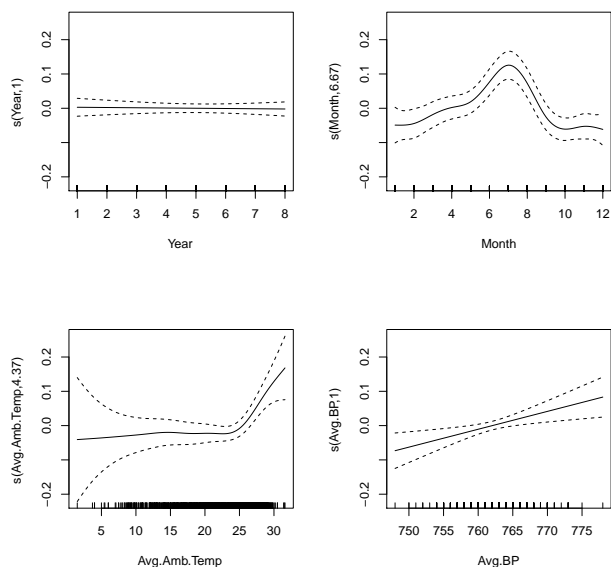
Figure 2.35: Non parametric estimates of trend ($p = 0.797$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p = 0.004$) effect of mixture 2 in Tampa

As it was the case for the trend, several mixtures share the same seasonality. Mixtures 9 is seasonal and see it lowest concentrations around September in all three regions. Mixtures 1, 2, 7, 8, 11, 13 and 14 are seasonal with highest concentrations observed in the summer in Tampa and Houston and lowest in September in Pittsburgh. Seasonal mixtures with high concentrations in the summer in Tampa and low concentrations around September in both Houston and Pittsburgh are observed with mixtures 4 and 5. Mixtures 6, 10 and 15 are seasonal with high concentrations during summer in all three regions. Mixtures 12 is seasonal with high concentrations in summer in Houston and Pittsburgh while mixture 18 sees low concentrations in Tampa and Houston in September. Seasonal with high concentrations in the summer and either low or undecided concentrations in September in Pittsburgh are observed with mixtures 3, 16 and 17.
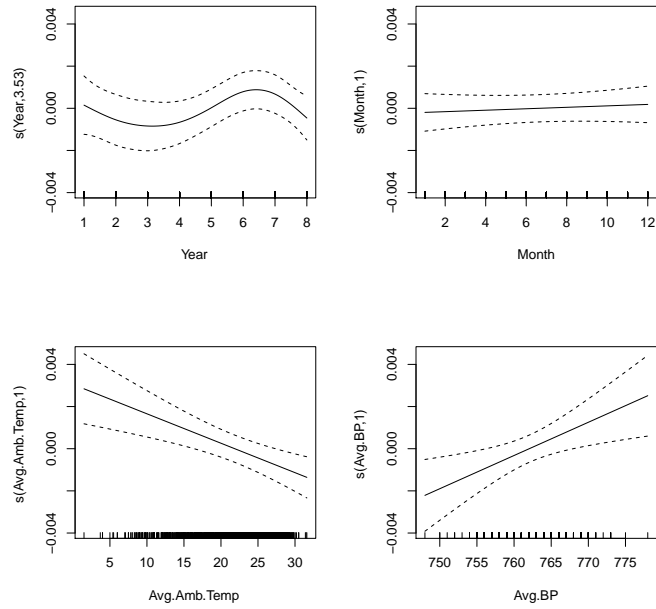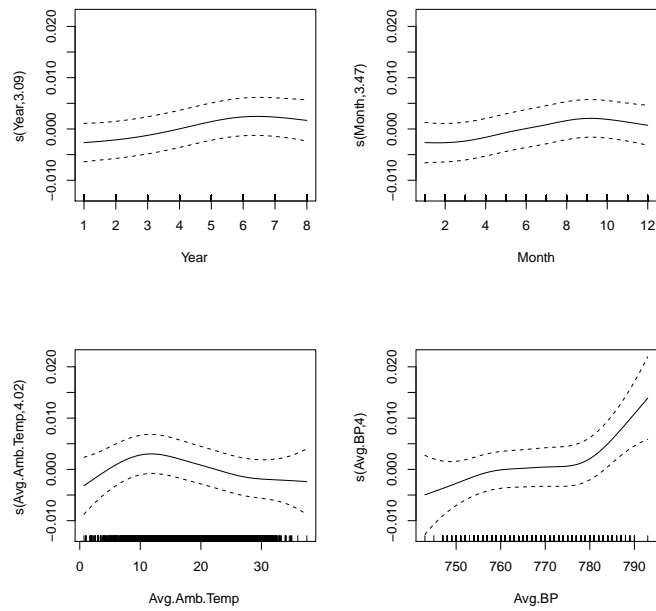
Temperature
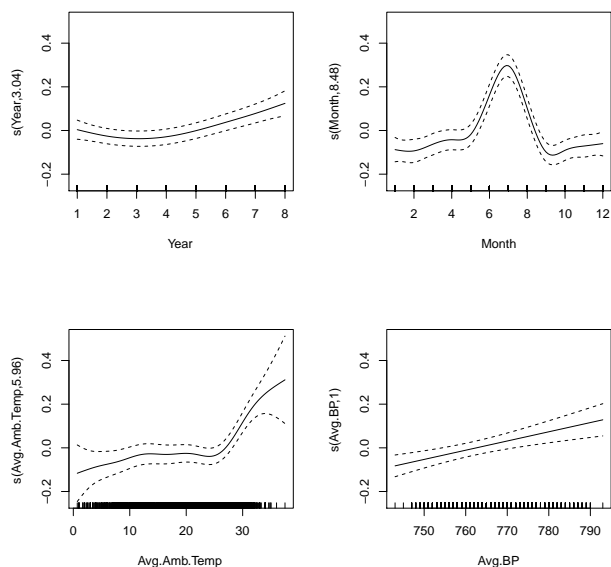
Figure 2.36: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure (p=0.003) effect of mixture 2 in Houston

Mixtures 2-5, 7, 11, 13, 14, and 17 have increasing concentrations with temperature in all three regions. Increasing concentrations with temperature in one or two regions with either non significant or undecided temperature effect in the remaining regions is observed with mixtures 1, 6, 8-10 and 15. Contrary to other mixtures, differences between regions in temperature effect is observed with mixture 18 which is increasing with temperatures in Tampa and Houston but decreasing with temperatures in Pittsburgh. Mixture 15 sees no temperature effect in all three regions. Mixtures with no or unclear temperature effect are mixtures 12 and 16.

Barometric Pressure

Barometric pressures have no effect with mixture 16 in all regions. Concentrations are increasing with barometric pressures in mixtures 2, 11, 13, 14 every where. Mixtures 9 and 17 see decreasing concentrations with barometric pressures in Tampa but not significant or undecided in the other two regions. Mixtures 1, 3-8, 10 and 15 have
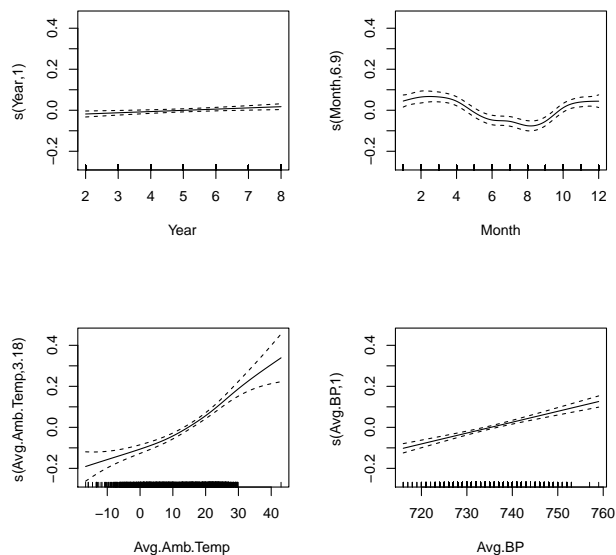
Figure 2.37: Non parametric estimates of trend ($p = 0.003$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of mixture 2 in Pittsburgh
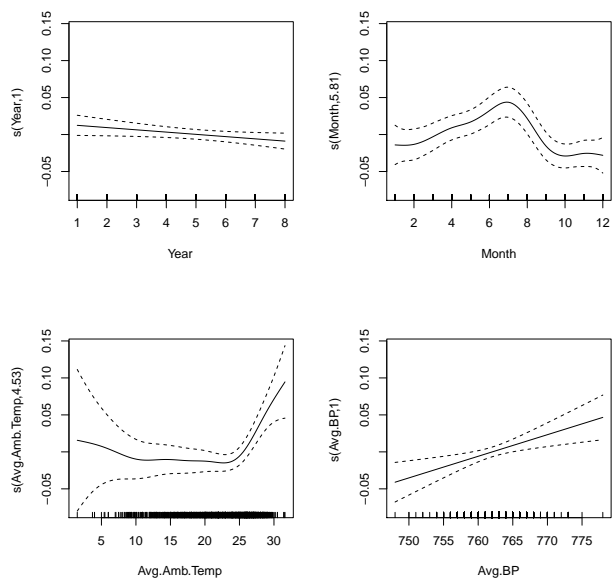
Figure 2.38: Non parametric estimates of trend ($p = 0.037$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p = 0.002$) effect of mixture 14 in Tampa
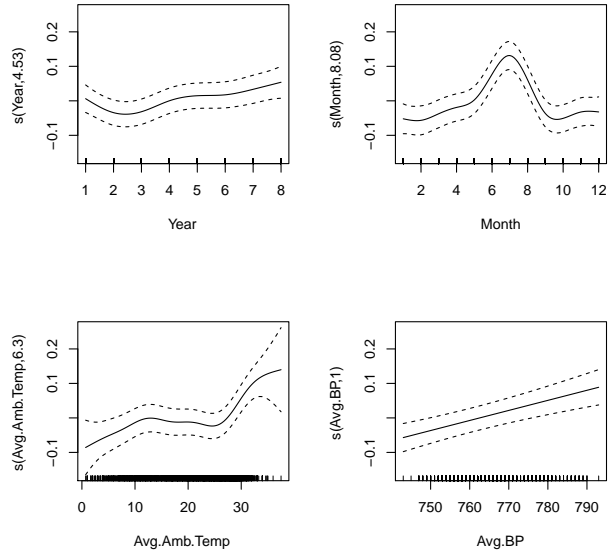
Figure 2.39: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of mixture 14 in Houston
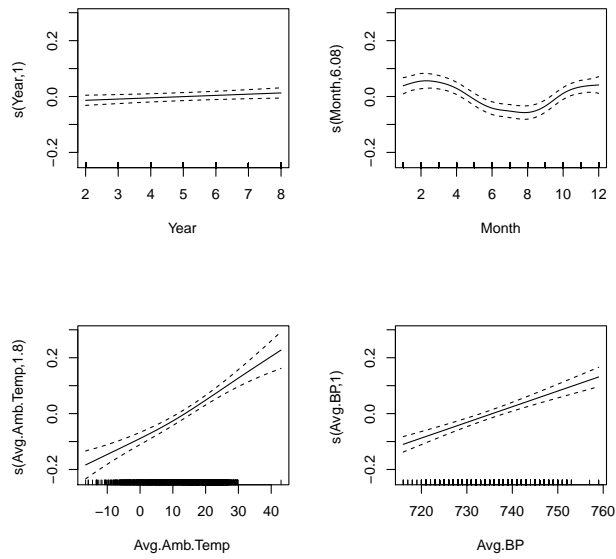
Figure 2.40: Non parametric estimates of trend ($p = 0.020$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of mixture 14 in Pittsburgh

Increasing concentrations with barometric pressures in one or two regions with either non significant or undecided pressure effects in the remaining regions. Concentrations from mixtures 9 and 17 are decreasing in Tampa and increasing in Pittsburgh with barometric pressures. Decreasing concentrations with barometric pressures in both Tampa and Houston are seen with mixture 18 while barometric pressures have no effects with mixture 16.

Overall we can note that mixtures 1, 2, 4, 5, 8 and 11 have similar trend, seasonality, temperature and barometric pressure effects in all three regions. They can be grouped in three categories. mixtures 1 and 8 belong to the first group, while 2, 11 are in the second group with 4 and 5 belonging to the third and last group. Mixture 18 show differences between regions especially on temperature effects. Looking at it composition we notice the presence of sodium in the mixture which is the most abundant metal in all regions. The metal also shows the same differences observed in the mixture. It is increasing with temperature in both Tampa and Houston but decreasing in Pittsburgh. Proximity with the sea might be the cause. Figures 2.32-2.43 present some mixtures in Tampa, Houston and Pittsburgh. They also show a clear spatial and temporal variations between regions. From the standard deviations of sites random effects and the autoregressive of order one correlation coefficient summarized in Table 2.12, we notice the absence of spatial variation in Tampa MSA. This is an evidence that spatial variation is mostly driven by emission sources which are absent is Tampa MSA as there are no industries in the region except maybe power plants. In Houston and Pittsburgh meanwhile we see the presence of spatial variation between sites within each region which is mainly due to the presence of industries as source emissions.

Different metals or mixtures behave differently depending on the regions. This is due primarily to the metals source origins which vary by regions. According to the US EPA, in 2005, the main sources of PM in Tampa MSA are variable and are distributed as follows by counties where PM monitoring stations are located. In Hills-
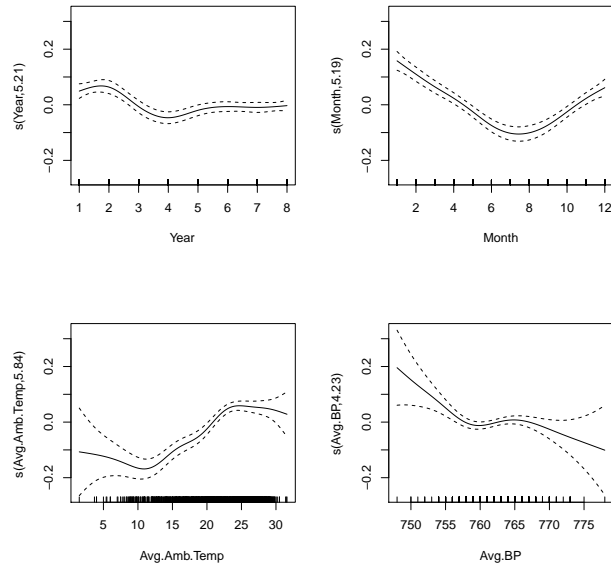
Figure 2.41: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p < 0.001$) effect of mixture 18 in Tampa
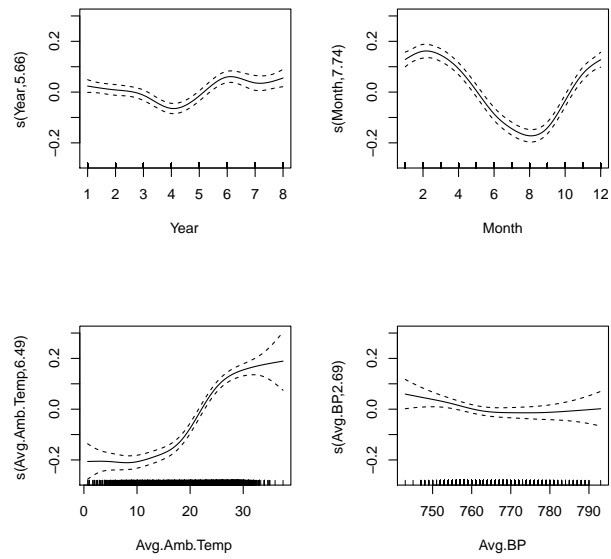
Figure 2.42: Non parametric estimates of trend ($p < 0.001$), seasonality ($p < 0.001$), temperature ($p < 0.001$) and barometric pressure ($p = 0.001$) effect of mixture 18 in Houston
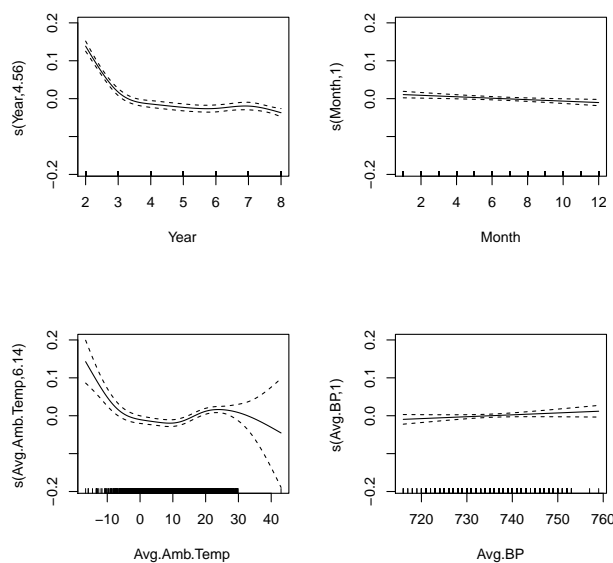
Figure 2.43: Non parametric estimates of trend ($p < 0.001$), seasonality ($p = 0.004$), temperature ($p < 0.001$) and barometric pressure ($p = 0.110$) effect of mixture 18 in Pittsburgh

borough county, the principal source is electricity generation (52.5%) followed by non road equipment (13.5%), road dust (9.9%), industrial processes (9.8%) and on road vehicles (5.1%). In Pinellas county, the main source of PM is fossil fuel combustion (27.6%), followed by non road equipment (17.6%), industrial processes (16.7%) and on road vehicles (9.9%). In Pittsburgh MSA the distribution of PM source varies also by county. In Allegheny county, the major sources are: Industrial processes (39.4%), fossil fuel combustion (16.6%), non road equipment (15.0%), residential wood combustion (8.4%) and electricity generation (7.8%). In Washington county, the major source is electricity generation (26.7%), followed by road dust (16.0%) and waste disposal (14.7%). In Westmoreland County, waste disposal (19.7%) is the major source of PM, followed closely by road dust (19.3%) and industrial processes (15.5%). Houston MSA also sees different source of PM by county. Harris County has miscellaneous (24.7%) as the major source, followed closely by industrial processes (23.7%), road dust (16.7%) and non road equipment (12.2%) while in Montgomery County, road

Table 2.13: Particulate matter source emission in percentages by counties where monitoring stations are located in Tampa, Houston and Pittsburgh MSAs, EPA 2005

Sources Type	Tampa MSA		Houston MSA		Pittsburgh MAS		
	Hills,	Pinellas	Harris,	Montg	Alleg,	Wash,	West
Elect. Generation	52.5	1.1	2.4	1.0	7.8	26.7	0.0
Non Road Equipment	13.5	17.6	12.2	3.5	15.0	6.3	9.4
Road Dust	9.9	9.3	16.7	66.9	2.4	16.0	19.3
Indust. Processes	9.8	16.7	23.7	2.6	39.4	11.2	15.5
On Road Vehicles	5.1	15.3	3.8	1.6	4.3	5.1	6.4
Res. Wood Combustion	2.8	9.9	2.5	1.3	8.4	4.9	8.0
Fossil Fuel Comb.	1.1	27.6	6.4	0.8	16.6	3.4	8.2
Waste Disposal	0.5	0.1	7.3	8.7	0.5	14.7	19.7
Fires	0.2	0.0	0.0	1.3	0.0	0.0	0.0
Solvent Use	0.2	0.0	0.3	0.0	0.3	0.0	0.5
Miscellaneous	4.3	2.5	24.7	12.3	5.3	11.6	12.9

dust (66.9%) is the leading PM source. The other sources are miscellaneous (12.3%), waste disposal (8.7%) and non road equipment (3.5%). As expected, industrial processes are not the major source of PM in Tampa MSA, while it play a major role in the other two MSAs. Also it is relevant to mention the important role play by non road equipments in all three regions. All PM sources and their contribution are summarized in Table 2.13.

2.5   Discussion

Air pollution in general and particulate matter in particular have been found to be associated with several diseases. A good understanding of their sources and variations will lead to major regulation by federal agency (EPA) and could help significantly reduce their health hazard. Several studies have found statistically significant association between PM and morbidity and mortality. It is suspected that PM and especially it metal components are responsible for several cancer disease and several studies are underway around the country to elucidate the association. Molinelli et al (2002) exposed a human airway epithelial cell line to aqueous extracts of PM col-

lected in the Utah Valley. In this study, part of the extract was treated to remove cations, including transition metals. Cells exposed to the untreated extract showed a concentration-dependent increase in the inflammatory mediator interleukin (IL)-8 compared to controls; cells incubated with the treated extract showed no such change. This suggests that the removal of metal cations attenuated cellular response to the aqueous extract, and supports a role for transition metal involvement in PM toxicity.

Based on their geographic location and more importantly the existing types of industries, three metropolitan statistical areas (MSA) in the US have been selected for our study: Pittsburgh MSA belongs to the Ohio valley region dominated by steel industries. Houston area in addition to being a metropolitan area where traffic is an important source of particulate matter, is in addition a region dominated by oil industries. Tampa Bay MSA, the third region is comparable to Houston as for its traffic and geographic location but yet very different in term of industries composition. Tampa Bay is extremely poor in industries with power plants for electricity generation as major industries in the region.

In order to carry out the spatio-temporal analysis of particulate matter metals, additive mixed models have been used because of the irregular pattern observed with the data which is unlikely to be captured by the usual parametric time series analysis models. Additive mixed models provide a good alternative in modeling correlated data such as the air pollution data. Furthermore, its flexibility allows the combination of both the parametric and non parametric components during data analysis. These models also provide a good possibility of handling correlation as various correlation structure found in longitudinal, hierarchical (multilevel) and spatial data could be incorporated into the models.

It is widely believe that in the natural environment, we are exposed to several chemicals at the same time, which constitute a mixture. To date we lack clear statistical methods of how to define mixtures. Little is known about chemical mixtures, their interaction and how they affect human health. In this study we investigated

mixture of metals derived based on mathematical concept of factor analysis.

Results of the analysis show difference in variation depending on metal types and regions with respect to trend, seasonality and weather conditions. Aluminum, calcium, sodium, iron, chromium, lead, manganese, magnesium and zinc show different trend, seasonality, temperature and pressure effect depending on regions. Tampa, Houston and Pittsburgh are three different regions as for their industries composition and geographic locations. Houston area is reach in oil industries, Pittsburgh is reach is still industries while Tampa dominant industries are power plant for electricity generation. Aluminum, iron and magnesium show no trend in Tampa. Chromium, lead, magnesium, zinc show no seasonality in both Tampa and Pittsburgh. This is an indication that they are not driven by seasonal sources. Temperature has no effect on chromium in Tampa and Houston, no effect on zinc in Pittsburgh. Pressure has no effect on magnesium in all three regions, no effect on aluminum in Houston and Pittsburgh, no effect on sodium in Pittsburgh, manganese in Houston and both chromium and lead in Tampa. These differences are explained by the differences in emission sources. Overall there is an improvement in air quality in all three regions. Following our analysis we may conclude that steel industries may not be a significant source of aluminum. High demand of electricity in Tampa and Houston in the summer could explain the seasonality of aluminum. Oil industries maybe a good source of particulate matter calcium and magnesium. Steel industries might not be a significant source of chromium and both oil and steel industries are significant sources of iron. Lead is decreasing everywhere but we still see a heavy presence in steel industries area. This is probably due to it usage by those industries.

3 Exposure to $PM_{2.5}$ speciation chemicals during pregnancy and Risk of Low Birth Weight

3.1 Introduction

Low Birth weight (LBW) or infants weighing less than $< 2500g$ is one of the major health issues in public health. Epidemiologic studies commencing in the 1990s to date have shown that exposure to ambient air pollution during the gestational or prenatal period could intensify the risk of low birth weight (LBW), small-for-gestational age (SGA) and preterm infants (Hwang et al. 2011, Lacasana et al. 2005, Maisonet et al. 2004, Ritz et al. 1999, Ritz et al. 2007). Studies done in different geographic regions have reported associations between air pollution and birth outcomes such as LBW, SGA and preterm delivery and increased infant morbidity and mortality (Rogers et al. 2006, Sram et al. 2005). Exposure to higher concentrations of carbon monoxise (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), total suspended particles (TSP) (Ha et al. 2001) and PM_{10} (Le et al. 2010) during the first trimester to mid pregnancy periods were associated with an increased risk of LBW. Several $PM_{2.5}$ chemicals such as aluminum, elemental carbon, nickel and titanium were found to be associated with LBW (Ebisu and Bell, 2012). Darrow et al. (2011) found that exposure to various concentrations of air pollutants in the later stages of pregnancy causes slight decreases in the birth weights of full term infants. There has been a strong association between PM and its subsequent effects on LBW and preterm birth. However, there is yet to be an agreement on the causative pollutants (Wilhelm et al. 2011). The

pathophysiological mechanisms that may contribute to effects of air pollution on birth outcomes remain uncertain even though various hypotheses exist. Particulate matter of aero-dynamic diameter less than 2.5 micrometers is a complex mixture of several chemicals, including metals of varying toxicity to humans. This requires relating the level of exposure to the particular chemical characteristics of $PM_{2.5}$ to individual health outcomes in the same locale, to identify which components are hazardous and which are not. Several epidemiologic studies have identified metals as PM chemicals associated with birth outcomes (Ebisu and Bell, 2012, Ibrahimou et al. 2014) and additionally it is believe that our environmental exposure is in form of mixtures. Our study examines the connection between level of exposure to $PM_{2.5}$ Mixture speciation metals during pregnancy and the risk of having LBW in offspring, by relating individual exposure to individual maternal outcome for each pregnant woman in our study.

3.2   Methods

Our data sources include the linked de-identified Hospital Inpatient Discharge (HID) data and the Florida vital statistics birth data file for Hillsborough and Pinellas Counties for the years 2004 to 2007, together with the Air Quality System (AQS) speciation data collected by the USEPA covering the same pregnancy period. The HID data were obtained from the Florida Agency for Health Care Administration (AHCA), and vital records were acquired from the Florida Department of Health (FDOH). The birth data contained birth-related variables together with antenatal information and any medical or labor complications experienced by the mother. Each observation in the birth data contained variables for the birth location by county, maternal characteristics (marital status, education, age, race, and maternal smoking during pregnancy), infant sex, gestational age in weeks, pre-pregnancy body mass index (BMI), and prenatal care. Our main outcome of interest was birth weight. The study population

included all singleton live births in Hillsborough and Pinellas counties, Florida. AQS speciation data covering the period 2003-2007, which include 35 different metals, was used to estimate the average daily exposure to speciation chemicals for each woman. Estimation of exposure assessment was made based on $PM_{2.5}$ speciation chemical concentrations collected from three stationary monitoring stations located in Hillsborough and Pinellas counties. Two exposure metrics were used. First, we allotted each mother's residential zip code to one of the three monitoring sites based on the site closest to her residential area. We relied on estimations of exposure assessment by the use of distances from residential areas to the monitor sites within an area, since assessing exposure per pollutant for each individual is difficult at the population level (Sagiv et al., 2005; Salihu et al., 2011). The distance between maternal residence at delivery and the three monitoring sites was calculated using the corresponding longitudes and latitudes (Parker et al., 2005). Using the closest station, residential average daily concentrations were calculated for each mother. Based on the gestational age and the date of birth, we generated average exposure estimates for the first trimester of pregnancy and the entire pregnancy period. Second we use the following personal exposure model to estimate mother exposure

$$E_{inh} = \frac{[C_{out}^d T_{out} V_{out} + C_{in}^d T_{in} V_{in} + C_{in}^d T_{sl} V_{sl}] \times R}{W \times (T_{out} + T_{in} + T_{sl})} \qquad (3.2.1)$$

where

- T_{out}, T_{in}, T_{sl} are durations in which the mother stayed outdoors, indoors (awake) and indoor (sleeping)

- V_{out}, V_{in}, V_{sl} are the corresponding ventilation rates

- C_{out}^d and C_{in}^d are zip code level daily average concentration outside and inside

- W weight of the mother

- R the particulate matter absorption rate

Average estimates of all the parameters in (3.2.1) are obtained from the Consolidated Human Activity Database (CHAD). For an adult woman for example, CHAD estimates national daily outdoor time to be approximately 1.8 hours (7.6%). Because we lack the indoor concentration exposure estimates, we assume that the indoor concentrations are a fraction of the outdoor concentrations, that $C_{in}^d = \alpha C_{out}^d$ where $0 < \alpha < 1$. We assign to each mother daily averages of $PM_{2.5}$ speciation metal mixtures for both the first trimester of pregnancy and the entire pregnancy period.

We fit a logistic regression model to the combined data to estimates the adjusted odds of low birth weight after exposure to $PM_{2.5}$ speciation metals during pregnancy.

3.3   Results

Metal mixtures were fitted one at the time to the model after controlling for known risk factors (giving in Table 3.1) such as preeclamsia and tobacco use. The results of the overall association between mixtures (Table 3.2) and the risk of low birth weight for exposure during the first trimester and the entire pregnancy period are presented in Table 3.3.

It shows that Mixture 6 (OR=5.08, p=0.006 and OR=3.22, p=0.001) and Mixture 7 (OR=2.04, p=0.012 and OR=4.69, p=0.003) are significantly associated with higher odds of having low birth weight babies when exposure happen both during the first trimester and the entire pregnancy period. We note that mixture 7 contains all the metals that are present in mixture 6. When exposure happen only during the first trimester, we see that mixture 1 (OR=3.51, p=0.001) and Mixture 8 (3.45, p=0.001) are associated with higher odds of having a low birth weight infant. No association is found when exposure happen during the entire pregnancy period (OR=2.99, p=0.095 and 2.63, p=0.156) respectively. In contrast mixture 4 (OR=3.17, p=0.043) and Mixture 9 (OR=3.07, p= 0.051) are associated with higher odds of having a low birth weight baby for exposure during the entire pregnancy but no association is found for

Table 3.1: Adjusted Odds Ratios and Confidence Interval for Low birth weight

Outcome Variables	OR(95% CI)
Education	0.750 (0.689-0.817)
Married	0.759 (0.705-0.812)
Pre-pregnancy BMI	0.958 (0.952-0.964)
Male	0.668 (0.625-0.714)
Gestation in weeks	0.445 (0.435-0.453)
Tobacco	1.941(1.749-2.154)
Abuse	1.709(1.202- 2.431)
Previa	1.667(1.251-2.223)
Abruption	2.547(2.028-3.199)
Preeclampsia	2.980(2.663-3.338)
Eclampsia	3.777(1.703-8.373)
G.Hypertention	1.642(1.442-1.870)
C.Hypertention	1.399(1.146-1.708)
Gestational Diabetes	0.734(0.630-0.854)
Diabetes Mellitus	0.751(0.550-1.026)
Infarction	2.138(1.474-3.100)
Black	1.887(1.740-2.046)
Anemia	0.794(0.709-0.889)

the first trimester exposure (OR=1.04, p=0.885 and OR=1.03, p=0.908) respectively. We also note that mixture 8 contains all the metals of mixture 1. None of mixtures 4 and 9 contains all the chemicals of the other but they share in common calcium, magnesium and titanium. Mixtures 2, 3, 5 and 10 are not associated with any risk of low birth weight both for the first trimester or the entire pregnancy period. We note that mixture 3 formed of tantalum and iridium is contains in mixture 10 and mixtures 2 and 5 share several metals together.

3.4   Discussion

Low birth weight is a serious health issue that could lead to severe morbidity and mortality. Several studies have shown association between $PM_{2.5}$ total mass and low birth weight. As PM is a complex mixture of several chemicals and not all of them being toxic to human health, it is important to separate between harmful and not

70

Table 3.2: $PM_{2.5}$ speciation metals mixtures

Metals/Mixtures	1	2	3	4	5	6	7	8	9	10
Aluminum		+		+			+	+		+
Barium	+				+			+	+	
Cadmium										+
Calcium		+		+	+	+	+		+	
Chromium					+		+			
Cobalt		+			+		+		+	
Copper	+				+		+	+		+
Cesium									+	
Gallium							+			+
Iron		+		+	+	+	+			+
Hafnium									+	+
Lead					+		+	+		
Indium										
Manganese		+			+	+	+		+	
Iridium			+						+	+
Molybdenum					+		+		+	+
Nickel					+	+	+			+
Magnesium	+			+	+			+	+	
Mercury										+
Gold									+	+
Lanthanum									+	
Niobium									+	
Tin					+				+	+
Titanium		+		+	+		+		+	
Scandium										
Vanadium		+			+		+	+		
Silver									+	
Zinc					+	+	+			
Strontium	+				+			+		
Tantalum			+						+	+
Rubidium										
Potassium	+				+		+	+		
Yttrium									+	
Sodium					+		+		+	+
Zirconium									+	

Table 3.3: Odds ratios and p-value of low birth weight for selected mixtures of metals for the entire pregnancy and the first trimester pregnancy.

Mixtures	Entire Pregnancy		First Trimester	
	OR	P-value	OR	P-value
Mixture 1	2.99	0.095	**3.51**	**0.001**
Mixture 2	2.35	0.083	1.04	0.866
Mixture 3	4.60	0.099	1.01	0.980
Mixture 4	**3.17**	**0.043**	1.04	0.885
Mixture 5	2.35	0.085	1.04	0.865
Mixture 6	**5.08**	**0.006**	**3.22**	**0.001**
Mixture 7	**4.69**	**0.003**	**2.04**	**0.012**
Mixture 8	2.63	0.156	**3.45**	**0.001**
Mixture 9	**3.07**	**0.051**	1.03	0.908
Mixture 10	5.25	0.065	1.06	0.891

harmful chemicals. Some metals such as lead, nickel and manganese for example are known to be harmful and because our exposure always happen in mixtures, we attempt to find out mixtures of metals that could be associated with low birth weight. Our findings show that most of the mixtures that are associated with low birth weight contains calcium, iron, aluminum, copper, manganese, magnesium and titanium.

4 Bayesian Factor Analysis for Temporally Correlated $PM_{2.5}$ Speciation Mixture Data

4.1 Introduction

Air pollution in general and particulate matter of aerodynamic less than 2.5 micrometer of diameter ($PM_{2.5}$) in particular had been found to be harmful to human health (Glinianaia et al., 2004, Donaldson and MacNee W., 2001, Gilboa et al. 2005, Dadvand et al. 2013). $PM_{2.5}$ is a complex mixture of extremely small particles and liquid droplets. They comes from many different sources including stationary sources such as factories, power plants, dry cleaners, mobile sources such as cars, buses, trucks, and trains and naturally occuring sources such as windblown dust, wildfires, and volcanic eruptions and even indoor combustion of cooking and heating. Particle pollution is made up of a number of components, including metals, acids (e.g. nitrates and sulfates), organic chemicals, and soil or dust particles. These different components are measured in some select monitoring stations around the country which formed the particulate matter speciation data collection network.

Given the complex composition of pollutants such as $PM_{2.5}$ in the environment, it is believed that our exposure is not limited to one chemical at the time but possibly to infinitely many possible mixtures with varying composition and mix ratio. These mixtures could be formed because they are emitted at the same time from the same source origin. It is suspected that these chemicals can be characterized by certain mixtures corresponding to emission source such as power plants, fixed or moving

73

vehicles, or ecological and environmental conditions. As a result, human are more susceptible to exposure to these mixtures as characterized by these latent factors. One statistical tool that could be used to characterize mixtures in the environment, is factor analysis.

Factor analysis is generally defined as statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called factors. It is performed by examining the pattern of correlations (or covariances) between the observed measures. Measures that are highly correlated (either positively or negatively) are likely influenced by the same factors and can be therefore grouped together under that factor(DeCoster, 1998).

Most proposals in the factor analysis literature assume that the data represent random, independent samples from a multivariate distribution (lawley, 1940). This cannot apply necessarily for all type of multivariate data, such as time series data where observations appear in certain order that cannot be changed without fundamentally changing the outcome. Air pollution data is an example of one such data.

The US EPA particulate matter speciation data are large datasets containing more than 100 different chemicals measured daily or measured every three to six days from the environment with different source origins. Because concentrations are measured daily (or every 3 to 6 days), PM speciation data is a typical example of a temporally correlated data which are likely to be correlated as compared to independent data for which the traditional factor analysis were developed.

To the best of our knowledge factor analysis is so far only developed under some conditions such as stationarity for temporally correlated (time series) data as an extension of the traditional factor analysis which was developed for cross-sectional data where the assumptions are reasonable. Most notably the traditional factor analysis assumes observations are independent and identically distributed (i.i.d.). Pena and Box extended the traditional factor analysis to time series data where they discussed methods of identifying a simplifying structure in time series under the assumption of

stationarity (Pena & Box, 1987). Gilbert, in his turn developed a theory where there is no need of stationarity condition as long as difference data satisfied some weak stationarity condition that he called weak boundedness condition (Gilbert, 2005). Our objective in this work are two-folds. First, to extend the traditional factor analysis for independent data to (spatio-)temporally correlated air pollution $PM_{2.5}$ chemicals speciation data via Bayesian statistics with no stationarity assumption. Second, making use of the source apportionment models and factor loadings obtained from the previous analysis, to convert latent factors into mixtures. This is done by utilizing loadings to develop mixture coefficients. This is the first time source apportionment models and factor analysis are combined to define chemical mixtures. The Bayesian approach allows the introduction of a temporal factor model within the covariance matrix by using Kronecker product (Rowe 1998). First we begin by introducing the factor analysis model.

4.2   Factor Analysis

Suppose there are p continuous variables (x) that are manifestation of m latent constructs or factors (f); f are difficult to observe but x are observable; $m << p$. Let $(x_1, \cdots, x_i, \cdots x_N)$ $(i = 1, \cdots, N)$ be a sample of x. A factor analysis model is given by

$$
\begin{array}{ccccccc}
(x_i | \mu, \Lambda, f_i, m) & = & \mu & + & \Lambda & f_i & + & \epsilon_i \\
(p \times 1) & & (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1)
\end{array}
\tag{4.2.1}
$$

where

μ is a p - dimension vector of unobserved population mean,

f_i is an m - dimension vector of unobservable "common" factor scores at the ith occasion,

Λ is a $p \times m$ matrix of unobserved constants called the factor loadings; $\Lambda = (\lambda_1', \cdots, \lambda_p')'$;

and

ϵ_i is a p - dimension vector of errors or disturbance for the ith measurement, and it is independent of f_i.

Factor analysis is commonly used for dimension reduction of data when we assume the p - vector x are surrogates of f, or f represent certain common characteristics among x. When the assumption is reasonable, the factors f explain the underlying covariation seen in the observed values of x. Thus the goal of factor analysis is to determine and quantify the extent to which a set of m distinct underlying factors f can describe the covariation among x. Within the context of $PM_{2.5}$ x are individual metal constituents, and the factors f could be distinct emission sources such as traffic, industries, power plants etc, because each emission source tends to release distinct mixtures of pollutants at a given point in time.

Without loss of generality we assume f_i follows a distribution with mean 0 and variance matrix R. The covariance between f_i and the observable data x_i is given by

$$
\begin{aligned}
cov(x_i, f_i) &= E(x_i f_i') - E(x_i)E(f_i') \\
&= \Lambda E(f_i f_i') \\
&= \Lambda R.
\end{aligned}
$$

Ortho-normalization of f makes $R = I_m$, and the factor loading matrix Λ can be interpreted as a matrix of covariances (correlations) between the variables x and factors f.

After stacking the observation vectors $x_1, \cdots, x_N$ into a single $Np \times 1$ vector x^*,

model (4.2.1) can be written as

$$
\begin{array}{cccccccc}
(x^*|\mu, \Lambda, f^*, m) & = & \mathbf{1} \otimes \mu & + & (I_N \otimes \Lambda) & f^* & + & \epsilon \\
(Np \times 1) & & (Np \times 1) & & (N \times N \otimes pm) & (Nm \times 1) & & (Np \times 1)
\end{array}
$$
$$(4.2.2)$$

where $\otimes$ is the Kronecker product, and

$x^* \equiv (x_1', \cdots, x_N')'$;

$\mathbf{1} = N$ - dimension 1-vector;

$f^* \equiv (f_1', \cdots, f_N')'$ an Nm vector of unobserved "common" factor scores;

and $\epsilon \equiv (\epsilon_1', \cdots, \epsilon_N')'$ is an Np vector of "specific" error.

Assuming the data can be centered around its mean (or detrended, after removing the trend from the data by applying loess smoothing for example), we can remove the mean parameter μ from the model:

$$
\begin{array}{cccccc}
(x^*|\Lambda, f^*, m) & = & (I_N \otimes \Lambda) & f^* & + & \epsilon. \\
(Np \times 1) & & (N \times N \otimes pm) & (Nm \times 1) & & (Np \times 1)
\end{array}
$$
$$(4.2.3)$$

In the classic setting of factor analysis, we assume the sample $x_1, \ldots, x_N$ are independent conditioning on the unobservable factor scores. That is under normality

$$
\epsilon = [(x^*|\Lambda, f^*, m) - (I_N \otimes \Lambda)f^*] \sim N(0, I_N \otimes \Psi).
$$

where $\Psi > 0$ is the variance-covariance of ϵ_i.

The likelihood function given the latent factors is of the form

$$
p(x^*|f^*, \Lambda, \Psi, m) \propto |I_N \otimes \Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}[x^* - (I_N \otimes \Lambda)f^*]'(I_N \otimes \Psi)^{-1}[x^* - (I_N \otimes \Lambda)f^*]}.
$$
$$(4.2.4)$$

In matrix notation $X \equiv (x_1, \cdots, x_N)$ and $F \equiv (f_1, \cdots, f_N)$, the likelihood can

also be written as

$$p(X|F, \Lambda, \Psi, m) \propto |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr \Psi^{-1} (X - \Lambda' F)' (X - \Lambda' F)}, \quad \Psi > 0 \qquad (4.2.5)$$

We will use this matrix notation in the remainder of this paper because it makes computation easier.

4.3 Bayesian Factor Analysis

While likelihood-based inference has been the main stream for factor analysis models (Lawley, 1940, Pena & Box, 1987, Gilbert, 2005), Bayesian inference has also been developed (Press and Shigemasu, 1989). Following Press and Shigemasu (1989), we use generalized natural conjugate families of prior distributions for the parameters F, Λ, Ψ. The number of latent factors m is an important parameter especially in exploratory factor analysis where m cannot be predetermined. We will consider m as fixed in the present paper. As in Rowe (1998) the factor scores F are assumed to be independent of the factor loadings Λ and the disturbance covariance matrix Ψ; Λ depends on Ψ. This leads to the following joint prior distribution for F, Λ, and Ψ:

$$p(F, \Lambda, \Psi | m) = p(\Lambda | \Psi, m) p(\Psi) p(F | m) \qquad (4.3.1)$$

where we may assume

$$p(\Lambda | \Psi, m) \quad \propto \quad |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1} (\Lambda - \Lambda_0) H (\Lambda - \Lambda_0)'}, \qquad (4.3.2)$$

$$p(\Psi) \quad \propto \quad |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B}, \quad \nu > 2p \qquad (4.3.3)$$

$$p(F | m) \quad \propto \quad e^{-\frac{1}{2} \mathrm{tr} F' F} \qquad (4.3.4)$$

with $H(m \times m) > 0$, $B(p \times p) > 0$, and $\Lambda_0(p \times m)$ are hyperparameters associated with the prior distribution, and need to be estimated. In particular, B can be

made a diagonal matrix. Conditioning on Ψ, the elements of Λ are jointly normally distributed; Ψ^{-1} follows a Wishart distribution, with the hyperparameters (ν, B) to be estimated; The factor scores $f_i s$ are independent and normally distributed under ortho-normalization.

Under the Bayes rule, the posterior distribution of the factor scores F given the factor loadings Λ, the disturbance covariance matrix Ψ and the data are normally distributed:

$$p(F|\Lambda, \Psi, X, m) \propto e^{-\frac{1}{2}tr((F-\tilde{F})'(I_m+\Lambda'\Psi^{-1}\Lambda)(F-\tilde{F}))} \tag{4.3.5}$$

where $\tilde{F}' = X'\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$. The conditional posterior distribution of the factor loadings Λ, given the factor scores F and the disturbance covariance matrix Ψ is also multivariate normal:

$$p(\Lambda|F, \Psi, X, m) \propto e^{-\frac{1}{2}tr(\Psi^{-1}(\Lambda-\tilde{\Lambda})(H+FF')(\Lambda-\tilde{\Lambda})')} \tag{4.3.6}$$

where $\tilde{\Lambda} = (XF'+\Lambda_0 H)(H+FF')^{-1}$. Given the factor scores, the factor loadings, and the data, the conditional posterior distribution of the disturbance covariance matrix is that of an inverted Wishart:

$$p(\Psi|F, \Lambda, X, m) \quad \propto \quad |\Psi|^{-\frac{N+m+\nu}{2}} e^{-\frac{1}{2}\text{tr}(\Psi^{-1}U)} \tag{4.3.7}$$

where $U = (X - \Lambda\tilde{F})(X - \Lambda\tilde{F})' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

Note that the preceding Bayes FA model also assumes independent observations. However, serial $PM_{2.5}$ data violate the data independence assumption. Ignoring the serial correlation can result in misleading results from the factor analysis models. Therefore as in Rowe (1998) we propose to incorporate serial correlation into the Bayesian factor analysis model for the analysis of temporal $PM_{2.5}$ data.

4.4 Bayesian Models for Temporally Correlated Data

4.4.1 Model

We start with model as in (4.2.2), and considered centered data and the residuals of the detrended data after applying loess smoothing to remove the time trend. Rowe (1998) extended the work of Press and Shigemasu (1989) of Bayes factor analysis to allow for correlated observations. A key innovation in his work was the introduction of variance components of within- and between-observation covariances. Our approach here follows Rowe's idea of separating the between- and within-measurement covariance seen in the $PM_{2.5}$ speciation data.

Instead of using the error covariance matrix $I_N \otimes \Psi$ as in (4.2.2), consider the more general form

$$\epsilon \sim N(0, \Omega),$$

where we can let $\Omega = \Phi \otimes \Psi$, $\Phi > 0$, $\Psi > 0$, where Ψ remains the within-observation covariance of x_i, and Φ is covariance matrix between x_i and x_j. This special form is possible only if we assume a constant correlation between any pair of elements one from each vector. Specifically Ω can be expressed as

$$\mathbf{\Omega} = \Phi \otimes \Psi \equiv \begin{pmatrix} \phi_{11}\Psi & \phi_{12}\Psi & \cdots & \phi_{1N}\Psi \\ & \phi_{22}\Psi & & \\ & & \ddots & \vdots \\ & & & \phi_{NN}\Psi \end{pmatrix} \tag{4.4.1}$$

from which we see that

$$var(x_i|\Phi, \Psi, m, f, \lambda) = \phi_{ii}\Psi = \Psi$$

and the covariance between any two observation vectors is

$$cov(x_i, x_j | \Phi, \Psi, m, f, \lambda) = \phi_{ij} \Psi,$$

and ϕ_{ij} is the common correlation between any pair of elements, one from x_i and one from x_j. Letting Φ be the identity matrix result in the classic case of independent data. Letting $\phi_{ij} = \phi$ leads to a constant correlation between any pair of vectors. Other forms such as auto-regressive correlation $\phi_{ij} = \phi^{|j-i|}$ may also be considered.

Under this serial correlation structure, $\Omega = \Phi \otimes \Psi$, the likelihood function becomes

$$p(X|F, \Lambda, \Phi, \Psi, m) \propto |\Phi|^{-\frac{p}{2}} |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr(\Psi^{-1}(X-\Lambda'F)'\Phi^{-1}(X-\Lambda'F))}. \qquad (4.4.2)$$

Natural conjugate families of prior distributions for Φ, Ψ, Λ and F are developed similarly:

$$p(\Phi, \Psi, F, \Lambda|m) = p(\Psi)p(\Phi)p(F|\Phi, m)p(\Lambda|\Psi, m),$$

where

$$p(\Lambda|\Psi, m) \quad \propto \quad |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} \text{tr}(\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)')}, \qquad (4.4.3)$$

$$p(\Psi) \quad \propto \quad |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr\Psi^{-1}B} \quad \nu > 2p, \qquad (4.4.4)$$

$$p(F|\Phi, m) \quad \propto \quad |\Phi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr\Phi^{-1}F'F}. \qquad (4.4.5)$$

where $H(m \times m) > 0$, and $B(p \times p) > 0$ are as defined before. Note that the prior distributions for Λ and Ψ remain the same as in (4.3.2) and (4.3.3). This indicates that the dependance among observations x_i does not alter prior knowledge about factor loadings and within-observation covariance. What's new is that the prior distribution on F has changed from (4.3.4) to the one that depends on the between-observation covariances Φ. This reflects the belief that the between-observation dependence is in part resulted from the dependence of the latent factor scores F.

We have not yet discussed the prior distribution for Φ, but will discuss it in the following section. Postponing the discussion is because we may be able to utilize highly structural serial correlation to simplify subsequent computation.

The joint posterior distribution for the unknown parameters of interest is conceptually simple to obtain as given below:

$$p(F, \Lambda, \Psi, \Phi | X, m) \propto p(\Phi)|\Phi|^{-\frac{p+m}{2}}|\Psi|^{-\frac{N+m+\nu}{2}}|H|^{\frac{p}{2}}e^{-\frac{1}{2}\text{tr}(\Psi^{-1}U_1)}e^{-\frac{1}{2}\text{tr}(\Phi^{-1}F'F)} \quad (4.4.6)$$

where

$$U_1 = (X - \Lambda F)\Phi^{-1}(X - \Lambda F)' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B.$$

The conditional posterior density of the factor loadings given the factor scores, the disturbance covariance matrix, and the data is again normally distributed and given by

$$p(\Lambda | F, \Psi, \Phi, X, m) \propto e^{-\frac{1}{2}tr(\Psi^{-1}(\Lambda-\tilde{\Lambda})'(H+F\Phi^{-1}F')(\Lambda-\tilde{\Lambda}))} \quad (4.4.7)$$

where $\tilde{\Lambda} = (X\Phi^{-1}F' + \Lambda_0 H)(H + F\Phi^{-1}F')^{-1}$. The difference between (4.4.7) and (4.3.6) reveals the effect of incorporating the dependance between observations into the analysis on the factor loadings. The introduction of the temporal correlation does affect the posterior distribution of factor loadings as Φ is incorporated in both the covariance matrix and its central tendency.

The conditional posterior density of the disturbance covariance matrix given the factor scores, the factor loadings, and the data is an inverted Wishart density

$$p(\Psi | F, \Lambda, \Phi, X, m) \propto |\Psi|^{-\frac{N+m+\nu}{2}}e^{-\frac{1}{2}tr(\Psi^{-1}U_1)} \quad (4.4.8)$$

where U_1 is given above.

Finally, the conditional posterior distribution for the factor scores given the cor-

relation matrix Φ, the disturbance covariance matrix Ψ, the number of factors, the factor loadings and the data is normally distributed as given below

$$p(F|\Lambda, \Psi, \Phi, X, m) \quad \propto \quad e^{-\frac{1}{2}\text{tr}(\Phi^{-1}(F-\tilde{F})'(I_m+\Lambda'\Psi^{-1}\Lambda)(F-\tilde{F}))}, \qquad (4.4.9)$$

where $\tilde{F}' = X'\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$ remains the same as in (4.3.5), but the density function differs from (4.3.5) with the leading Φ^{-1} in the trace function.

We now consider the posterior distributions for Φ. Common serial correlation structures may be utilized to capture the between-measurement correlation. For instance, the first order autoregressive correlation

$$\Phi = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{N-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{N-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \rho^{N-4} & \cdots & 1 \end{pmatrix} \qquad (4.4.10)$$

is relatively simple to use, which is somewhat implicated from our preliminary data analysis. In this case Φ reduces to a single correlation parameter ρ that is to be computed. A natural choice of prior distribution is a beta distribution (Rowe, 1998):

$$p(\rho) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\rho^{\alpha-1}(1 - \rho)^{\beta-1}, \qquad (4.4.11)$$

where $\Gamma(\cdot)$ is the gamma function and $\alpha, \beta > 0$ and $0 \leq \rho \leq 1$.

The conditional posterior density for ρ is then

$$p(\rho|\Psi, m, F, \Lambda, X) \propto |\Psi|^{-\frac{(p+m)}{2}}\rho^{\alpha-1}(1 - \rho)^{\beta-1}e^{-tr\Phi^{-1}C} \qquad (4.4.12)$$

where

$$C = (X - \Lambda F)'\Psi^{-1}(X - \Lambda F) + F'F.$$

Because

$$|\Phi| = (1 - \rho^2)^{N-1} \tag{4.4.13}$$

and the fact

$$\mathbf{\Phi}^{-1} = \begin{pmatrix} 1 & -\rho & & & & 0 \\ -\rho & (1+\rho^2) & -\rho & & & \\ & \ddots & \ddots & \ddots & & \\ & & & (1+\rho^2) & -\rho \\ 0 & & & -\rho & 1 \end{pmatrix} \tag{4.4.14}$$

equation (4.4.12) becomes

$$p(\rho|\Psi, m, F, \Lambda, X) \propto p(\rho)(1 - \rho^2)^{-\frac{(N-1)(p+m)}{2}} e^{-\frac{(k_1 - k_3\rho + k_2\rho^2)}{2(1-\rho^2)}} \tag{4.4.15}$$

where $k_1 = tr(C)$, $k_2 = tr(M_2 C)$ and $K_3 = tr(M_3 C)$ with M_2 and M_3 given as follows

$$\mathbf{M_2} = \begin{pmatrix} 0 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & 0 \end{pmatrix}$$

and

$$\mathbf{M_3} = \begin{pmatrix} 0 & 1 & & & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 0 & 1 \\ 0 & & & 1 & 0 \end{pmatrix}$$

4.4.2   Computation

Hyperparameters

To define the hyperparameters, H, B, Λ_0 as well as α and β from equations (4.3.2)-(4.3.4) and (4.4.11), we follow the process described in Rowe (2000, 2002, 2003ab). H is by definition a positive definite matrix, it assumes the form $H = n_H I_m$ for simplicity and ease of computation, where n_H is assessed below using the method attributed to Hayashi (Hayashi, 1997). The maximum likelihood (regular factor analysis) estimates $\hat{\Lambda}$ and $\hat{\Psi}$ of Λ and Ψ and their variance $Var(\hat{\Lambda})$ and $Var(\hat{\Psi})$ will be used. We will choose the values of hyperparameters by equalling the moments (means and variances) of the prior distributions with their corresponding estimates obtained from the regular factor analysis using training data, which is the first half of the data.

Estimation of Λ_0 and H

The choice of Λ_0 is straightforward and is chosen to be $\hat{\Lambda}$. Letting $\lambda = vec(\Lambda) = (\lambda_{11}, \lambda_{21}, \cdots, \lambda_{p1}, \cdots, \lambda_{1m}, \lambda_{2m}, \cdots, \lambda_{pm})$ we have

$$Var(\lambda|\Psi) = \Psi \otimes H^{-1}$$

and

$$Var(\lambda) = E(\Psi) \otimes H^{-1} \tag{4.4.16}$$

In equation (4.4.16), we replace $E(\Psi)$ by $\hat{\Psi}$ to get

$$Var(\lambda) = \hat{\Psi} \otimes H^{-1}, \tag{4.4.17}$$

and equate to Hayashi's approximation

$$Var(\lambda) = c^{-1}Var(\hat{\lambda}) \tag{4.4.18}$$

where the constant c is to be estimated below. From (4.4.17) and (4.4.18), we obtain

$$c^{-1}Var(\hat{\lambda}) = \hat{\Psi} \otimes H^{-1}. \tag{4.4.19}$$

Let $H = (h_j)$ be a diagonal matrix. Applying trace to each non-zero block diagonal matrix involving h_j we obtain

$$tr(\hat{\Psi})h_j^{-1} = \sum_{i=1}^{p} c^{-1}Var(\hat{\lambda}_{ij}), \quad j = 1, \cdots, m \tag{4.4.20}$$

or

$$h_j = \frac{c \times tr(\hat{\Psi})}{\sum_{i=1}^{p} Var(\hat{\lambda}_{ij})}, \quad j = 1, \cdots, m \tag{4.4.21}$$

Following Hayashi (1997), using A optimality (use of trace operator)(Shah & Sinha, 1989), he showed that the constant c can be expressed as

$$c = \frac{n}{N} tr(\hat{F}'\hat{F})(tr(\hat{\Psi}))^{-1} \left(\sum_{j=1}^{m} (\sum_{i=1}^{p} Var(\hat{\lambda}_{ij}))^{-1} \right)^{-1}$$

where n and N are the training and the actual data sample size respectively. Now replacing c above, we get the estimator of the diagonal hyperparameter matrix $\hat{H}$ whose j^{th} diagonal element is

$$\hat{h}_j = (n) \left\{ \frac{1}{m} tr(\frac{\hat{F}'\hat{F}}{N}) \right\} \frac{(\frac{1}{p}\sum_{i=1}^{p} \hat{V}ar(\hat{\lambda}_{ij}))^{-1}}{\frac{1}{m}\sum_{j=1}^{m}(\frac{1}{p}\sum_{i=1}^{p} \hat{V}ar(\hat{\lambda}_{ij}))^{-1}} \quad j = 1, ..., m.$$

To eliminate the terms involving $Var(\lambda_{ij})$ in $\hat{h}_j$ we also assume that

$$\sum_{i=1}^{p} Var(\hat{\lambda}_{ij}) = \sum_{i=1}^{p} Var(\hat{\lambda}_{ik}), \quad j \neq k$$

along with the large sample approximation under the ortho-normal condition of the

factors,

$$\frac{FF'}{n} \approx I_m,$$

and $\hat{h}_j$ reduced simply to n, the sample size of the training data. As a result, we assess the hyperparameter $n_H = n$ as in Hayashi (Hayashi, 1997).

Estimation of B and ν

Following Rowe (Rowe, 2000) $B = b_0 I_p$. Therefore the mean of any diagonal element of the disturbance matrix Ψ_{ij} under the prior distribution given in (4.4.4) is

$$E(\Psi_{ii}) = \frac{b_0}{\nu - 2p - 2}, \quad i = 1, \cdots, p. \tag{4.4.22}$$

From the classical factor analysis model we have $\Sigma = \Lambda\Lambda' + \Psi$ where Σ is the covariance matrix for the observations. Substituting the training sample covariance matric $\hat{\Sigma}$ and the priori mean for the factor loadings into the above equation we have $\Psi_0 = \hat{\Sigma} - \Lambda_0\Lambda_0'$. Then we take the average of the diagonal elements

$$\frac{1}{p}tr(\Psi_0) = \frac{1}{p}tr(\hat{\Sigma} - \Lambda_0\Lambda_0') \tag{4.4.23}$$

as our prior mean for the diagonal element $E(\Psi_{ii})$ of the disturbance covariance matrix. Equaling (4.4.22) and (4.4.23) we have

$$\frac{b_0}{\nu - 2p - 2} = \frac{1}{p}tr(\hat{\Sigma} - \Lambda_0\Lambda_0')$$

thus $b_0 = \frac{\nu - 2p - 2}{p}tr(\hat{\Sigma} - \Lambda_0\Lambda_0')$ and $\hat{\Sigma}$ is the training data covariance matrix.

For the hyperparameter ν we follow Rowe (2003) using the method due to Hayashi (Hayashi 1997). Starting with the Bayes estimator for the disturbance covariance matrix

$$\hat{\Psi} = \frac{\hat{U}}{N + m + \nu - 2p - 2}$$

where

$$\hat{U} = (X - \hat{\Lambda}\hat{F})(X - \hat{\Lambda}\hat{F})' + (\hat{\Lambda} - \Lambda_0)'H(\hat{\Lambda} - \Lambda_0) + \hat{B}.$$

we can consider $\hat{\Psi}$ as a weighted average of the three terms in $\hat{U}$. The scalar values associated with the terms are n, m, and $\nu - 2p - 2$ respectively. Because we consider first and third terms as representing the current and training data, we equate $\nu - 2p - 2$ with n to obtain $\nu = n + 2p + 2$.

Estimation of α and β

For parameters α and β, if no priors belief is assumed, then an uninformative prior which correspond to $\alpha = \beta = 1$ could be chosen. If the structure is assumed to be the first order Markov as above then α and β are chosen such that the prior mean

$$\frac{\alpha}{\alpha + \beta}$$

is made equal to the estimated correlation value from the training data. We may also assess that by pure subjective prior belief of the correlation parameter (Rowe, 1998).

Gibbs sampling

For Gibbs estimation of the posteriors, we start with initial values for the parameters Λ, ρ, Ψ and F say $\overline{\Lambda}_0$, $\overline{\rho}_{(0)}$, $\overline{\Psi}_{(0)}$ and $\overline{F}_{(0)}$.

Then for a given m, we cycle through

$$
\begin{aligned}
\overline{\Lambda}_{(i+1)} &\equiv \text{ a random sample from } p(\Lambda | \overline{\rho}_{(i)}, \overline{F}_{(i)}, \overline{\Psi}_{(i)}, m, X) \\
\overline{\Psi}_{(i+1)} &\equiv \text{ a random sample from } p(\Psi | \overline{\rho}_{(i)}, \overline{F}_{(i)}, \overline{\Lambda}_{(i+1)}, m, X) \\
\overline{F}_{(i+1)} &\equiv \text{ a random sample from } p(F | \overline{\rho}_{(i)}, \overline{\Psi}_{(i+1)}, \overline{\Lambda}_{(i+1)}, m, X) \\
\overline{\rho}_{(i+1)} &\equiv \text{ a random sample from } p(\rho | \overline{\Psi}_{(i+1)}, \overline{F}_{(i+1)}, \overline{\Lambda}_{(i+1)}, m, X)
\end{aligned}
$$

to have a sequence

$$(\overline{\rho}_{(1)}, \overline{\Psi}_{(1)}, \overline{F}_{(1)}, \overline{\Lambda}_{(1)})$$

$$\vdots$$

$$(\overline{\rho}_{(s)}, \overline{\Psi}_{(s)}, \overline{F}_{(s)}, \overline{\Lambda}_{(s)})$$

$$(\overline{\rho}_{(s+1)}, \overline{\Psi}_{(s+1)}, \overline{F}_{(s+1)}, \overline{\Lambda}_{(s+1)})$$

$$\vdots$$

$$(\overline{\rho}_{(s+t)}, \overline{\Psi}_{(s+t)}, \overline{F}_{(s+t)}, \overline{\Lambda}_{(s+t)}).$$

During preliminary analysis, an autoregressive correlation of order 1 with $\rho = 0.05$ have been observed and this value serve as initial value ρ_0 for ρ. We choose the initial values Λ_0 and F_0 to be respectively the factor loadings and factor scores obtained from the regular factor analysis of the data. As discussed in subsection (4.4.2), the initial value for Ψ is chosen to be $\Psi_0 = \Sigma - \Lambda_0 \Lambda_0'$ where Σ is the data covariance matrix.

We run 100000 iteration with 50000 burn. The first s random samples called "burn" (here $s = 50000$) are discarded and the remaining t samples are used for the posterior estimates which will be given by their means as follow:

$$
\begin{aligned}
\overline{F} &= \tfrac{1}{t}\sum_{k=1}^{t} \overline{F}_{s+k} \\
\overline{\Lambda} &= \tfrac{1}{t}\sum_{k=1}^{t} \overline{\Lambda}_{s+k} \\
\overline{\Psi} &= \tfrac{1}{t}\sum_{k=1}^{t} \overline{\Psi}_{s+k} \\
\overline{\rho} &= \tfrac{1}{t}\sum_{k=1}^{t} \overline{\rho}_{s+k}.
\end{aligned}
$$

In the following section we consider an application to Air pollution data where we will discuss mixtures of air pollutants.

Table 4.1: Average daily concentration in nanogram per cubic meter (standard error), autocorrelation, percentage of undetectable values (USEPA, 1998), 75th and 95th percentiles of $PM_{2.5}$ speciation metals measured from February, 2000 to December, 2007 in Tampa MSA.

metal	mean (SE)	auto-corr	% undetect	75th percentile	90th percentile
Al	44.33(0.086)	0.141	35.52	31.30	95.16
Ba	32.84(0.028)	0.289	15.01	50.30	73.87
Cd	1.95(0.003)	-0.051	53.06	02.90	05.97
Ca	62.51(0.031)	0.182	00.21	76.50	105.92
Cr	1.52(0.003)	0.051	22.62	01.87	03.23
Co	0.15(0.000)	0.001	74.84	00.04	00.61
Cu	3.90(0.010)	0.052	11.62	03.85	08.69
Cs	5.14(0.008)	-0.101	51.79	07.91	17.52
Ga	0.84(0.001)	0.464	48.41	01.27	02.68
Fe	66.22(0.055)	0.241	00.00	71.30	123.00
Hf	3.38(0.004)	0.051	63.42	03.76	12.24
Pf	3.76(0.003)	0.027	15.43	04.83	06.75
In	2.54(0.003)	0.110	50.52	03.63	07.37
Mn	1.64(0.001)	0.157	20.93	02.44	03.72
Ir	0.67(0.001)	-0.001	42.70	02.74	05.11
Mo	1.15(0.001)	-0.024	55.18	01.74	03.77
Ni	4.83(0.017)	0.005	11.21	02.80	05.00
Mg	9.93(0.030)	-0.007	59.83	13.16	29.78
Hg	1.32(0.002)	0.142	41.22	02.23	04.08
Au	1.62(0.001)	0.073	42.07	02.53	04.77
La	11.02(0.012)	0.066	50.53	14.30	39.00
Nb	0.63(0.001)	0.167	58.56	00.93	02.33
Sn	11.19(0.007)	0.317	16.28	17.80	22.89
Ti	6.93(0.007)	0.217	03.59	07.53	13.33
Sc	0.26(0.000)	-0.037	72.73	00.11	01.08
V	2.79(0.002)	0.275	12.68	03.77	05.80
Ag	2.75(0.002)	0.015	36.15	04.55	07.60
Zn	4.71(0.005)	0.056	07.82	06.53	09.55
Sr	1.05(0.004)	-0.006	44.40	01.28	02.49
Ta	11.92(0.007)	0.345	22.83	20.30	30.98
Rb	0.34(0.001)	0.102	57.08	00.47	01.14
K	89.11(0.282)	0.010	00.00	80.81	121.24
Y	0.43(0.001)	-0.005	65.75	00.47	01.64
Na	124.00(0.113)	0.137	28.75	194.00	312.20
Zr	1.05(0.003)	-0.006	58.77	00.89	02.34

4.5 Applications

4.5.1 Data

$PM_{2.5}$ is a complex mixtures of extremely small particles and liquid droplets. It is made up of a number of different components, including metals, acids and organic chemicals which form the $PM_{2.5}$ chemicals speciation. $PM_{2.5}$ chemicals speciation data and other air quality data such as Carbon Monoxide (CO), Lead (pb), Nitogen Dioxide (NO_2), Ozone (O_3), Sulfur Dioxide (SO_2), Particulate Matter (PM_{10}, $PM_{2.5}$) are housed in the US EPA Air Quality System (AQS) data base. AQS is divided into several groups around the continental US called Metropolitan Statistical Area (MSA) with each MSA containing several stations. To protect people and the environment, the Clean Air Act, which was last amended in 1990, requires EPA to set National Ambient Air Quality Standards (NAAQS) for pollutants considered harmful to public health and the environment. EPA has set NAAQS for six principal pollutants, which are called "criteria" pollutants: CO, pb, NO_2, O_3, SO_2 and PM. Chemicals speciation data from Tampa MSA will be used in this analysis to identify metal mixtures. Unlike other air pollutants such as CO, pb and PM which are collected daily from the environment, chemicals speciation data, a large dataset containing more than 100 different chemicals including 35 metals are measured every three to six days from the environment with different source origins. This data is a typical example of a time series data which are likely to be temporally correlated as compared to independent data for which the traditional factor analysis were developed. There are three stations in Tampa MSA and data from one station will be used in this application. Table 4.1 presents average daily concentration, standard error in nanogram per cubic meter, autocorrelation of PM speciation metals from the selected station in Tampa MSA, the proportion of non detectable concentration values and the 75th and 90th percentiles of all speciation metals. Aluminum, calcium, iron, potassium and sodium are the most abundant metals, while cobalt, niobium and rubidium are the least abundant in

91

the area. We also note a strong autocorrelation with some metals such as: barium, gallium, iron, tin, titanium, vanadium and tantalum while the remaining metals show moderate autocorrelation.

4.5.2 Preliminary Analysis

Computations have been carried out using R statistical software version 3.1.0. We use data from one station in Tampa MSA for application. From Table 4.1, we note that 17 (48.57%) of metals have less than 40% of proportion of undetectable values. Correlation coefficients between $PM_{2.5}$ speciation metals shows a strong positive correlations between certain metals such as aluminum and iron, titanium, calcium or manganese. Calcium is highly correlated with iron. Copper is strongly correlated with lead, magnesium, strontium and potassum. Titanium is highly correlated with aluminum, iron, calcium and potassium.

Without taking into account the temporal correlation between observations, the traditional factor scores are given in Table 4.2. We see from Table 4.2 that factor 1 is dominated with barium, copper, lead, magnesium, titanium, vanadium, strontium and potassium. Factor 2 is mainly composed of aluminum, calcium, iron, manganese, and titanium. Factor 3 has tantalum, iridium, nickel and tin as dominant elements. Cadmium, hafnium, indium, titanium, tin and sodium are leading Factor 4. Not taking into account the serial correlation when one exist could lead to misleading factor estimates. As seen in Table 4.1, autocorrelation do exist among metals. Therefore there is a need to account for those correlations and the Bayesian correlated factor analysis will play an important role in that direction. Additionally there is a lack of consistent criteria of how to determine factors and its components especially in the context of mixture of chemicals. Defining mixtures by selecting metals with high loadings as in Table 4.5 may not account well for all chemicals contribution. Instead Factor analysis in conjunction with source apportionment to be introduced below will

be used.

4.5.3 Bayesian Correlated Factors

To account for correlation, we apply the Bayesian correlated factor analysis model developed above to the residuals data after removing the trend through smoothing and to the raw data (without smoothing). The results of the loadings after the smoothing are presented in Table 4.3 and the one for the raw data are in Table 4.4. No major difference in loadings was observed between the two loadings results. This means that, removing trend has no much effect on the factor loadings. But compared to Table 4.2 and after addition of the temporal correlation, we observe different results. Table 4.3 shows factor loadings together with their 95% credibility intervals after applying the Bayesian correlated factor theory on the residuals. From Table 4.3, choosing metals with loading of at least 0.10 in absolute value, we see that the first factor is mostly determined by aluminum, barium, cesium, iron, niobium, tin, titanium and tantalum. The second factor is constituted mostly by cadmium, calcium, copper, cesium, gallium, hafnium and strontium. The third factor is determined by nickel, gold, lead, iridium, molybdenum and aluminum, while the fourth factor is associated mostly with aluminum, cobalt, cesium, lead, molybdenum, mercury, scandium, zinc and zirconium. The difference observed between Tables 4.2 and 4.3 highlight the importance of appropriately identifying and taking into account the data correlation.

4.5.4 Source Apportionment

Determining mixture of chemicals is one of the main goal of this study. As we have said earlier, there is a lack of consistent criteria of how to determine factors and its components especially in the context of mixture of chemicals. Choosing chemicals based on hypothetical high loadings might not be appropriate (see Table 4.5). Instead we will use source apportionment methods in conjunction with factor analysis. We

93

Table 4.2: Regular Factor Loadings ignoring Serial Correlation

	Factor1	Factor2	Factor3	Factor4
aluminum	.14	.81	.12	-.22
barium	.84	-.00	.23	-.10
cadmium	-.00	.02	-.05	.37
calcium	-.06	.87	.08	.11
chromium	-.02	.10	.08	.21
cobalt	-.07	-.07	-.21	-.00
copper	.83	-.01	-.01	.23
cesium	.15	-.04	.09	-.07
gallium	-.06	-.02	.26	-.22
iron	.03	.99	.12	-.03
hafnium	.02	-.07	.02	.38
lead	.71	.01	.12	.13
indium	-.00	-.02	-.09	.36
manganese	.24	.52	-.00	-.14
iridium	-.02	.01	.47	-.11
molybdenum	-.00	-.04	-.04	.27
nickel	-.00	-.01	.36	.05
magnesium	.87	.00	-.05	-.19
mercury	-.03	-.12	.12	-.02
gold	-.03	-.11	.03	.21
lanthanum	.10	-.06	.30	-.24
niobium	-.02	.09	.01	-.08
tin	.04	.02	.37	.35
titanium	.60	.71	.06	-.11
scandium	-.05	-.04	-.12	-.10
vanadium	.55	.04	.05	-.08
silver	-.02	-.00	-.11	.13
zinc	.41	.14	-.13	.17
strontium	.97	.08	-.10	-.00
tantalum	.01	.05	.77	-.09
rubidium	.16	.09	.04	.08
potassium	.99	.05	-.10	-.10
yttrium	-.04	.09	-.05	.05
sodium	.01	-.05	.19	-.38
zirconium	-.01	-.01	.00	.08

Table 4.3: Temporally Correlated Factor Loadings and 95% credibility intervals of the residuals after smoothing

	Factor1	Factor2	Factor3	Factor4
aluminum	.73(.61,.85)	.02(-.07,.11)	-.11(-.20,-.02)	.58(.47,.69)
barium	.10(.03,.18)	-.05(-.12,.02)	.08(.00,.15)	-.02(-.09,.05)
cadmium	.04(-.06,.14)	.76(.64,.88)	-.06(-.15,.04)	-.01(-.12,.09)
calcium	-.01(-.10,.09)	.45(.35,.56)	.03(-.07,.13)	.05(-.05,.15)
chromium	.03(-.03,.10)	.02(-.04,.09)	-.05(-.12,.01)	-.05(-.11,.02)
cobalt	-.02(-.11,.07)	-.02(-.11,.07)	-.02(-.11,.07)	.71(.58,.82)
copper	.01(-.09,.10)	.16(.07,.26)	.03(-.06,.12)	.04(-.06,.14)
cesium	.11(.03,.18)	.11(.04,.18)	.06(-.01,.12)	.10(.02,.18)
gallium	.03(-.06,.11)	-.18(-.25,-.08)	.03(-.06,.12)	.06(-.03,.15)
iron	.21(.11,.30)	-.03(-.11,.06)	.00(-.09,.09)	-.01(-.10,.07)
hafnium	-.02(-.09,.04)	.11(.04,.19)	.01(-.06,.07)	.03(-.04,.10)
lead	.04(-.06,.13)	-.03(-.12,.05)	.28(.18,.37)	.10(.01,.19)
indium	.09(.02,.16)	.00(-.07,.07)	.07(.00,.13)	-.01(-.07,.06)
manganese	.02(-.06,.10)	.01(-.07,.08)	-.01(-.09,.07)	-.01(-.09,.07)
iridium	-.03(-.12,.05)	.03(-.05,.11)	.21(.12,.30)	-.03(-.12,.05)
molybd	.04(-.05,.13)	-.01(-.09,.07)	.23(.14,.32)	.60(.47,.71)
nickel	-.02(-.11,.07)	-.03(-.12,.05)	.70(.55,.82)	-.05(-.14,.04)
magnesium	-.09(-.16,-.02)	-.01(-.08,.05)	.04(-.03,.11)	.01(-.06,.08)
mercury	-.09(-.16,-.02)	-.05(-.12,.02)	-.01(-.08,.06)	-.19(-.26,-.12)
gold	-.00(-.09,.08)	.06(-.02,.14)	.54(.43,.64)	.00(-.08,.09)
lanthanum	-.01(-.07,.06)	-.03(-.09,.04)	.05(-.01,.12)	-.05(-.12,.02)
niobium	.65(.52,.76)	.01(-.08,.09)	-.00(-.09,.08)	-.06(-.15,.02)
tin	.20(.13,.28)	-.00(-.07,.06)	.05(-.02,.12)	.05(-.02,.12)
titanium	.13(.04,.22)	.03(-.06,.12)	-.04(-.13,.04)	-.05(-.14,.04)
scandium	-.04(-.12,.05)	-.07(-.15,.01)	.05(-.04,.14)	.53(.42,.63)
vanadium	-.01(-.08,.06)	.01(-.07,.09)	.08(.01,.15)	.00(-.07,.07)
silver	-.06(-.13,.00)	-.02(-.09,.05)	.01(-.06,.08)	.07(.00,.13)
zinc	.01(-.09,.11)	.02(-.06,.11)	-.01(-.10,.08)	.76(.62,.87)
strontium	-.01(-.11,.09)	-.11(-.21,-.02)	.00(-.10,.11)	.03(-.07,.13)
tantalum	.53(.41,.63)	.01(-.08,.09)	.04(-.05,.13)	-.05(-.14,.05)
rubidium	.02(-.05,.08)	-.02(-.09,.04)	-.02(-.08,.05)	-.04(-.10,.03)
potassium	.05(-.06,.16)	-.05(-.16,.06)	-.07(-.18,.04)	.05(-.06,.16)
yttrium	-.01(-.08,.07)	-.05(-.11,.01)	.03(-.03,.10)	-.05(-.11,.02)
sodium	.06(-.01,.13)	.08(.00,.15)	-.05(-.13,.02)	-.05(-.12,.02)
zirconium	.08(.00,.16)	.09(.03,.15)	-.06(-.12,.01)	.11(.05,.18)

Table 4.4: Temporally Correlated Factor Loadings of the raw data without smoothing

	Factor1	Factor2	Factor3	Factor4
aluminum	.72	.03	-.10	.57
barium	.11	-.05	.08	-.03
cadmium	.04	.75	-.05	-.00
calcium	-.01	.45	.02	.04
chromium	.04	.01	-.06	-.05
cobalt	-.02	-.01	-.00	.69
copper	.01	.19	.02	.04
cesium	.11	.11	.06	.10
gallium	.01	-.15	.00	.05
iron	.20	-.03	.00	-.02
hafnium	-.03	.14	.01	.03
lead	.03	-.03	.29	.10
indium	.10	.01	.06	-.01
manganese	.01	.01	-.00	-.01
iridium	-.04	.03	.20	-.03
molybdenum	.04	-.01	.23	.60
nickel	-.01	-.04	.70	-.05
magnesium	-.09	-.04	.04	.09
mercury	-.11	.01	-.02	-.30
gold	.07	.12	.45	.02
lanthanum	.11	.07	.04	-.10
niobium	.50	-.01	.05	-.05
tin	.17	-.03	.05	.05
titanium	.29	.06	-.07	-.07
scandium	-.07	-.08	.02	.63
vanadium	-.06	.02	.04	.05
silver	-.06	-.05	-.01	.08
zinc	-.00	.02	.03	.75
strontium	-.01	-.14	-.02	.02
tantalum	.50	.00	.03	-.10
rubidium	.01	-.03	-.02	-.05
potassium	.07	-.05	-.08	.01
yttrium	-.04	-.04	.07	-.03
sodium	.06	.04	-.18	-.05
zirconium	.04	.09	-.05	.11

would like to compute a mixture of metals by converting a latent factor to an explicit mixture in which we do not have to decide which metal to keep and which to omit.

To better understand and control PM emission in the atmosphere, major sources of PM need to be identified and understood. Source apportionment methods have thus been developed to serve that purpose. Several methods that include positive matrix factorization (PMF) (Oh et al., 2011; Gugamsetty et al., 2012), principal component analysis (PCA) (Lee and Hieu, 2011), chemical mass balance (CMB) (Ni et al., 2012), UNMIX (Murillo et al., 2012) have been developed. Below we introduce PMF model developed by US EPA and which is one of the most widely used model for source apportionment.

Giving any $(N \times p)$ data matrix X, it can be factorized into two matrices $F(N \times m)$ and $\Lambda'(m \times p)$, and the residual matrix, E as follows:

$$X = F\Lambda' + E \tag{4.5.1}$$

or

$$x_{ij} = \sum_{k=1}^{m} f_{ik}\lambda'_{kj} + e_{ij}$$

where F is an $N \times m$ matrix of source contributions that describe the temporal variation of the source strengths and Λ' an $m \times p$ matrix of source chemical compositions, or source profiles (Yu et al., 2013). The process is carried out by minimizing the expression

$$R = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{e_{ij}^2}{s_{ij}^2}$$

where e_{ij} are the residuals and s_{ij} are the error estimates of the data point.

Solving for F in (4.5.1) we define the mixture of chemicals as

$$F = (X - E)\Lambda(\Lambda'\Lambda)^{-1}. \tag{4.5.2}$$

97

From equation (4.5.2) we see that a latent factor F has been converted to an explicit mixture. This is expressed as a product of $(X - E)$ and a mixture coefficients obtained from the utilization of the loadings as

$$\alpha = \Lambda(\Lambda'\Lambda)^{-1}. \tag{4.5.3}$$

For computing mixture coefficients α, we use the following two criteria to simplify the loading coefficients Λ,

- We keep all significant loadings and all non significant loadings bigger than 0.15.

- If a loading is not significant and is between 0.10 and 0.15, it will be kept if the 75th percentile value is bigger than 200% of the corresponding metal detection limit (Table 4.9). Otherwise it is replaced by 0.

Application of the above simplification on Table 4.3 of loadings, yielded the simplified loadings matrix giving in Table 4.6. Additionally we also apply equation (4.5.3) directly on the obtained loadings from Table 4.3 and set very small coefficients of the obtained mixture coefficients α to zero. The two approaches yielded very similar results and those obtained after application of the former method is presented in Table 4.7.

From Table 4.7, we see that mixture 1 is composed of aluminum, iron, tin and tantalum. It is not surprising to see that this mixture contains both tantalum and niobium as they are generally found together (Mineral Information Institute, 2011). This mixture may comes from traffic as aluminum and iron are traffic emitted metals (Lough et al. 2005). The second mixture which contains the associated metals of cadmium and copper (ATSDR, 2012) also contain calcium and hafnium. This may comes from miscellaneous sources. The third mixture composed of lead, iridium, molybdenum, nickel, gold may comes from power plants as it contained nickel, one of the main power plant emitted metals (Wang et al. 2010). Mixture 4, which contains

Table 4.5: Correlated metals at four different correlation levels

Loadings	Factor 1	Factor 2	Factor 3	Factor 4
0.1	aluminum barium cesium iron niobium tin titanium tantalum	cadmium calcium copper cesium gallium hafnium strontium	aluminum lead iridium molybdenum nickel gold lead	aluminum cobalt cesium lead molybdenum mercury scandium zinc, zirconium
0.2	aluminum iron niobium tin tantalum	cadmium calcium	lead iridium molybdenum nickel gold	aluminum cobalt molybdenum scandium zinc
0.3	aluminum tantalum niobium	cadmium calcium	nickel gold	aluminum cobalt molybdenum scandium zinc
0.4	aluminum tantalum niobium	cadmium calcium	nickel gold	aluminum cobalt molybdenum scandium zinc

aluminum, cobalt, molybdenum, scandium and zinc may originates from industrial plants as the main sources of atmospheric pollution for cobalt (one of the main metal of the mixture) are industrial plants such as incinerators and chemical plants. It could also come from the burning of fossil fuels (Vouk and Piver, 1983; EIP, 2013, Wang et al. 2010) and agriculture. Figure 4.1 shows the time series plots of the four mixtures. Mixture 1(black) and mixture 2(red) show a common variability while mixture 3(green) seems to vary similarly as mixture 4(blue).

Table 4.6: Simplified correlated loading Matrix to be used for computing mixture coefficients.

	Factor1	Factor2	Factor3	Factor4
aluminum	.73↑	.00	-.11↓	.58↑
barium	.10↑	.00	.00	.00
cadmium	.00	.76↑	.00	.00
calcium	.00	.45↑	.00	.00
chromium	.00	.00	.00	.00
cobalt	.00	.00	.00	.71↑
copper	.00	.16↑	.00	.00
cesium	.11↑	.11↑	.00	.10↑
gallium	.00	-.18↓	.00	.00
iron	.21↑	.00	.00	.00
hafnium	.00	.11↑	.00	.00
lead	.00	.00	.28↑	.10↑
indium	.09↑	.00	.07↑	.00
manganese	.00	.00	.00	.00
iridium	.00	.00	.21↑	.00
molybdenum	.00	.00	.23↑	.60
nickel	.00	.00	.70↑	.00
magnesium	-.09↓	.00	.00	.00
mercury	-.09↓	.00	.00	-.19↓
gold	.00	.00	.54↑	.00
lanthanum	.00	.00	.00	.00
niobium	.65↑	.00	.00	.00
tin	.20↑	.00	.00	.00
titanium	.13↑	.00	.00	.00
scandium	.00	.00	.00	.53↑
vanadium	.00	.00	.08	.00
silver	-.06↓	.00	0.00	.07↑
zinc	.00	.00	.00	.76↑
strontium	.00	-.11↓	.00	.00
tantalum	.53↑	.00	.00	.00
rubidium	.00	.00	.00	.00
potassium	.00	.00	.00	.00
yttrium	.00	.00	.00	.00
sodium	.00	.08↑	.00	.00
zirconium	.08↑	.09↑	.00	.11↑

Table 4.7: Mixture coefficients.

	Mixture1	Mixture2	Mixture3	Mixture4
aluminum	.46	-.02	-.09	.19
barium	.09	-.05	.08	-.03
cadmium	-.01	.83	-.02	-.01
calcium	-.03	.50	.05	.03
chromium	.03	.02	-.05	-.02
cobalt	-.11	-.02	-.05	.35
copper	-.01	.18	.04	.02
cesium	.06	.12	.06	.03
gallium	.02	-.18	.02	.02
iron	.16	-.04	.01	-.04
hafnium	-.03	.13	.01	.02
lead	.03	-.03	.27	.03
indium	.07	.00	.07	-.02
manganese	.01	.00	-.01	-.01
iridium	-.01	.04	.21	-.02
molybdenum	-.04	.00	.20	.28
nickel	.03	-.01	.69	-.05
magnesium	-.07	-.01	.03	.02
mercury	-.04	-.05	-.01	-.08
gold	.02	.08	.54	-.02
lanthanum	.01	-.03	.05	-.03
niobium	.50	-.03	.03	-.12
tin	.15	-.01	.06	-.01
titanium	.10	.02	-.03	-.04
scandium	-.10	-.07	.02	.26
vanadium	-.01	.02	.08	.00
silver	-.05	-.02	.00	.04
zinc	-.10	.03	-.04	.37
strontium	-.00	-.12	-.00	.02
tantalum	.40	-.02	.07	-.10
rubidium	.02	-.03	-.01	-.02
potassium	.03	-.06	-.07	.02
yttrium	.01	-.05	.03	-.02
sodium	.05	.08	-.04	-.03
zirconium	.04	.09	-.05	.05
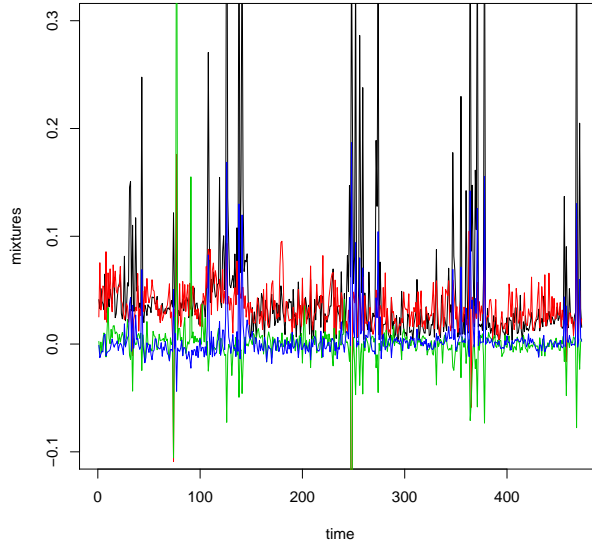
Figure 4.1: Plots of Mixtures 1(black), 2(red), 3(green) and 4(blue)

4.6 Extension to Spatially and Temporally Correlated Data

Particulate matter speciation data are generally collected from several stations within a given metropolitan statistical area (MSA). In addition to the temporal correlation within each station, data from adjacent stations might also be spatially correlated. This additional correlation structure needs also to be taken into account when using factor analysis to avoid inaccurate estimates. In this section, we extend the Bayesian factor analysis for temporally correlated data discussed in previous section to cover data that are both spatially and temporally correlated.

We start with the more general covariance matrix form

$$\epsilon \sim N(0, \Delta)$$

as in (4.4.1), where we can let $\Delta = \Theta \otimes \Psi$, $\Theta > 0$, $\Psi > 0$. We maintain the same separation structure introduced earlier. But here Θ is assume to be of the form $\Theta = \Phi_T \otimes \Phi_S$ where Φ_T is the temporal correlation matrix as above and Φ_S is the

matrix accounting for the spatial correlation. If we let Φ_S be identity I_S and especially when $s = 1$, then $\Theta = \Phi_T$ and we have the Bayesian factor analysis for temporally correlated data described above.

as before the temporal correlation Φ_T could be assumed to be an auto regressive of order 1 giving by

$$\Phi_T = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n_s-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n_s-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n_s-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{n_s-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n_s-1} & \rho^{n_s-2} & \rho^{n_s-3} & \rho^{n_s-4} & \cdots & 1 \end{pmatrix} \tag{4.6.1}$$

where n_s is the station sample size.

We may also assume that Φ_S the between station correlation is a fix matrix and the correlation between stations could be an autoregressive of order 1 or inversely proportional to the distance between stations. For simplification and considering three stations for application purpose, we assume that Φ_S is also giving by

$$\Phi_S = \begin{pmatrix} 1 & \eta & \eta^2 & \eta^3 \\ \eta & 1 & \eta & \eta^2 \\ \eta^2 & \eta & 1 & \eta \\ \eta^3 & \eta^2 & \eta & 1 \end{pmatrix}.$$

Table 4.8, shows the factor loadings without accounting for spatial correlation (left) and with account of partial correlation (right) between three stations in Tampa MSA. We can note some differences between the loadings which highlight the importance of properly taking into account the data correlation.

Table 4.8: Loadings without spatial correlation (left) and with spatial correlation (right).

	Factor1	Factor2	Factor3	Factor4	Factor1	Factor2	Factor3	Factor4
Al	0.768	0.025	-0.082	0.430	**0.777**	0.026	-0.082	0.430
Ba	0.009	0.012	0.030	-0.069	0.008	0.011	0.031	-0.072
Cd	-0.242	0.808	-0.065	0.067	-0.241	**0.819**	-0.067	0.069
Ca	-0.152	0.428	-0.250	-0.077	-0.154	0.431	-0.252	-0.079
Cr	0.032	0.221	-0.024	-0.104	0.033	0.222	-0.026	-0.106
Co	0.365	-0.231	-0.154	0.531	0.365	-0.230	-0.157	0.536
Cu	-0.135	-0.010	0.094	0.102	-0.135	-0.009	0.096	0.106
Cs	0.049	-0.235	0.331	0.064	0.046	-0.236	0.337	0.06
Ga	-0.012	-0.230	-0.109	0.117	-0.018	-0.239	-0.108	0.112
Fe	0.429	-0.0432	-0.043	-0.060	0.437	-0.045	-0.044	-0.059
Hf	-0.025	0.174	0.076	0.014	-0.021	0.177	0.077	0.015
Pd	-0.018	-0.110	0.376	0.263	-0.019	-0.110	0.382	0.270
In	0.197	0.049	0.196	0.039	0.198	0.047	0.198	0.041
Mn	-0.021	0.118	0.026	-0.023	-0.019	0.122	0.029	-0.022
Ir	-0.112	-0.012	0.525	-0.055	-0.116	-0.015	**0.538**	-0.063
Mo	0.131	0.049	0.347	0.776	0.133	0.052	**0.358**	0.794
Ni	-0.014	-0.091	0.779	-0.062	-0.011	-0.094	**0.796**	-0.059
Mg	-0.136	-0.077	0.115	-0.073	-0.136	-0.077	0.119	-0.074
Hg	-0.267	0.033	0.015	-0.234	-0.271	0.031	0.016	-0.239
Au	-0.206	0.100	0.766	0.142	-0.204	0.104	**0.774**	0.145
La	-0.056	-0.175	0.169	-0.037	-0.056	-0.176	0.171	-0.034
Nb	0.672	-0.000	0.024	-0.046	0.671	-0.005	0.019	-0.047
Sn	0.238	-0.298	-0.053	0.172	0.236	-0.302	-0.053	0.174
Ti	0.038	0.084	0.050	-0.148	0.041	0.084	0.053	-0.149
Sc	-0.072	0.110	0.121	0.680	-0.072	0.113	0.121	**0.688**
V	0.028	0.061	0.125	-0.223	0.030	0.060	0.128	-0.218
Ag	-0.220	0.045	-0.119	-0.006	-0.222	0.045	-0.122	-0.004
Zn	0.692	0.130	-0.030	-0.014	0.697	0.127	-0.032	-0.013
Sr	0.168	0.135	0.119	-0.007	0.168	0.134	0.121	-0.008
Ta	-0.020	-0.011	-0.190	0.006	-0.025	-0.014	-0.192	0.000
Rb	0.054	0.634	0.014	-0.064	0.059	**0.645**	0.013	-0.061
K	0.044	0.139	-0.046	0.110	0.046	0.139	-0.044	0.109
Y	0.018	-0.107	-0.038	-0.081	0.017	-0.104	-0.038	-0.080
Na	0.102	-0.097	0.143	-0.140	0.100	-0.099	0.140	-0.137
Zr	-0.070	-0.106	0.368	-0.031	-0.068	-0.102	0.370	-0.032

4.7   Discussion

Factor analysis is a powerful tool that could be used in order to group temporally correlated air pollution data in general and $PM_{2.5}$ speciation metals in particular according to theirs source origins. As pollution data is correlated, the traditional factor model is not adequate for their analysis. Omitting to account for the correlation if it exists could lead to a serious misclassification and inadequate estimates. In our development above, we illustrate the use of autoregressive correlation, but any other correlation as dictated by available data could be used just my modifying the correlation matrix Φ. As introduced in section 6, the theory could easily be extended to several correlated or uncorrelated stations. Φ will then be written as $\Phi = \Phi_T \otimes \Phi_S$, where S will represent the number of stations. Although we consider Φ_S to be fixed here, a more probabilistic approach could be taken. In the case that an autoregressive is assumed a distribution of η similar to the one for ρ need to be found.

Many possible extension of the temporal and spatial correlation could be considered. For spatial correlation for example we could consider the correlation between stations to be inversely proportional to the distance between them. For both temporal and spatial correlation, uniform correlation model such as $(1 - \rho)I + \rho J$ where I is the identity matrix and J a matrix where all the elements are 1. Also an exponential correlation model giving by $\sigma^2 \rho^{|j-k|}$ where $\rho = exp(-\phi d)$ and d is the difference between two time points, to name a few could be considered.

An important issue to be faced is the number of factors. We select the number of factors by previous believe that four main pollution sources: traffic, power point, industries and miscellaneous are the main sources. A probabilistic approach could be taken. Defining $p(m)$, a prior on m, by Bayes' Rule it can be computed that the probability of each of the number of factors given the parameters is given by

$$p(m|\mu, \Lambda, F, \Psi, X) \sim p(m)p(\mu)p(\Lambda|\Psi, m)p(F|m)p(\Psi)p(X|\Lambda, F, \Psi, m)$$

and we can then determine the most probable number of factors. As this is not our main goal in this paper, we did not elaborate further.

Table 4.9: Minimum detection Limit of Air Filter Samples for Different Analytical Methods in ng/m^3

Species	INAA	XRF	PIXE	Flame AAS	Flame AAS
Ag	0.14	7	NA	5	0.006
Al	26	6	14	36	0.01
As	0.2	1.0	1	120	0.2
Au	NA	2	NA	25	0.1
Ba	7	30	NA	10	0.05
Be	NA)	NA	NA	2	0.06
Br	0.5	0.6	1	NA	NA
Ca	113	2	5	1	0.06
Cd	5	7	NA	1	0.004
Ce	0.07)	NA	NA	NA	NA
Cl	6	6	10	NA	NA
Co	0.02	0.5	NA	7	0.02
Cr	0.2	1	2	2	0.01
Cs	0.04	NA	NA	NA	NA
Cu	36	0.6	1	5	0.02
Eu	0.007	NA	NA	25	NA
Fe	5	0.8	2	5	0.02
Ga	0.6	1.1	1	62	NA
Hf	0.01	NA	NA	2,400	NA
Hg	NA	1	NA	600	25
In	0.007	7	NA	37	NA
K	29	4	6	2	0.02
La	0.06	36	NA	2,400	NA
Mg	360)	NA	24	0.4	0.005
Mn	0.14	1.0	2	1	0.01
Mo	NA	1	6	37	0.02
Na	2	NA	72	0.2	< 0.06
Ni	NA	0.5	1	6	0.1
Pb	NA	1	4	12	0.06
Rb	7	0.6	2	NA	NA
Sc	0.001	NA	NA	60	NA
Sn	NA	10	NA	37	0.2
Sn	22	0.6	2	5	0.2
Ta	0.02	NA	NA	2,400	NA
Ti	78	2	4	114	NA
V	0.7)	1	4	62	0.2
Y	NA	0.7	NA	360	NA
Zn	4	0.6	1	1	0.001
Zr	NA	1.0	4	1,200	NA

References

[1] Agency for Toxic Substances and Disease Registry (ATSDR). 2012. Toxicological Profile for Cadmium. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.

[2] ALA 2001: Selected key studies on particulate matter and health: 1997 - 2001

[3] Andersen, Z. J., Wahlinc P., Raaschou-Nielsend, O., Scheikea, T. and Lof, S.(2007) Ambient particle source apportionment and daily hospital admissions among children and elderly in Copenhagen. J. of Exp. Sci. and Envir. Epidemiology (2007) 17, 625-636.

[4] Breslow, N.E. and Clayton D.G. (1993) Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association. Vol; 88:421:9-25.

[5] Brown, P. E., Diggle, P. J., Lord, M. E. and Young, P. C. (2001) Space-time calibration of radar rainfall data. Appl. Statist., 50, 221-241.

[6] Carrero, J. A., I. Arrizabalaga, N. Goienaga, G. Arana, J. M. Madariaga. Traffic related metal distribution profiles and their impact on urban soils *Urban Environment, Proceedings of the 10th Urban Environment Symposium*, Springer, 2012

[7] Choi J, Fuentes M, and Reich B J: Spatial-temporal association between fine particulate matter and daily mortality Comput Stat Data Anal. 2009 June 15; 53(8): 2989-3000.

[8] Cressie, N. A. C. (1993). Statistics for Spatial Data, revised edition. New York: Wiley.

[9] Cressie, N. (1994) Comment on "An approach to statistical spatial-temporal modeling of meteorological fields" by M. S. Handcock and J. R. Wallis. J. Am. Statist. Ass., 89, 379-382.

[10] Dadvand P,. Figueras F., Basagana., Beelen R., Martinez D., Cirach., Schembari A., Koek G., Brunekreef B., and Nieuwenhuijsen M. J.(2013) Ambient Air Pollution and Preeclampsia: A Spatiotemporal Analysis. Environmental Health Perspectives. 121:1365-1371.

[11] Darrow, L. A., Klein, M., Strickland, M. J., Mulholland, J. A., & Tolbert, P. E. (2011). Ambient air pollution and birth weight in full-term infants in Atlanta, 1994-2004. Environmental Health Perspectives, 119 (5), 731-737.

[12] DeCoster, J. (1998). Overview of Factor Analysis. Retrieved 12-17-2012 from http://www.stat-help.com/notes.html

[13] 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. New England Journal of Medicine, 329:1753-9.

[14] Donaldson K, MacNee W.(2001). Potential mechanisms of adverse pulmonary and cardiovascular effects of particulate air pollution (PM10) Int J Hyg Environ Health., 203:411-415

[15] Ebisu, K., Bell, M.L., Airborne PM2.5 Chemical Components and Low Birth Weight in the Northeastern and Mid-Atlantic Regions of the United States. Environ Health Perspect. 2012;120 (12): 1746-1752.

[16] Environmental Integrity Project (EIP) (Jan 2013) The Toxic Ten: Top Power Plant Emissions of Mercury, Toxic Metals, and Acid Gases in 2011

[17] Forsberg B, Hansson HC, Johansson C, Areskoug H, Persson K, Jarvholm B. Comparative health impact assessment of local and regional particulate air pollutants in Scandinavia. Ambio 2005; 34: 11-19.

[18] Forsberg, C., M. Hovland, S. Planke, A. Rike, J. Strout, G. Svano, H. Svensen, and T. Tjelta. (2005). Pockmarks with methane derived authigenic carbonate in the Norwegian Trench, Northern North Sea. 2005. Eos, Transactions, American Geophysical Union 86(52):A4.

[19] Gilboa, S. M., Mendola, P., Olshan, A. F., Langlois, P.H., Savitz, D. A., Loomis, D., Herring, A. H., & Fixler, D. E. (2005). Relation between Ambient Air Quality and Selected Birth Defects, Seven County Study, Texas, 1997-2000. American Journal of Epidemiology, 162 (3), 238-252.

[20] Gilbert, P.D. Time Series Factor Analysis with an Application to Measuring Money, University of Groningen, Research School SOM, Research Report 05f10, Nov. 2005

[21] Goodall, C. and Mardia, K. V. (1994) Challenges in multivariate spatio-temporal modeling. In Proc. 17th Int. Biometric Conf., Hamilton, Aug. 8th-12th, pp. 1-17. Hamilton: McMaster University Press.

[22] Graham AJ, Davies R, Xuereb J, Halliday GM, Kril J, Creasey H, et al. Pathologically proven frontotemporal dementia presenting with severe amnesia. Brain (2005) 128:597-605.

[23] Grahame T, Schlesinger R. Evaluating the health risk from secondary sulfates in eastern North American regional ambient air particulate matter. Inhal Toxicol 2005; 17: 15-27.

[24] Gu, C. (2002), Smoothing Spline ANOVA Models, New York Springer-Verlag.

[25] Gugamsetty, B., Wei, H., Liu, C.N., Awasthi A., Hsu, S.C., Tsai, C.J., Roam, G.D., Wu, Y.C. and Chen, C.F. (2012). Source Characterization and apportionment of PM_{10}, $PM_{2.5}$ and PM0.1 by using Positive Matrix Factorization. Aerosol Air Qual. Res. 12: 476-491.

[26] Gurgueira, S. A., Lawrence, J., Coull, B., Murthy, G. G., and Gonzalez- Flecha, B. 2002. Rapid increases in the steady-state concentration of reactive oxygen species in the lungs and heart after particulate air pollution inhalation. Environ. Health Perspect. 110:749-755.

[27] Ha, E. H., Hong, Y. C., Lee, B. E., Woo, B. H., Schwartz, J., Christiani, D. C. (2001). Is air pollution a risk factor for low birth weight in Seoul? Epidemiology, 12, 643-648.

[28] Hastie, T. and Tibshirani, R. (1990) Generalized Additive Models. London: Chapman and Hall.

[29] Hayashi, K. The Press-Shigemasu Bayesian Factor Analysis with Estimated Hyperparameters. PhD thesis, University of North Carolona, Chapel Hill, 1997

[30] Hwang, B-F., Lee, Y. L., Jaakkola, J. J. K (2011). Air pollution and still birth: A population- based case-control study in Taiwan. Environmental Health perspectives, 119 (9), 1345-9.

[31] Ibrahimou, B., Salihu, H.M., Muktar, A., & Anozie, C.(2014), Risk of Preeclampsia from Exposure to Particulate Matter (PM2.5) Speciation Chemicals During Pregnancy, Journal of Occupational and Environmental Medicine, 2014 DOI: 10.1097/JOM.0000000000000317

[32] Kent, J. T. and Mardia, K. V. (1994) The link between Kriging and thin-plate splines. In Probability, Statistics and Optimisation (ed. F. P. Kelly), pp. 324-329. New York: Wiley.

[33] Krewski, D., Burnett, R.R., Goldberg, M.S., Hoover, K., Siemiatycki, J., Jerrett, M., Abrahamowicz, M., White, W.H, and Others (2000). Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Health Effects Institute, July, 2000.

[34] Kyriakidis, P. C. and Journel, A. G. (1999) Geostatistical space-time models: a review. Math. Geol., 31, 651-684.

[35] Laden, F. Neas, L.M., Dockery, D.W., and Schwartz, J. Association of Fine Particulate Matter from Different Sources with Daily Mortality in Six U.S. Cities. Environmental Health Perspectives 108:941- 947, October 2000.

[36] Lacasana, M., Esplugues, A., & Ballester, F. (2005). Exposure to ambient air pollution and Prenatal and early childhood health effects. Environmental Journal of Epidemiology, 20, 183-199.

[37] Laud, P. W. and Ibrahim, J. G. (1995) Predictive model selection. J. R. Statist. Soc. B, 57, 247 262.

[38] Le, N. D., Li, S., & Zidek, J. V. (2010). Air pollution. Chronic Diseases in Canada, 29 (suppl 2), 144-63.

[39] Lee, B.K. and Hieu, N.T. (2011). Seasonal Variation and Sources of Heavy Metals in Atmospheric Aerosols in a Residential Area of Ulsan, Korea. Aerosol Air Qual. Res. 11: 679-688.

[40] Lin, X. and Breslow, N.E. (1996) Bias Corrections in Generalized Linear Mixed Models. Journal of the American Statistical Association.Vol 91;

[41] Long, J. F., Waldman, W. J., Kristovich, R., Williams, M., Knight, D., and Dutta, P. K. 2005. Comparison of ultrastructural cytotoxic effects of carbon and carbon/iron particulates on human monocyte-derived macrophages. Environ. Health Perspect. 113:170-174.

[42] Lough, G., Schauder, Soopark, J., Shafer, M.M. Deminter, J. T., and Weinstein, J.P.() Emissions of Metals Associated with Motor Vehicle Roadways. *Environ. Sci. Technol.* (2005), 39, 826-836

[43] Maisonet, M., Correa, A., Misra, D, Jaakkola, J. J. K (2004). A review of the literature on the effects of ambient air pollution on fetal growth. Environmental Research, 95, 106-115.

[44] Mardia, K. V. and Goodall, C. (1993) Spatial-temporal analysis of multivariate environmental monitoring data. In Multivariate Environmental Statistics (eds G. P. Patil and C. R. Rao), pp. 347-386. Amsterdam: Elsevier.

[45] Mardia, K. V., Goodall, C., Redfern, E. J. and Alonso, F. J. (1998) The Kriged Kalman filter (with discussion). Test, 7, 217-252.

[46] Mezzetti, M. & Billari, F.C Bayesian correlated factor analysis of socio-demographic indicators. *Stat Methods Appl* (2005) 14:223-241

[47] Michelle L. Bell, Francesca Dominici, Keita Ebisu, Scott L. Zeger, and Jonathan M. Samet: Spatial and Temporal Variation in PM2.5 Chemical Composition in the United States for Health Effects Studies Environmental Health Perspectives, vol 115 7 (2007)

[48] Mineral Information Institue. (2011) www.eoearth.org/view/article/156409

[49] Molinelli, A. R., Madden, M. C., McGee, J. K., Stonehuerner, J. G., and Ghio, A. J.(2002). Effect of metal removal on the toxicity of airborne particulate matter from the Utah Valley. Inhalation toxicology. 14; 1069-1086.

[50] Murillo, M.J., Ramos, A.C., Carcia, F.A., Jimenez, S.B., Cardenas, B. and Mizohata, A. (2012). Chemical Composition of $PM_{2.5}$ Particles in Salamanca, Guanajuato Mexico: Source Apportionment with Receptor Models. Atmos. Res. 107: 31-41.

[51] National Research Council. 2004. Research priorities for airborne particulate matter, IV. Continuing research progress. Washington, DC: National Academies Press.

[52] Ni, T., Han, B. and Bai, Z.P. (2012). Source Apportionment of PM_{10} in Four Cities of Northeastern China. Aerosol Air Qual. Res. 12: 571582.

[53] Oh, M.S. T. J. Lee and D. S. Kim (2011). Quantitative Source Apportionment of Size-segregated Particulate Matter at Urbanized Local Site in Korea. Aerosol Air Qual. Res. 11: 247264.

[54] Ostro B, Feng W Y, Broadwin R, Green, S and Lipsett M The Effects of Components of Fine Particulate Air Pollution on Mortality in California Environmental Health Perspectives vol 115 7 2007

[55] Pena, D. and G. E. P. Box. Identifying a Simplifying Structure in Time Series. *Journal of the American Statistical Association*, 1987, vol 82(399), 836-843

[56] Peng RD, Dominici F, Pastor-Barriuso R, Zeger SL, Samet JM. Seasonal analyses of air pollution and mortality in 100 US cities. Am J Epidemiol 2005; 161: 585-94.

[57] Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E., and Heath, C.W. Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults. American Journal of Respiratory Critical Care Medicine, 1995:151:669-74.

[58] Press, S.J. and K. Shigemasu. Bayesian inference in factor analysis. In *Contributions to Probability and Statistics*, chapter 15. Springer-Verlag, 1989

[59] Press, S.J. and K. Shigemasu. Bayesian inference in factor analysis-Revised. Technical Report No. 243, Department of Statistics, University of Clifornia, Riverside, April 1994.

[60] Ritz, B. & Yu, F. (1999). The effect of ambient carbon monoxide in low birth weight among children in Southern California between 1989 and 1993. Environmental Health Perspectives, 107, 17-25.

[61] Ritz B., Wilhelm, M., Hoggart, K. J., & Ghosh, J. K. C. (2007). Ambient air pollution And preterm birth in the UCLA environment and pregnancy outcomes study. American Journal of Epidemiology, 166, (9), 1045-52.

[62] Rogers, J. F., & Dunlop, A. L. (2006). Air pollution and very low birth weight infants: A target population? Pediatrics, 118, 156. DOI:10.1542/peds.2005-2432.

[63] Rowe, D.B. Correlated Bayesian Factor Analysis. Ph.D. Dissertation, University of Clifornia, Riverside, December 1998.

[64] Samet, J.M., Dominici, F., Zeger,S.L., Schwartz, J., and Dockery, D.W. The National Morbidity, Mortality, and Air Pollution Study. Part I: Methods and Met hodologic Issues. Health Effects Institute Research Report 94, Part I, May 2000.

[65] Samet, J.M., Zeger, S.L., Dominici, F., Curriero, F., Coursac, I., Dockery, D.W., Schwartz, J., and Zanobetti, A. The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity, Mortality and Air Pollution in the United States. Health Effects Institute Research Report 94, Part II, June 2000.

[66] Schauer, J.J, G. C. Lough, M.M. Shafer, W. F. Christensen, M. F, Arndt, J. T. DeMinter, and J. S. Park: Characterization of Metals Emitted from Motor Vehicles. *Research Report* March (2006) 133

[67] Schwartz, J. 2004. Is the association of airborne particles with daily deaths confounded by gaseous air pollutants? An approach to control by matching. Environ. Health Perspect. 112(5):557-561.

[68] Schwarze, P.E J vrevik, M Lag, M Refsnes, P Nafstad, RB Hetland and E Dybing: Particulate matter properties and health effects: consistency of epidemiological and toxicological studies Human & Experimental Toxicology (2006) 25: 559-579

[69] Schlesinger, R. B., and Cassee, F. 2003. Atmospheric secondary inorganic aerosols: The toxicological perspective as a basis for health effects risk assessment. Inhal. Toxicol. 15:197-235.

[70] Seong-Gil Kim, Jung-Hoon Jee, Ju-Chan Kang: Cadmium accumulation and elimination in tissues of juvenile olive flounder, Paralichthys olivaceus after sub-chronic cadmium exposure Environmental Pollution 127 (2004) 117-123

[71] Serensen, M., Daneshvar, B., Hensen, M., Dragsted, L. O., Hertel, O., Knudsen, L., and Loft, S. 2003. Personal PM2.5 exposure and markers of oxidative stress in blood. Environ. Health Perspect. 111:161-165.

[72] Shah, K.R.,& Sinha, B.K. (1989) Theory of optimal design. New York: springer-Verlag.

[73] Sram, R. J., Binkova, B., Dejmel, J., Bobak, M. (2005). Ambient air pollution and Pregnancy outcomes: A review of the literature. Environmental Health Perspectives, 113, 375-382.

[74] Stone, C.J. (1985),"Additive Regression and Other Nonparametric Models", Annals of Statistics, 13, 689-705.

[75] Stroud, J. R., M uller, P. and Sans o, B. (2001) Dynamic models for spatiotemporal data. J. R. Statist. Soc. B, 63 673-689.

[76] Sujit K. Sahu: A Bayesian kriged Kalman model for short-term forecasting of air pollution levels, Appl. Statist. (2005) 54, Part 1, pp. 223-244

[77] U.S. Environmental Protection Agency. 1998. Guideline on Speciated Particulate Monitoring: U.S. Environmental Protection Agency. EPA August 1998

[78] U.S. Environmental Protection Agency. 2004. Air quality criteria for particulate matter. Washington, DC: U.S. Environmental Protection Agency. EPA 600/P-99/002aF-bF.

[79] Vouk, V.B. and Piver, W.T. (1983). Metallic elements in fossil fuel combustion products: amounts and forms of emissions and evaluation of carcinogenicity and mutagenicity. Environmental Health Perspectives.47 . 201 225.

[80] Wang,Y.F, H. r. Chao, L. c. wang, G. P. C. Chien and T. C. Tsou. Characteristics of Heavy Metals Emitted from a Heavy Oil-Fueled Power Plant in Northern Taiwan. *Aerosol and Air Quality Research* (2010), 10: 111-118

[81] Wikle, C. K. and Cressie, N. (1999) A dimension-reduced approach to space-time Kalman filtering. Biometrika, 86, 815-829.

[82] Wilhelm, M., Ghosh, J., Su, J., Cockburn, M., Jerrett, M. & Ritz, B. (2011). Traffic-related air toxics and term low birth weight in Los Angeles County, California. Environmental Health Perspectives, 120 (1), 132-138.

[83] Wittig, A. E., Anderson, N., Khlystov, A. Y., Pandis, N. S., Davidson, C., Robinson, A. L. (2004) Pittsburgh air quality study overview. Atmospheric Environment 38, 3107-3125.

[84] Wood, S. N. (2000) Modeling and smoothing parameter estimation with multiple quadratic penalties. J. R. Statist. Soc. B 62, part 2, 413-428.

[85] Wood, S. N. (2006) Generalized Additive Models: An Introduction with R. Cahpman & Hall/ CRC, Boca Raton, FL 33487.

[86] Wikle, C. K., Berliner, L. M. and Cressie, N. (1998) Hierarchical Bayesian space-time models. Environ. Ecol. Statist., 5, 117-154.

[87] Yu, L., G. Wang, R. Zhang, L. Zhang, Y. Song, B. Wu, X. Li, K. An, and J. Chu: Characterization and Source Apportionment of PM2.5 in an Urban Environment in Beijing, Aerosol and Air Quality Research, 2013, 13: 574-583

[88] Zeger S.L. and Diggle, P.J. (1994) Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics.50:3; 689-699.

[89] Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998) Semiparametric stochastic mixed models for longitudinal data. Journal of American Statististical Association. Vol 93; 710-719.

[90] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. and Smith, G. M. Mixed Effects Models and Extensions in Ecology with R. Springer 2009