12-20-2013

Intraurban Variability of Ambient Particulate Matter

Varun Yadav
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Energy, Environmental and Chemical Engineering

Dissertation Examination Committee:
Jay Turner, Chair
Pratim Biswas
Benjamin de Foy
Rudolf Husar
Brent Williams
Allison Willis

Intraurban Variability of Ambient Particulate Matter

by

Varun Yadav

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2013
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Several projects in this dissertation research involved substantive collaborations. The funding sources and the people involved with each project are acknowledged in the individual chapters.

I acknowledge Prof. Jay R. Turner for instilling a keen sense of integrity; meticulous work ethic and dedication that now inspired every aspect of my life. The problem solving approaches, both academic and otherwise, that he inspired will be my greatest assets for handling the challenges ahead. I also extend my gratitude towards my committee members, Professors Pratim Biswas, Benjamin de Foy, Rudolf Husar, Brent Williams and Allison Willis for their support and guidance through the entire process. I am also grateful for the support from my friends and the department staffs for making this stage a truly memorable experience.

I am greatly in debt for the inspiration and support of my family, both in India and US. This work is dedicated to Timothy Mohney and Danny, who encouraged me even when I couldn't see the way ahead and inspire me every day to push further.

Varun Yadav

Washington University in St. Louis

December 2013

ABSTRACT OF THE DISSERTATION

Intraurban Variability of Ambient Particulate Matter

by

Varun Yadav

Doctor of Philosophy in Energy, Environmental & Chemical Engineering

Washington University in St. Louis, 2013

Professor Jay R. Turner, Chair

An understanding of spatial and temporal variability in ambient particulate matter (PM) is important for effective air quality management and for assessing potential exposure misclassification in epidemiological and exposure studies used to support health-based standards. Spatiotemporal variability of PM in urban areas can be influenced by many factors, such as local sources of primary PM; source locations and their emission profiles; topographic barriers; meteorological patterns; behavior of semi-volatile components; and measurement errors. Intraurban variability is often gauged by conducting measurements at a network of monitoring stations across the region of interest. While certain statistical metrics are commonly used and interpreted in a relative sense across site-pairs, there is no standardized framework for analyzing such datasets.

This dissertation presents systematic data analysis approaches applicable to a variety of monitoring networks for assessing intraurban variability in PM and its components. Interpreting patterns in statistical metrics for a network with a large number of sites can be particularly challenging, and calculating these metrics for each site with respect to a reference concentration

time series may better reveal the variability. In the absence of a representative background site, the network itself can be utilized to generate baseline and site-specific excess concentration time series to semi-quantitatively differentiate urban- and larger-scale contributions from local-scale emissions. Utilizing this approach to interpret patterns in the statistical metrics provides insights into the factors influencing the baseline and the monitoring sites displaying greater variability.

Apportionment of measured concentrations at each site into baseline and site-specific excess concentrations towards refined application of wind regression tools for estimating local emission source regions are also discussed. The approach is also utilized for identifying meteorological and geographic factors that modulate the spatial and temporal PM trends. It also provides a weight-of-evidence to conventional source apportionment tools used for estimating local and regional source impacts. The strengths and limitations of the proposed approaches are discussed for a variety of networks measuring PM and/or its components on varying spatial and temporal scales. Issues regarding measurement uncertainty estimation and precision in data reporting which can influence interpretation of variability are also discussed.

Chapter 1 :Introduction

1.1. Introduction

An understanding of ambient particulate matter (PM) spatiotemporal variability is important for effective air quality management. Such management is often practiced on the urban scale, and intraurban variability affects the spatial zone that is represented by observational data collected at a given site. Control strategy development requires an understanding of both the spatial zones represented by the monitoring network and the spatial zones of influence for emission sources. Spatiotemporal variability influences the value of each monitor in a network and, when considered in the context of the network objectives, this information can be used to optimize the network design. In addition to air quality management, epidemiological and exposure studies used to support health-based standards can also be affected by exposure misclassification from inadequate accounting of spatial and/or temporal variability (Wilson et al., 2005). It is commonly assumed that the spatial distribution of certain pollutants is homogeneous within large urban areas and ambient pollutant burdens at a central monitoring site are representative for the spatial extent of the study area. Many studies supported this assumption and indeed found well correlated distributions of PM within urban areas (Burton et al., 1996 ($PM_{2.5}$ and PM_{10}); Pope et al., 2002 ($PM_{2.5}$); Roosli et al., 2001 (PM_{10})). In other cases, however, studies examining datasets on finer spatial and temporal scales suggest greater variation within urban areas than previously characterized (Ito et al., 2004 ($PM_{2.5}$); Pinto et al., 2004 ($PM_{2.5}$); Kim et al., 2005 ($PM_{2.5}$)).

Spatial variability of PM mass and its constituents in urban areas can be influenced by several factors such as those identified by Pinto et al. (2004):

- local sources of primary PM emissions;
- transient emission events;
- topographic barriers that isolate sub-regions of the urban area;
- meteorological phenomena that vary on spatial scales within the urban area;
- differences in behavior of semi-volatile components; and
- measurement errors.

Influences from local sources have been well studied and documented. Watson and Chow (2001) summarized spatial scales of influence for emission sources (Table 1-1**)**. These operationally defined scales can also be viewed from the perspective of the receptor sites where the observational data are collected. Micro-, middle-, neighborhood-, and urban-scale emission sources will necessarily drive intraurban variability. Topographical barriers and meteorological phenomena acting on spatial scales finer than the urban area can also modulate spatial variability. While mountains and valleys are visually evident topographical barriers, urban canyons and water bodies can also influence airflow patterns. Spatial variability in the environmental conditions such as meteorological conditions, temperature and humidity can also induce variability in the ambient particulate concentrations (Seinfeld and Pandis, 2006). Humidity, fog, as well as temperature differences due to urban-heat island effect can alter the PM characteristics.

Table 1-1. Spatial scales of emission source influence (Watson and Chow (2001)).

Scale	Spatial range
Micro	~10 m
Middle	~0.1-1 km
Neighborhood	~ 1-5 km
Urban	~5-50 km
Regional	~50-1000 km
Continental	~1000-5000 km
Global	>5000 km

Transient emissions events that exert intraurban variability are typically less well characterized, especially if they are rare events within the dataset. These exceptional events are often excluded from data analyses depending on the study objectives. For example, holiday fireworks can result in large spikes in PM mass but are viewed as anomalies and are often excluded from source apportionment analyses. Exceptions are transient events that lead to exceedances of a National Ambient Air Quality Standard (NAAQS); there is considerable motivation to quantify the impact from such events to petition for having such data excluded from the determination of whether a monitor violates the NAAQS. Spatial differences in semi-volatile components are becoming better understood primarily due to advances in the measurement methods such as continuous monitors that can classify ambient PM mass into operationally defined volatile and nonvolatile components. Finally, measurement error can also lead to artificial spatial variability in observational datasets. In addition to these physical factors that modulate intraurban variability, analyses must consider the spatial extent of the region, temporal scales, and PM attributes of interest. Depending on the analysis objectives, the time scale of interest may include climatological norms, annual, seasonal, day-of-week, diurnal or even hourly time scales.

Intraurban variability studies are conducted to meet a variety of study objectives; in many cases, however, the objectives are not well defined *a priori* and the data analysis proceeds without a clear roadmap. One common objective – whether or not it is explicitly stated – is to delineate the relative contributions from sources exerting influence on different scales (e.g., local versus regional sources). Long-range transport of pollutants can lead to spatial homogeneity or uniformity in species concentrations across urban networks. Depending on the location and strength of the local sources, meteorological conditions, and various other factors described earlier, local emission can disrupt this homogeneity. Because of the numerous factors affecting variability, there are no general guidelines for data analysis and studies have used one-or-more different methods to gauge variability in PM and its constituents. A few such metrics and approaches are listed in Table 1-2 with their advantages and limitations.

1.1.1. Metrics for gauging variability

PM uniformity within urban areas is examined using several metrics, none of which alone can capture all aspects of variability. A commonly used statistical metric is the Pearson's correlation coefficient (r or PCC) which is defined as the ratio of the covariance between two datasets to the product of their respective standard deviations. The coefficient is often used to gauge correlation in concentrations between two monitoring sites (Burton et al., 1996; DeGaetano et al., 2004). High intersite correlation indicates spatially consistent temporal variability but not necessarily spatial uniformity in mass concentrations (Pinto et al., 2004). The coefficient of variation, which is a normalized measure of dispersion and defined as the ratio of the standard deviation to the mean value of the dataset expressed as a percentage, has also been used for characterizing variability. It often utilized for examining temporal variability in concentration time series

Table 1-2. Various methods used for intraurban variability studies with their advantages and disadvantages.

Method	Advantages	Disadvantages	Examples
Pearson Correlation Coefficient	Measures linear relationship between two datasets. Tracks temporal variation.	Does not inform about spatial variability. Sensitive to outliers	Burton et al. (1996); Pinto et al. (2004)
Coefficient of Variation	Informs about dispersion within a dataset.	Very sensitive to outliers.	Houthuijs et al. (2001); Martuzevicius et al. (2004)
Coefficient of Divergence	Tracks spatial homogeneity.	Does not inform about temporal variability.	Wongphatarakul et al. (1998); Pinto et al. (2004)
Absolute Concentration Difference	Provides direct measure of concentration gradient across sites.	Prone to large propagated uncertainty,	Blanchard et al. (1999); Röösli et al. (2004)
Ratio-of-Ratios	Indicates if the quantities are impacted by similar sources.	Undermines drivers of variability at the background site.	Schwab et al. (2004)
Cumulative Probability Function	Relates concentration distribution to wind direction.	Dependent on concentration distribution. Different time scales of mass and wind datasets can cause smearing.	Kim et al. (2003); Begum et al. (2004); Wang et al. (2011)
Nonparametric Wind Regression	Relates concentrations to wind directions. Independent of concentration distribution.	Smoothing parameter used can either add noise or overlook minor sources.	Henry et al. (2002); Wang et al. (2011)
Nonparametric Trajectory Regression	Relates concentrations to air mass locations along the near-field back trajectory	Requires short-term variability in winds and high-time resolution data.	Henry et al. (2011)

(either in its entirety or windowed) at a single site and can be used for examining spatial variability across simultaneously operated monitoring stations for a given sampling event (Houthuijs et al., 2001; Martuzevicius et al., 2004). However, this metric is very sensitive to the extreme values and is typically used to gauge dispersion within a distribution (either in concentration time series of a site or for a sampling event across multiple sites). Coefficient of Divergence (COD), defined as the root mean square of the ratio of sample-specific difference to its sum, is also used as a scale to gauge the concentration variability (Wongphatarakul et al., 1998). It is usually compared to other COD values to gauge spatial variability in a relative sense (Pinto et al., 2004). COD < 0.20 are often associated with spatial homogeneity while higher values are associated with relative spatial heterogeneity (Wilson et al., 2005). However, a context for the absolute COD value has not been provided and thus the extent of variability cannot be inferred from its value alone.

1.1.2. Other approaches for interpreting variability

In addition to the aforementioned metrics, various other approaches have been used for gauging intraurban variability. A common approach is to calculate inter-site concentration differences as they provide a direct measure of excess at one site compared to the another (Blanchard et al., 1999; Noble et al., 2003; Nerriere et al., 2005). Kim et al. (2005) observed significant spatial variability in source contributions identified from the receptor modeling of $PM_{2.5}$ mass across ten monitoring sites in St. Louis, even though $PM_{2.5}$ mass was fairly well correlated. However, comparing average concentration differences alone disregards valuable information in the distribution of differences that can be used to examine the underlying drivers for variability. Species contributed by local sources can display discernible variation but the interpretation of variability in species dominated by regionally transported contributions can be confounded by

the often relatively large propagated uncertainty for the concentration differences (Roosli et al., 2001).

Schwab et al. (2004) examined seasonal variability across urban and rural sites in the New York state by defining a "ratio-of-ratios" (RR) metric. For a given species, the ratio of seasonal average at a given site to the seasonal average at a rural site was defined as the "base ratio". The site- and season-specific base ratios for each species were divided by the corresponding base ratio for $PM_{2.5}$ mass to gauge variability in site- and season-specific species concentrations relative to $PM_{2.5}$ mass. Species with significant local influences would be expected to have RR values greater than unity, while species dominated by regional transport would be less than unity. By interpreting RR values in a relative sense, the study examined spatial uniformity of sulfates and seasonal variability of nitrates. While RR values outside 0.6-2 were deemed significantly different from unity, values within 0.6-2 were more challenging to interpret.

Goldman et al. (2010) assessed spatial variability of twelve $PM_{2.5}$ pollutants in the Metropolitan Atlanta area by calculating modified semi-variances, defined as the ratio of one-half of the variance of the two inter-site concentration differences normalized by the average of those two observations for each species. For correlated observations the modified semi-variances approached zero and unity for uncorrelated observations. By plotting the logarithmic semi-variances against the corresponding distance between the two monitoring sites, the estimated range for secondary and primary pollutants were determined to be consistent with the respective zone of influence.

1.1.3. Identifying emission sources using meteorological datasets

The coupling between observed concentrations with surface and synoptic winds can provide considerable insight into the impact of emission sources. Even though the wind direction and speed often vary throughout the sampling duration, various tools building upon the simple 'pollution rose' have been developed to take advantage of meteorological data to determine the directionality of possible sources. Calm winds (operationally, wind speeds below the starting threshold of the sensor such as ~1 m/s) are usually removed from the analysis since stagnation or sloshing of air masses leads to elevated concentrations which cannot be tracked back to any particular source through the surface winds data. Long-range transport of pollutants, i.e. contributions from emission sources located at large distances compared to the spatial scale of an urban-scale network, is expected to result in uniform contributions at all monitoring sites. Concentration variation with synoptic winds are often examined using air mass back trajectories to identify emission source regions (Davis et al., 2003; Allen and Turner, 2008; Robinson et al., 2011). In general, each observed/modeled concentration at the receptor location is assigned to every grid in the spatial domain along the corresponding back trajectory and combined using some statistical measure to obtain a conditional probability field based on the residence time of air masses (Stohl, 1998). Various approaches have been used to identify the source regions for concentrations over continental scales (Stohl, 1996; Keeler and Samson, 1989; Polissar et al., 2001; Zhou et al., 2004; Scheifinger and Kaiser, 2007).

A variety of approaches are utilized to identify the directionality of local sources leading to high concentrations at a receptor site. One such tool is the Conditional Probability Function (CPF) plot, which displays the fraction of samples from each discrete wind sector that have concentrations at the receptor site higher than a predetermined threshold concentration (Kim et

al. (2003) based on a conditional probability framework described by Ashbaugh et al. (1985)). The threshold concentration is typically chosen based on the distribution of receptor concentrations such as the top 25th percentile. The putative source bearings obtained from CPF plots can be smeared across wind sectors if there is mismatch between the time resolution of the meteorological data and pollutant data (e.g., 24-hour integrated samples with hourly winds). In the ideal case, the pollutant samples would be collected with time direction shorter than the characteristic time for persistent changes in the wind direction. Pollutant and concurrent meteorological measurements at fine resolution, if available, provide robust identification of the influential sources (Davidson et al., 2009). While such synchronization is often not practical, CPF plots have been used to identify source locations in various PM data analysis and source apportionment studies (Kim and Hopke, 2004; Begum et al., 2004).

Non-parametric wind regression (NWR) is a method of quantifying the relationship between the dependent (concentration) and predictor (wind direction) variables without making any assumptions about the functional form of the relationship or the statistical distribution of the data (Henry et al., 2002). By using NWR with Gaussian kernel as a non-subjective alternative to the usual discrete wind sector method, the directionality of peak concentrations at the receptor sites can be estimated. The optimized width of the kernel function (the smoothing parameter) plays a significant role in determining the bearings of the concentration peaks. A large smoothing parameter results in smoother curves and some peaks corresponding to real sources may be lost or not resolved, while a small smoothing parameter can lead to peaks dominated by noise and/or large peaks being resolved into multiple false peaks. Confidence intervals for NWR plots, generated using a bootstrap method, are recommended to examine the validity of the NWR peaks. NWR plots are often utilized for determining the directionality of the emission sources

impacting a receptor site (Henry et al., 2002; Yu et al., 2004; Henry 2008; Wang et al., 2011). Henry et al. (2002) used the estimated peak locations in NWR plots of cyclohexane concentrations measured at two sites to triangulate the likely source *location*. However, such triangulation has not been attempted in an area with a complex set of emission sources. Similar to CPF plots, NWR plots can also be noisy if the time resolution of the meteorological dataset and the measured mass are not same. Henry et al. (2011) proposed a nonparametric trajectory regression (NTA) to utilize high time resolution concentrations and local wind trajectories to estimate emission source locations, but its applicability to various scenarios is yet to be tested.

1.1.4. Variability from measurement error

Measurement error results from differences between measured and true ambient concentrations and is inherent to ambient air pollution monitoring. Errors can occur because of instrument and/or spatial errors which is the inability of a single time-series to accurately represent the ambient levels throughout the study area (Wade et al., 2006). For example, similar time-averaged concentrations at two sites may suggest spatial uniformity but if the concentration differences on finer time resolutions cannot be explained by measurement uncertainty, then there could be spatial variability between the sites which is masked by the time averaging (Turner, 2008). Furthermore, observed inter-site concentration differences that can be explained by precision estimates of the measured concentrations may not lead to robust estimation of spatial and temporal variability. Thus, datasets that capture a range of concentrations with high precision in data reporting are preferred for such analysis to be representative of true variability. Uncertainty estimates of the species concentrations, when available, enable better understanding of concentration differences that can prevent misclassification of variability.

1.2. Thesis objectives and structure

Air quality studies performed during the course of this dissertation are summarized in Appendix A. The main body of this dissertation is structured to focus on the unifying theme of this work which is the development and evaluation of strategies for assessing intraurban spatial variability in PM and its components that can be applied to urban- and finer-scale monitoring networks. In particular, three networks with markedly different characteristics are examined (Table 1-3). The common approach for analyzing these datasets is to evaluate variability in concentrations across the monitoring sites by utilizing one-or-more metrics/approaches and then to identify the probable emission source locations/zones that contribute to the variability across the network. The approach has been iterative with lessons learned from one study used to inform reanalysis of the other studies. A systematic approach to identifying and interpreting variability in concentrations from a network of urban- and finer-scale monitoring sites has emerged and its advantages and limitations are examined for a variety of networks.

Table 1-3. Characteristics of monitoring networks examined in this dissertation.

Network			Measurements		
Location	Spatial scale	Sites	Parameters	Duration	Resolution
St. Louis, MO, USA	Urban	4	PM_{10} air toxic metals	One year (2008)	24-hour integrated 1-in-3 days
Dearborn, MI, USA	Neighborhood	3	$PM_{2.5}$ EC/OC	Three months (Sep-Dec, 2008)	Hourly continuous
Hong Kong, China	Urban	14	PM_{10} Mass	Ten years (1998-2008)	Hourly continuous
		10	PM_{10} Species	Ten years (1998-2008)	24-hour integrated 1-in-6 days
		4	$PM_{2.5}$ Species	Three years (2000-01; 2004-05; 2008-09)	24-hour integrated 1-in-6 days

The first objective of this dissertation was to develop a systematic approach to combine statistical metrics used for gauging spatiotemporal variability in concentrations. In Chapter 2, air toxic metals measured across a four site urban-scale network in St. Louis, MO, are utilized to introduce a graphical tool for interpreting spatiotemporal variability using correlation coefficient and coefficient of divergence. Next, an approach is proposed to apportion the measured concentrations at each site into a network-wide baseline and site-specific excess concentrations to semi-quantitatively differentiate local-scale emission source contributions from sources exerting influence over larger spatial scales. Finally, concentration-wind direction relationships are examined for estimating the emission source locations/zones influencing the observed variability in concentrations. Strengthens and limitations of these approaches are examined using various air toxic metals species that were measured.

Chapter 3 extends this methodology to PM carbon concentrations measured at high time resolution across a three site neighborhood-scale network in an industrial area in Dearborn, MI. This area has several emission sources and the emissions are likely intermittent. These factors complicate the identification of putative sources driving the observed variability. Chapter 4 presents the assessment for an urban-scale air quality network in Hong Kong with multiple PM mass and speciation datasets collected over eleven years at fourteen sites. The methodology contributes to a weight-of-evidence approach for interpreting spatial and temporal variability occurring over multiple spatial scales with well characterized concentration-source contribution linkages obtained from conventional source apportionment tools and in some cases evaluated using an emission inventory. The methodology is also utilized to identify the various meteorological and geographic factors that modulate the spatial and temporal PM trends over Hong Kong. Additional information about the source apportionment of Hong Kong speciation

datasets performed during this dissertation work is presented in Appendices B and C. Issues regarding measurement uncertainty estimation and precision in data reporting that can influence interpretation of variability are discussed in Chapter 5 using the collocated species data collected in Hong Kong. Finally, the overall contributions of this dissertation work are summarized in Chapter 6.

The work performed in this doctoral research was the result of collaborations with state, regional and national government agencies, equipment manufacturers, and also academic institutions. Most dissertation chapters are academic papers which are in various stages of peer-review and publication. The content within each manuscript is/will be largely the same as that found in the dissertation chapters and appendices. The ambient PM samples used in Chapter 2 were collected by a four-site network managed by the Missouri Department of Natural Resources (MDNR). The candidate was responsible for the data collection at the Washington University in St. Louis (WUSTL) monitoring site; analytical analysis of samples collected at all four sites; the data analysis; interpretation and writing of the chapter. Chapter 2, co-authored with Prof. Jay Turner, was submitted to *Atmospheric Environment* in October 2013 for the peer-review publication process. The data used in Chapter 3 was collected by the Michigan Department of Environmental Quality (MDEQ) and the candidate was responsible for the data analysis, interpretation and writing of the chapter. Chapter 3, co-authored with Prof. Jay Turner, will be submitted for peer-review publication after minor modifications.

The datasets used in Chapter 4, Chapter 5, Appendix B and Appendix C were collected by the Hong Kong Environment Protection Department (HKEPD). The overall data analysis project was a collaboration between Prof. Alexis Lau Dr. Zibing Yuan of Hong Kong University of Science and Technology (HKUST), the candidate and Prof. Jay Turner of Washington University

in St. Louis, and Dr. Peter Louie of HKEPD. The candidate was responsible for the modeling, interpretation and writing of Chapter 4, Chapter 5 and Appendix C (which is the Supplementary Information for Chapter 4). HKEPD and HKUST collaborated by providing their insights/comments on the interpretation of the analysis performed by the candidate. Appendix B was published in the journal *Atmospheric Environment* in 2013 with Dr. Zibing Yuan (HKUST) as the first author, who conducted Positive Matrix Factorization (PMF) modeling of the PM_{10} dataset and took the lead in writing the paper. The candidate also conducted PMF modeling of the PM_{10} data in parallel to HKUST to support the collaboration, and conducted all of the modeling presented in the Supplementary Information section of Appendix B. The candidate conducted all of the analyses in Chapter 4 (with Appendix C) and Chapter 5 which are in the review process with the co-authors at HKEPD and HKUST and will be submitted for peer-review publication. In accordance with the Washington University of St. Louis standards for completion of the thesis, the candidate is the first author and primary contributor to all four papers embodied by Chapters 2-5 and is the secondary author to the published work presented in Appendix B. The funding sources and other people involved with each project are acknowledged within each chapter.

1.3. References

Allen, D.T. and Turner, J.R.; Transport of atmospheric fine particulate matter: Part 1 – Findings from recent field programs on the extent of regional transport within North America, *Journal of Air and Waste Management Association*, **2008**, 58, 254-264.

Ashbaugh, L.L.; Malm, W.C.; Sadeh, W.Z.; A residence time probability analysis of sulfur concentrations at Grand Canyon National Park, *Atmospheric Environment*, **1985**, 19(8), 1263–1270.

Begum, A.B.; Kim, E., Biswas, S.K.; Hopke, P.K.; Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh, *Atmospheric Environment*, **2004**, 38, 3025–3038.

Blanchard, C.L.; Carr, E.L.; Collins, J.F.; Smith, T.B.; Lehrman, D.E.; Michaels, H.M.; Spatial representativeness and scales of transport during the 1995 integrated monitoring study in California's San Joaquin Valley, *Atmospheric Environment*, **1999**, 33, 4775–4786.

Burton, R.M.; Suh, H.H; Koutrakis, P.; Spatial variation in particulate concentrations within metropolitan Philadelphia, *Environmental Science and Technology*, **1996**, 30, 400–407.

Davidson, C.I.; Chu, N.; Kadane, J.B.; Identifying Likely $PM_{2.5}$ Sources on days of elevated concentration: A simple statistical approach, *Environmental Science and Technology*, **2009**, 43, 2407–2411.

Davis, R.E.; Normile, C. P.; Sitka, L.; Hondula, D.M.; Knight, D.B.; Gawtry, S.P.; Stenger, P.J.; A comparison of trajectory and air mass approaches to examine ozone variability, *Atmospheric Environment*, **2010**, 44, 64-74.

DeGaetano, A.T.; Doherty O.M.; Temporal, spatial and meteorological variations in hourly $PM_{2.5}$ concentration extremes in New York city, *Atmospheric Environment*, **2004**, 38, 1547–1558.

Goldman, G.T.; Mulholland, J.A.; Russell, A.G.; Srivastava, A.; Strickland, M.J.; Klein, M.; Waller, L.A.; Tolbert, P.E.; Edgerton, E.S.; Ambient air pollutant measurement error: Characterization and impacts in a time-series epidemiologic study in Atlanta, *Environmental Science and Technology*, **2010**, 44, 7692-7698.

Henry, R.C.; Changa Y.; Spiegelman, C.H.; Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction, *Atmospheric Environment*, **2002**, 36, 2237-2244.

Henry, R.C.; Locating and quantifying impact of local sources of air pollution, *Atmospheric Environment*, **2008**, 48, 358-363.

Henry, R.C.; Vette, A.; Norris, G.; Vedantham, R.; Kimbrough, S.; Shores, R.C.; Separating the air quality impact of a major highway and nearby sources by nonparametric trajectory analysis, *Environmental Science and Technology*, **2011**, 45, 10471-10476.

Houthuijs, D.; Breugelmans, O.; Hoek, G.; Vaskovi, E.; Mihalikova, E. Pastuszka, J.S.; Jirik, V.; Sachelarescu, S.; Lolova, D.; Meliefste, K.; Uzunova, E.; Marinescu, C.; Volf, J.; de Leeuw, F.; Van de Wiel, H.; Fletcher, T.; Lebret, E. and Brunekreef, B.; PM_{10} and $PM_{2.5}$ Concentrations in Central and Eastern Europe: Results from the Cesar Study, *Atmospheric Environment*, **2001**, 35, 2757–2771.

Ito, K; Xue, N.; Thurston, G.; Spatial Variation of $PM_{2.5}$ Chemical Species and Source-apportioned Mass Concentrations in New York City, *Atmospheric Environment*, **2004**, 38, 5269–5282.

Keeler, G. J.; Samson, P. J., Spatial representativeness of trace element ratios, *Environmental Science & Technology*, **1989**, 23,1358-1364.

Kim, E.; Hopke, P.K.; Edgerton, E.S.; Source Identification of Atlanta Aerosol by Positive Matrix Factorization, *Journal of Air and Waste Management Association*, **2003**, 53, 731–739.

Kim, E.; Hopke, P.K.; Improving Source Identification of Fine Particles in a Rural Northeastern U.S. Area Utilizing Temperature-Resolved Carbon Fractions, *Journal of Geophysical Research*, **2004**, 109, D09204.

Kim, E.; Hopke, P.K; Pinto, J.P.; Wilson, W.E; Spatial Variability of Fine Particle Mass, Components, and Source Contributions during the Regional Air Pollution Study in St. Louis, *Environmental Science and Technology*, **2005**, 39, 4172–4179.

Martuzevicius, D.; Grinshpun, S.A.; Reponen, T.; Gorny, R.L.; Shukla, R.; Lockey, J.; Hu, S.H.; McDonald, R.; Biswas, P.; Kliucininkas, L. and LeMasters, G.; Spatial and Temporal Variations of $PM_{2.5}$ Concentration and Composition Throughout an Urban Area with High Freeway Density—The Greater Cincinnati Study, *Atmospheric Environment*, **2004**, 38, 1091–1105.

Nerriere, E.; Zmirou-Navier, D.; Blanchard, O.; Momas, I.; Ladner, J.; Moullec, Y.Le.; Personnaz, M.B.; Lameloise, P.; Delmas, W.; Target, A. and Desqueyroux, H; Can we use fixed ambient air monitors to estimate population long-term exposure to air pollutants? The case of spatial variability in the Genotox ER study, *Environmental Research*, **2005**, 97, 32–42.

Noble, C.A.; Mukerjee, S.; Gonzales, M.; Rodes, C.E.; Lawless, P.A.; Natarajan, S.; Myers, E.A.; Norris, G.A.; Smith, L.; Ozkaynak, H.; and Neas, L.M.; Continuous measurement of fine and ultrafine particulate matter, criteria pollutants and meteorological conditions in urban El Paso, Texas, *Atmospheric Environment*, **2003**, 37, 827–840.

Pinto, J.P.; Lefohn, A.S; Shadwick, D.S; Spatial variability of $PM_{2.5}$ in urban areas in the United States, *Journal of the Air & Waste Management Association*, **2004**, 54, 440–449.

Polissar, A.V.; Hopke, P.K.; Harris, J.M.; Source regions for atmospheric aerosol measured at Barrow, Alaska, *Environmental Science and Technology*, **2001**, 35, 4214-4226.

Pope, C.A.; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito K.; Thurston, G.D.; Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *Journal of the American Medical Association*, **2002**, 287, 1132–1141.

Robinson, N. H. ; Newton, H. M.; Allan, J. D.; Irwin, M.; Hamilton, J. F.;Flynn, M.; Bower, K. N.; Williams, P. I.; Mills, G.; Reeves, C. E.; McFiggans, G.; Coe, H.; Source attribution of Bornean air masses by trajectory analysis during the OP3 project, *Atmospheric Chemistry and Physics*, **2011**, 11, 9605-9630.

Röösli, M.; Theis, G.; Kunzli, N.; Staehelin, J.; Mathys, P.; Oglesby, L.; Camenzind, M.; and Braun-Fahrlander, C; Temporal and spatial variation of the chemical composition of PM_{10} at urban and rural sites in the Basel area, Switzerland, *Atmospheric Environment*, **2001**, 35, 3701–3713.

Schwab, J.J.; Felton, H.D.; Demerjian, K.L.; Aerosol chemical composition in New York from integrated filter samples: Urban/rural and seasonal contrasts, *Journal of Geophysical Research Atmosphere*, **2004**, 109, D16S05.

Scheifinger, H.; Kaiser, A.; Validation of trajectory statistical methods, *Atmospheric Environment*, **2007**, 41, 8846-8856.

Seinfeld, J.H.; Pandis, S.N.; Atmospheric chemistry and physics: From air pollution to climate change, *Wiley-Interscience*, **2006**.

Stohl, A.; Trajectory statistics - A new method to establish source receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, *Atmospheric Environment*, **1996**, 30, 579-587.

Stohl, A.; Computation, accuracy and applications of trajectories – A review and bibliography, *Atmospheric Environment*, **1998**, 32, 947-966.

Turner, J.R.; Measurement error as a context for assessing intraurban variability in chemical speciation network data, *Proceeding of the 101st Annual Meeting of Air & Waste Management Association*, **2008**, Paper No. 664.

Wade, K.S.; Mulholland, J.A.; Marmur, A.; Russell, A.G.; Hartsell, B.; Edgerton, E.S.; Klein, M.; Waller, L.A.; Peel, J.L.; Tolbert, P.E.; Effects of instrument precision and spatial variability on the assessment of the temporal variation of ambient air pollution in Atlanta, Georgia, *Journal of Air & Waste Management Association*, **2006**, 56(6), 876-888.

Watson, J.G.; Chow, J.C.; Estimating middle- neighborhood-, and urban-scale contributions to elemental carbon in Mexico City with a rapid response aethalometer, *Journal of Air & Waste Management Association*, **2001**, 51, 324–333.

Wang, G.; Hopke, P.; Fu, G.; Identification of major sources of $PM_{2.5}$ in St. Louis Missouri USA, *Journal of Ocean University of China*, **2009**, 8, 101-110.

Wilson, J.G.; Kingham, S.; Pierce, J.; Struman, A.P.; A review of intraurban variations in particulate air pollution: Implications for epidemiological research, *Atmospheric Environment*, **2005**, 39, 6444-6462.

Wongphatarakul, V.; Friedlander, S.K.; Pinto, J.P.; A Comparative study of $PM_{2.5}$ ambient aerosol chemical databases, *Environmental Science and Technology*, **1998**, 32, 3926-3934.

Zhou, L.; Hopke, P.K.; Wei Liu, W.; Comparison of two trajectory based models for locating particle sources for two rural New York sites *Atmospheric Environment*, **2004**, 38, 1955–1963.

Yu, K.N.; Cheung, Y.P.; Cheung, T.; Henry, R.C.; Identifying the impact of large urban airports on local air quality by nonparametric regression, *Atmospheric Environment*, **2004**, 38, 4501–4507.

Chapter 2 :Gauging intraurban variability of ambient particulate matter arsenic and other air toxic metals from a network of monitoring sites

2.1. Abstract

A four site monitoring network was established in the Missouri portion of Metropolitan St. Louis during 2008 to characterize spatiotemporal patterns in PM_{10} arsenic. Arsenic measured at two urban sites in the City of St. Louis was typically higher than arsenic at two suburban sites. Spatiotemporal variability in arsenic is examined by plotting the Pearson correlation coefficient (PCC) against the coefficient of divergence (COD) for each site-pair to merge the temporal tracking ability of PCC with COD's ability to gauge spatial homogeneity. Arsenic measured across the network is apportioned into a network-wide baseline and site-specific excess concentrations to semi-quantitatively differentiate local-scale emission source contributions from sources exerting influence over larger spatial scales. Comparing measured concentrations at each site against the network-wide baseline concentration using a scattergram of PCC and COD emphasizes the impact of local sources on intraurban variability. Conditional probability function (CPF) plots constructed using site-specific measured arsenic and surface winds identify a broad emission source region towards the east, but mask the bearings of local sources in the urban core. CPF plots using site-specific arsenic in excess of the baseline concentrations provide better resolution of local emission source bearings and are triangulated to identify a likely arsenic emission source zone along the industrialized Mississippi Riverfront. Additional air toxic metals

measured in this study (selenium, manganese and lead) are also investigated to examine the efficacy of this methodology to characterize intraurban variability.

2.2. Introduction

The United States Clean Air Act Amendments of 1990 identified 187 hazardous air pollutants (HAPs, or Air Toxics) based on their potential for serious health effects. Acute exposure to air toxic metals such as arsenic can produce multiple organ toxicity, while long-term exposure can be carcinogenic (Duker et al., 2005; USEPA, 2004). The St. Louis Community Air Project (CAP), conducted over the period 2001-05, featured detailed measurements of air toxics in an urban residential neighborhood and identified arsenic as an air toxic metal of concern for 1 in 100,000 increased cancer risks from a 70-year exposure sustained at the observed annual average ambient concentration (USEPA, 2005). However, these ambient particulate matter metals data were derived from Chemical Speciation Network (CSN) protocol measurements for $PM_{2.5}$ and time-average concentrations were sensitive to the method used to impute concentration values below the minimum detection limit. Annual average PM_{10} arsenic measured at Blair St., a National Air Toxics Trends Station (NATTS) site in the City of St. Louis, is of the same magnitude as other such sites across the United States (USEPA, 2007). However, maximum 24-hour arsenic concentrations measured at Blair St. are frequently higher than those measured at the other sites. The poorly characterized influence of local sources and the need for improved arsenic detectability to estimate robust time-average concentrations motivated the examination of short-term spatiotemporal variability of arsenic to identify local sources in the St. Louis region using sampling and analytical methods with higher sensitivity than those provided by the CSN.

Ambient arsenic can originate from natural sources (e.g. soil erosion and volcanic emissions) and anthropogenic sources (e.g. smelting of metals, use of pesticides, and combustion of fuels including coal-fired power plants). However, contributions from anthropogenic sources are estimated to be almost three times more than that from natural sources (WHO, 2001). Air samples collected over a decade at 14 counties along the Texas-Mexico border indicated a rise in ambient arsenic concentrations with rapid industrialization in the region (Shields, 1991). An arsenic speciation study during 2001-02 at the highly industrialized Huelva region in southwestern Spain identified peak arsenic concentrations to occur at the urban background site for winds from the direction of a metal smelting facility (Sánchez-Rodas et al., 2007). $PM_{1.3}$ air toxic metals such as arsenic, selenium and lead, were measured at 30 minute resolution in East St. Louis (USA) during 2001-02 using the Semi-continuous Elements in Aerosol System (SEAS). Wind directions during periods of high species concentration were examined to identify potential emission source regions (Wang et al., 2011).

Many air quality studies have assumed that the spatial distribution of the pollutant is homogeneous within large urban areas and that concentrations at a central site are representative of the entire study area (Burton et al., 1996; Roosli et al., 2001). However, several recent studies examining datasets on finer spatial and temporal scales often indicate greater variation within urban areas than was previously characterized (Pinto et al., 2004; Kim et al., 2005). The spatial zone of representation for observational data collected at a site depends on the spatial scale of influence exerted by emission sources and thus is species-specific. Inadequate accounting of spatial and/or temporal variability can lead to exposure misclassification in epidemiological and exposure studies used to support health-based standards (Wilson et al., 2005).

Summary statistics such as the mean and median concentrations are commonly used to describe variability. Another commonly used metric is the Pearson correlation coefficient (*r* or PCC) which gauges the strength of association between two distributions. PCC is defined in equation (2-1) where, x_{ij} is the concentration for sample *i* at site *j*; $\bar{x}_j$ is the average concentration at site *j*; *j* and *k* are the two sampling sites; and, *p* is the number of paired observations. PCC is the ratio of covariance between the two datasets to the product of their respective standard deviations and is bounded by [-1, +1]. The coefficient gauges association between concentrations at two sites over the entire sampling duration, with greater correlation indicated by values closer to ±1 and low correlation expressed by values closer to zero (DeGaetano et al., 2004; Sajani et al., 2004). While being a good reference metric, it is not a robust measure of spatial variability as it tracks temporal similarity between paired sites, and can have a weak association with the actual spatial homogeneity of concentrations (Pinto et al., 2004). Coefficient of Divergence (COD), defined in equation (2-2), is another commonly used statistical metric (Wongphatarakul et al., 1998). COD utilizes daily inter-site differences to gauge the concentration variability between two sites. COD < 0.20 is often associated with spatial homogeneity while higher values are associated with spatial heterogeneity (Wilson et al., 2005).

$$PCC_{jk} = \frac{\sum_{i=1}^{p}(x_{ij}-\bar{x}_j)\times(x_{ik}-\bar{x}_k)}{\sqrt{\sum_{i=1}^{p}(x_{ij}-\bar{x}_j)^2 \times \sum_{i=1}^{p}(x_{ik}-\bar{x}_k)^2}} \quad (2\text{-}1)$$

$$COD_{jk} = \sqrt{\frac{1}{p}\sum_{i=1}^{p}\left\{\frac{(x_{ij}-x_{ik})}{(x_{ij}+x_{ik})}\right\}^2} \quad (2\text{-}2)$$

PCC and COD can be interpreted either individually or together to understand temporal and spatial variations in species concentrations (Krudysz et al., 2008). A background monitoring

station with ideally no contributions from local source(s) could provide a reference concentration time series to gauge variability across a network. In the absence of a well characterized background site, variability using PCC and COD is typically interpreted in a relative sense between site-pairs (Pinto et al., 2004). These metrics are impacted by several factors influencing the measured concentrations at a site, including the mix of source emissions occurring on one or more spatial scales (e.g. local point sources, urban-scale sources such as traffic, and regionally transported emissions). If concentration levels are dominated by regional transport, there might be only minor variations in the metrics between site-pairs. While this does capture the true concentration relationship between the sites, it also obscures the use of such metrics to gain deeper insights into the drivers for the differences that do exist.

The increasing influence of local source contributions could result in greater variability in concentrations between sites within an urban region. To constrain the geographic location of the local sources, surface winds are typically utilized to identify the bearings of emission source regions. Graphical tools, such as the Conditional Probability Function (CPF) and Nonparametric Wind Regression (NWR) plots, have been used for identifying point source locations in various studies (Henry et al., 2002; Kim et al., 2003; Begum et al., 2004; Lee et al., 2006; Turner 2008). The CPF plot graphs the fraction of samples having a measured concentration higher than a predetermined threshold concentration for each discrete wind sector (Wang et al., 2011). It provides insight into the bearings of emission sources impacting the monitoring site from a perspective that is not biased by the prevailing wind patterns.

To examine the spatial variability in arsenic and other air toxic metals across the Missouri portion of Metropolitan St. Louis, a four site monitoring network was established during the year 2008 under the United States Environmental Protection Agency (USEPA) funded Community

Air Toxics program. In this study (hereafter called the Air Toxics Study), the concentration data are examined to identify local point source(s) contributing to spatiotemporal variability. Exploratory tools are investigated to take advantage of PCC's ability to track temporal similarity and COD's ability to gauge concentration homogeneity between site-pairs. Concentrations measured across the network are utilized to construct a time series for a network-wide baseline concentration to differentiate the relative contribution of local sources from larger-scale sources that impact the entire network. Site-specific peak arsenic concentrations are then combined with meteorological data to identify the local emission source regions using CPF plots, and are evaluated against known arsenic source locations listed in the 2008 National Emission Inventory (NEI) (USEPA, 2008). Other air toxic metals measured during the study are also examined to investigate the strengths and limitations of the proposed methodology for gauging spatiotemporal variability in species concentration and to identify emission source regions using a network of sites.

2.3. Method

2.3.1. Sample Collection and Analysis

Figure 2-1 shows the four site monitoring network where PM_{10} samples were collected at one-in-three day frequency during 2008. Two urban sites, Blair St. and Hall St., were located in the City of St. Louis at ~3 and 7 km north of the urban core, respectively, near the industrialized Mississippi Riverfront. The two suburban sites, Washington University in St. Louis (WUSTL) and Arnold, were located ~10 km west and 25 km southwest of the urban core, respectively. 24-hour integrated ambient particulate matter was collected onto 8"×10" quartz fiber filter sheets (Whatman®, P/N QMA) using PM_{10} HiVol samplers (Graseby Andersen GMW Model 1200) operated at 1.13 m^3/min. Filters folded in half with the deposit side facing inwards were mailed

to WUSTL with no special handling (e.g. no cold shipping) and stored in freezers at the laboratory for analytical analysis. A 1"×8" filter strip sectioned from each filter was immersed in a 70 mL polypropylene vial (Capitol Vial, Inc., Auburn, AL, P/N 03EDM33) with 16 mL of the extraction solution (16% nitric acid and 4% hydrochloric acid) and digested using a MODBlock™ (CPI, Santa Rosa, CA) operated at 95°C for 90 minutes. Samples were cooled for 30 minutes with 10 mL deionized water (DI) and the extract was subsequently diluted to 50 mL. Approximately 12 mL was filtered using nylon syringes with Acrodisk® 25 mm syringe filters (0.45µm nylon membrane, Pall Corporation, P/N 4438) into 15 mL polystyrene conical tubes for analysis. Calibration standards were prepared by diluting concentrated stock solutions (P/N 4400-ICP-MSCS, CPI International) with reagent water (5% nitric acid and 1.25% hydrochloric acid in DI water).

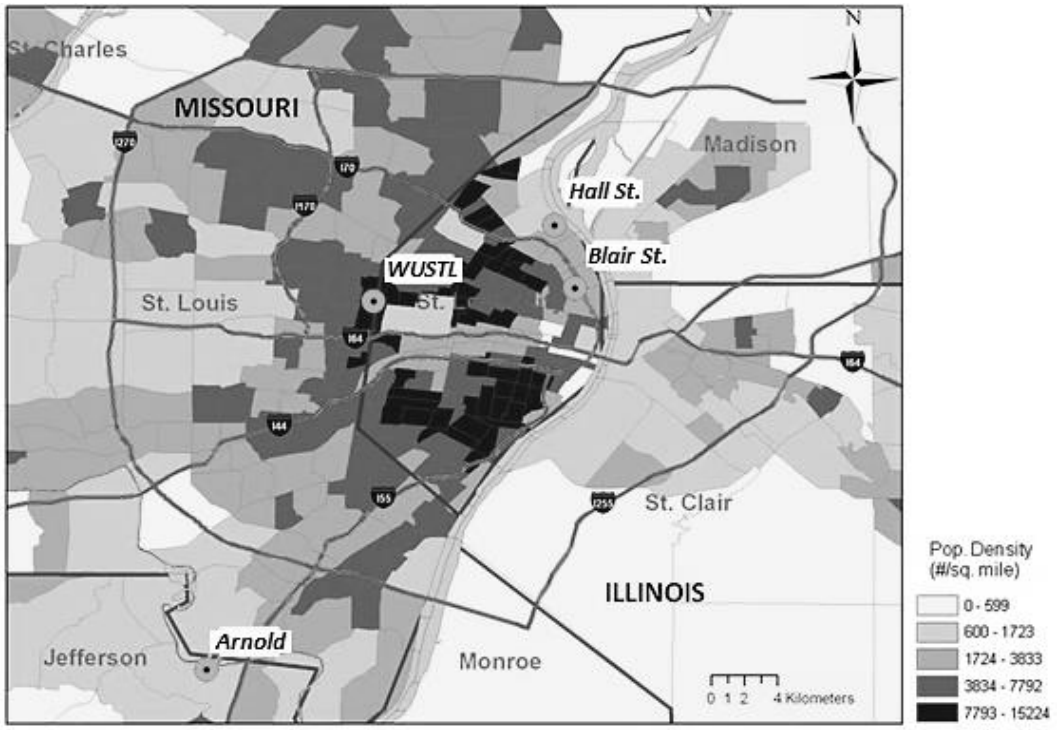


Figure 2-1. The four site monitoring network for the Air Toxics Study during 2008 across the St. Louis Metropolitan area (population density map from 2000).

The extracts were analyzed using an Agilent 7500ce Inductively Coupled Plasma/Mass Spectrometer (ICP/MS). ICP/MS analyses were performed using three quantitative modes with carrier gases Ar/He, Ar/H_2, and Ar only. Interferences by polyatomic ions originating from the carrier gases and sample matrices were corrected as appropriate (Agilent, 2007). For example, arsenic quantification can be confounded by isobaric polyatomic ion interference from argon chloride, with the chlorine originating from complex sample and/or extraction acid matrices (Brown et al., 2004). The ICP/MS instrument used in this work is equipped with a reaction cell (Octopole Reaction System) to suppress such interferences. Quality control measures included field blanks collected every month; method blanks (clean filters); reagent blanks for each batch of samples digested; calibration verification using NIST-traceable multi-element standards from a source separate from the calibration standards; check standards run every ten samples; and internal standards. Ambient PM_{10} elemental concentrations were calculated from the solution concentrations measured by ICP/MS using the total air volume sampled and are reported as ng/m^3 at standard conditions of 25°C and 1 atm. While the analytical method was optimized for arsenic, other air toxic metals (antimony, beryllium, cadmium, chromium, cobalt, lead, manganese, nickel and selenium) measured by ICP/MS were also investigated.

Method recoveries were examined by analyzing NIST standard reference material (SRM) for urban particulate matter (1648a) and coal fly ash (1633b) for air toxic metals measured by ICP/MS. With coal fly ash, arsenic and selenium had 100-105% recoveries, while other metals had recoveries from 45 to 75%. Recoveries for urban particulate matter mass loadings, consistent with the concentration range of the ambient samples, ranged between 115 and 120% for arsenic. The higher recoveries for arsenic could not be explained but the potential for multipath complex reactions cautions that the reaction cell may be subject to a new breed of

isobaric interferences (Baranov et al., 1999). Chromium, cobalt and nickel are excluded from further analysis because of extremely low or unrealistically high recoveries, while other air toxic metals had 90-105% recoveries. Beryllium is not quantified in either SRM and therefore, is also excluded. Instrument detection limits (IDL), used as proxy for the method detection limit, were calculated as three times the standard deviation of seven-to-ten replicate spiked reagent blanks analyzed by ICP/MS on different sample analysis days. Concentration values below the IDL were replaced by 1/2 IDL unless otherwise noted.

2.3.2. Data Characteristics

Routinely collected NATTS air toxic metals data from Blair St., at 1-in-6 day frequency using a separate HiVol sampler, provided an independent check of our data. Analysis of these samples at the Eastern Research Group, Inc. involved filter extraction with 4% nitric acid via sonication for 3 hours followed by ICP/MS (USEPA, 2009). Table 2-1 tabulates the Reduced Major Axes (RMA) regression statistics for metals from the Air Toxics Study against the corresponding

Table 2-1. Reduced Major Axes (RMA) regression statistics for the Air Toxics Study dataset against NATTS dataset analyzed by ICP/MS with 95% confidence intervals for slope and intercept (total 60 concentration pairs). Intercepts and detection limits are in ng/m^3

	Slope	Intercept	R^2	NATTS		Air Toxics Study	
	± 95% C.I.	± 95% C.I.		MDL[a]	N < MDL	IDL	N < IDL
As	1.02 ± 0.12	0.29 ± 0.15	0.80	0.15	0	0.04	0
Cd	1.50 ± 0.10	-0.15 ± 0.14	0.94	0.09	0	0.99	44
Mn[b]	1.13 ± 0.13	0.50 ± 1.58	0.81	0.14	0	0.20	0
Pb	0.95 ± 0.04	0.47 ± 0.79	0.98	0.19	0	1.87	2
Se	1.26 ± 0.07	-0.03 ± 0.08	0.96	-	-	0.03	0
Sb	1.30 ± 0.06	0.24 ± 0.18	0.97	-	-	0.66	11

[a] NATTS nation-wide average species MDL (USEPA, 2010)

[b] Excludes one high concentration pair to prevent bias
(Air Toxics Study, Mn = 584 ng/m^3; NATTS, Mn = 734 ng/m^3)

NATTS samples. For this comparison, values below the MDL (NATTS) and IDL (Air Toxics Study) were used as reported. Despite statistically significant biases in the slopes and intercepts (based on 95% confidence intervals) for many elements, high coefficients of determination (R^2) between the two independently evaluated datasets validates the sample-to-sample variability in the Air Toxics Study data which are exploited in this analysis for examining local source impacts.

All sites had at least 95% data completeness for the 122 scheduled sampling events. Table 2-2 tabulates for each element the site-specific number of valid samples, annual mean and median concentrations, the estimated IDL, and the number of samples below one and three times the IDL. Arsenic, manganese and selenium concentrations at all sites are above 3×IDL for all samples and about 65% of the lead concentrations are above 3×IDL. In contrast, about 77% of antimony concentration values are below 3×IDL and nearly 95% of the cadmium concentrations are below 3×IDL. For cadmium, the few high concentration values coincided with high lead concentrations which suggest a similar emission source or source region(s). The remainder of the analysis focuses on elements with at least 50% of data greater than 3×IDL, i.e. arsenic, lead, manganese and selenium.

2.3.3. Spatiotemporal Analysis

Variability in air toxic metals across St. Louis is examined by comparing summary statistics of concentrations between the sites. Inter-site spatial and temporal relationships are further investigated by calculating PCC and COD. However, calculating these metrics for species with concentrations near or below the detection limit downgrades PCC and inflates COD because of the large uncertainty associated with such concentrations. Further, inclusion of such samples as imputed concentrations breaks the spatial and temporal correlations across the sites and can

Table 2-2. Summary statistics of metal measurements from the Air Toxics Study. All concentrations are in ng/m^3.

			Arsenic (As)				Selenium (Se)			
		Count	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)
Urban	Blair St.	120	1.34 ± 0.81	1.15	0	0	1.21 ± 0.85	0.99	0	0
	Hall St.	119	1.05 ± 0.56	0.89	0	0	1.28 ± 0.80	1.15	0	0
Suburban	Arnold	122	1.01 ± 1.20	0.68	0	0	0.95 ± 0.55	0.74	0	0
	WUSTL	116	0.80 ± 0.63	0.62	0	0	1.00 ± 0.68	0.80	0	0
			Lead (Pb)				Manganese (Mn)			
		Count	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)
Urban	Blair St.	120	14.50 ± 16.00	8.96	2 (2)	25 (21)	19.30 ± 55.40	10.06	0	0
	Hall St.	119	10.40 ± 10.40	7.41	5 (4)	40 (34)	15.10 ± 18.30	11.04	0	0
Suburban	Arnold	122	14.00 ± 21.90	6.67	8 (7)	49 (40)	6.81 ± 4.86	4.98	0	0
	WUSTL	116	8.72 ± 10.10	6.01	15 (13)	53 (46)	8.28 ± 7.13	6.25	0	0
			Cadmium (Cd)				Antimony (Sb)			
		Count	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)	Mean ± Std. Dev.	Annual Median	# < IDL (%)	# < 3xIDL (%)
Urban	Blair St.	120	1.02 ± 1.64	0.50	89 (74)	115 (96)	2.32 ± 3.07	1.36	14 (12)	78 (65)
	Hall St.	119	1.04 ± 1.48	0.50	90 (76)	111 (93)	2.10 ± 3.12	1.50	22 (18)	83 (70)
Suburban	Arnold	122	1.47 ± 3.99	0.50	91 (75)	115 (94)	1.18 ± 2.12	0.90	39 (32)	110 (90)
	WUSTL	116	1.03 ± 2.36	0.50	98 (85)	110 (95)	1.50 ± 1.33	1.16	25 (22)	96 (83)

result in biased values of such metrics. Consequently, only days with concentrations above the IDL at all four sites are utilized in this analysis. This corresponds to 113 sample days for arsenic, selenium and manganese, and 95 sample days for lead for the inter-site comparisons.

If species concentrations across the network are impacted by large-scale sources located far away from the network with no significant local sources within/near the network, the spatiotemporal homogeneity between site-pairs will be indicated by high PCC and low COD. In contrast, the presence of local sources within/near the network leads to varying site-specific contributions and the resulting spatiotemporal heterogeneity will be represented as lower PCC and/or higher COD between site-pairs. However, the interpretation of such inter-site metrics for a network with a large number of monitoring sites can be cumbersome because a network with *n* sites (*n*>1) will result in $\sum_{i=1}^{n}(i-1)$ site-pairs. Calculating these metrics for each site with respect to a reference concentration time series results in a single value per site for each metric and elucidates the spatiotemporal variability across the network relative to the selected reference.

The selected reference time series influences the PCC and COD values and their interpretation and hence needs to be well understood. One approach is to select a central monitoring site that is commonly used to represent air quality conditions across the area. Another approach is to select a well-characterized background monitoring site, if available, to focus on the network behavior relative to a site that presumably captures regional- and larger-scale impacts. In the absence of a representative background site, the network itself can be utilized to generate a baseline concentration to semi-quantitatively differentiate local-scale emission sources from urban- and larger-scale contributions. If emission sources located far away from the network domain equally impact all sites, additional concentrations measured at each site can be attributed to local

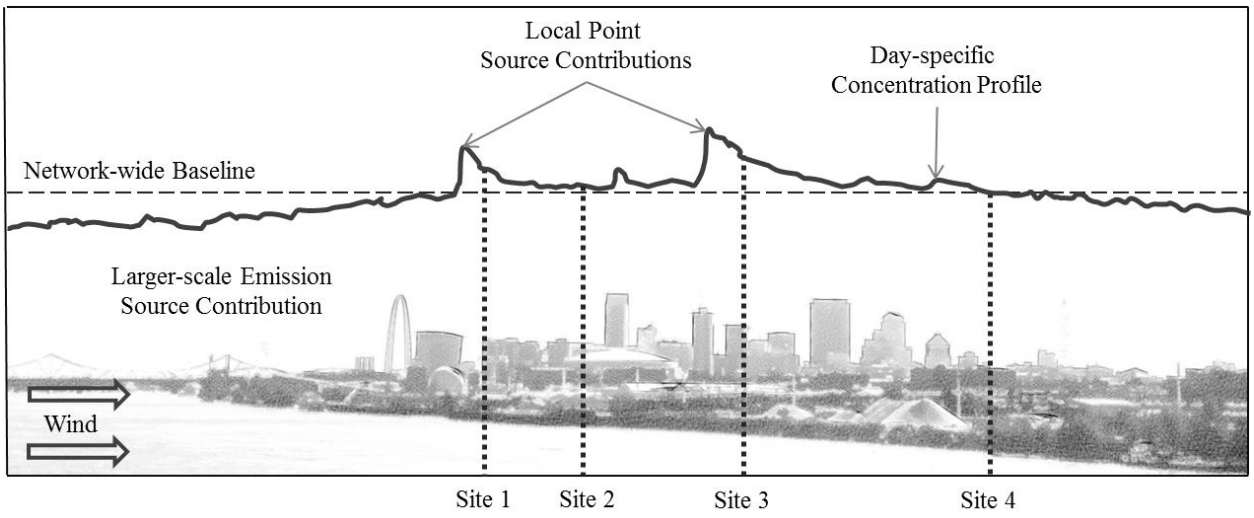


Figure 2-2. Conceptual representation of spatial variability over St. Louis as captured by a four site monitoring network on a particular day. Total concentrations measured at the sites (dotted line) capture contributions from local point sources in addition to the network-wide contributions from larger-scale emission sources. Defining the minimum concentration (Site 4) across the network as baseline (dashed line) enables estimation of local source contributions by at the other sites (Sites 1-3).

sources (Figure 2-2). In principle, local and/or area sources could impact all sites on any day and result in an overestimation of urban/larger-scale contributions, but emissions of air toxic metals are typically associated with point sources. Hence, a *baseline* concentration can be subjectively defined as the minimum or x^{th} minimum concentration across the sites based on the network characteristics to best capture well-understood/interpretable trends in measured concentrations.

The difference between the measured daily concentration at each site and the corresponding baseline (i.e. daily *excess*) is now attributed to contributions from local source(s) impacting the site on that particular day. Hence, site-specific measured or *total* concentrations are apportioned into baseline and site-specific excess concentrations, which semi-quantitatively represent the upper bound for urban/larger-scale contributions and the lower bound for local-scale contributions, respectively. This baseline-excess apportionment approach is applied to the four

site Air Toxics Study network by defining the daily minimum concentration as the baseline concentration for that day. Comparing PCC and COD computed between the baseline and site-specific total concentrations now results in a single pair of metrics for each site in the network.

A scattergram of PCC and COD (i.e., *PCC-COD plot*) can be utilized as a graphical tool for gauging spatiotemporal variability in species concentrations - either between site-pairs (a site-site plot) or for site-specific concentrations against a reference time series (a site-baseline plot). This tool combines the temporal tracking ability of PCC with COD's ability to gauge spatial variability across site-pairs. General characteristics of the *PCC-COD plot* are shown in Figure 2-3. The presented analysis is confined to PCC ≥ 0 for visual clarity because negative correlations were not observed in this dataset, though the methodology can be extended to include negative correlations. A relative measure of inter-site spatiotemporal variability can be gauged based on the sites' position away from absolute spatiotemporal homogeneity (PCC = 1

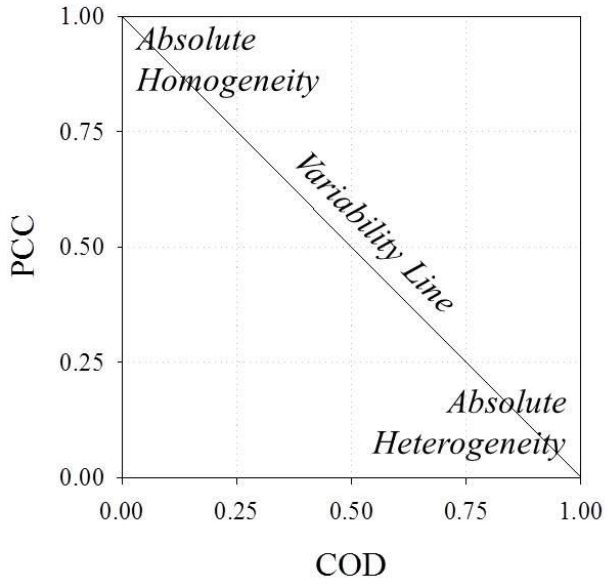


Figure 2-3. A *PCC-COD plot* highlighting the key characteristics of the plot: *variability line*, which spans absolute spatiotemporal homogeneity (PCC = 1 and COD = 0) and absolute spatiotemporal heterogeneity (PCC = 0 and COD = 1).

and COD = 0), in reference to the line spanning absolute homogeneity and absolute spatiotemporal heterogeneity (PCC = 0 and COD = 1) i.e. the '*variability line*'. If the concentration time series at two sites are dominated by larger-scale contributions, then the site-pair will be positioned closer to the homogeneity end of the plot. The position of the sites that are impacted by local sources can depend on many factors, but in general will be further away from the homogeneity end. For example, concentrations measured at two sites impacted simultaneously by the same sources will display higher correlation but the COD will increase with increasing spatial separation between the sites; thereby placing the sites to the right of the variability line. In contrast, if the sources are located between the two sites, the concentrations measured will display lower correlation with COD varying based on the concentration differences. Overall, local source contributions to the ambient concentrations (site-specific total, excess or even baseline) depend on many factors including but not limited to, the number of sources, their spatial extent of influence, emission patterns of the sources, location of monitoring sites and, meteorological conditions (Pinto et al., 2004). Thus, the positioning of the sites on the *PCC-COD plot* represents the admixing of such influences, while the *variability line* provides a reference for gauging inter-site variability in the context of absolute spatiotemporal homogeneity and heterogeneity.

The site-baseline *PCC-COD plot* enables a particularly manageable inter-site comparison because each site in the network is now represented by a single data point on the plot. Further, this plot also provides insight into the factors defining the baseline concentrations and thus lends to meaningful interpretation of spatiotemporal variability relative to the baseline. For species characterized by impacts from large-scale sources located far away from the network, the site with least impact from local sources within/near the network will chiefly define the baseline

concentrations and hence will be positioned closer to the homogeneity end of the *variability line*. Sites with greater variability in species concentrations, e.g. resulting from the increasing influence of local source contributions, will be positioned further towards the heterogeneity end of the *variability line* and can be used to estimate the location of emission sources.

2.3.4. Source Region Identification

To identify the bearings of local sources of air toxic metals, CPF plots are constructed using site-specific total concentration and compared with the corresponding plots for excess concentrations. The threshold for defining a "high concentration day" is chosen as the top 25^{th} percentile of the concentration distribution. By using data from the top 25^{th} percentile, CPF plots avoid the noise associated with including data closer to detection limit as well as the increased uncertainty in excess concentration resulting from the difference of two measurements. Hourly metrological data for Saint Louis Lambert International Airport was obtained from the National Climatic Data Center (NCDC/NOAA). For computing the conditional probability, the daily averaged species concentration value is assigned to each hourly wind direction for that respective day. Such assignment and confounding by high concentrations caused by stagnant conditions can result in smearing of a given source impact across off-plume wind directions that also occur on days with high concentration plume impacts. Nonetheless, these plots can be a powerful tool to determine the bearing of sources leading to high concentration days. The CPF plots are geo-referenced in Google Earth® to triangulate the source emission regions and are evaluated against known source locations of air toxic metals from the 2008 NEI.

2.4. Results and Discussions

2.4.1. Arsenic

Annual average arsenic at the two urban sites is generally higher than the suburban sites (Table 2-2). The annual mean arsenic at Blair St. is ~67% higher than at WUSTL and ~30% higher than at Hall St. and Arnold. While the annual medians at the urban sites are also higher than that the suburban sites, the medians at the suburban sites are similar with 0.62 ng/m^3 at WUSTL and 0.68 ng/m^3 at Arnold. Such differences, which lead to different interpretations of spatial variability, are investigated in Figure 2-4 (a) by plotting site-specific excess arsenic against the baseline concentrations. Excess arsenic at the urban sites is considerably higher than at the suburban sites and is also uniformly distributed across the entire range of baseline concentrations. Nearly 50% of the annual average excess arsenic at WUSTL arises from only five high concentration days. Despite the relatively large geographic separation between the suburban sites, arsenic at Arnold and WUSTL is well correlated (PCC = 0.79) and primarily defines the baseline concentrations. In contrast, arsenic concentrations at the Blair St. and Hall St. urban sites, which are separated by only ~4 km, display greater variability (PCC = 0.49). These trends are also captured in the site-site *PCC-COD plot* with the suburban site-pair positioned closer to the homogeneity end than the urban pair (Figure 2-5 (a)). However, variations between the site-pairs can be meaningfully interpreted from these metrics for any three sites only in reference to the fourth site. For instance, with respect to WUSTL, Arnold has higher PCC (0.80) and lower COD (0.19) than the corresponding values for Blair St. (PCC = 0.43 and COD = 0.36) and Hall St. (PCC = 0.65 and COD = 0.25), indicating greater homogeneity between WUSTL and Arnold than between WUSTL and the urban sites. However, with respect to Hall St., the other three sites have similar CODs (0.25-0.26) and varying

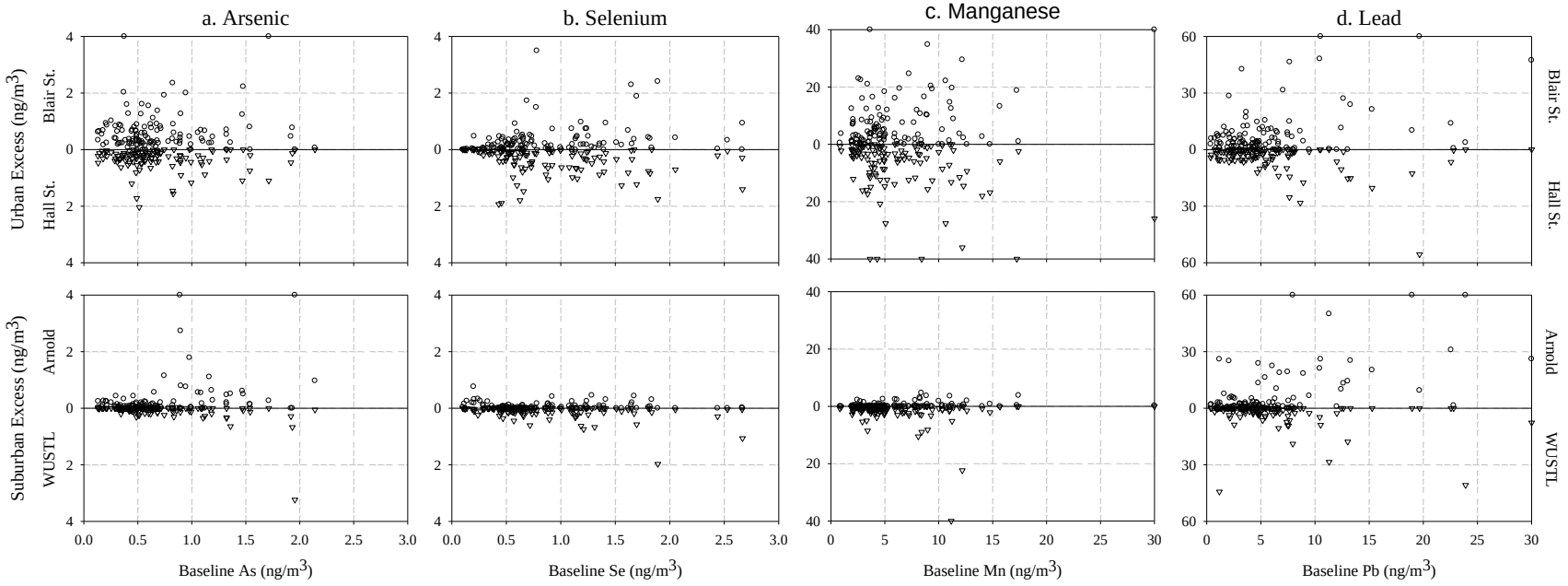


Figure 2-4. Distribution of site-specific excess concentrations with respect to corresponding baseline concentrations at urban (top panel) and suburban (bottom panel) sites for (a) arsenic, (b) selenium, (c) manganese and (d) lead. Concentrations beyond the axis range are plotted on the graph edges.

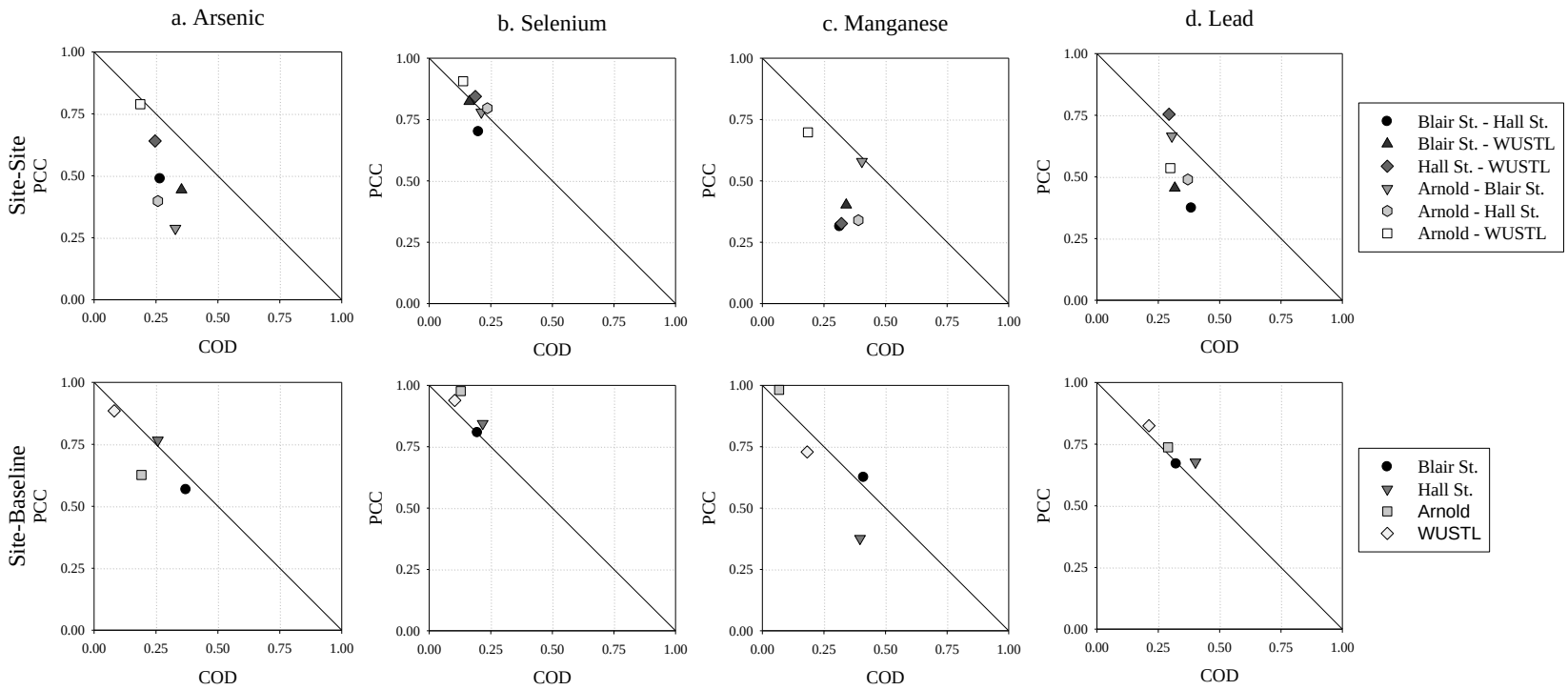


Figure 2-5. *PCC-COD plots* for concentrations between the site-pairs (top panel) and for site-specific total concentrations against the baseline concentrations (bottom panel) for (a) arsenic, (b) selenium, (c) manganese and (d) lead.

correlation, confounding the interpretation of arsenic variability across the network. In contrast, the site-baseline *PCC-COD plot* provides a more accessible understanding of spatiotemporal variability. WUSTL is positioned closest to the homogeneity end of the *variability line*, as the baseline is primarily defined by arsenic measured at WUSTL. Arnold, Hall St. and Blair St. are incrementally positioned further away from the homogeneity end of the variability line, indicating higher spatiotemporal variability from the increasing influence of local arsenic emission sources at these sites.

CPF plots of baseline and site-specific excess arsenic are shown in Figure 2-6 (a). CPF plots of total arsenic at each site (urban sites shown in Figure 2-7 (a)) collectively indicate emission source bearings to the east of St. Louis Metropolitan area, consistent with the regional transport of arsenic from the eastern United States including coal-fired power plants located along the Ohio River Valley (Allen et al., 2008). These eastward bearings are also captured by the CPF plot of the baseline concentration, which accounts for almost 90% of the arsenic at WUSTL, but only 50-70% at the remaining sites. CPF plots constructed using excess concentrations provide better resolution of putative local arsenic emission source bearings. For example, Figure 2-7 shows the CPF plots of total and excess arsenic at the two urban sites. While the CPF plots for total arsenic at the these sites are dominated by bearings consistent with regional transport, excess concentrations at the urban sites triangulate to an emission source zone near the industrialized Mississippi Riverfront that is north/northeast of Blair and southeast of Hall. Arsenic sources consistent with this zone, and listed in the 2008 NEI, include a chemical manufacturing plant with a coal-fired industrial boiler and a municipal wastewater treatment plant with a sludge incinerator. Another resolved emission source zone - northwest of Blair St.

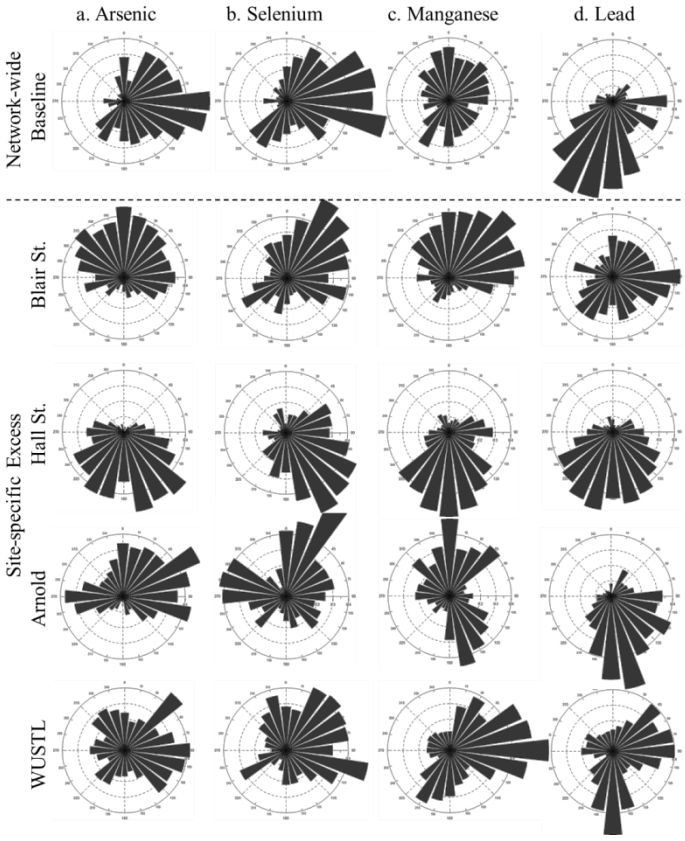


Figure 2-6. CPF plots of the network-wide baseline and site-specific excess concentrations for (a) arsenic, (b) selenium, (c) manganese and, (d) lead.

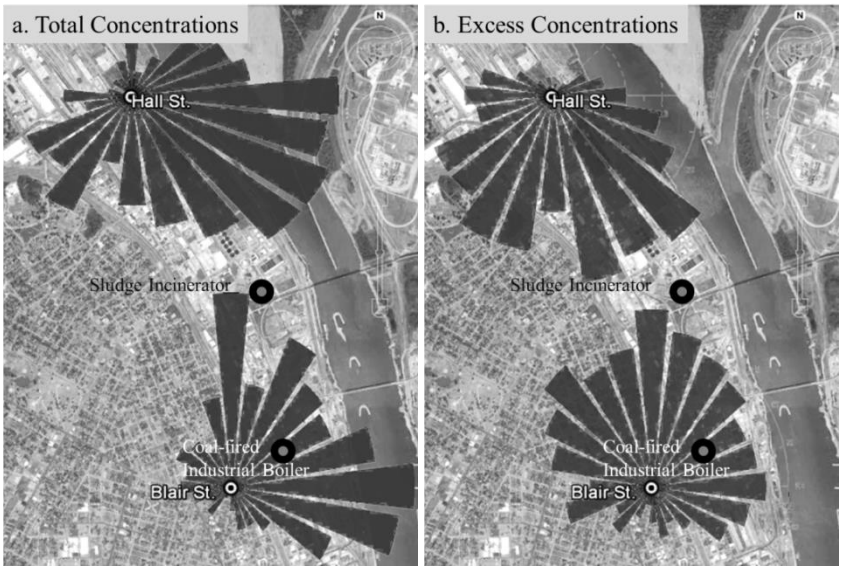


Figure 2-7. CPF plots of (a) total and (b) excess arsenic at the urban sites. The plots and arsenic point sources listed in 2008 NEI are geo-referenced in Google Earth™.

and southwest of Hall St. - is not matched to any known arsenic source in the 2008 NEI and will be discussed in Section 2.4.3. Additional bearings from the CPF plot of excess arsenic at Arnold are also consistent with a large coal-fired electric utility power plant located ~40 km west of Arnold and listed as an arsenic source in the 2008 NEI.

2.4.2. Selenium

Arsenic and selenium emitted during coal pyrolysis and combustion are often used as markers for long-range transport of emissions from coal-fired power plants (Zeng et al., 2001). Similar to the arsenic variability across the network, annual median selenium at Blair St. and Hall St. urban sites are ~30% and 50% higher than at the suburban sites, respectively (Table 2-2). The excess selenium at the urban sites are also uniformly scattered across the entire range of baseline concentrations which are primarily defined by selenium measured at the suburban sites (Figure 2-4 (b)). But unlike arsenic, the relatively high PCC (0.7-0.9) and low COD (0.15-0.25) between selenium across all site-pairs leads to tight clustering towards the homogeneity end of the *variability line* in the site-site *PCC-COD plot* (Figure 2-5 (b)). This pattern suggests a dominant impact from urban- and larger-scale emission sources which masks the subtle differences between the urban and suburban sites. In the site-baseline *PCC-COD plot*, the positions of the suburban sites towards the homogeneity extreme of the *variability line* further confirms that the selenium baseline is predominantly defined by the suburban sites. However, relatively higher variability at the urban sites is now emphasized by their positions farther away from the homogeneity end of the *variability line*, indicating the presence of local selenium source(s) in the urban core.

CPF plots of baseline and excess selenium at all four sites are shown in Figure 2-6 (b). Similar to arsenic, CPF plots of total and baseline selenium concentrations again indicate bearings

consistent with the regional transport from the eastern United States. Selenium apportioned to the baseline is ~90% of the selenium measured at the suburban sites but only ~70% at the urban sites. CPF plots using excess concentrations at the Blair St. and Hall St. sites again provide better resolution of local emission source bearings that triangulate to the zone with the coal-fired industrial boiler and the municipal wastewater treatment plant incinerator. While arsenic and selenium share similar emission source regions, differences in species emission profiles from individual sources could lead to the observed differences in spatiotemporal variability. Overall, the uniformity in selenium concentrations across the network provides insight into the efficacy of *PCC-COD plots* at capturing the behavior of regionally transported species, as well the advantage of using the baseline-excess apportionment approach for identifying local emission source regions.

2.4.3. Manganese

Manganese exhibits a different intraurban variability pattern than that of arsenic and selenium which have discernible contributions from both local and larger-scale sources. Annual median manganese concentrations at Blair St., Hall St. and WUSTL are higher than at Arnold by ~102%, 121% and 25%, respectively (Table 2-2). Excess manganese concentrations at the urban sites and WUSTL are higher and more uniformly distributed than at Arnold, with maximum excess at the urban sites exceeding 100 ng/m^3 in contrast to a maximum excess of 5 ng/m^3 at Arnold (Figure 2-4 (c)). In addition, Arnold defines the baseline for ~70% of the sampling days which suggests it is effectively the reference site for manganese measurements across this network. This behavior is again better captured by the site-baseline *PCC-COD plot* than the site-site *PCC-COD plot* (Figure 2-5 (c)). With respect to the network-wide baseline, Arnold is positioned at the homogeneity extreme of the *variability line* (PCC = 0.98, COD = 0.07) with increased

variability at WUSTL and the urban sites. Similar COD but varying PCC between the two urban sites indicates these sites experience temporally different but on average spatially similar impacts from local manganese source(s).

Figure 2-6 (c) shows the CPF plots of baseline and site-specific excess manganese. The CPF plots of baseline and total manganese at Arnold lack clear directionality and do not provide discernible bearings for larger-scale emission contributions. While the CPF plots of excess concentration at Arnold has distinct features, this excess is on average only 7% higher than the baseline and thus its interpretation is confounded by the relatively high propagated measurement uncertainty. CPF plots constructed using the site-specific excess manganese at the remaining sites indicate the same source bearings as the corresponding total concentrations because high manganese contributions at these sites are governed primarily by local emission sources. Bearings from the CPF plots at Blair St. and Hall St. triangulate towards Granite City, IL which is the location of a large steelworks facility and other associated emission source(s). Additionally, both manganese and arsenic CPF plots for the urban sites also triangulate a region north/northwest of Blair St. and south/southwest of Hall St., though no corresponding sources are listed in 2008 NEI for the region. Perhaps one or more small industrial facilities, such as a metals workshop, located in the triangulated zone sporadically emit metals but may not have sufficient annual emissions to be captured in the NEI. Further, CPF plots constructed using 24-hour integrated mass against hourly meteorological dataset results in smearing of impacts along the bearings corresponding to changing wind direction on days with peak metal concentrations. Using high time resolution concentration measurements (e.g. hourly) can reduce the smearing associated with variations in the surface winds for better estimation of local source bearings. For example, two-hour integrated arsenic measurements at Blair St. also identified this

north/northwest bearing with impacts only during daytime hours on weekdays (Yadav et al., 2009).

2.4.4. Lead

The urban sites also display elevated lead concentrations with the lowest values observed at WUSTL. However, total lead at Arnold are comparable to those at the two urban sites, with the excess at Arnold frequently higher than at the urban sites (Figure 2-4 (d)). The site-site *PCC-COD plot* demonstrates a narrow range of spatial variability and a broad range of temporal variability, but lacks the distinguishing feature exhibited by other air toxic metals of greater spatiotemporal homogeneity between the suburban site-pair than the urban site-pair (Figure 2-5 (d)). However, the site-baseline *PCC-COD plot* - where the baseline defined by ~50% samples from WUSTL and 15-25% each from the remaining three sites - aligns the sites along the *variability line* and provides insight into the minor spatiotemporal variations in lead across this network. WUSTL now lies relatively closer to the homogeneity end of the *variability line*, while the remaining sites display marginally higher heterogeneity. Further, the CPF plots of concentrations at all four sites (total, excess or baseline; the latter two are shown in Figure 2-6 (d)) indicate a major lead source south of the network, consistent with the location of a large primary lead smelter listed in the 2008 NEI. It is among the world's largest lead smelting facilities and its location only ~20 km south of Arnold (i.e., ~30 km south of the St. Louis urban core) leads to a spatial gradient in its contribution across the network. Depending on meteorological conditions, the smelter plume impacts one or more sites with high lead concentrations which are captured by both the baseline and site-specific total concentrations. Hence, CPF plots using excess lead at all sites are still dominated by the southward bearings of the smelter and could mask the bearings of local sources in the urban core. Additional bearings

at Blair St. and WUSTL indicate another source due east of the network, but its bearings at Hall St. are masked by the southward bearings of the lead smelter. In regards to this network, high lead contributions from the smelter are neither near enough to be characterized as local nor far enough from the network to result in network-wide uniform contributions. This case violates the assumptions inherent to the baseline-excess apportionment approach and prevents identification of any local lead source(s) in the urban core.

2.5. Conclusions

24-hour integrated PM_{10} samples were collected at four sites across the St. Louis Metropolitan region during 2008, on a 1-in-6 day schedule and analyzed for air toxic metals. Spatiotemporal variability in arsenic and other air toxic metals was examined by developing intuitive tools for data analysis. The *PCC-COD plot* graphically merges the temporal and spatial tracking ability of these metrics. The cautions that should be applied to the calculation and interpretation of PCC and COD also apply to these plots. For example, concentration values near or below the detection limit must be carefully handled because they can dramatically influence these metrics and result in misinterpretation. Relative spatiotemporal variability between the sites is gauged by their positions on the plot in reference to the *variability line*, which spans absolute homogeneity to absolute heterogeneity. While the PCC and COD for these site-pairs are the intrinsic measures of variability, a network of *n* sites results in $\sum_{i=1}^{n}(i-1)$ site-pairs which complicates their interpretation. This dimensionality was reduced to one data point per site by calculating the PCC and COD for the concentrations measured at each site compared to a reference concentration time series. Concentrations measured at each site were apportioned into daily network-wide baseline and site-specific excess concentrations to semi-quantitatively represent contributions from urban- and larger-scale emission sources and local-scale emission

sources, respectively. It is crucial to have an understanding of the network-wide baseline to meaningfully interpret the drivers of variability at each site compared to the baseline. The resulting site-baseline *PCC-COD plot* provides a framework for interpreting inter-site spatiotemporal uniformity relative to this baseline time series. For most air toxic metals, the suburban sites were closer to the homogeneity end of the *variability line* than the urban sites. Greater spatiotemporal variability at the urban sites with respect to the baseline reflects the stronger influences of local sources at these sites.

This apportionment, together with the CPF plots, demonstrated that arsenic and selenium in St. Louis are dominated by urban- and larger-scale influences such as regional transport from coal-fired power plants in the eastern United States. Large contributions from urban/larger-scale influences confounded the use of CPF plots constructed using measured concentrations to identify the bearings of local sources. However, the bearings of local sources were better resolved using CPF plots based on the excess concentrations and triangulated to an area along the industrialized Mississippi Riverfront as a source of arsenic and selenium emissions. The baseline-excess apportionment approach was less useful for lead because a very large point source located outside the network boundary causes spatial gradient in lead contributions across the network and masks the bearings of any local lead sources. Resolving local emission source zones could be further refined by using measurements at high time resolution (e.g., hourly) to reduce the smearing associated with changing wind direction. Overall, the monitoring network enabled better understanding of intraurban variability in arsenic and other air toxic metals over the St. Louis region that could not be characterized by measurements at a single site.

2.6. Acknowledgement

The authors gratefully acknowledge the Missouri Department of Natural Resources (MDNR) Air Pollution Control Program, Air Quality Analysis Section and particularly Jerry Downs and Terry Rowles, and also the field monitoring support provided by the MDNR Environmental Services Program staff and especially Celeste Koon. The assistance of Timothy Herschbach and James Wang (WUSTL) is also acknowledged. Although this research has been funded wholly or in part by the USEPA through grant XA987912-01 to the MDNR, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

2.7. References

Agilent Technologies; *ICP-MS 3700ce Operating Manual*, **2007**.

Allen, D.T.; Turner, J.R.; Transport of atmospheric fine particulate matter: Part 1—Findings from recent field programs on the extent of regional transport within North America, *Journal of the Air and Waste Management Association*, **2008**, 58, 254-264.

Baranov, V. I.; Tanner, S. D.; A dynamic reaction cell for Inductively Coupled Plasma Mass Spectrometry (ICP-DRC-MS) Part I. The rf-field energy contribution in thermodynamics of ion–molecule reactions, *Journal of Analytical Atomic Spectrometry*, **1999**, 14, 1133.

Begum, A.B.; Kim, E.; Biswas, S.K.; Hopke, P.K.; Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh, *Atmospheric Environment*, **2004**, 38, 3025-3038.

Brown, R.J.C.; Yardley, R.E.; Brown, A.S.; Milton, M.J.T.; Sample matrix and critical interference effects on the recovery and accuracy of concentration measurements of arsenic in ambient particulate samples using ICP-MS, *Journal of Analytical Atomic Spectrometry*, **2004**, 19, 703-705.

Burton, R.M.; Suh, H.H.; Koutrakis, P.; Spatial variation in particulate concentrations within Metropolitan Philadelphia, *Environmental Science and Technology*, **1996**, 30, 400–407.

DeGaetano, A.T.; Doherty O.M.; Temporal, spatial and meteorological variations in hourly $PM_{2.5}$ concentration extremes in New York City, *Atmospheric Environment*, **2004**, 38, 1547-1558.

Duker, A.A.; Carranza, E.J.M.; Hale, M.; Arsenic geochemistry and health, *Environment International*, **2005**, 31, 631–641.

Henry, R.C.; Spiegelman, C.H.; Chang, Y.S.; Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction, *Atmospheric Environment*, **2002**, 36, 2237–2244.

Kim, E.; Hopke, P.K.; Edgerton, E.S.; Source identification of Atlanta aerosol by Positive Matrix Factorization, *Journal of Air and Waste Management Association*, **2003**, 53, 731–739.

Kim, E.; Hopke, P.K; Pinto, J.P.; Wilson, W.E; Spatial variability of fine particle mass, components, and source contributions during the regional air pollution study in St. Louis, *Environmental Science and Technology*, **2005**, 39, 4172–4179.

Krudysz, M.A.; Froines, J.R.; Fine, P.M.; Sioutas, C.; Intra-community spatial variation of size-fractionated PM mass, OC, EC, and trace elements in the Long Beach, CA area, *Atmospheric Environment*, **2008**, 42, 5374-5389.

Lee, J.H.; Hopke, P.K.; Turner, J.R.; Source identification of airborne PM-2.5 at the St. Louis – Midwest Supersite, *Journal of Geophysical Research*, **2006**, 111, D10S10.

Pinto, J.P.; Lefohn, A.S; Shadwick, D.S; Spatial variability of $PM_{2.5}$ in urban areas in the United States, *Journal of the Air and Waste Management Association*, **2004**, 54, 440–449.

Röösli, M.; Theis, G.; Kunzli, N.; Staehelin, J.; Mathys, P.; Oglesby, L.; Camenzind, M.; Braun-Fahrlander, C; Temporal and spatial variation of the chemical composition of PM_{10} at urban and rural sites in the Basel area, Switzerland, *Atmospheric Environment*, **2001**, 35, 3701–3713.

Sajani, S.Z.; Scotto, F.; Lauriola, P.; Galassi, F.; Montanari, A.; Urban air pollution monitoring and correlation properties between fixed-site stations, *Journal of the Air and Waste Management Association*, **2004**, 54, 1236–1241.

Sánchez-Rodas, D.; Sánchez de la Campa, A.M.; de la Rosa, J.D.; Oliveira, V.; Gómez Ariza, J.L.; Querol, X.; Alastuey, A.; Arsenic speciation of atmospheric particulate matter (PM_{10}) in an industrialized urban site in southwestern Spain, *Chemosphere*, **2007**, 66, 1485–1493.

Shields, J.; Ambient air arsenic levels along the Texas-Mexico border, *Journal of the Air and Waste Management Association*, **1991**, 41, 827-831.

Turner, J.R.; Measurement error as a context for assessing intraurban variability in chemical speciation network data, *Proceedings of the 101st Annual Meeting of the Air and Waste Management Association*, Paper No. 664, Portland, OR, June 24-27, **2008**.

United States Environmental Protection Agency (USEPA); Issue paper on the human health effects of metals, **2004**
(http://www.epa.gov/raf/publications/pdfs/HUMANHEALTHEFFECTS81904.PDF; accessed Sep 25, 2013).

United States Environmental Protection Agency (USEPA); St. Louis Community Air Project Air Toxics Risk Characterization, **2005** (http://www.epa.gov/oaqps001/aqmp/pdfs/2005StLouisCAPReport.pdf; accessed Sep 25, 2013).

United States Environmental Protection Agency (USEPA); AirData: Access to air pollution data, **2007** (http://www.epa.gov/airdata/ad_reports.html; accessed Sep 25, 2013).

United States Environmental Protection Agency (USEPA); National Emission Inventory, **2008** (http://www.epa.gov/ttn/chief/net/2008inventory.html; accessed Sep 25, 2013).

United States Environmental Protection Agency (USEPA); Technical Assistance Document for the National Air Toxics Trends Stations Program, **2009** (http://www.epa.gov/ttnamti1/files/ambient/airtox/nattsTADRevision2_508Compliant.pdf; accessed Sep 25, 2013).

United States Environmental Protection Agency (USEPA); National Air Toxics Trends Stations Quality Assurance Annual Report - Calendar Year 2008, **2010** (http://www.epa.gov/ttnamti1/files/ambient/airtox/NATTS2008QAAnnualReport.pdf; accessed Sep 25, 2013).

Wang, G.; Hopke, P.K.; Turner, J.R.; Using highly time resolved fine particulate compositions to find particle sources in St. Louis, MO, *Atmospheric Pollution Research*, **2011**, 2, 219-230.

World Health Organization (WHO); Air quality guidelines for Europe Second Edition, **2001** (http://www.euro.who.int/__data/assets/pdf_file/0005/74732/E71922.pdf; accessed Sep 25, 2013).

Wilson, J.G.; Kingham, S.; Pierce, J.; Struman, A.P.; A review of intraurban variations in particulate air pollution: Implications for epidemiological research, *Atmospheric Environment*, **2005**, 39, 6444-6462.

Wongphatarakul, V.; Friedlander, S.K.; Pinto, J.P.; A comparative study of $PM_{2.5}$ ambient aerosol chemical databases, *Environmental Science and Technology*, **1998**, 32, 3926-3934.

Yadav, V.; Turner, J.R.; Downs, J.; Rowles, T.; High Time Resolution PM_{10} Metals by the Cooper Environmental Services Ambient Metals Monitor (Xact 620): Field Performance Evaluation and Data Trends for St. Louis, *Proceedings of the 102nd Annual Meeting of the Air & Waste Management Association*, Paper No. 2009-A-146-AWMA, Detroit, MI, June 16-19, **2009**.

Zeng, T., Sarofim, A.F.; Vaporization of arsenic, selenium and antimony during coal combustion, *Combustion and Flame*, **2001**, 126, 1714–1724.

Chapter 3 :Estimating local emission source zones using high-time resolution carbon measurements across a monitoring network in a multi-source industrial area.

3.1. Abstract

Carbon measurements conducted at a three site monitoring network in an industrial area of Dearborn, MI during fall 2008 were examined for identifying local emission sources. Elemental carbon (EC) concentrations, measured using a Sunset Labs ECOC Analyzer, at the two sites Dearborn and Miller, separated by ~400 m, were on average more than 50% higher than at the Ten Eyck site located ~6.5 km southwest of the Dearborn site. Organic carbon (OC) concentrations were generally uniform across the network and only 8-9% higher at the two sites compared to those at Ten Eyck. For better estimation of the directionality of the local EC sources contributing to such variability, nonparametric wind regression (NWR) plots were constructed for EC concentrations at the two sites, Dearborn and Miller in excess of that at Ten Eyck using hourly wind directions measured at the Dearborn site. These plots were fitted with a series of Gaussian curves to estimate the locations of the concentration peaks in the NWR plots. The bearings estimated were assigned an uncertainty corresponding to the optimized soothing parameter used for the NWR plots. Triangulation of the resolved bearings of peak concentrations from the two sites, in some cases, identified multiple emission zones corresponding to each resolved bearing for a site. Such ambiguities can result from various factors such as the presence of multiple nearby sources or sources aligned in corridors, intermittent source emissions, and a lack of sufficient data for certain wind sectors. Other

challenges in estimating emission source zones using high time resolution ambient measurements across a network of monitoring sites in a multi-source industrial area are also investigated. Despite these limitations, the approach does to some extent constrain the location of putative emission sources.

3.2. Introduction

A multi-county area in Southeast Michigan was in nonattainment of the annual-average $PM_{2.5}$ National Ambient Air Quality Standards (NAAQS) in 2008, with a compliance monitoring station at Dearborn, MI exhibiting the highest design values for both the 24-hour and annual average $PM_{2.5}$ NAAQS (MDEQ, 2008). Compared to other monitoring stations in the area, similar concentrations were observed for species characterized by regional transport (sulfate, nitrate, and ammonium) but higher concentrations at Dearborn station were observed for carbonaceous particulate matter and metal oxides that could result from local emission sources (Buzcu-Guven et al., 2007; Gildemeister et al., 2007; Turner, 2008). Hourly PM carbon measurements at two additional temporary monitoring sites within 6.5 km of the Dearborn station were commissioned to provide insights into the spatial variability and source locations of carbon fractions at neighborhood- and finer-scales (operationally < 5km spatial scale, Watson and Chow, 2001). This dataset provided an opportunity to evaluate strategies to estimate the local emission source locations using high time resolution concentration data collected across a monitoring network in an industrial area with multiple emission sources.

A variety of approaches are used to identify the directionality of the sources resulting in high concentrations at the receptor sites. One such tool is the Conditional probability Function (CPF) plot, which displays the fraction of samples from each discrete wind sector (usually set to 15°)

that lead to concentrations at the receptor site to be higher than a predetermined threshold concentration (Kim et al., 2003) based on a condition probability framework described by Ashbaugh et al. (1985). These plots have been used to identify source locations in various ambient particulate matter and source apportionment studies (Kim et al., 2003; Begum et al., 2004). Another commonly used approach is the Nonparametric Wind Regression (NWR) plot, which quantifies the relationship between wind direction and concentrations without making any assumptions about the functional form of the relationship or the statistical distribution of the dataset (Henry et al., 2002). NWR plots utilize a Gaussian smoothing kernel as a non-subjective alternative to the discretized wind sectors for estimating the directionality of the sources resulting in peak concentrations at the receptor site. NWR plots are often used for resolving the directionality of the sources estimated from source apportionment and monitoring studies (Kim and Hopke, 2004; Wang et al., 2011). For example, Henry et al. (2002) identified a major source of cyclohexane by triangulating the NWR bearings from two sites located within 6 kilometer of the source. However, these plots were constructed for isolating the impacts of the particular source by selectively restricting the dataset to wind speeds greater than 5 mile/hour where direct plume hits should be most evident. The challenges in extending such a triangulation approach to ambient concentrations measured at receptor sites located in areas with intermittently emitting multiple sources are explored in this chapter.

3.3. Datasets

3.3.1. Field campaign

As part of the Midwest Rail Study, the Michigan Department of Environmental Quality (MDEQ) conducted air quality measurements at three sites in Dearborn, MI (Turner et al., 2009). The Dearborn air monitoring station (USEPA AQS ID 26-163-0033) is an MDEQ compliance

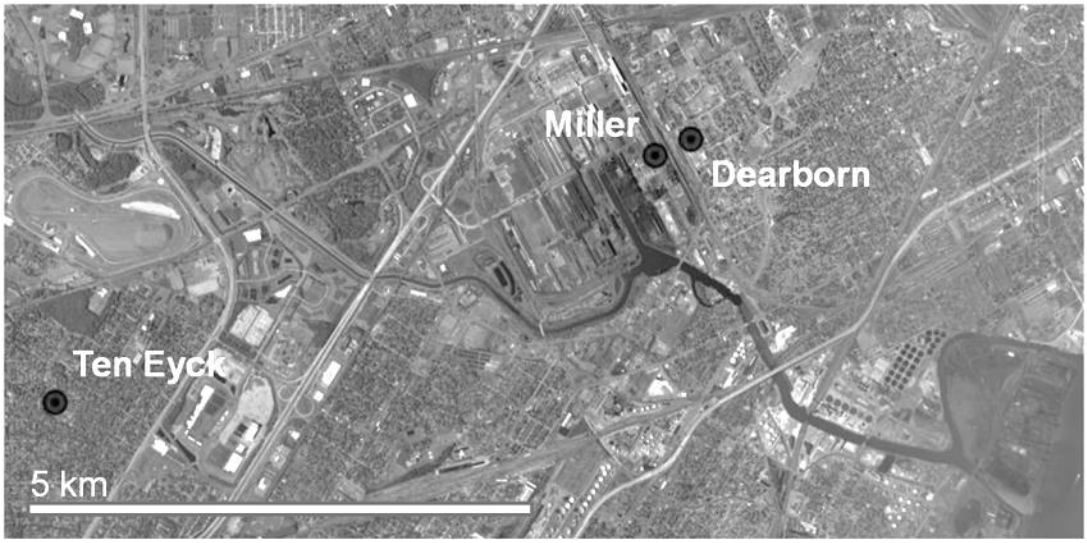


Figure 3-1. Site map of the three-site neighborhood-scale monitoring network for hourly EC and OC measurements in Dearborn, MI.

monitoring station and includes routine semi-continuous measurement of thermal-optical organic carbon (OC) and elemental carbon (EC) by the Sunset Laboratory Semi-continuous Carbon Aerosol Analyzer, and black carbon (BC) by the Magee Scientific Aethalometer. Two special purposes monitoring stations within the City of Dearborn were commissioned, shown in Figure 3-1, for the measurement of OC, EC, and BC during a three month field campaign in 2008. The Miller site (USEPA AQS ID 26-163-0044) was only 400 meters southwest of the Dearborn station with a rail yard and public roadway between these sites. The Ten Eyck site (USEPA AQS ID 26-163-0043) was located ~ 6.5 km southwest of the Dearborn station in a residential neighborhood with a major steelmaking plant located along the path between these sites.

The three monitoring sites were aligned in an approximate straight line oriented ~240/60°N (from Ten Eyck in the southwest to Dearborn in the northeast). Two-wavelength Aethalometers and Field OCEC Analyzers were operated from 9/12/2008 to 12/16/2008 at all three sites. At the end of the field campaign all of the instruments were collocated at the Dearborn station from 12/16/2008 to 1/20/2009 to document measurement precision and identify any instrument-to-

instrument biases. Meteorological measurements (hourly wind speed and wind direction) used in this study were also collected at the Dearborn site. The analysis presented in this chapter is based on the EC and OC data collected by the Sunset OCEC Analyzers

3.3.2. Data characteristics

The Sunset OCEC analyzer reports total carbon (TC) from the thermal analysis, and carbonic and EC from the thermal-optical analysis (ThOC, ThEC). In addition, the OptEC/TC mode of operation reports EC from a purely optical (light transmission) measurement (OptEC) and also thermal TC. OptOC is calculated as the difference between TC and OptEC. ThOC and ThEC are the conventionally reported carbon fractions, but the thermal-optical EC/OC split was poorly identified for a large number of the samples in this study with virtually all of the carbon assigned to ThOC. 24-hour averages of the hourly OptEC measurements were compared to routinely collected 24-hour integrated EC measurements from the Chemical Speciation Network (CSN) sampler at Dearborn station. The filter-based on-site CSN sampler, which operates on a 1-in-6 day schedule, utilizes the IMPROVE_A thermal-optical protocol for EC measurements. OptEC measurements from this study were in excellent agreement with IMPROVE_A EC measured by thermal-optical transmittance, TOT (Reduced Major Axes, RMA (Bohonak, 2004) slope: 0.99 ± 0.08 and intercept: 0.02 ± 0.06 $\mu g/m^3$ at 95% confidence interval). The agreement in OptEC measurements from this study and EC/TOT from the filter-based speciation network data adds confidence to the data quality and confirms the observed sample-to-sample variability in carbon measurements which are utilized in this study. OptEC and OptOC were utilized in this study and are used interchangeably with the terms OC and EC in the remainder of the chapter.

To facilitate comparison of hour-specific differences across the three sites, collocated datasets were used to examine and remove bias between instruments. Such bias can result in systematic

errors between the two concentration time series and confound the interpretation of spatiotemporal variability between the sites. The ECOC analyzer at Dearborn was used as the reference instrument as it is a routine monitoring station and continues to operate beyond this study period. Miller and Ten Eyck instrument datasets were adjusted to remove any bias relative to the Dearborn instrument data using RMA regression which accounts for measurement error in both datasets. Table 3-1 summarizes the performance of the collocated EC and OC data. For both Miller and Ten Eyck measurements against Dearbon OC, the slopes were statistically different from unity and intercepts were statistically different from zero at 95% confidence interval. For EC, both slopes were also statistically different from unity but both intercepts were statistically indistinguishable from zero. Miller EC was biased high compared to Dearborn EC, while Ten Eyck EC was biased low compared to Dearborn EC. The collocated precision, after adjusting for instrument-to-instrument bias, was 9-14 % for OC and 5-6 % for EC.

Table 3-1. Summary statistics for collocated hourly data collected by Sunset OCEC field analyzer used at Miller and Ten Eyck stations with the instrument at Dearborn station.

	EC		OC	
	Miller	Ten Eyck	Miller	Ten Eyck
Number of samples	244	259	185	254
Slope[a]	1.08 ± 0.01	0.81 ± 0.01	0.93 ± 0.05	0.87 ± 0.05
Intercept[a] ($\mu g/m^3$)	0.00 ± 0.01	-0.01 ± 0.01	-0.31 ± 0.16	-0.28 ± 0.18
R^2	0.99	0.98	0.88	0.75
Collocated precision[b] ($\mu g/m^3$)	0.024	0.028	0.29	0.42
Collocated precision[c] (%)	4.9	6.0	9.4	13.5

a. Coefficients for the reduced major axis regression (Bohonak, 2007) of Miller or Ten Eyck data on Dearborn data. Slopes and intercepts are reported along with their 95% confidence intervals.

b. Collocated precision after transforming the Miller and Ten Eyck data using the regressions.

c. Absolute collocated precision divided by the arithmetic mean concentration.

EC and OC concentrations measured at the three site monitoring network during the study period are summarized in Table 3-2. Out of a total 2275 hours of measurements, EC and OC concentrations were measured at 71-77% and 61-77% completeness respectively, depending on the site. High hourly concentrations at all sites were frequently measured during calm conditions i.e., wind speeds of ≤ 1 m/s. Interpretation of inter-site variability can be confounded by higher concentrations resulting from reduced dispersion of ambient PM matter during calm conditions (Bathmanabhan et al., 2010). Thus, concentrations measured only during advective conditions (i.e., wind speed ≥ 1 m/s) were retained in this analysis. The dataset was further screened to include only the hours when concentrations were concurrently measured at all three sites for evaluating inter-site differences across the network and resulted in a total of 694 hours (31%) of EC and OC measurements for the study period.

Table 3-2. Summary statistics for hourly EC and OC concentrations (Total 2275 hours) collected at the three sites, after adjusting measurements at Miller and Ten Eyck against Dearborn measurements to remove sampler-to-sampler bias. All concentrations are in $\mu g/m^3$.

	EC			OC		
	Number of samples (%)	Mean ± Std. Dev.	Median	Number of samples (%)	Mean ± Std. Dev.	Median
	Measurements during field campaign					
Dearborn	77	0.74 ± 0.76	0.50	77	3.97 ± 2.26	3.31
Miller	71	0.78 ± 0.82	0.53	61	3.71 ± 2.50	3.22
Ten Eyck	74	0.56 ± 0.60	0.36	74	4.03 ± 3.12	3.15
	Measurements during calm conditions (wind speed ≤ 1 m/s)					
Dearborn	26	1.10 ± 1.06	0.74	26	5.05 ± 2.75	4.22
Miller	19	1.29 ± 1.21	0.95	18	4.78 ± 3.83	4.09
Ten Eyck	23	0.93 ± 0.82	0.68	23	5.72 ± 4.19	4.41
	Concurrent measurements during advective conditions (wind speed > 1 m/s)					
Dearborn		0.55 ± 0.41	0.43		3.23 ± 1.70	2.75
Miller	31	0.63 ± 0.50	0.48	31	3.25 ± 1.47	2.90
Ten Eyck		0.38 ± 0.37	0.26		2.99 ± 1.88	2.33

3.4. Methodology

3.4.1. Spatiotemporal variability

Spatial variability was examined by comparing summary statistics of concurrently measured concentrations during advective conditions. Inter-site variability in species concentrations were also examined using the methodology presented in Yadav et al. (2013) for gauging spatiotemporal variability using data from a network of monitoring sites. Based on this approach, concentrations measured at each site were apportioned into hourly baseline and site-specific excess concentrations to semi-quantitatively represent contributions from urban- and larger-scale emission sources (spatial range > ~5 km) and neighborhood- and finer-scale (or local-scale) emission sources (spatial range < 5 km). The minimum hourly concentration observed across the three sites in this network was defined as the hourly baseline concentration and concentrations at each site in excess of the baseline were defined as the hourly site-specific excess concentrations. The defined baseline does not strictly represent the urban- and larger-scale influences because it can capture neighborhood- and finer-scale contributions simultaneously occurring at all three sites. For example, sources located outside the network that are aligned with this linear monitoring network and are sufficiently close to the study area for plume dispersion can result in differential impact across all three sites. Therefore, excess concentrations do not necessarily provide an estimate of the absolute source contributions, but utilize the spatial variability in concentrations across the network that results from local source impacts. The scattergram of Pearson's correlation coefficient (PCC) and coefficient of divergence (COD) calculated between the site-specific measured concentrations and the baseline concentrations (i.e., *PCC-COD plot*) can provide insight into the sites contributing to the baseline and hence the sites with least impact from local sources (Yadav et al., 2013). Sites with

greater variability in concentrations compared to the baseline, resulting from the increasing influence of local source contributions, will be positioned further away from the homogeneity end (PCC = 1 and COD = 0) of the plot and can be utilized for identification of local source regions.

3.4.2. Estimating local emission source zones

The NWR plots were constructed using hourly concentration time series with the hourly wind direction to estimate the directionality of the emission source zones. The kernel regression in NWR utilizes a Gaussian fit to find the bearings of peaks in ambient concentrations (Henry et al., 2002). The locations of the peaks in these NWR plots are sensitive to the chosen smoothing parameter ($\Delta\theta$ or width of the discrete wind angle) and thus the peaks cannot be estimated better than $\pm\Delta\theta$. A larger smoothing parameter provides a smoother curve and can result in peaks being merged or unresolved while a smaller smoothing parameter can result in multiple peaks that are caused by either measurement noise or of the splitting of a single larger peak into multiple false peaks. The cross validation method for optimizing smoothing parameter, described in Henry et al., 2002, was utilized to obtain the smoothing parameters. Confidence intervals for NWR plots generated based on the asymptotic normal distribution of the kernel estimate are recommended by Henry et al., 2002 to examine the validity of the NWR peaks.

NWR plots provide the wind direction bearings that result in elevated average concentrations at the receptor site. To refine the estimation of emission source zones, the NWR curves were fitted with a series of Gaussian curves using the Multi-peak Fitting feature in IGOR Pro™ software. The goodness of multi-peak curve fitting was gauged by utilizing residual plots. The uncertainty in number and amplitude of peaks resolved is subject to the noise and peak-fitting smoothing parameter utilized by the auto peak finder tool. Since the locations of the resolved peaks are

subject to a maximum uncertainty of NWR smoothing parameter, each resolved bearing was represented by a triangular emission source zone radiating out from the receptor site. The resolved triangular zones for each bearing at a particular site were matched with triangular zones resolved for other sites to triangulate the probable emission source zones. There can be ambiguity in estimating such zones as a bearing from one site can match to one-or-more bearings from another site. Despite such limitations, this approach enables narrowing down the likely locations of emission sources which can then be validated with sources listed in archived emission inventories when such information is available.

3.5. Results and Discussion

3.5.1. Elemental carbon

Summary statistics of concurrently measured EC concentrations during advective conditions, shown in Table 3-2, indicate that the lowest concentrations across this network are observed at Ten Eyck. EC concentrations at Dearborn and Miller are on average 45 and 66%, respectively, higher than at Ten Eyck. The *PCC-COD plot* constructed using site-specific measured concentrations against the baseline concentrations indicates that EC baseline is chiefly defined by concentrations at Ten Eyck (Figure 3-2 (a)) which is consistent with 78% of the EC measurements at Ten Eyck being apportioned to the baseline. Only 14 and 7% of the EC measurements at Dearborn and Miller, respectively, contribute to the baseline and hence display relatively greater variability than Ten Eyck with respect to this baseline in the plot. The distribution of median hourly baseline and site-specific excess EC concentrations over for the study period, with 24 to 35 concentration values for each hour, are shown in Figure 3-3 (a). EC at Ten Eyck is almost exclusively captured by the baseline for all hours of the day resulting in Ten Eyck being an effective background site for EC measurements across this network. The

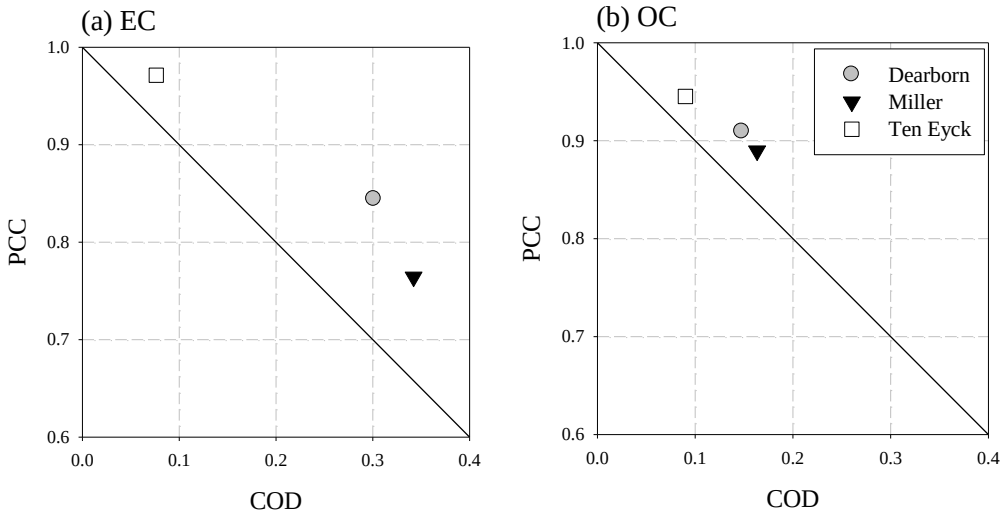


Figure 3-2. The PCC-COD plot or scattergram of Pearson's correlation coefficient (PCC) and coefficient of divergence (COD) calculated for (a) EC and (b) OC concentrations measured at each of the three sites with respect to the baseline concentrations.

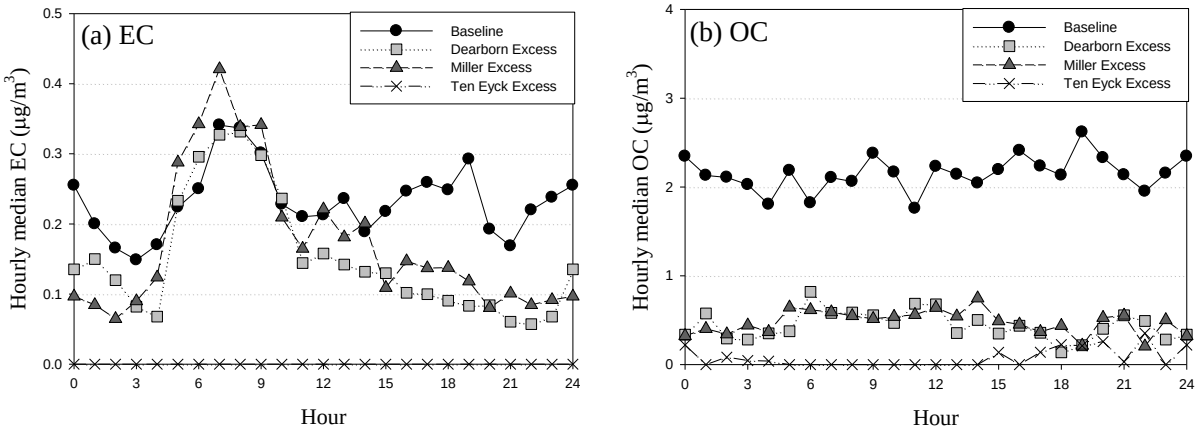


Figure 3-3. Hourly median baseline and site-specific excess (a) EC and (b) OC concentrations trends over the entire study period.

baseline EC varies between 0.15 to 0.35 μg/m^3, with highest EC observed between 6-11 am and 3-8 pm. In addition, excess EC at Dearborn and Miller also peak during the morning hours and remain relatively low during mid-to-late afternoon. The EC trends at these sites are consistent with their respective proximity to roadways with substantial traffic volumes but other local emissions sources could also contribute to such inter-site differences. The EC concentrations at

Dearborn and Miller in excess of Ten Eyck will be further examined for local source influences that arise from sources in the vicinity of these sites.

Figure 3-4 shows NWR plots for hourly EC concentrations constructed using hourly wind direction from Dearborn station with an optimized smoothing parameter of 8°. The plots of site-specific EC concentrations also confirm the higher EC burdens at Dearborn and Miller compared to Ten Eyck (Figure 3-4 (a), (b) and (c)). Mean EC concentrations ≥ 1 $\mu g/m^3$ were observed at Dearborn for winds from ~190° and at Miller for winds from ~190° and ~240°. The large confidence intervals observed in these plots will be discussed later in this section. NWR plots for hourly EC concentrations at Dearborn and Miller in excess of Ten Eyck are shown in Figure 3-4 (d) and (e), respectively. The NWR curves based on difference of two measurements may be subject to higher uncertainty. However, maximum inter-site EC concentration differences are at least an order of magnitude higher that the propagated absolute uncertainty of ~0.04 $\mu g/m^3$ from collocated precision. Thus, the larger peaks should be relatively insensitive to noise from the propagated uncertainties. The locations of peaks resolved from the peak-fitting exercise suggest that multiple sources may be influencing the excess concentrations observed at the receptor sites. Further, the three peaks identified for winds from southwest (180°-270°) using this approach could not be resolved from the site-specific NWR plots. Presence of multiple sources and/or intermittent emission profiles of the sources can result in large variations in the measured concentrations at the receptor site and can lead to relatively large confidence intervals observed in the NWR plots of site-specific concentrations. Thus, confidence intervals of bearings in NWR plots may not be a robust metrics for evaluating if the peaks are real in this dataset.

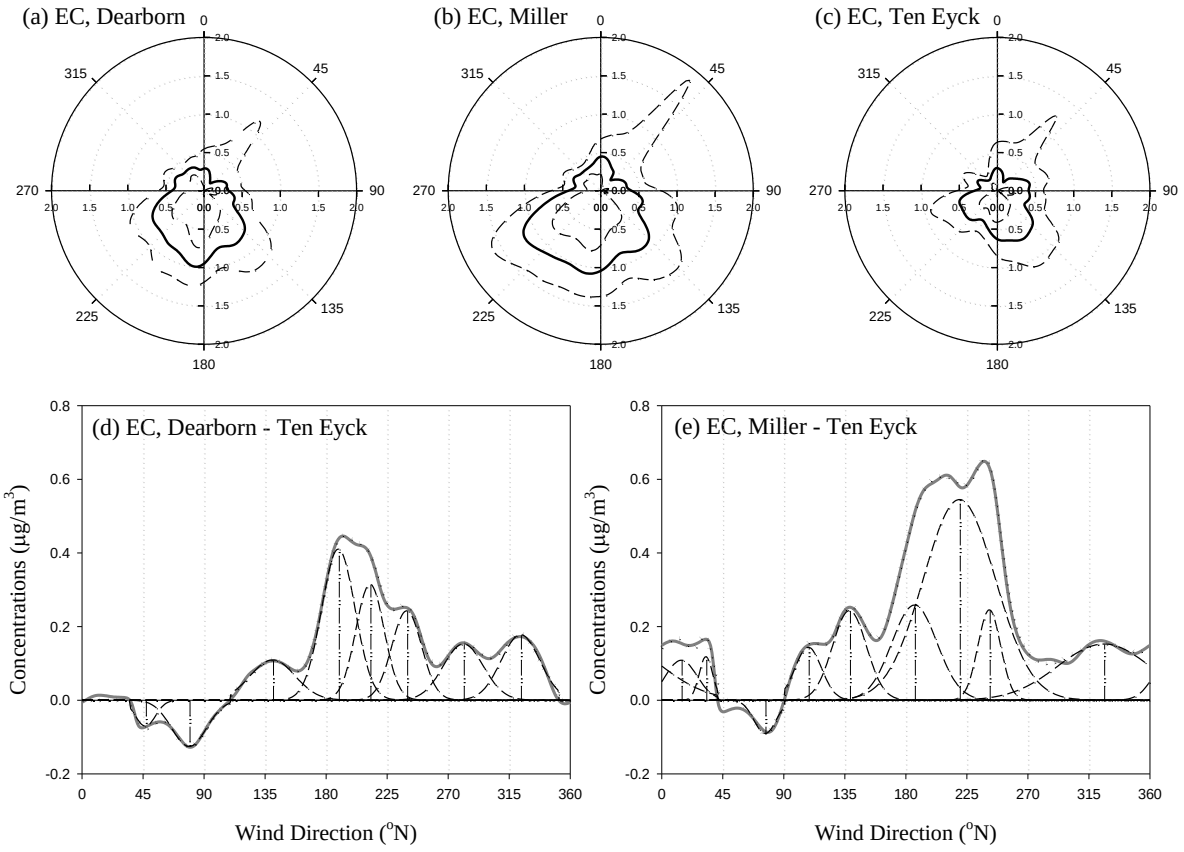


Figure 3-4. 1-D nonparametric wind regression (NWR) plots for EC concentrations at (a) Dearborn, (b) Miller and (c) Ten Eyck with confidence intervals (dashed lines). All radial axes range from 0-2 μg/m^3. NWR plots of EC at (d) Dearborn and (e) Miller in excess of Ten Eyck (solid grey line), along with individual Gaussian peaks (dashed lines) and reconstructed NWR plots (dotted line) obtained from peak-fitting of inter-site NWR plots. The bearings of the individual peaks obtained are also shown (dash-dot-dot). NWR for all EC concentration series were performed using smoothing parameters of 8°. Positive peaks correspond to excess at either Dearborn or Miller while negative peaks correspond to excess at Ten Eyck.

Peaks resolved from peak-fitting exercise of EC concentrations at Dearborn in excess of Ten Eyck that converge with the corresponding peaks at Miller are tabulated in Table 3-3 (a). Emission zones are triangulated for the converging bearings by utilizing the smoothing parameter of 8° for EC as the maximum uncertainty in NWR peaks. Almost all the bearings converged to two-or-more potential emission zones and the sources

Table 3-3. Wind direction bearings (°N) of peaks resolved from peak-fitting of NWR plots of (a) EC and (b) OC concentrations at Dearborn and Miller in excess of Ten Eyck for the triangulated zones (DXMY), corresponding to the convergence of Dearborn bearing, DX with Miller bearing, MY. Emission sources identified in the triangulated zones are also listed. Peaks and sources identified for wind sectors with sparse datasets are italicized.

(a) EC

Dearborn	Miller	Zone	Sources in triangulated zone
141°	*108°*	*D1M3*	*Roadway*
	139°	*D1M4*	*Industrial area near Zug Island, MI*
189°	*108°*	*D2M3*	*Rail yard*
	139°	*D2M4*	*Rail yard and Power generating station*
	186°	D2M5	Industrial area south of Rouge River, MI
212°	*108°*	*D3M3*	*Rail yard*
	139°	*D3M4*	*Integrated steelworks*
	186°	D3M5	Integrated steelworks
240°	*108°*	*D4M3*	-
	139°	*D4M4*	-
	186°	D4M5	-
	220°	D4M6	Integrated steelworks
	242°	D4M7	Integrated steelworks
281°	*14°*	*D5M1*	*Rail yard*
	33°	*D5M2*	*Rail yard*
	326°	*D5M8*	*Roadway*
323°	*14°*	*D6M1*	*Roadway*
	33°	*D6M2*	*Roadway*
	326°	*D6M8*	-

(b) OC

Dearborn	Miller	Zone	Sources in triangulated zone
50°	*52°*	*D1M1*	*Industrial area northeast of Dearborn, MI*
198°	*98°*	*D2M2*	*Rail yard*
	198°	D2M3	Industrial area south of Rouge River, MI
233°	*98°*	*D3M2*	*Rail yard*
	198°	D3M3	Integrated steelworks
	233°	D3M4	Integrated steelworks
287°	*52°*	*D4M1*	-
	319°	D4M5	Vehicle manufacturing plant
329°	*52°*	*D5M1*	-

identified in these zones are also listed in Table 3-3 (a). Some of the triangulated zones did not correspond to known EC emission sources and could be rejected. Depending on the location of the emission zones and meteorological conditions, the hours of EC impact at Dearborn may be different from those at Miller. Ambiguities in the estimation of emission source locations can also result from factors such as the presence of multiple sources in the same area or corridor, and intermittent source emissions. The triangulation approach requires that both receptor sites be impacted by EC burdens from the same source which may not occur if the data are sparse for certain wind directions or the impact at a receptor site is caused by an exceptional event that may not occur during hours when the wind direction can cause simultaneous impacts on both sites.

Some of the nearby emission zones, labeled as DXMY corresponding to the resolved bearings of excess concentrations at Dearborn (DX) and Miller (MY) compared to Ten Eyck, are also shown in Figure 3-5 (a). Bearings to the southwest of both Dearborn and Miller sites triangulate to various sectors of the integrated steelworks such as zones D3M5, D4M6, D4M7 and may correspond to different EC point sources within the facility. Southeastern bearings (D2M5, not shown) triangulate to the industrial corridor south of Rouge River which includes a refinery and other industrial sources. Triangulated zones on the rail yard footprint correspond to known locations of switcher operations (D2M3, D5M1 and D5M2) and refueling (D2M4 and D3M3) where high PM emissions can occur from idling of locomotives (Feinberg et al., 2012), but such peaks cannot be validated due to sparse concentrations dataset with the associated wind sectors. There are only 1 to 11 concentration values in the 10° sectors for 0°-140° and 330°-360° in contrast to 14 to 44 values in the 10° sectors for 140° to 330°. Thus, the large

confidence intervals for 0-140° wind sectors in the NWR plots of site-specific concentrations (Figure 3-4 (a), (b) and (c)) results from lack of sufficient frequency in concentrations from these directions during the study period and prevents meaningful interpretation of the wind direction-concentration relationship for these bearings. Despite such limitations, the peak-fitting exercise enabled narrowing down the probable EC emission source zones that can be further examined through other approaches such as dispersion modeling and/or a micro-emissions inventory.

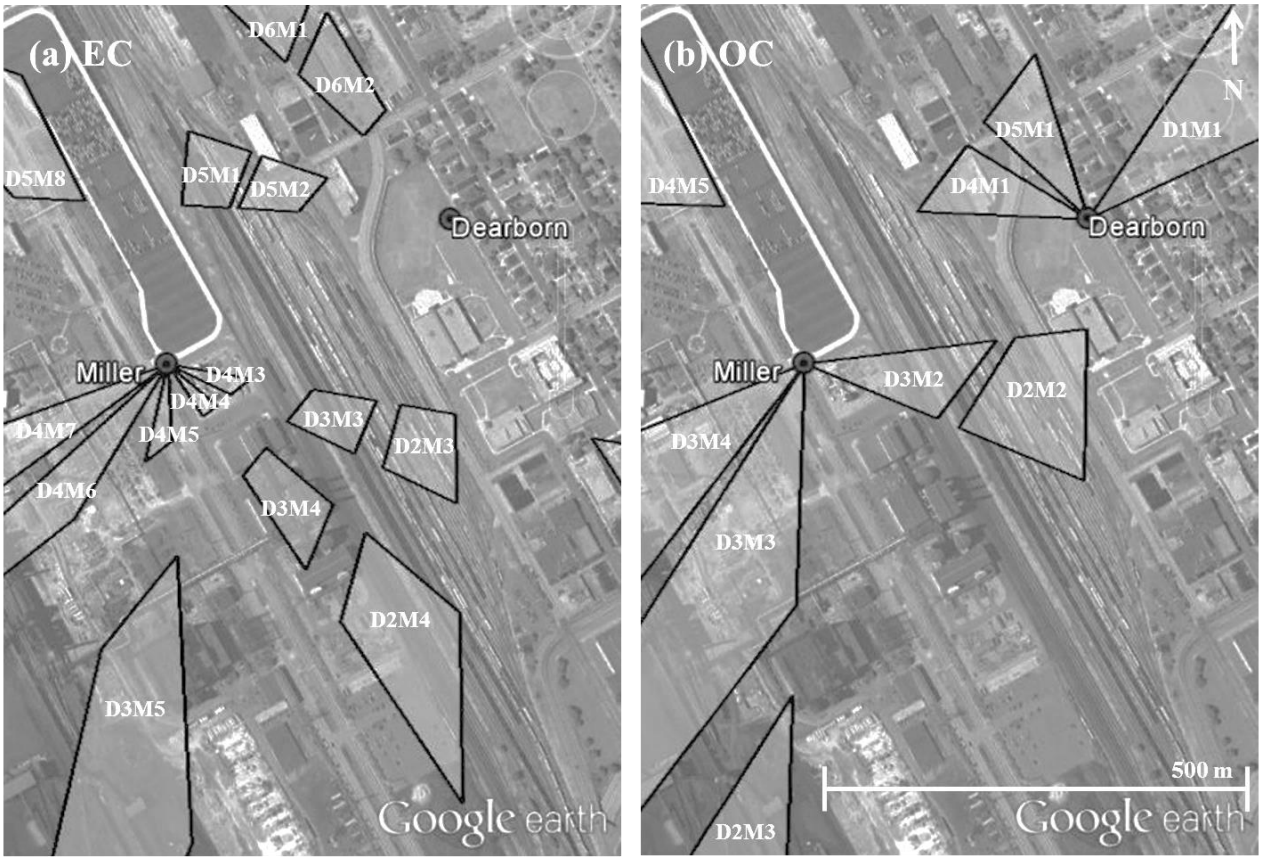


Figure 3-5. Estimated emission source zones based on triangulation of bearings for (a) EC and (b) OC concentrations at the Dearborn and Miller in excess of Ten Eyck. Emission zone DXMY represents the area triangulated by convergence of Dearborn bearing, DX and Miller bearing, MY with the maximum allowed variation in bearings based on the NWR smoothing parameter of 8° and 15° for EC and OC, respectively.

3.5.2. Organic carbon

OC concentrations at Dearborn and Miller are only 8-9% higher than OC at Ten Eyck (Table 3-2). The general uniformity in OC concentrations across the network is also captured by the tight clustering of all three sites in the *PCC-COD plot* (Figure 3-2 (b)) towards the homogeneity end of the plot (PCC=1 and COD=0). 53% of measurements from Ten Eyck and 23-24% from each of the remaining two sites contributes to the OC baseline, which on average ranges between 2-4 $\mu g/m^3$. OC excess concentrations at Dearborn and Miller display marginally greater variability than at Ten Eyck, but these variations could be confounded by the higher measurement uncertainty associated with using the difference of two concentrations (propagated absolute uncertainty of ~0.5 $\mu g/m^3$ from collocated precision). The distribution of median hourly site-specific excess and baseline OC concentrations for the study period are also shown in Figure 3-3 (b). The median hourly OC baseline typically ranged between 2 to 2.5 $\mu g/m^3$ and accounts for ~79-83% of the OC measured across the network, with no clear hourly trend. Excess OC at Dearborn and Miller are typically ~0.6-0.9 $\mu g/m^3$ higher than at Ten Eyck during morning to mid-afternoon. Thus, OC concentrations at Dearborn and Miller in excess of Ten Eyck are examined for local emission source influences that lead to the marginally higher variability at these sites.

The NWR plots for hourly OC concentrations constructed using hourly wind directions from the Dearborn station with optimized smoothing parameter of 15° are shown in Figure 3-6. The general uniformity of OC concentrations across this network prevents identification of local sources using NWR plots of site-specific OC concentrations (Figure 3-6 (a), (b) and (c)). Despite the low measurement precision and higher

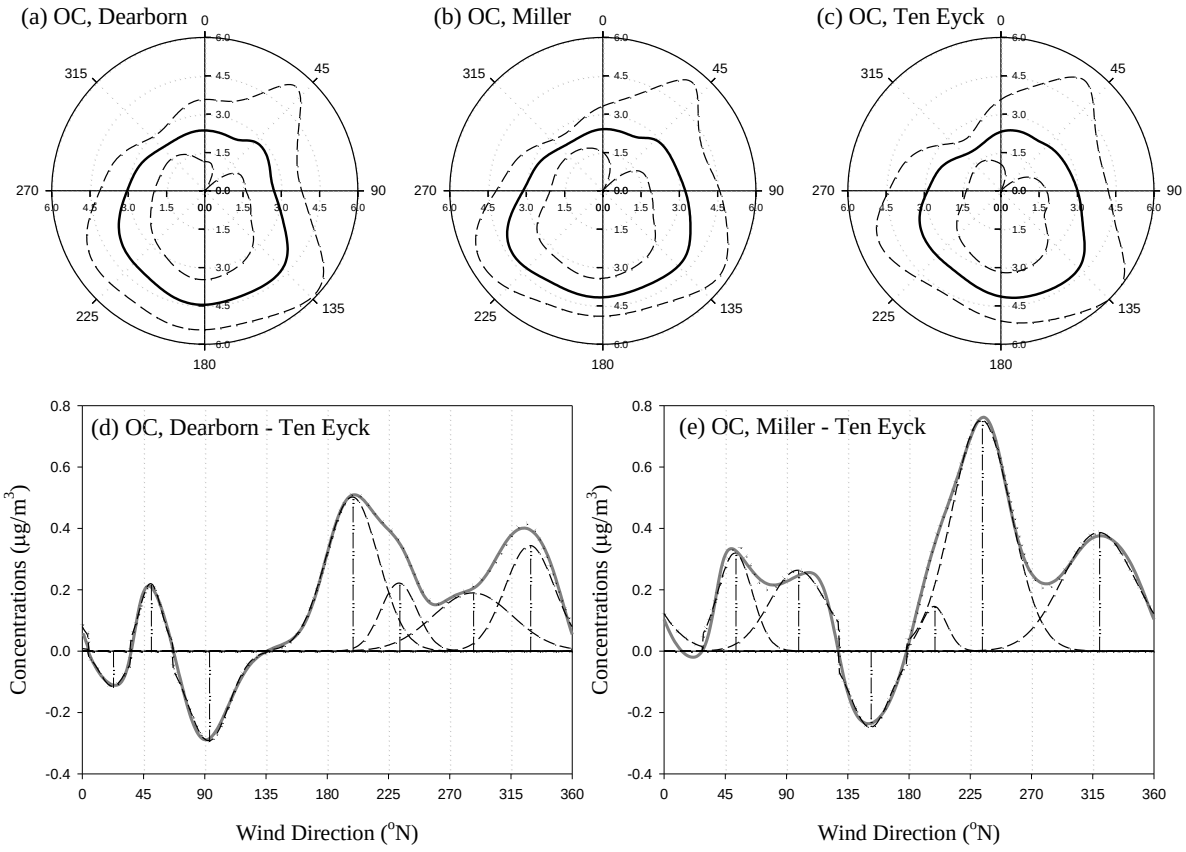


Figure 3-6. 1-D nonparametric wind regression (NWR) plots for OC concentrations at (a) Dearborn, (b) Miller and (c) Ten Eyck with confidence intervals (dashed lines). All radial axes range from 0-6 μg/m^3. NWR plots of OC at (d) Dearborn and (e) Miller in excess of Ten Eyck (solid grey line), along with individual Gaussian peaks (dashed lines) and reconstructed NWR plots (dotted line) obtained from peak-fitting of inter-site NWR plots. The bearings of the individual peaks obtained are also shown (dash-dot-dot). NWR for all OC concentration series were performed using smoothing parameters of 15°. Positive peaks correspond to excess at either Dearborn or Miller while negative peaks correspond to excess at Ten Eyck.

uncertainty of inter-site differences, the peak-fitting of NWR plots for OC concentrations at Dearborn and Miller in excess of Ten Eyck (Figure 3-6 (d) and (e)) identified bearings of OC impacts that could explain the marginally greater variability at these sites. Peaks resolved from peak-fitting of OC concentrations at Dearborn in excess of Ten Eyck that converge with the corresponding peaks at Miller are tabulated in Table 3-3 (b).

Triangulation of the converging bearings, subject to uncertainty of 15°, results in broader emission zones for OC than EC. A few emission zones triangulated in the vicinity of Dearborn, shown in Figure 3-5 (b), include various sections of the integrated steelworks (D3M3 and D3M4) and the rail yard (D2M2 and D2M3). Similar to EC, the lack of sufficient frequency in OC concentrations from wind directions corresponding to the rail yard prevent meaningful interpretation of the wind direction-concentration relationship for these bearings. Thus, a larger dataset with sufficient frequency in concentrations associated with all wind directions is needed for better characterization of the emission source zones.

3.6. Conclusions

Hourly EC and OC concentrations, collected using ECOC Sunset Analyzer during a three month study period, at a three site monitoring network in the industrial area in Dearborn, MI were examined for local emission source influences. While OC concentrations were generally uniform across the network, EC concentrations at the Dearborn and Miller sites were on average 50% higher than at Ten Eyck. Wind direction dependence of hourly carbon concentrations at Dearborn and Miller sites in excess of that at Ten Eyck were examined using NWR plots. Peak-fitting of the NWR plots with a series of Gaussian peaks enabled estimation of locations of peak concentrations. While the representativeness of the bearings resolved from the peak-fitting of NWR plots needs to be evaluated, for the preliminary triangulation of emission source zones, the estimated bearings for each site were given conservative uncertainty estimates based on the optimized NWR smoothing parameter. The resolved bearings from the two sites triangulated to multiple emission zones for each bearing at a site. While few zones with

no known sources could be rejected, elimination of triangulated zones with known sources through deductive reasoning is challenging. Various sectors of the integrated steelworks facility, located in the vicinity of Dearborn and Miller monitoring sites, identified from the triangulation approach could correspond to various intermittently emitting point sources in the industrial facilities. Other emission sources zones identified, such as the rail yard located between Dearborn and Miller sites and various roadways, are subject to higher uncertainty in the NWR plots because of lack of sufficient concentration data from wind sectors corresponding to their bearings. A larger data with sufficient frequency in concentrations associated with all wind directions could enable better characterization of the emission zones. Furthermore, the presence of multiple or intermittent sources can also result in relatively large confidence intervals in the NWR plots. Henry et al. (2002) proposed using the confidence intervals to evaluate whether peaks are real, but this approach is valid only if the source(s) are continuously emitting at a nearly constant rate. Despite such limitations, this data-driven approach assists in narrowing down the spatial extent of emission zones that could lead to elevated receptor concentrations.

3.7. Acknowledgement

This study was funded by a RARE Grant from USEPA ORD (Office of Research and Development) to USEPA Region 5. We gratefully acknowledge the assistance from Loretta Lehrman, Monica Paguia and other USEPA Region 5 staff. The Washington University in St. Louis (WUSTL) data analysis and dispersion modeling project funded under the study was administered by the Lake Michigan Air Directors Consortium (LADCO). Donna Kenski, Michael Koerber and Mark Janssen (LADCO), Amy

Robinson and the late Mary Ann Heindorf (Michigan DEQ) and Stephen N. Fienberg (WUSTL) are also gratefully acknowledged. Although the research described in this manuscript has been funded wholly or in part by the USEPA, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

3.8. References

Ashbaugh, L.L.; Malm, W.C.; Sadeh, W.Z.; A residence time probability analysis of sulfur concentrations at Grand Canyon National Park, *Atmospheric Environment*, **1985**, 19(8), 1263–1270.

Bathmanabhan, S.; Madanayak, S.N.S; Analysis and interpretation of particulate matter – PM_{10}, $PM_{2.5}$ and PM_1 emissions from the heterogeneous traffic near an urban roadway, *Atmospheric Pollution Research*, **2010**, 1, 184-194.

Begum, A.B.; Kim, E., Biswas, S.K.; Hopke, P.K.; Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh, *Atmospheric Environment*, **2004**, 38, 3025–3038.

Bohonak, A.J.; RMA: Software for Reduced Major Axis Regression, Version 1.17; San Diego State University, **2004**; http://www.bio.sdsu.edu/pub/andy/rma.html (accessed December 8, 2013).

Buzcu-Guven, B.; Brown, S. G.; Frankel, A.; Hafner, H.R.; Roberts, P.T.; Analysis and apportionment of organic carbon and fine particulate matter sources at multiple sites in the Midwestern United States, *Journal of the Air & Waste Management Association*. **2007**, 57:5, 606-619.

Feinberg, S.N.; Yadav, V.; Heiken, J.G.; Turner, J.R.; "Midwest Rail Study: Modeled Near-Field Impacts of PM2.5 Emissions from Railyard Activities", *Transportation Research Record*, **2012**, 106-114.

Gildemeister, A.E.; Hopke, P.K.; Kim, E.; Sources of fine urban particulate matter in Detroit, MI, *Chemosphere*, **2007**, 69 (7), 1064-1074.

Henry, R.C.; Changa Y.; Spiegelman, C.H.; Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction, *Atmospheric Environment*, **2002**, 36 (13), 2237-2244.

Henry, R.; Norris, G.A.; Vedantham, R.; Turner, J.R.; Source region identification using kernel smoothing, *Environmental Science and Technology*. **2009**, 43 (11) 4090-4097.

Kim, E.; Hopke, P.K.; Edgerton, E.S.; Source Identification of Atlanta Aerosol by Positive Matrix Factorization, *Journal of Air and Waste Management Association*, **2003**, 53, 731–739.

Kim, E.; Hopke, P.K.; Improving Source Identification of Fine Particles in a Rural Northeastern U.S. Area Utilizing Temperature-Resolved Carbon Fractions, *Journal of Geophysical Research*, **2004**, 109, D09204.

Michigan Department of Environmental Quality (MDEQ); State implementation plan submittal for fine particulate matter ($PM_{2.5}$), **2008** (http://michigan.gov/documents/deq/deq-aqd-air-aqe-sip-pm25-1-14-08_223446_7.pdf; accessed December 8, 2013).

Turner, J.R.; A conceptual model for ambient fine particulate matter over Southeast Michigan: High concentration days, prepared for the Southeast Michigan Council of Governments, **2008**. (http://www.semcog.org/uploadedfiles/Programs_and_Projects/Air/SEMOS_HighPM_FinalReport_Version1.0.pdf; accessed December 8, 2013).

Turner, J.R.; Yadav, V.; Feinberg, S.N.; Data analysis and dispersion modeling for the Midwest Rail study – Final Report, prepared for Lake Michigan Air Directors Consortium, **2009**. (http://www.ladco.org/reports/general/new_docs/WUSTL_MidwestRailStudy_FinalReport.pdf; accessed December 8, 2013)

Wang, G.; Hopke, P.K.; Turner, J.R.; Using highly time resolved fine particulate compositions to find particle sources in St. Louis, MO, *Atmospheric Pollution Research.* **2011**, 2 (2), 219-230.

Watson, J.G.; Chow, J.C.; Estimating middle-, neighborhood-, and urban-scale contributions to elemental carbon in Mexico City with a rapid response aethalometer, *Journal of Air and Waste Management Association*, **2001**, 51 (11):1522-1528.

Yadav, V.; Turner, J.R.; Gauging intraurban variability of ambient particulate matter arsenic and other air toxic metals from a network of monitoring sites, *Submitted to Atmospheric Environment*, **October 2013**.

Yu, K.N.; Cheung, Y.P.; Cheung, T.; Henry, R.C.; Identifying the impact of large urban airports on local air quality by nonparametric regression, *Atmospheric Environment*, **2004**, 38, 4501–4507.

Chapter 4 :A weight of evidence approach using network datasets to characterize drivers of ambient particulate matter air quality in Hong Kong.

4.1. Abstract

Despite numerous emission control measures implemented within Hong Kong during 1998-2008, air quality conditions remained relatively constant or degraded across the region. This motivated a detailed examination of various particulate matter (PM) datasets to understand the drivers of observed air quality trends. PM datasets collected by the Hong Kong Environmental Protection Department at multiple sites included 24-hour integrated PM_{10} and $PM_{2.5}$ speciation mass concentrations and also hourly PM_{10} mass measured by Tapered Element Oscillating Microbalance (TEOM) monitors. This chapter extends the source apportionment of the PM_{10} speciation data presented in Yuan et al. (2013) to provide insights obtained from the other datasets. Positive Matrix Factorization (PMF) on the $PM_{2.5}$ and PM_{10} speciation datasets yield consistent source contribution estimates for the two dominant factors – vehicle exhaust and secondary sulfate. Spatiotemporal variability in TEOM mass is examined by developing day-specific semi-quantitative estimates of regional- and larger-scale contributions that uniformly impact most of the network. PMF-resolved source contributions for the speciation datasets provide a context to understand the observed spatial and temporal patterns for PM_{10} TEOM mass. Temporal variations in TEOM mass and PMF-resolved source contributions are examined by categorizing air mass back trajectories into trajectory

classes. The decrease in vehicle exhaust contributions over the decade is demonstrated by improvement or stasis of air quality during summer months at monitoring stations located even in heavily urbanized regions of Hong Kong. In contrast, such improvements from reductions in vehicular emissions are being offset by the increased transport of particulate matter from mainland China during the winter months, resulting in the stasis in air quality. Analysis of the distinct PM datasets provide a weight of evidence for the drivers leading to recent trends in Hong Kong PM air quality.

4.2. Introduction

In 1987, Hong Kong Air Quality Objectives (AQO) was established for seven widespread air pollutants. Respirable Suspended Particulates (RSP or PM_{10}) objectives include an annual arithmetic mean concentration of 55 $\mu g/m^3$ and a 24-hour concentration of 180 $\mu g/m^3$ not to be exceeded once a year. The annual average PM_{10} AQO was achieved at all ten neighborhood-scale monitoring stations in 2009 with highest value of 51 $\mu g/m^3$ (93% of AQO) observed at Yuen Long (YL). In contrast, this AQO was not met at the three roadside monitoring stations with highest value of 71 $\mu g/m^3$ (129% of AQO) observed at Causeway Bay (CB). Proposed revisions to the Hong Kong AQO, expected to take effect in 2014, would decrease the PM_{10} AQO and establish AQO for fine PM ($PM_{2.5}$). Based on the PM_{10} levels in 2009, all sites within Hong Kong will be in exceedance of the proposed annual average PM_{10} AQO of 50 $\mu g/m^3$. While PM and pollutant concentrations across the region have decreased by almost 13% between 2006 to 2011 (Zhong et al., 2013), in the decade prior to 2009 the PM_{10} concentrations remained relatively constant at most monitoring sites in Hong Kong despite numerous emission control measures implemented within the Hong Kong Special Administrative

Region (HKSAR). For instance, Figure 4-1 shows average concentrations over the five-year periods 1999-2003 and 2004-2008. Significant reductions were observed at only the CB roadside station while the remaining stations exhibited virtually no change or even an increase in PM_{10} mass concentration. The drivers leading to such stasis/degradation of air quality conditions are examined though a detailed analysis of various particulate matter (PM) datasets routinely collected across the Hong Kong region.

Hong Kong is located within one of the most densely populated urban regions in the world with nearly 50 million people residing in the Pearl River Delta (PRD) region. Located along the southeastern coastline of China, the region has transformed from once agricultural hinterland to one of the largest manufacturing and shipping hubs in the world. For development of efficient air quality control strategies, a refined understanding of the emission sources, meteorology and atmospheric processes modulating the observed

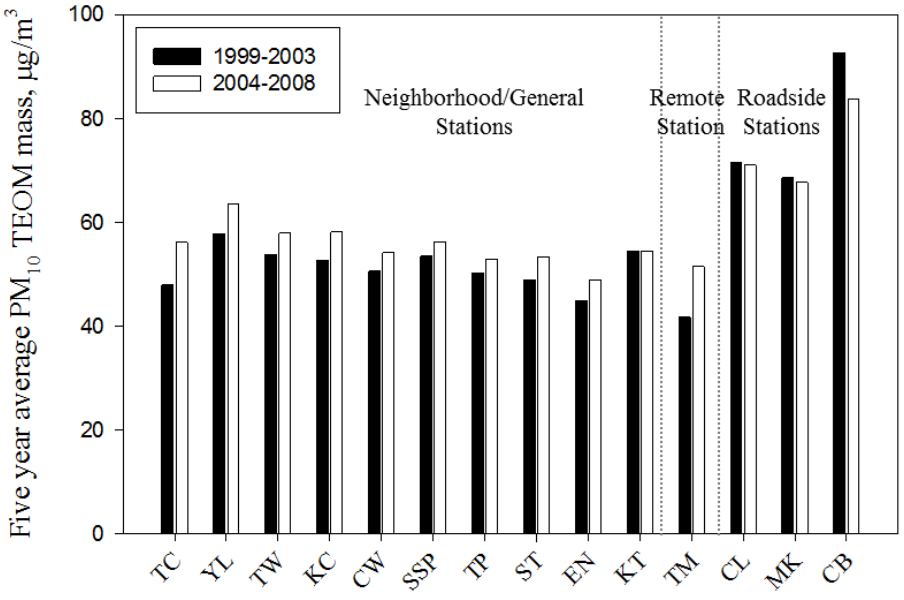


Figure 4-1. Average PM_{10} TEOM mass concentrations for the five-year periods 1999-2003 and 2004-2008 by monitoring sites spanning west to east of Hong Kong.

particulate matter burdens is needed. The Hong Kong Environmental Protection Department (HKEPD) has operated a network of monitoring stations collecting PM mass and speciation data to gauge air quality over the region (Figure 4-2). Studies using speciation data collected prior to 1998 (Fung and Wong, 1995; Qin et al., 1997, Lee et al., 1999; Qin et al., 2002) excluded key species such as carbon fractions/nitrates and/or included species with measurement/analysis artifacts (Yuan et al., 2006). Adjustments were made to the speciation methods in 1998 and later studies have typically focused on 1998-and-later data. These contemporary datasets include more than ten years of 24-hour integrated PM_{10} speciation mass collected at up to ten sites; three years of 24-hour

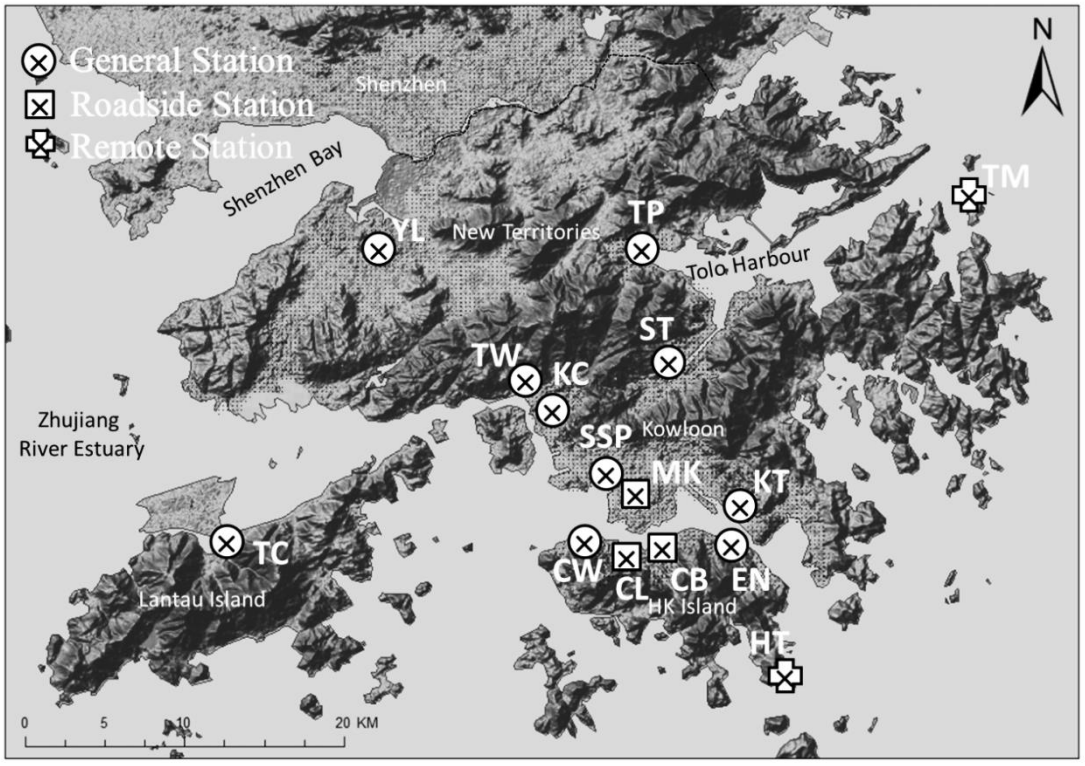


Figure 4-2. Geographical distribution of the Air Quality Monitoring Stations (AQMS) in Hong Kong.

integrated $PM_{2.5}$ speciation mass collected at up to four sites; and more than ten years of hourly PM_{10} mass concentrations collected at up to fourteen sites. Subsets of these datasets have been analyzed and updated as more data was made available (Ho et al., 2003b; Yu et al., 2004; Hagler et al., 2006; Huang et al., 2009; Cheng et al., 2010).

Ho et al. (2003a) examined the characteristics of PM_{10} and $PM_{2.5}$ species collected during November 2000 to February 2001 across three sites in Hong Kong and attributed most of the PM_{10} mass to $PM_{2.5}$ (60-80% depending on the site). From the yearlong $PM_{2.5}$ campaign data, Louie et al. (2005a) deduced carbonaceous species to be the largest contributor to $PM_{2.5}$ mass (50–75% at urban sites) followed by ammonium sulfate. Highest $PM_{2.5}$ concentrations were observed during winter for air masses arriving from north/northeast of HKSAR (Louie et al., 2005b). Guo et al. (2009) applied Principal Component Analysis (PCA) to $PM_{2.5}$ speciation datasets collected during 2000-2001 and 2004-2005. Secondary sulfate and vehicle exhaust factors were identified as the major contributors to the $PM_{2.5}$ mass. However, the vehicle exhaust factor was admixed with species associated with biomass burning and nitrate was smeared over various factors. Yuan et al. (2006) applied Positive Matrix Factorization (PMF) and Unmix receptor models to 24-hour integrated PM_{10} speciation data collected at eleven sites during 1998-2002 to identify sources and quantify their contributions.

Source apportionment of the PM_{10} speciation data was updated to include the time period from 1998 to 2008 (Yuan et al., 2013). Nine source categories resolved through PMF were split into local (operationally within HKSAR jurisdiction) and non-local sources, although the close proximity of the highly industrialized PRD region to HKSAR could prevent these ‘non-local’ sources from exhibiting characteristics of a regional source (i.e.

homogeneous impacts across the HKSAR). Vehicle exhaust, residual oil combustion, fresh and aged sea salt were categorized as local sources, while non-local sources were comprised of secondary sulfate, secondary nitrate, coal combustion/biomass burning, trace metals and soil/dust. Contributions from the local and non-local source categories were similar during summer, but during the winter the non-local source contributions were higher (77%) than the local source contributions (22%). From 1998 to 2008, the annual average contributions from secondary sulfate and nitrate, the dominant non-local sources, increased by ~ 8 $\mu g/m^3$. At the same time, annual average contributions from vehicle emissions, the dominant local source, decreased by ~ 6 $\mu g/m^3$ and were linearly correlated with the HKEPD road transport PM emission inventory, providing evidence for the reduction of vehicle emissions from the numerous emission controls implemented within the HKSAR.

In addition to the PM speciation data, HKEPD has also collected hourly PM_{10} mass since 1997 using Tapered Element Oscillating Microbalance (TEOM) monitors at a network of fourteen sites across HKSAR. Man and Shih (2001) used TEOM data collected at 11 sites during 1997-1999 to analyze seasonal trends in PM mass by classifying air masses into seven distinct patterns. The highest observed concentrations were attributed to wintertime air masses originating over the Asian continent and lowest for oceanic air masses in summer. Such seasonality of the air masses in the region is governed by the East Asian Monsoon and characterized by a long winter and summer with a relatively short spring and fall (Murakami and Nakazawa, 1985). In summer, the rising temperature over the continent develops a low-pressure system and drives the south/southwestern cyclonic winds to bring oceanic air masses to Hong Kong. During

winter, the prevailing north/northeastern anti-cyclonic winds over the continent bring air masses from mainland China to the Hong Kong region. The transition seasons of spring and fall are characterized by weak winds over the coastal region.

Synoptic-scale air mass transport patterns have been frequently utilized to identify and interpret relationships between observed air quality parameters and air mass history (Ashbaugh et al., 1985; Cheng and Lin, 2001; Lupu and Maenhaut, 2002; Hsu et al., 2003; Louis et al., 2005b). In general, each observed concentration at a given receptor location is assigned to every grid in the spatial domain along the corresponding back trajectory and combined using some statistical measure to obtain a conditional probability field based on the residence time of air masses (Stohl, 1998). Various approaches have been used to identify source regions for particulate matter over continental scales (Stohl, 1996; Keeler and Samson, 1989; Zhou et al., 2004; Scheifinger and Kaiser, 2007). Application of such tools to the Hong Kong region can be challenging because the heavily industrialized PRD region in Guangdong Province, China, is adjacent to the HKSAR. The spatial dimension of this region (~200 km) is on the order of the spatial resolution of meteorology data typically used for trajectory generation (~100 km). Hence, PRD-scale impacts cannot be resolved from larger-scale transport impacts. Air mass trajectory fields at much finer spatial resolution than routinely available are needed to meaningfully gauge impacts from source regions located at such close proximity but beyond HKSAR jurisdiction.

In this chapter, source apportionment results for the HKEPD $PM_{2.5}$ speciation dataset are compared and contrasted with the PM_{10} source apportionment presented in Yuan et al. (2013). In addition, hourly PM_{10} TEOM mass data, collected continuously at fourteen

sites, is used to examine spatiotemporal variability by developing day-specific semi-quantitative estimates of regional- and larger-scale contributions that uniformly impact most of the network. The site-specific seasonal and temporal patterns are interpreted by applying the methodology for analyzing monitoring network data presented in Yadav et al. (2013). PMF-resolved source contributions are used to interpret the spatial and temporal variability in PM_{10} TEOM mass collected across the network. Air mass back trajectories are clustered and used to interpret the trends in the TEOM data and PMF-resolved source contribution estimates. These analyses of the $PM_{2.5}$ speciation data and PM_{10} TEOM mass data, together with the PM_{10} source apportionment by Yuan et al. (2013), collectively provide a weight of evidence for the drivers leading to recent trends in Hong Kong PM air quality.

4.3. Datasets

Figure 4-2 shows the network of Air Quality Monitoring Stations (AQMS) collecting multiple particulate matter datasets across the HKSAR. Monitoring stations are categorized as: general stations, located in urban/commercial/residential areas to gauge neighborhood exposure; roadside stations, located at vehicle traffic-heavy intersections; and remote stations, located in regions typically isolated from urban development. Table 4-1 lists the valid sampling periods for each dataset from 1998 to 2009. Additional details on the monitoring network and data characteristics are presented in Section C.1 of the Appendix C.

PM_{10} speciation data were collected at up to ten sites from 1998 to 2008. 24-hour integrated samples were collected at 1-in-6 day frequency onto quartz filters and analyzed

Table 4-1. Overview of Hong Kong air quality monitoring stations, PM datasets, and the valid sampling periods used in this study.

Site Characteristics	Monitoring Station		PM_{10} (24 hour integrated)		$PM_{2.5}$ (24 hour integrated)		TEOM PM_{10} (hourly)	
			Sampling Period	Count	Sampling Period	Count	Sampling Period	Count
Urban/ Residential/ Commercial	Central/Western	CW	Jan. 1998 - Dec. 2008	641	-	-	Jan. 1998 - Dec. 2008	92427
	Eastern	EN	-	-	-	-	Jan. 1999 – Dec. 2008	86512
	Kwai Chung	KC	Jan. 1998 - Dec. 2000	178	-	-	Jan. 1999 - Dec. 2008	86292
	Kwun Tong	KT	Jan. 1998 - Dec. 2008	598	-	-	Jan. 1998 - Dec. 2008	87871
	Sham Shui Po	SSP	Jan. 1998 - Dec. 2008	586	-	-	Jan. 1998 - Dec. 2008	92680
	Sha Tin	ST	Jan. 1998 - Dec. 2000	173	-	-	Jan. 1998 - Dec. 2008	93434
	Tung Chung	TC	Apr. 1999 - Dec. 2008	575	-	-	Apr. 1999 - Dec. 2008	83523
	Tai Po	TP	Jan. 1998 - Dec. 2000	181	-	-	Jan. 1998 - Dec. 2008	87768
	Tsuen Wan	TW	Jan. 1998 - Dec. 2008	631	Nov. 2000 - Oct. 2001 Nov. 2004 - Oct. 2005 Dec. 2008 - Dec. 2009	184	Jan. 1998 - Dec. 2008	90750
	Yuen Long	YL	Jan. 1998 - Dec. 2008	649	Nov. 2004 - Oct. 2005 Dec. 2008 - Dec. 2009	129	Jan. 1998 - Dec. 2008	94446
Roadside	Causeway Bay	CB	-	-	-	-	Jan. 1998 - Dec. 2008	93839
	Central	CL	-	-	-	-	Jan. 1999 - Dec. 2008	84430
	Mong Kok	MK	Jan. 1998 - Dec. 2008	640	Nov. 2000 - Oct. 2001 Nov. 2004 - Oct. 2005 Dec. 2008 - Dec. 2009	183	Jan. 2001 - Dec. 2008	69280
Remote	Hok Tsui	HT	-	-	Nov. 2000 - Oct. 2001 Nov. 2004 - Oct. 2005 Dec. 2008 - Dec. 2009	180	-	-
	Tap Mun	TM	-	-	-	-	Apr. 1998 - Dec. 2008	90829

for gravimetric mass, elements, ions, organic carbon and elemental carbon. $PM_{2.5}$ speciation data were collected at up to four sites during three one-year campaigns (2000-2001, 2004-2005 and 2008-2009). Samples were collected at 1-in-6 day frequency onto quartz and Teflon filters that were analyzed for gravimetric mass and the aforementioned components. In addition to these PM speciation datasets, TEOM monitors have been operated at fourteen sites to provide hourly PM_{10} mass. Data from 1998 to 2008 are used in this study. The TEOM units were operated at 50^{o}C to remove particle-bound water and consequently other volatile species, thereby representing the non-volatile component of the ambient particulate matter. Hourly mass data were converted to daily averages with data gaps of one-to-two hours imputed by linear interpolation. For each site, only those days with 24 hours of measured/imputed data were retained resulting in 84-95% daily data completeness depending on the site.

Air mass back trajectories were obtained from www.sharedair.org, which includes an archive for several cities worldwide, as compiled by the University of Michigan for U.S. Environmental Protection Agency (USEPA). The seven-day air mass back trajectory dataset used in this study spans the period January 1, 2000 through December 31, 2009 and includes four trajectories per day (arrival times of 00, 06, 12, and 18 UTC) generated using Revision 4.9 of the NOAA HYSPLIT4 Model (Draxler, 1999). Meteorological data sets used to generate the trajectories were FNL (2000-2003) and GDAS1 (2004-2009; 1° resolution which corresponds to ~100 km x 100 km). The trajectory arrival height was fixed at 500 m and back trajectory endpoints were generated for every two hours for up to seven days duration. Trajectories generated for the Central/Western (CW) station on Hong Kong Island were utilized in this study with 99.6% completeness (14,546

trajectories out of a possible 14,608 trajectories). 70% of the trajectories spanned the entire seven day duration while the reminder exited the modeled spatial domain prior to seven days.

4.4. Methodology

4.4.1. Source Apportionment of PM Speciation Dataset

Receptor modeling of the PM speciation datasets was performed to gain insights into the emission sources and atmospheric processes driving air quality conditions. Several receptor models were used and the results obtained using EPA PMF 3.0 are utilized in this chapter. Source apportionment of the Hong Kong PM_{10} speciation dataset from 1998 to 2008 was presented in Yuan et al. (2013). Source apportionment of the $PM_{2.5}$ dataset was performed by combining data from the three one-year campaigns and the four sites. Details are presented in Section C.2 of the Appendix C. Sample-specific uncertainties were developed using collocated samples as described by the Supplementary Information in Yuan et al. (2013) and these uncertainty estimates were used in the source apportionment modeling. Solutions for six to ten factors were examined and the eight factor solution was deemed optimal and mapped to source categories based on their tracer species (Yuan et al., 2013).

4.4.2. Examining spatiotemporal variability in TEOM mass

The extensive hourly PM_{10} TEOM mass collected at up to 14 sites from 1998 to 2008 was used to examine spatiotemporal variability by developing day-specific semi-quantitative estimates of regional- and larger-scale (i.e. synoptic) contributions that uniformly impact most of the network. Similar approaches have been utilized for multi-

site monitoring networks to understand spatiotemporal trends in high PM days and to identify putative local source locations (Turner, 2008; Turner, 2009; Yadav et al., 2009; Yadav et al., 2013). Depending on the monitoring station locations, the sites experience differential PM impacts from urban- and finer-scale (i.e. local) sources in addition to the presumed homogeneous regional- and larger-scale impacts. Non-local PM impacts in this study are collectively labeled as synoptic-scale even though synoptic-scale (i.e. spatial domain > 1000 km) PM impacts may be admixed with regional-scale (~50-1000 km) PM burdens from PRD source emissions. The PM mass data measured across the network can be utilized to construct a time series for a network-wide baseline to differentiate the network-wide, uniform behavior from the site-specific PM impacts. Despite the limitations posed by this simplified representation of a complex system, this conceptualization dampens fluctuations in synoptic-scale impacts arising from year-to-year variations in synoptic weather patterns and thereby enables a refined interpretation of the spatiotemporal patterns.

The 5^{th}-lowest daily average TEOM mass across the 14 site network was chosen as the daily baseline in this study to capture the behavior broadly representative of the general stations. Daily average mass at each site in excess of the baseline was assigned as the site-specific excess mass. The daily excess mass with respect to the defined baseline can be negative at some sites yet are operationally referred to as excess as well in this chapter. Additional details on this approach, including the rationale and application to the TEOM monitoring network in Hong Kong, are presented in Section C.3 of the Appendix C. Homogeneity in mass measured across a network of monitoring sites is gauged using a scattergram of Pearson's correlation coefficient (PCC) and the coefficient

of divergence (COD) i.e., a PCC-COD plot, for the site-specific daily average TEOM mass against the network-wide baseline (Yadav et al., 2013). Annual and monthly distributions in daily average baseline and site-specific excess TEOM mass were examined towards understanding the drivers of PM variability across the region.

4.4.3. Clustering of Air Mass Trajectories

To understand the drivers of spatiotemporal patterns in both TEOM mass and PMF-resolved source contributions, synoptic-scale air mass back trajectories are clustered into groups with similar transport patterns. Air masses with similar transport histories are presumably exposed to similar emission fields but could have different atmospheric processing capacity, depending on various factors such as moisture content, pollutant composition, etc. The clustering approach used in this study assigns each trajectory to an air mass transport pattern independent of the air quality data parameter, enabling examination of any parameter of interest such as concentration metrics or modeled source contribution estimates. By simply relating the observed concentrations to air mass transport patterns, this approach makes no inference whether the pollutant is regionally transported or the transport pattern is a proxy for local meteorological conditions that modulate local source impacts (e.g. atmospheric ventilation).

Air mass trajectories for the Hong Kong region spanning 2000 to 2009 are grouped into clusters based on refinements to the methodology of Dorling et al. (1992) as described in Section C.4 of the Appendix C. A subset of N trajectories is chosen as seeds – each representing one cluster – and the individual trajectories are matched to the seed which has the minimum trajectory-seed separation distance along the trajectory path. After the first iteration the seeds are replaced with the cluster-mean trajectories and the entire set of

trajectories are iteratively reassigned to clusters until there is no change in the trajectory-cluster matching with each update. The cluster set for which an increase in number of clusters does not appreciably reduce a global measure of separation distances between the trajectories and their respective cluster-mean trajectories is chosen as the final set of clusters. A ten trajectory subset (N = 10 seeds) was used to obtain a five cluster solution with each individual trajectory assigned to a particular cluster.

The five resolved air mass transport patterns (i.e. clusters), shown in Figure 4-3, have the following characteristics: (1) Slow ECC – relatively slow moving air masses transported from the north with the cluster-centroid trajectory located along the Eastern Coast of China (ECC); (2) Fast ECC – relatively fast moving air masses transported from the north with the cluster-centroid trajectory located along the ECC; (3) Stagnant/circulating – air masses residing over the HKSAR region for much of the seven day period; (4) S/SW – south/southwesterly air masses arriving from the Indian Ocean across Southeast Asia; and (5) East – air masses arriving from the Pacific Ocean due east of Hong Kong.

The trajectory clusters also corroborate the seasonal patterns observed over the region as described in Section C.4 of the Appendix C. Synoptic conditions during winter months are dominated by northerly air masses transecting the eastern region of mainland China. Summer monsoon months are characterized by prevailing south/southwesterly and easterly oceanic air masses. The transitions between these two seasons feature a mix of air masses with higher frequency of the stagnant/circulating air masses. These seasonal behaviors are generally consistent over the ten year period examined in the study, with 30-40% annual frequency of Slow ECC except in 2004 (44%). The annual frequency of oceanic air masses (S/SW and East) is ~30% with highest frequency in 2005 (34%)

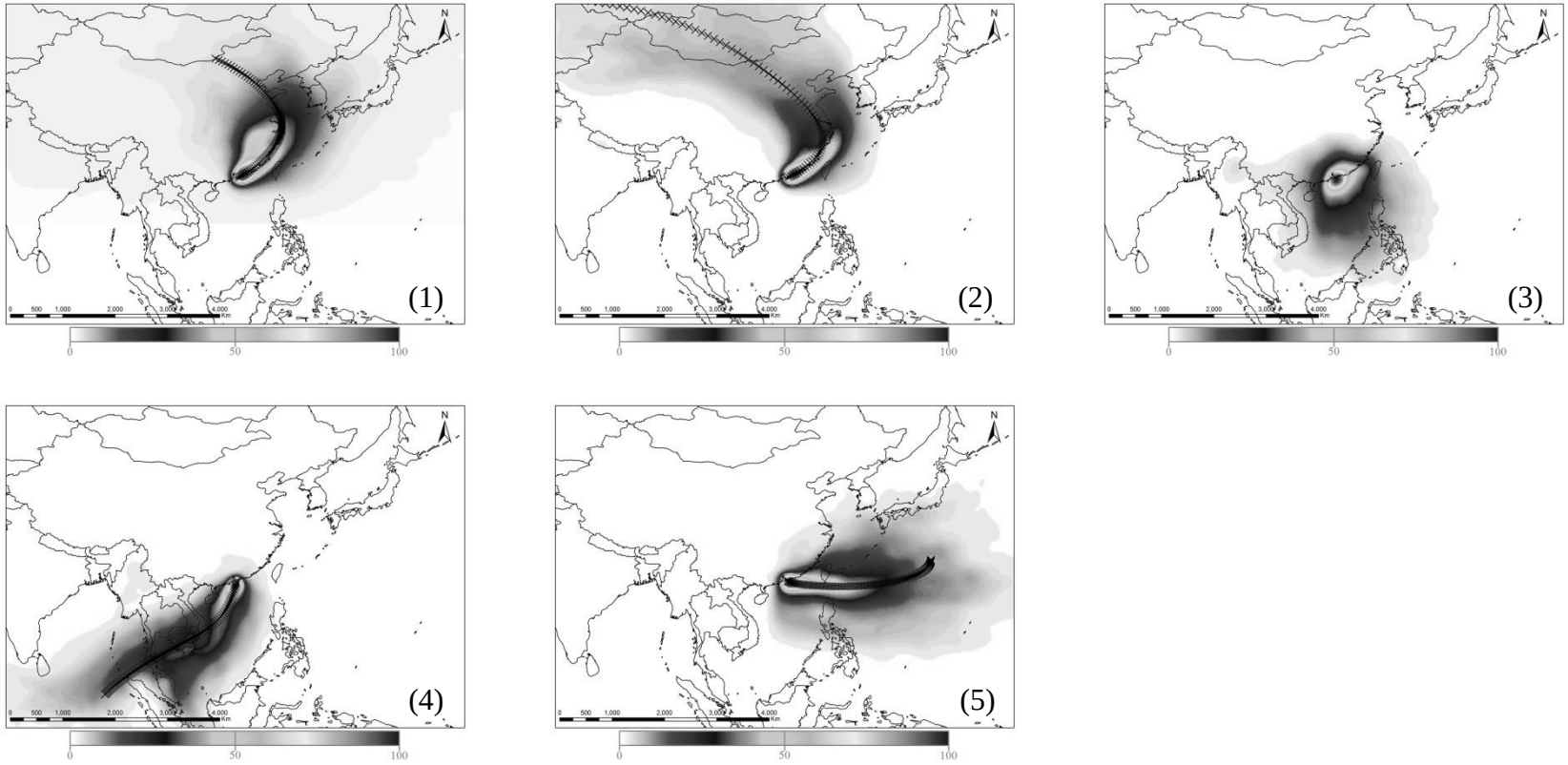


Figure 4-3. Air mass patterns resolved by cluster assignment of seven-day air mass back trajectories from 2000-2009 for Hong Kong: (1) Slow ECC, (2) Fast ECC, (3) Stagnant/circulating (4) S/SW and, (5) East. The graphs show the normalized number density of trajectories passing over each grid (on a scale of 0-100) and the cluster-mean trajectories for every cluster except the Stagnant/circulating cluster.

and lowest in 2004 (21%). Such year-to-year variability in the relative frequency of the air mass transport patterns could influence the temporal trends observed for air quality parameters. To associate daily averaged mass or species concentrations with the representative air mass trajectory cluster, a single cluster is selected to represent the day if at least three of the four daily trajectories are assigned to the same cluster. The remaining days are deemed to have transitional conditions and no air mass patterns are assigned to such days. This approach results in an air mass cluster designation for 82% of days over 2000 to 2009 which are used to interpret temporal trends in PM burdens.

4.5. Results and Discussions

4.5.1. Source Apportionment of $PM_{2.5}$ Speciation Dataset

Source apportionment modeling results for the $PM_{2.5}$ speciation dataset are presented in detail in Section C.2 of Appendix C. Secondary sulfate and vehicular exhaust factors were the dominant contributors to total $PM_{2.5}$ mass at 32% and 26%, respectively. Secondary nitrate and biomass burning factors contribute 13% and 11%, respectively, while contributions from each of the remaining factors – trace metals, residual oil combustion, crustal/soil and chlorine/fresh sea salt – were generally less than 7% of the total $PM_{2.5}$ mass. The secondary sulfate factor shows consistent mass contributions at all sites irrespective of the site characteristics, with its contribution to $PM_{2.5}$ mass increasing from 24% in 2001 to 33% in 2009. Vehicular exhaust contributions, although highest at the MK roadside station and lowest at the TM remote station, consistently decreased by ~50-58% depending on the site. Secondary nitrate factor contributions also decreased at all sites over the decade, but the highest contributions were observed at the YL general station and the MK roadside station.

4.5.2. Comparison between $PM_{2.5}$ and PM_{10} factor contributions

PMF-resolved source contribution estimates (SCE) from source apportionment of the $PM_{2.5}$ (this study) and PM_{10} (Yuan et al., 2013) datasets are compared in Figure 4-4. Annual average SCEs are calculated using the matched $PM_{2.5}$-PM_{10} sampling days for TW and YL. $PM_{2.5}$ and PM_{10} sampling days were not synchronized at MK. Thus the annual average SCEs for MK were calculated using the respective year-long field campaign data and could explain the higher $PM_{2.5}$ annual SCEs than PM_{10}.

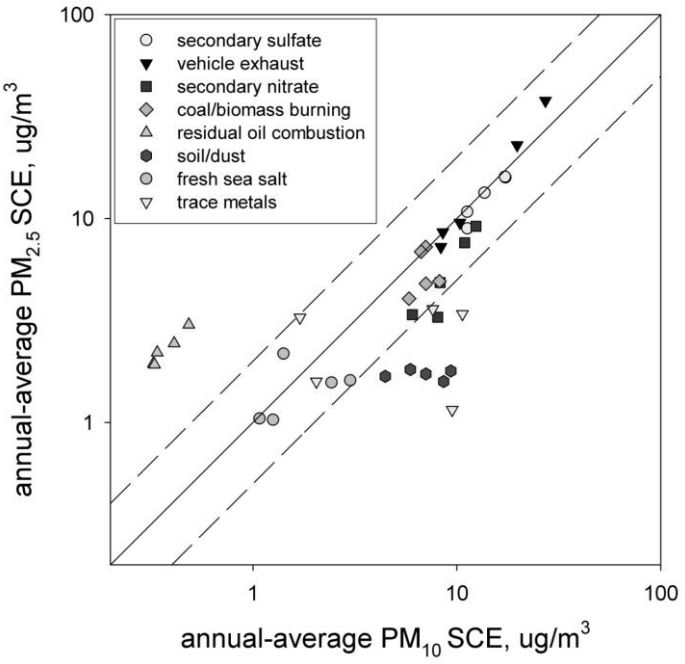


Figure 4-4. Annual average $PM_{2.5}$ and PM_{10} source contribution estimates (SCEs) for each of the eight PMF-resolved factors by site and year: TW (2001, 2005), YL (2005), and MK (2001, 2005). The solid line is the 1:1 line; the dashed lines are 2:1and 1:2 lines.

In general, there is good quantitative agreement between the $PM_{2.5}$ and PM_{10} SCEs for both the secondary sulfate and vehicle exhaust factors, which is consistent with emissions from these source categories being dominated by fine PM. $PM_{2.5}$ secondary nitrate contributions are about 50% of the PM_{10} factor, suggesting that considerable nitrate is in the coarse particle fraction, although the nitrate data might be prone to sampling artifacts. The unexplained mass in the $PM_{2.5}$ and PM_{10} secondary nitrate factors are nearly identical and thus cannot explain the difference. Soil/dust contributions are predominantly in the coarse particle size range, as expected. $PM_{2.5}$ coal/biomass burning contributions are about 30% of the PM_{10} factor possibly because of re-suspension of soil associated with biomass burning. Fresh sea salt contributions are scattered about the 1:1 line but the $PM_{2.5}$ factor contributions are likely unreliable as sodium data was excluded from modeling. Trace metals contributions are scattered possibly because of the large unexplained mass in the PM_{10} factor (9% and 68% for $PM_{2.5}$ and PM_{10}, respectively). Residual oil combustion contributions to $PM_{2.5}$ are almost twice the corresponding PM_{10} factor because of the higher loading of EC and OC onto the $PM_{2.5}$ factor profile. Overall, despite differences in sampling and analysis techniques, the PM_{10} and $PM_{2.5}$ apportionments are generally consistent with how source contributions are expected to distribute between fine and coarse fractions with all of the secondary sulfate and vehicle exhaust in fine PM.

Temporal changes in $PM_{2.5}$ SCEs are not clearly apparent with the dataset limited to three one-year campaigns. However, temporal trends in PMF-resolved $PM_{2.5}$ and PM_{10} secondary sulfate and vehicular exhaust factors, shown in Figure 4-5, are consistent. The annual average $PM_{2.5}$ secondary sulfate factor SCEs track the increase in annual average

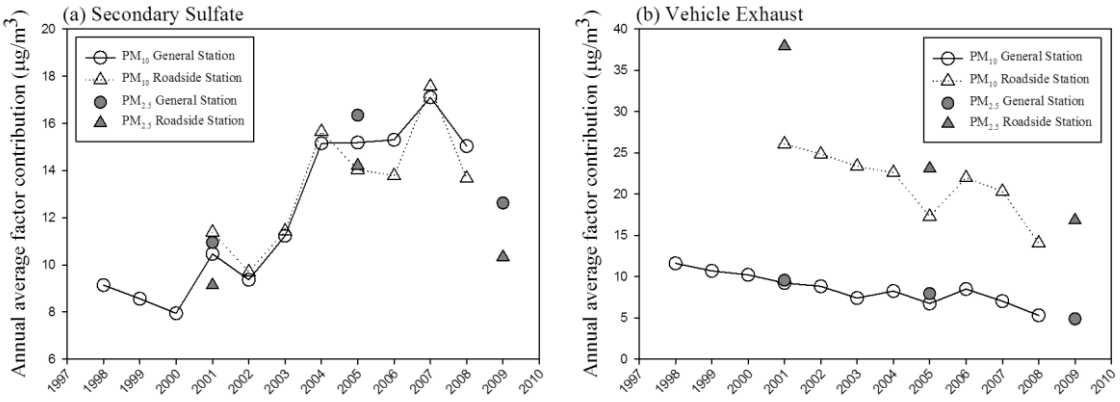


Figure 4-5. Temporal trends for annual average PM_{10} and $PM_{2.5}$ factor contributions for general and roadside stations for PMF-resolved (a) secondary sulfate and (b) vehicle exhaust factors.

PM_{10} sulfate factor SCEs. The drastic decrease in secondary sulfate contributions in 2009, observed only for $PM_{2.5}$ because 2009 PM_{10} data were not available when this analysis was conducted, corroborates the trend in PM burdens observed across the PRD region during 2006-2012 (Zhong et al., 2013). The decrease could result from the decrease in economic activity following the global recession in 2008 as well as from implementation of stringent pollution control policies in PRD region such as desulfurization measures in power plants (Zhong et al., 2013).

Similar secondary sulfate SCEs at general and roadside stations further strengthens its non-local emission scale characteristic. On the contrary, vehicle emissions contributions at the roadside station are significantly higher than at general stations and both have monotonically decreased for $PM_{2.5}$ and PM_{10}. Roadside station $PM_{2.5}$ vehicle exhaust SCEs are greater than the PM_{10} contributions possibly due to the aforementioned mismatch in $PM_{2.5}$ and PM_{10} sampling days at MK roadside station and/or higher EC/OC ratio in $PM_{2.5}$ factor which might be driven by differences in EC measurement method between the two datasets (see Section C.1 of the Appendix C). Overall, source

apportionment of the $PM_{2.5}$ data corroborates the PM_{10} source apportionment by Yuan et al. (2013) and adds to the weight of evidence for the sources influencing PM air quality in Hong Kong.

4.5.3. Spatial and Temporal Trends in TEOM PM_{10} mass

Spatiotemporal variability in the PM_{10} TEOM network data is first examined by comparing the site-specific daily average mass against the network-wide baseline. This approach could not be taken for the PM_{10} speciation data because speciation sampling days were not synchronized across the network. Figure 4-6 shows the PM_{10} TEOM data PCC-COD plot where high PCC and low COD for a site with respect to the baseline indicate spatial/temporal homogeneity between the two time series. Tight clustering of most general stations (with the exception of TC and YL) towards the homogeneity end of the plot (i.e., towards PCC = 1 and COD = 0) indicates that the baseline is indeed

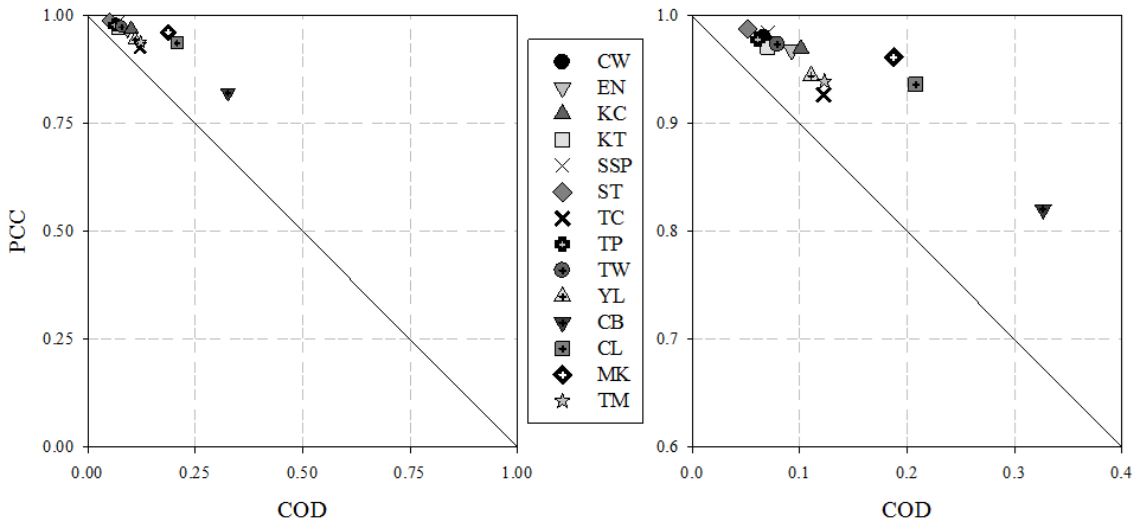


Figure 4-6. Scatter plot of Pearson's correlation coefficient and coefficient of divergence (PCC-COD plot) for the daily-averagePM_{10} concentration time series at each site and the daily-average baseline concentration time series. The right panel is an expanded view of the left panel.

characterized by the mass measured at these general stations. In contrast, the positioning of the TC and YL general stations, the TM remote station, and the three roadside stations on the PCC-COD plot indicates greater variability and hence different or additional drivers for PM burdens than at the other general stations such as local sources, meteorology and/or geographic factors. The factors contributing to such site-specific PM burdens are explored by examining patterns for the daily average excess mass at each site.

To meaningfully interpret the patterns in site-specific excess mass, it is pertinent to understand the trends in baseline mass against which the excess are quantified. Figure 4-7 shows the annual and monthly mass distributions for the baseline along with the excess mass at two general stations – TW and YL; each characteristic of general stations defining the baseline (CW, EN, KC, KT, SSP, ST, TP and TW) and distinct from the baseline (TC and YL). Over the entire eleven year period (Figure 4-7 (a)), the bottom 5^{th} and 25^{th} percentiles of the baseline, representative of the cleanest days, are nearly the same. In contrast, the higher concentration metrics ($95^{th}$and 75^{th} percentiles, mean and median values) are greater for 2004-2008 than 1999-2003. These metrics are more than twofold higher during winter months compared to the summer months (Figure 4-7 (b)). Thus, PM levels on the cleanest days – usually occurring during summer months – have remained unchanged, but in recent years there has been an increased frequency of higher wintertime PM observed across the entire network. The excess mass at the general stations which contribute to defining the baseline, such as TW, display no discernible annual (Figure 4-7 (c)) or seasonal (Figure 4-7 (d)) patterns. Hence, the baseline in this

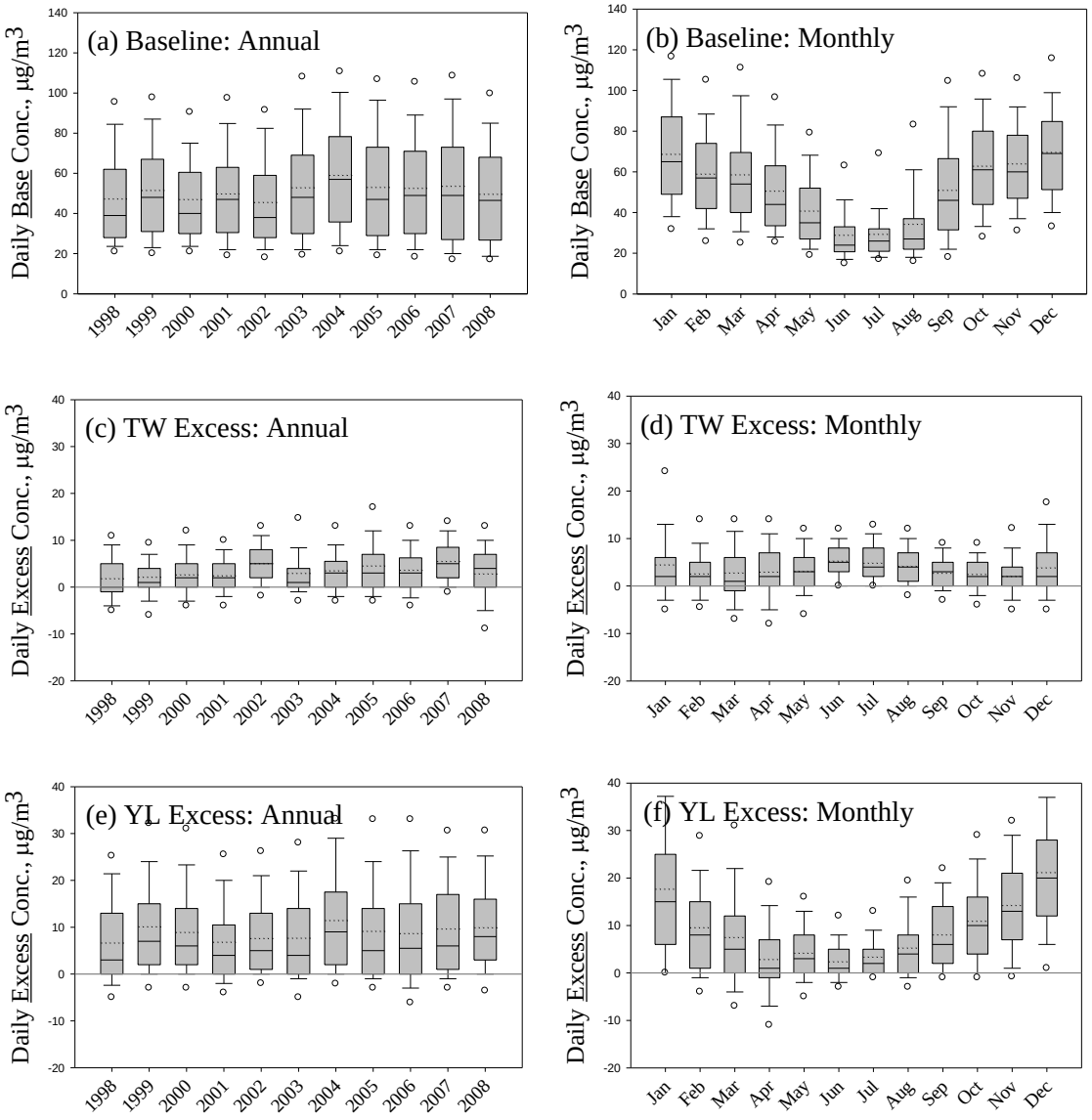


Figure 4-7. Annual and monthly distributions of daily-averaged PM_{10} TEOM mass concentrations for baseline (left column) and for excess mass concentrations at the TW and YL general stations. All box plots in this chapter are formatted as: open circles are 5th and 95th percentile values; the bottom and top of the boxes are 25th and 75th percentile values; and the interior dotted and solid lines are mean and median values, respectively.

study is again confirmed to broadly represent the mass measured at general stations that could result from the combined influence of synoptic -scale PM impacts as well as the PM burdens from increasing urbanization in Hong Kong. The contrasting PM trend at the YL (and TC) general station is characterized by higher annual excess mass (Figure 4-7 (e)) – particularly during winter months (Figure 4-7 (f)) – than at other general stations and are discussed later in this section.

PM trends for the CB roadside station and TM remote station, shown in Figure 4-8, further illustrate the advantage of using site-specific excess mass to understand the factors impacting spatiotemporal variability. The measured mass at CB (Figure 4-8 (a)) decreased during 1998-2003 and then remained generally constant during the latter half of the decade albeit with some year-to-year variability. With the baseline removed (Figure 4-8 (b)), these two patterns become sharper and the year-to-year variability is suppressed. Measured PM_{10} mass at TM (Figure 4-8 (d)) follows the modestly increasing temporal pattern exhibited by the baseline. In contrast to all other stations, excess mass at TM (Figure 4-8 (e)) monotonically increased over the decade. Similar to the seasonal patterns for the baseline (Figure 4-7 (b)), there is greater variability in excess mass stratified by month across 1998 to 2008 at TM during winter months compared to the summer months (Figure 4-8 (f)). This indicates that the monotonic increase primarily results from the long-term (inter-annual) increase in PM burdens from synoptic- and/or local-scale source contributions occurring during winter months. Relationships between seasonal air mass patterns and site-specific excess mass are discussed further in Section 4.5.4.

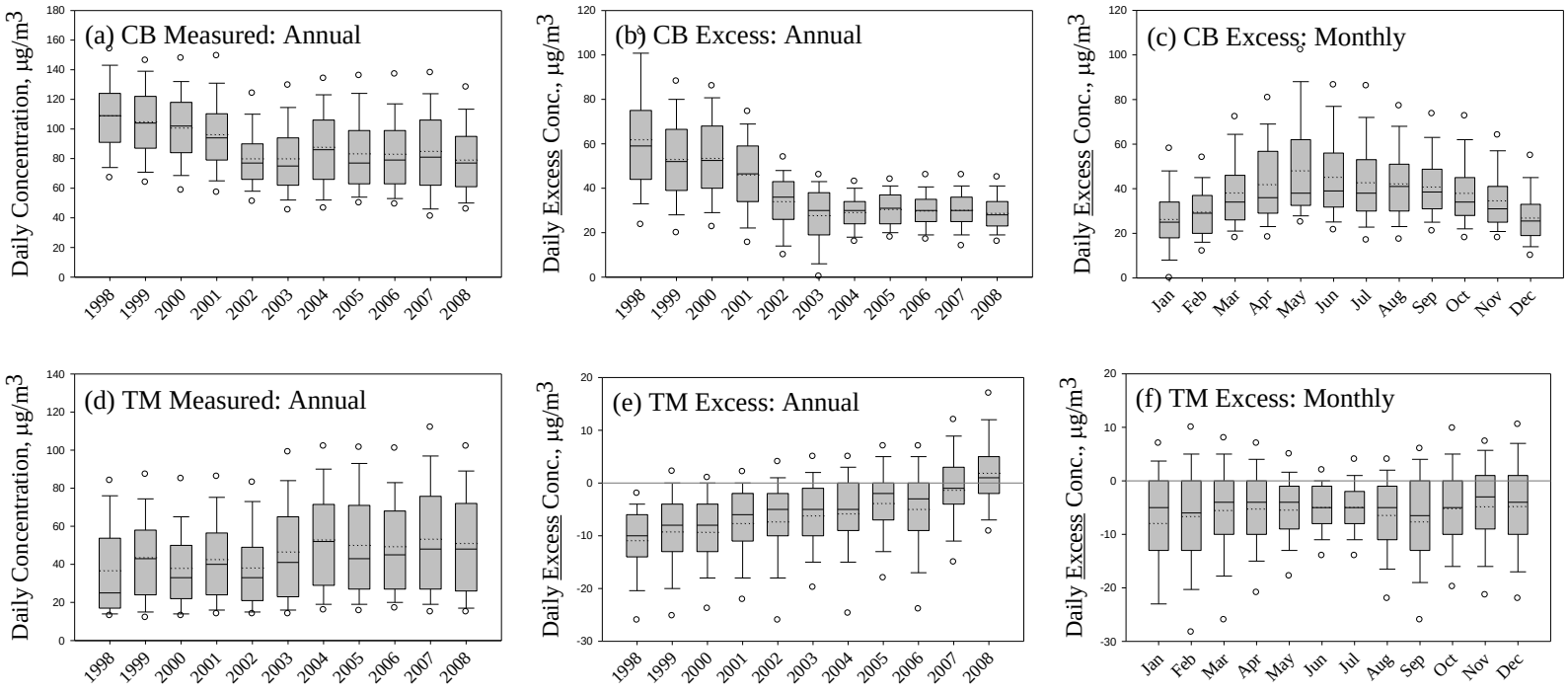


Figure 4-8. Daily-averaged PM_{10} mass concentration distributions at the CB roadside station (top) and TM remote station (bottom) for: measured mass by year (left); excess mass by year (center); and excess mass by month (right).

Figure 4-9 shows the seasonal distribution of PM_{10} excess mass at each site for the five year periods 1999-2003 and 2004-2008 during peak summer (June-July) and winter (Nov-Dec) months. The three categories of monitoring stations are arranged in the order of increasing distance from west to east of Hong Kong. Excess mass at the roadside stations are insensitive to season and have decreased from the earlier to latter half of the decade. Excess mass distributions at general stations display distinct patterns based on their location. General stations located on the southeastern side of the mountains ranging from northeast to southwest of Hong Kong i.e., TW, KC, CW, SSP, TP, ST, EN and KT general stations do not display discernible spatial patterns. The excess mass distributions are nearly identical across the two five-year periods and the site-to-site variability likely reflects the spatially varying impacts of local sources. In contrast, excess mass at sites on the western side of the mountains i.e., TC and YL general stations are significantly higher during winter months than summer for both the five-year periods. During winter of 2004 (the only year common among all three datasets for YL), the highest TEOM mass among the general stations was also observed at YL, which is also consistent with the higher contributions of PMF-resolved secondary nitrate, soil/dust and biomass burning factors at YL in both PM_{10} and $PM_{2.5}$ datasets.

Higher PM burdens at YL (and TC) could result from the combined influence of various factors discussed in this paragraph. The YL station, located on the northwestern edge of The New Territories, is exposed to the heavily industrialized PRD region and the Port of Shekou to the northwest and has 800-1000 m high mountains immediately to the east and south. Likewise, the TC station is located on the northwestern edge of Lantau Island and is exposed to the PRD region to northwest with 700-950 m high mountains to the east and

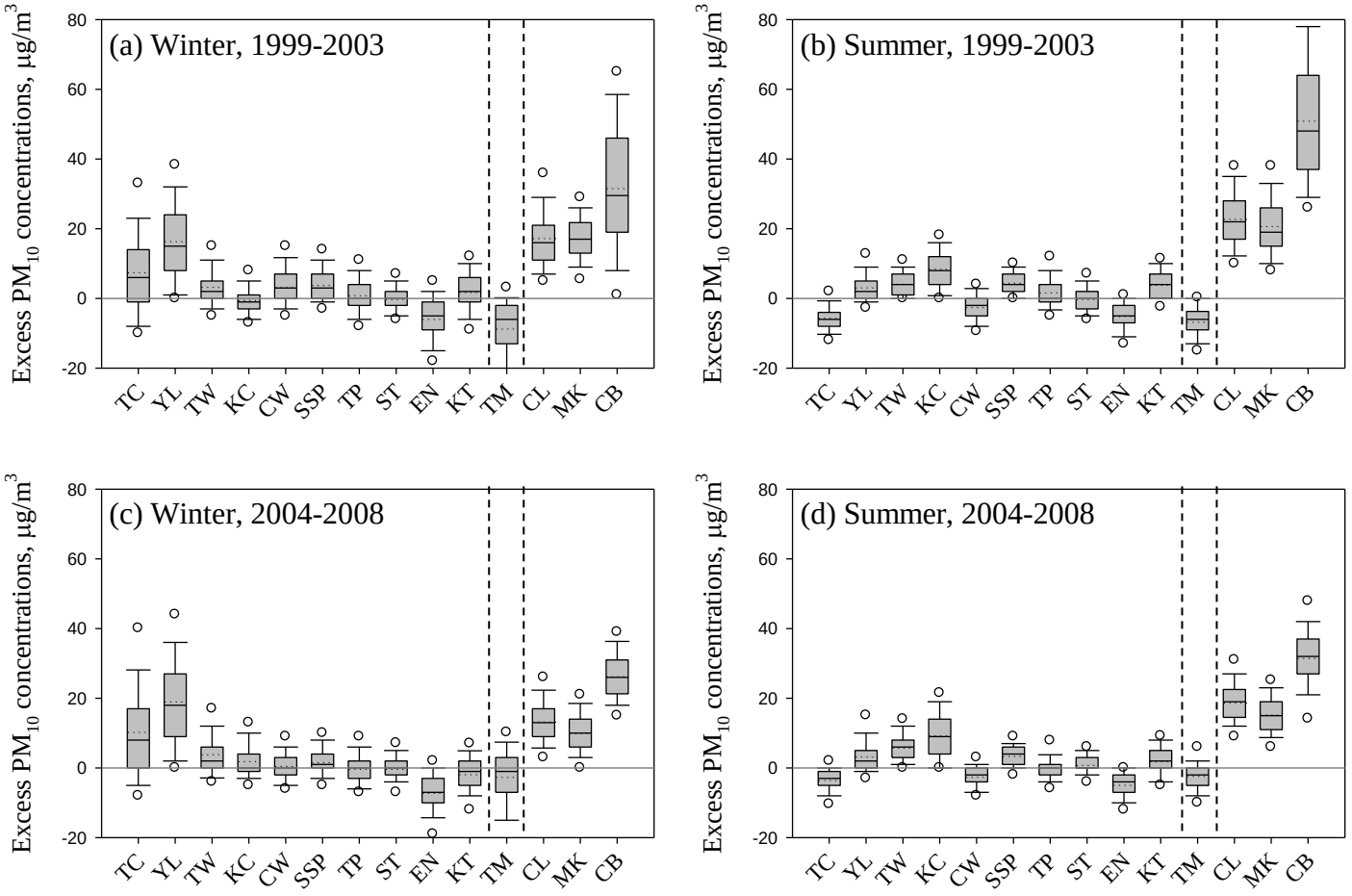


Figure 4-9. Spatial distributions of daily–average PM_{10} excess mass concentrations during winter (Nov-Jan) and summer (Jun-Jul) months over two five year periods (1999-2003 and 2004-2008) for the three categories of monitoring stations spanning west to east across Hong Kong.

south. The influence of such geographical features, seasonal air masses and location of major emission zones in close proximity to these sites are further examined by contrasting the seasonal distribution of hourly mass measured at YL in excess of that at TW, a general station located on the southeastern side of the mountains, over the two five year periods 1999-2003 and 2004-2008 (Figure 4-10). During the winter months (Nov-Jan) when air masses from north/northeast transect the PRD region, the median hourly mass observed between 7 am to 1 pm are ~67% higher than at other hours for both periods. In contrast, mass measured at YL and TW are similar during the summer months (June-July) when oceanic air masses prevail. Additionally, mixing layer height measurements at YL by Lidar at 30 minute resolution over the six year period 2003-2009 (Yang et al., 2013) indicates that the mixing layer depth grows from ~0.6 km at 8 am to ~1.0-1.2 km at noon during summer and from ~0.5 km to ~0.9-1.0 km during winter months. The combined influence of these factors could result in the larger excess mass observed at YL and TC than at general stations located on the southeastern side of the mountains. While PRD source region contributions to PM burdens in the HKSAR cannot be accurately quantified because of its proximity, they can influence the spatial variability of PM across Hong Kong.

4.5.4. Relationships between PM Burdens and Synoptic Air Mass Patterns

The temporal trend in PM datasets are further interpreted using the daily air mass back trajectory cluster assignments. Site-specific PM_{10} TEOM mass have high daily data completeness and thus the trajectory assignments are made for each site in the network. $PM_{2.5}$ speciation samples collected at 1-in-6 day frequency at three-to-four sites over three one-year campaigns account for only 5% of days overlapping with cluster

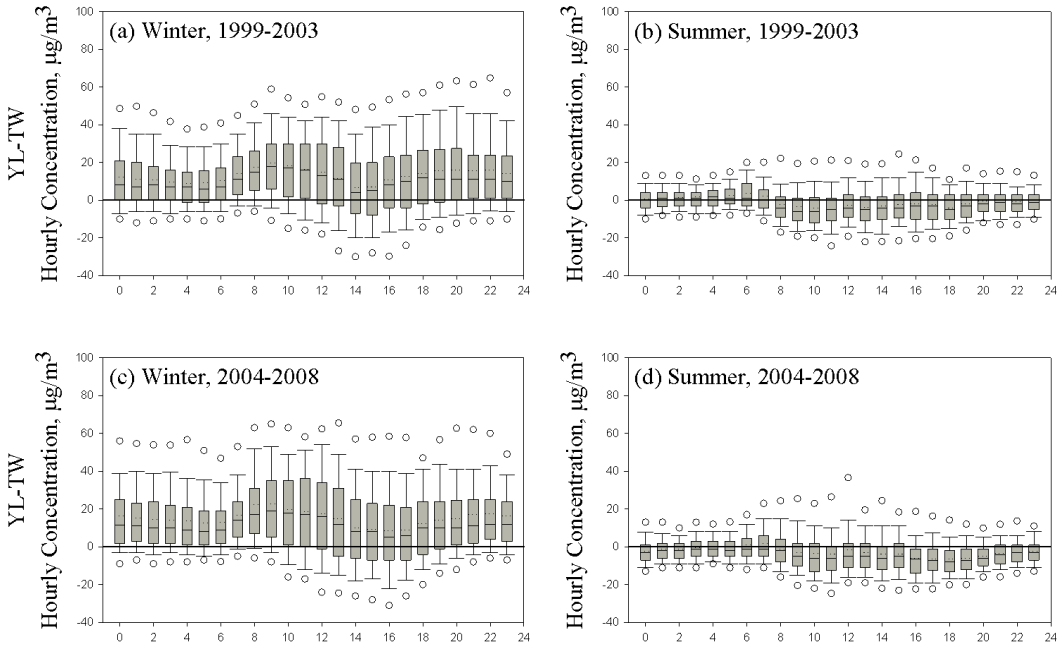


Figure 4-10. Hourly distributions of PM_{10} mass at YL in excess of TW (i.e. YL-TW) during winter (Nov-Jan) and summer (Jun-Jul) for two five-year periods: 1999-2003 and 2004-2008.

assignments for the period 2000-2008 and thus the $PM_{2.5}$ speciation dataset is excluded from this analysis. PM_{10} speciation samples were collected at 1-in-6 day frequency across the entire period 2000-2008. Sampling across the network of ten sites is not synchronized but rather occurs on different days of the week. Given that the SCEs for PMF-resolved factors are largely consistent across the general stations, especially for the secondary sulfate and vehicle exhaust factors, the non-synchronized sampling schedule is exploited to construct a time series with maximum data completeness. In this case, the PMF-modeled factor SCEs for concurrently sampling general stations are averaged to represent a network-wide SCE for each day. This approach is also consistent with the 1° resolution of the meteorological dataset used to generate the trajectories and yields a daily time series with 67% completeness over 2000-2008 as compared to a maximum

completeness of 17% for a single station. Key results are presented in the remainder of this section with additional details provided in Section C.5 of the Appendix C.

The daily network-average SCEs derived from the PM_{10} PMF modeling are grouped by cluster and scaled such that, over the entire dataset, the average SCE for each factor is unity. This scaling facilitates comparisons across different source categories. For each air mass cluster the scaled annual median factor contributions are regressed on year using a linear-least squares regression. Thus, the regression slope for each cluster represents the corresponding linearized rate of change in scaled factor contribution per year over 2000-2008. The slopes are presented in Table 4-2 for those cases where the change is statistically distinguishable from zero at the 95% confidence interval. The vehicle exhaust factor exhibits a statistically significant decrease for each of the air mass clusters. For example, the median scaled vehicle exhaust contributions decreased by 0.12 per year, i.e. 0.71$\mu g/m^3$ per year for the Slow ECC cluster. This pattern is consistent with a local source that has been the focus of emission reduction programs. Variations in the rate of change across the air mass transport patterns for this local source factor likely reflects the coupling between synoptic weather and the local atmospheric ventilation conditions. Secondary sulfate contributions increased for the Slow ECC and Fast ECC clusters and are statistically indistinguishable from zero for the remaining clusters. This pattern of a significant change for only the ECC clusters is consistent with increasing impacts from the eastern portion of China. Further, for such air masses the increase in secondary sulfate factor contributions is nearly identical to the decrease in the vehicle exhaust contributions.

Table 4-2. Annualized rate of change in the scaled annual median PM_{10} source contributions obtained from PMF modeling. The reported change is from a linear least-squares regression against year for each air mass transport cluster. Values are presented only for those cases with changes that are statistically distinguishable from zero at the 95% confidence level.

Air Mass	Vehicle Exhaust	Crustal/Soil	Smelting (Zn)	Secondary Sulfate
Slow ECC	-0.12 ± 0.06	-0.04 ± 0.03		0.14 ± 0.06
Fast ECC	-0.11 ± 0.05			0.10 ± 0.05
Stagnant	-0.06 ± 0.05			
S/SW	-0.08 ± 0.04		-0.02 ± 0.01	
East	-0.08 ± 0.04			

The site-specific daily average PM_{10} TEOM mass data are also stratified by air mass cluster and for each cluster the annual median mass is regressed on year using a linear-least squares regression. Thus, the regression slope represents a linearized rate of site-specific change in PM_{10} mass per year over 2000-2008. Figure 4-11 shows the temporal changes in PM_{10} mass across the network for the Slow ECC and East air mass clusters. These two clusters highlight the association of seasonal air mass transport patterns with PM burdens across the region. For air masses arriving from the eastern portion of China, which predominantly occurs during the wintertime, only those sites located along the periphery of the HKSAR (TC, YL, and TM) have experienced a degradation of air quality. At YL and TC, the compounding effects of PRD-scale emissions from close proximity and the wintertime accumulation due to geographic barriers also contribute to the increase in PM_{10} mass at these sites. The TM remote station, located on an island to the northeast of Hong Kong, is directly impacted by winter air masses arriving from the

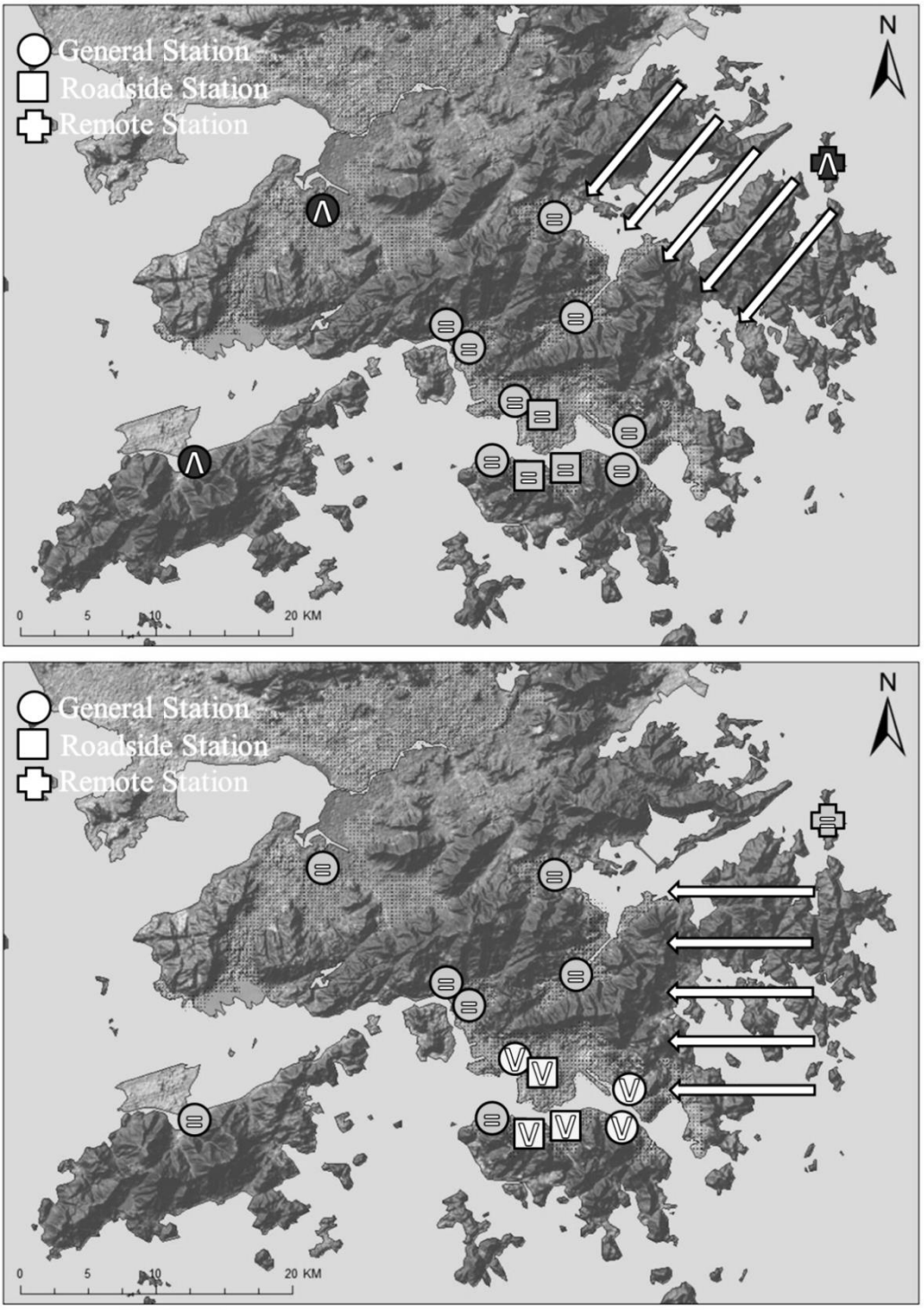


Figure 4-11. Site-specific rate of change in annual median PM_{10} mass over 2000-2008 for: (a) slow moving air masses along the Eastern coast of China (Slow ECC); and (b) air masses from the Pacific Ocean (East). Changes represented as: Black (∧) – Degradation of air quality; Grey (=) – Statistically indistinguishable change; and, White (∨) – Improvement in air quality.

eastern portion of China. Additionally, local marine activities at the Yantian Port, located 10 km northwest of TM have also increased over the study period (Ng et al., 2013). The observed degradation of air quality at TM corroborates with the increase in PM burdens on multiple spatial scales and raise concern about the representativeness of TM as a remote station. At sites located in the urban core of Hong Kong (SSP, KT, EN, MK, CL and CB), the increase in synoptically transported PM offsets the reduction from vehicle exhaust contributions resulting in overall statistically indistinguishable changes in air quality. However, when summertime oceanic air masses impact the region, such as the East trajectory cluster, an improvement in air quality over 2000-2008 is observed at all sites in the urban core, whether general or roadside. Hence, consistent with vehicle emission control measures, urban-scale motor vehicle contributions to PM_{10} reduced over the decade while the synoptically transported secondary sulfate and associated PM from mainland China have increased resulting in the observed stasis/degradation of air quality in Hong Kong.

4.6. Conclusions

Source apportionment of $PM_{2.5}$ speciation data from three one year field campaigns at three-to-four sites across Hong Kong was performed using PMF and compared to previously-published PMF modeling results for PM_{10} speciation data from 1998 to 2008. In both cases the secondary sulfate and vehicle exhaust factors are the main contributors to PM mass burdens observed in Hong Kong. For each of these factors the contributions to $PM_{2.5}$ and PM_{10} are quite similar which demonstrates the PM is in the fine fraction. Further, PMF-resolved $PM_{2.5}$ and PM_{10} secondary sulfate factor contributions increased over 2000-2008 irrespective of the site characteristics. In contrast, vehicle emission

contributions are significantly higher at the roadside station than at the general stations and have monotonically decreased for both $PM_{2.5}$ and PM_{10}. Overall, the $PM_{2.5}$ source apportionment results corroborate the PM_{10} source apportionment results.

Analysis of PM_{10} TEOM hourly mass measured at fourteen sites across the HKSAR from 1998-2008 complements the findings from the PM speciation source apportionments. The intuitive model based on daily average baseline and site-specific excess mass provides insight into the temporal behavior of PM burdens. The baseline as defined in this analysis captures the temporal patterns observed at most of the general stations with influences from synoptic-scale contributions and also urban-scale contributions resulting from increasing urban development within Hong Kong over the decade. Removal of the baseline amplifies the key trends in site-specific measured mass and enables gauging PM variability across the network. The monotonic decrease in excess TEOM mass at all roadside stations validates the observed decrease in PMF-resolved vehicle exhaust contributions. The increase in contributions from non-local sources arriving with air masses from the eastern portion of China during winter months, such as the PMF-resolved secondary sulfate factor, is also captured by the temporal increase in excess mass measured at the TM remote station.

The proximity of the heavily industrialized PRD region limits the use of air mass trajectories generated at ~100 km spatial resolution to isolate long-range transport impacts. However, the cluster-specific PM spatial and temporal patterns do provide insights into source contributions. For each cluster, the temporal variation in the PMF-resolved factor contributions composited across all general stations and in the site-specific PM_{10} TEOM mass provides complementary information. Air masses arriving

from the eastern portion of China exhibit an increase in PMF-resolved secondary sulfate contributions over the past decade that is nearly identical to the decrease in the vehicle emission contribution. Air quality has degraded at stations located along the HKSAR periphery for the two air mass clusters arriving from eastern portion of China, while the changes have been insignificant for the other three air mass transport patterns. Admixing of PRD-scale emissions with larger-scale contributions add to the spatial variability of PM across the region. The geographic locations of the TC and YL monitoring stations exposes them most strongly to PRD-scale impacts from north/northwest and the local topography promotes seasonal accumulation of pollutants. The change in PM concentrations over 2000-2008 at sites located in the urban core of Hong Kong are insignificant for air masses arriving from the eastern region of mainland China because the decrease in PM from Hong Kong motor vehicle emission control programs is offset by the increase in long-range transport of secondary sulfate and associated PM sources. This weight of evidence analysis of long term trends in the various PM datasets provides insight towards development of air pollution control strategies, both within the Hong Kong territory and in collaboration with other regions including but not limited to Guangdong Province.

4.7. Disclaimer

The content of this chapter does not necessarily reflect the views and policies of the Hong Kong Special Administrative Region (HKSAR) Government, nor does mention of trade names or commercial products constitute an endorsement or recommendation of their use.

4.8. Acknowledgments

We acknowledge the support from HKEPD (Tender Ref AS09-056). This work was also sponsored by the Research Grant Council of Hong Kong, China (615406), the University Grant Committee (GMGS07/08.EG03) and the Hong Kong University of Science and Technology (HKUST) Fok Ying Tung Graduate School (NRC06/07.SC01). We thank Damgy H.L. Chan and David Heinzerling for their contributions to this project.

4.9. References

Ashbaugh, L.L.; Malm, W.C.; Sadeh, W.Z.; A residence time probability analysis of sulfur concentrations at Grand Canyon National Park, *Atmospheric Environment*, **1985**, 19(8), 1263–1270.

Cheng, M., Lin, C.; Receptor modeling for smoke of 1998 biomass burning in Central America, *Journal of Geophysical Research*, **2001**, 106, 22871-22886.

Cheng, Y.; Lee, S.C.; Ho, K.F.; Chow, J.C.; Watson, J.G.; Louie, P.K.K.; Cao, J.J.; Hai, X.; Chemically-speciated on-road $PM_{2.5}$ motor vehicle emission factors in Hong Kong, *Science of the Total Environment*, **2010**, 408, 1621–1627.

Dorling, S.R.; Davies, T.D.; Pierce, C.E.; Cluster analysis: A technique for estimating the synoptic meteorological controls on air and pollution chemistry, *Atmospheric Environment*, **1992**, 26A, 2575-2581.

Draxler, R.R.; HYSPLIT4 user's guide. NOAA Tech. Memo. ERL ARL-230, NOAA Air Resources Laboratory, Silver Spring, MD, **1999**.

Fung, Y.S.; Wong, L.W.Y.; Apportionment of air pollution sources by receptor modeling in Hong Kong, *Atmospheric Environment*, **1995**, 29, 2041‑2048.

Guo, H.; Ding, A.J.; So, K.L.; Ayoko, G.; Li, Y.S.; Hung, W.T.; Receptor modeling of source apportionment of Hong Kong aerosols and the implication of urban and regional contribution, *Atmospheric Environment*, **2009**, 43, 1159–1169.

Hagler, G.S.W.; Bergin, M.H.; Salmon, L.G.; Yu, J.Z.; Wan, E.C.H.; Zheng, M.; Zeng, L.M.; Kiang, C.S.; Zhang, Y.H.; Lau, A.K.H.; Schauer, J.J.; Source areas and chemical composition of fine particulate matter in the Pearl River Delta region of China, *Atmospheric Environment*, **2006**, 40, 3802–3815.

Ho, K.F.; Lee, S.C.; Chow, J.C.; Watson, J.G.; Characterization of PM_{10} and $PM_{2.5}$ source profiles for fugitive dust in Hong Kong, A*tmospheric Environment*, **2003a**, 37,1023-1032.

Ho, K.F.; Lee, S.C.; Chan, C.K.; Yu, J.C.; Chow, J.C.; Yao, X.H.; Characterization of chemical species in PM2.5 and PM10 aerosols in Hong Kong, *Atmospheric Environment*, **2003b**, 37, 31–39.

Hsu, Y.K.; Holsen, T.M.; Hopke, P.K.; Comparison of hybrid receptor models to locate PCB sources in Chicago, *Atmospheric Environment*, **2003**, 37, 545–562.

Huang, X.F.; Yu, J.Z.; Yuan, Z.; Lau, A.K.H.; Louie, P.K.K.; Source analysis of high particulate matter days in Hong Kong, *Atmospheric Environment*, **2009**, 43, 1196–1203.

Keeler, G. J.; Samson, P. J.; Spatial representativeness of trace element ratios, *Environmental Science & Technology*, **1989**, 23,1358-1364.

Lee, E.; Chan, C.K.; Paatero, P.; Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, *Atmospheric Environment*, **1999**, 33, 3201–3212.

Louie, P.K.K.; Chow, J.C.; Chen, A.; Watson, J.G.; Leung, G.; Sin, D.W.M.; $PM_{2.5}$ chemical composition in Hong Kong: Urban and regional variations, *Atmospheric Environment*, **2005a**, 338, 267-281.

Louie, P.K.K.; Watson, J.G.; Chow, J.C.; Chen, A.; Sin, D.W.M.; Lau, A.K.H.; Seasonal characteristics and regional transport of $PM_{2.5}$ in Hong Kong, *Atmospheric Environment*, **2005b**, 39, 1695-1710.

Lupu, A.; Maenhaut, W.; Application and comparison of two statistical trajectory techniques for identification of source regions of atmospheric aerosol species, *Atmospheric Environment*, **2002**, 36, 5607–5618.

Man, C.K.; Shih, M.Y.; Identification of sources of PM_{10} aerosols in Hong Kong by wind trajectory analysis, *Journal of Aerosol Science*, **2001**, 32, 1213–1223.

Murakami, T.; Nakazawa, T.; Transition from the Southern to Northern Hemisphere Summer Monsoon, *Monthly Weather Review*, **1985**, 113, 1470-1486.

Ng, S.K.W.; Loh, C.; Lin, C.; Booth, V.; Chan, J.W.M; Yip, A.C.K.; Li, Y.; Lau, A.K.H.; Policy change driven by an AIS-assisted marine emissions inventory in Hong Kong and the Pearl River Delta, *Atmospheric Environment*, **2013**, 76, 102-112.

Qin, Y.; Oduyemi, K.; Chan, L.Y.; Comparative testing of PMF and CFA models, *Chemometrics and Intelligent Laboratory Systems*, **2002**, 61, 75–87.

Qin, Y.; Chan, C.K.; Chan, L.Y.; Characteristics of chemical compositions of atmospheric aerosols in Hong Kong: spatial and seasonal distributions, *Science of the Total Environment*, **1997**, 206, 25–37.

Scheifinger, H.; Kaiser, A.; Validation of trajectory statistical methods, *Atmospheric Environment*, **2007**, 41, 8846-8856.

Stohl, A.; Trajectory statistics - A new method to establish source receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, *Atmospheric Environment*, **1996**, 30, 579-587.

Stohl, A.; Computation, accuracy and applications of trajectories – A review and bibliography, *Atmospheric Environment*, **1998**, 32, 947-966.

Turner, J.R.; Measurement error as a context for assessing intraurban variability in chemical speciation network data, *Proceedings of Air and Waste Management Association's 101st Annual Conference and Exhibition*, Portland, OR, June 24–27, **2008**, paper no.664.

Turner, J.R.; A conceptual model for $PM_{2.5}$ over southeast Michigan: Intraurban variability and high concentration days, *Proceedings from 102nd Annual Meeting of Air & Waste Management Association*, Detroit, MI, June 16-19, **2009**. Paper No. 2009-A-410-AWMA.

Yadav, V.; Turner, J.R.; Downs, J.; Rowles, T.; Intraurban variability of ambient particulate matter arsenic and other air toxics metals in St. Louis, *Proceedings from 102nd Annual Meeting of Air & Waste Management Association*, Detroit, MI, June 16–19, **2009**. Paper No. 2009–A–370–AWMA.

Yadav, V.; Turner, J.R.; Gauging intraurban variability of ambient particulate matter arsenic and other air toxics metals from a network of monitoring sites, *Submitted to Atmospheric Environment,* **October 2013**.

Yang, D.; Li, C.; Lau, A.K.H.; Li, Y.; Long-term measurement of daytime atmospheric mixing layer height over Hong Kong, *Journal of Geophysical Research: Atmospheres*, **2013**, 118, 1-12.

Yu, J.Z.; Tung, J.W.T.; Wu, A.W.M.; Lau, A.K.H.; Louie, P.K.K.; Fung, J.C.H.; Abundance and seasonal characteristics of elemental and organic carbon in Hong Kong PM_{10}, *Atmospheric Environment*, **2004**, 38, 1511–1521.

Yuan, B.; Lau, A.K.H.; Zhang, H.Y.; Yu, Z.; Louie, P.K.K.; Fung, J.C.H.; Identification and spatiotemporal variations of dominant PM_{10} sources over Hong Kong, *Atmospheric Environment*, **2006**, 40, 1803-1815.

Yuan, Z.; Yadav, V.; Turner, J.R.; Louie, P.K.K.; Lau, A.K.H.; Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies in Hong Kong over 1998-2008, *Atmospheric Environment*, **2013**, 76, 21-31.

Zhong, L.; Louie, P.K.K.; Zheng, J.; Yuan, Z.; Yue, D.; Ho, J.W.K.; Lau, A.K.H.; Science-policy interplay: Air quality management in the Pearl River Delta region and Hong Kong, *Atmospheric Environment*, **2013**, 76, 3-10.

Zhou, L.; Hopke, P.K.; Wei Liu, W.; Comparison of two trajectory based models for locating particle sources for two rural New York sites, *Atmospheric Environment*, **2004**, 38, 1955–1963.

Chapter 5 :Assessing measurement error using collocated ambient particulate matter speciation datasets from Hong Kong.

5.1. Abstract

Environmental data should be analyzed and interpreted within the context of the corresponding data quality. Some models that apportion air quality and other environmental data to emission sources or environmental processes can exploit, or even require, uncertainty estimates for these data. Various approaches have been taken to generate sample-specific uncertainties for source apportionment modeling. Collocated data offers the opportunity to integrate uncertainties from several aspects of the measurement process. In this chapter, collocated data collected for two ambient particulate matter (PM) speciation networks operated by the Hong Kong Environmental Protection Department (HKEPD) are assessed with emphasis on generating error structures for source apportionment modeling. The concentration dependence of measurement precision is evaluated including the influence of outliers and the number of digits retained in data reporting. Additive and multiplicative coefficients for the error structure are generated by first binning the data by concentration intervals and then using a weighted regression to estimate the coefficients. While the analysis of collocated data suggests the overall data quality is good, improvements to the data reporting could be made to permit a more robust assessment of measurement precision. For example, the current approach to retaining significant digits leads to a false improvement in the collocated precision. The analysis methodology presented in this chapter provides a framework for analyzing collocated data from other networks.

5.2. Introduction

Precision in environmental – and especially air quality – datasets in the form of sample-specific uncertainties are used in various multi-linear receptor models such as Positive Matrix Factorization (PMF) and Chemical Mass Balance (CMB) to identify emission sources and atmospheric processes contributing to air pollution impacts. The evolution of source apportionment modeling by tools such as PMF has driven the need for robust uncertainty estimates. A variety of equations have been employed to estimate uncertainties as functions of concentration, analytical uncertainty and/or detection limits (Reff et al., 2007). Propagated sampling and analytical uncertainties, when available, have been the convention to construct sample-specific uncertainty estimates. Uncertainty in concentrations measured by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) have been estimated based on propagation of uncertainty in instrument drift, bias in the calibration standard recovery, sample dilution etc. (Barwick et al., 1999; Karakas, 2007). In the absence of reported uncertainty estimates, such as for data from the early years of the U.S. ambient fine particulate matter ($PM_{2.5}$) Chemical Speciation Network (CSN; which includes the Speciation Trends Network, STN), Ito et al. (2004) assumed sample-specific uncertainties to be 5% of the concentration value to construct the inputs needed for PMF modeling. Kim et al. (2005) used more-recent STN data with reported sample-specific uncertainties to develop error structures for the early data. Uncertainties, when available, were determined to be 5-20% of the measured concentrations. For some species no uncertainty estimates were reported and in these cases fractional errors of 5-30% were tested to obtain interpretable PMF solutions. Sample-specific uncertainties for the early CSN data were eventually backfilled by the U.S. Environmental Protection Agency.

Uncertainty or precision in measurement is defined by the International Standardization Organization (ISO) as "the closeness of agreement between quantity values obtained by replicating measurements of a quantity, under specified conditions […] precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under specified conditions of measurement" (ISO, 2004). Ideally, to estimate precision in the measured value, repetitive measurements at the same concentration must be conducted to assess whether sample-specific deviations from the mean follows a normal distribution (Taylor, 1982). However, the expense involved with field measurements and laboratory analysis for air quality monitoring usually renders this approach practically infeasible. Hence, precision is often estimated from measurements with two collocated samplers at one or more sites over multiple time periods. Collocated data from the two nationwide U.S. $PM_{2.5}$ monitoring networks – CSN and the Interagency Monitoring of Protected Visual Environments (IMPROVE) network – have recently been examined, including comparisons of collocated precision to reported analytical uncertainty estimates (Flanagan et al., 2006; Hyslop and White, 2011). For both networks, X-Ray Fluorescence (XRF) is the analytical method used for trace elements. While this approach based on using performance parameters of standardized methods provides a comprehensive evaluation across all aspects of sampling, it could complicate the interpretation of the results because the absolute concentration difference can vary with concentration. However, the ratio of the concentration difference to the average of the two collocated measurements (i.e. the relative differences) are assumed to be homogenous across the concentration range and modeled by a normal distribution (Hyslop and White, 2009).

Precision estimates can also be strongly influenced by outliers that could arise from any number of sporadic errors such as contamination, mishandling, or analytical interference and from

exceptional events such as micro-scale emissions leading to actual differences in concentrations at the two samplers. The removal of outliers from precision calculations is controversial. Outliers do not conform to the normal distribution of errors that we seek to characterize using collocated data. Therefore, they may be rejected with the understanding that such errors will be present – and not characterized – in the overall data set. Systematic errors in the extraction (if warranted) and chemical analysis steps for samples processed in the same batch cannot be captured from the analysis of collocated data. Despite such limitations, it is a valuable quality assurance measure and an attractive yet underutilized approach to generating error structures for air quality measurements used in source apportionment applications.

Collocated ambient PM_{10} and $PM_{2.5}$ species mass datasets utilized in this study were collected by the Hong Kong Environmental Protection Department (HKEPD) as part of routine ambient air quality monitoring network. Routine PM_{10} speciation datasets are collected at up to ten sites across the Hong Kong Special Administrative Region on a 1-in-6 day schedule using High Volume PM_{10} samplers with quartz filters and analyzed for gravimetric mass, elements by ICP-OES, ions by Ion Chromatography (IC), and Elemental Carbon (EC) and Organic Carbon (OC) by the NIOSH Thermal/Optical Transmittance (TOT) method. Preliminary PMF modeling of 1998-2008 PM_{10} speciation data used relative expanded uncertainty values (NIST, 1994) estimated as a function of ambient concentration and laboratory reported method detection limit values reported for each species (Lau et al., 2010). HKEPD also collected $PM_{2.5}$ speciation datasets at three-to-four-sites for three one-year periods over the past decade. Samples were collected on a 1-in-6 day schedule using Low Volume $PM_{2.5}$ samplers with Teflon filters that were analyzed for gravimetric mass and elements by X-Ray Fluorescence (XRF); and quartz filters that were analyzed for gravimetric mass, ions by IC, and EC and OC by the IMPROVE

Thermal/Optical Reflectance (TOR) method. The record-specific uncertainties provided with the $PM_{2.5}$ dataset exhibited large variations between analysis batches for some species. For example, record-specific uncertainties provided for 2009-2010 $PM_{2.5}$ lead data, shown in Figure 5-1, stratifies into four groups that precisely align with the analysis batches, with the uncertainty for a given concentration varying by nearly a factor of five across the analysis batches. Such concerns motivated the use of alternative approaches to generate uncertainty estimates for source apportionment modeling of these datasets.

In this chapter, collocated ambient PM_{10} and $PM_{2.5}$ species mass data from the Hong Kong networks were analyzed to evaluate measurement precision and the covariance in measurement errors between species. The quality of the collocated datasets were assessed based on the

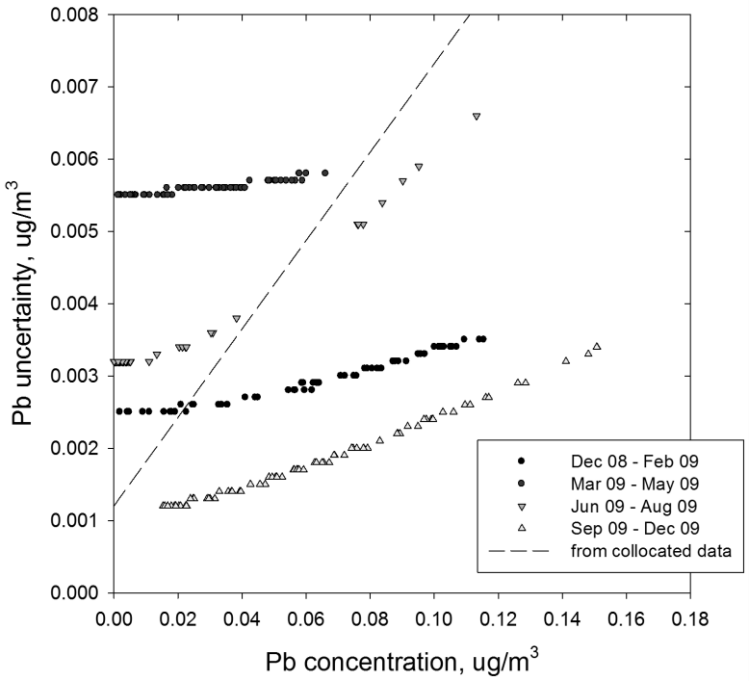


Figure 5-1. Error structures for Pb reported by the contracted analytical laboratory for the four analysis batches of 2009 $PM_{2.5}$ samples (markers) and the error structure derived from analysis of collocated precision data for 2009 (dashed line).

methodologies developed for examining collocated precision for the IMPROVE network dataset by Hyslop and White (2008); (2009); and (2011). The error structures from collocated data were generated using the methodology of Wade et al. (2008). With the species concentrations varying over multiple orders of magnitude, this dataset also enabled examination of concentration dependence of collocated precision.

The importance of quality data reporting and impact of outliers on constructing species-specific error structures representative of the entire concentration range are also discussed. A methodology for estimating sample-specific uncertainties is demonstrated using the PM_{10} dataset and is also applied to the $PM_{2.5}$ dataset. The estimated sample-specific uncertainties have been utilized for PMF (Yuan et al., 2013; Yadav et al., 2013) and CMB (Yuan et al., 2013) source apportionment modeling.

5.3. Datasets

Collocated PM_{10} samples collected from 1998 to 2008 at two sites in the Hong Kong – Sham Shui Po (SSP) and Tsuen Wan (TW) – were utilized in this study to assess measurement precision for the Hong Kong PM_{10} speciation network. Collocated species data for SSP (517 sample pairs) was used to generate error structures, and then evaluated using the corresponding TW dataset (517 most recent sample pairs for consistency with SSP dataset). Field blanks data was not available and thus the MDL values provided by the analytical laboratory could not be independently examined. SSP data characteristics for various PM_{10} species are listed in Table 5-1. MDLs for PM_{10} species in this study were estimated as: (i) the arithmetic mean of the actual concentration values reported below the MDL in SSP primary sampler data; (ii) for species with no concentration values below the MDL, the average of laboratory-reported

analytical detection limit provided by HKEPD; and (iii) for aluminum, the only species with no data below the MDL nor a laboratory-reported analytical detection limit, the MDL value used in earlier source apportionment study (Lau et al., 2010). Arsenic, cadmium, manganese, nickel, lead, vanadium and potassium ion had 28-60% of the data below MDL. Aluminum, chlorine, sodium and zinc each had less than 5% of the data below MDL, while all the concentration values for the remaining species were above MDL.

Table 5-1. Characteristics of the SSP site collocated PM_{10} dataset utilized for estimating error structures. Total number of sample pairs is 517. 'N < MDL (%)' represents the number of sample pairs with one-or-both values below minimum detection limit (percentage).

Species	Analysis Method	Reported MDL ($\mu g/m^3$)	N < MDL (%)
Al	ICP	0.0190[a]	1 (0)
As	ICP	0.0009[b]	164 (32)
Ca	ICP	0.0449[c]	0 (0)
Cd	ICP	0.0004[b]	182 (35)
Fe	ICP	0.0225[c]	0 (0)
Mg	ICP	0.0225[c]	2 (0)
Mn	ICP	0.0092[b]	188 (36)
Ni	ICP	0.0031[b]	224 (43)
Pb	ICP	0.0090[b]	144 (28)
V	ICP	0.0090[b]	310 (60)
Zn	ICP	0.0225[b]	39 (8)
Cl^-	IC	0.0276[b]	18 (3)
K^+	IC	0.3143[b]	234 (45)
Na^+	IC	0.2991[c]	40 (8)
NH_4^+	IC	0.0553[b]	7 (1)
NO_3^-	IC	0.0662[c]	0 (0)
SO_4^{2-}	IC	0.1496[c]	0 (0)
EC	TOT	0.2500[c]	0 (0)
OC	TOT	0.2500[c]	0 (0)

[a] Reported MDL value in Lau et al., 2010.

[b] MDL estimated as average of values reported as "<XX".

[c] Average of Laboratory MDL range reported by HKEPD.

Collocated $PM_{2.5}$ samples were collected in 2009 at all four sites of the $PM_{2.5}$ network, resulting in 10-13 collocated samples per site. Combining data from all four sites resulted in 47 sample pairs for elements and 40 sample pairs for ions and carbon fractions. Chromium, magnesium, strontium and tin had 40-85% of the data below MDL and were excluded from this analysis, while sodium and copper were excluded because of measurement artifacts associated with these species. Concentrations below the detection limit for the remaining $PM_{2.5}$ species were the values as reported by the analytical laboratory and were retained for the preliminary estimation of the error structures from this limited dataset. Characteristics of the collected $PM_{2.5}$ species data examined in this study are presented in Table 5-2.

Table 5-2. Characteristics of the multi-site collocated $PM_{2.5}$ dataset with estimated error structures coefficients (a_j,b_j) from the linear weighted regression after removing outliers. LQL is the lower quantifiable limit.

Species	Analysis Method	N	MDL ($\mu g/m^3$)	N < MDL (%)	LQL ($\mu g/m^3$)	N < LQL (%)	Rejected N (%)	Regression Coefficient	
								aj	*bj*
Al	XRF	47	0.0186	9 (19)	0.0262	14 (30)	0 (0)	0.0055	0.0859
Ca	XRF	47	0.0030	0 (0)	0.0139	0 (0)	3 (6)	0.0015	0.0470
Cl	XRF	47	0.0020	4 (9)	0.0115	11 (23)	1 (2)	0.0012	0.1255
Fe	XRF	47	0.0032	1 (2)	0.0126	2 (4)	1 (2)	0.0015	0.0362
K	XRF	47	0.0019	0 (0)	0.0049	0 (0)	0 (0)	0.0005	0.0295
Mn	XRF	47	0.0035	10 (12)	0.0032	10 (21)	4 (9)	0.0010	0.0365
Ni	XRF	47	0.0005	1 (2)	0.0007	1 (2)	0 (0)	0.0005	0.0265
Pb	XRF	47	0.0039	12 (26)	0.0031	12 (26)	0 (0)	0.0012	0.0612
Si	XRF	47	0.0150	3 (6)	0.0463	8 (17)	1 (2)	0.0013	0.0673
Ti	XRF	47	0.0014	7 (15)	0.0028	12 (26)	0 (0)	0.0005	0.1123
V	XRF	47	0.0003	0 (0)	0.0005	0 (0)	0 (0)	0.0005	0.0201
Zn	XRF	47	0.0016	1 (2)	0.0035	3 (6)	1 (2)	0.0008	0.0304
NH_4^+	IC	40	0.0623	1 (3)	0.0415	1 (3)	3 (8)	0.0493	0.0171
NO_3^-	IC	40	0.0623	0 (0)	0.1074	0 (0)	0 (0)	0.0380	0.0566
SO_4^{2-}	IC	40	0.0623	0 (0)	1.4417	0 (0)	8 (20)	0.3993	-0.0052
EC	TOR	40	0.0054	0 (0)	0.0634	0 (0)	0 (0)	-0.0144	0.0883
OC	TOR	40	0.2088	0 (0)	0.5241	0 (0)	3 (8)	0.0995	0.0456

5.4. Methodology and Results

5.4.1. Data quality assessment

The collocated PM_{10} data from SSP site was first examined for any systematic bias that could result from flow rate mismatch, miscalibration or other such differences between the two samplers, by plotting scattergrams (not shown) and performing Reduced Major Axes (RMA) regression (Bohonak, 2007). No significant sampler-to-sampler bias was detected from these metrics for PM_{10} species. To evaluate concentration dependence of measurement differences, scaled arithmetic differences (i.e. difference between sample pair concentrations divided by $\sqrt{2}$) and scaled relative differences (i.e. ratio of scaled arithmetic difference to the mean concentration for the sample pair) were plotted against the corresponding sample pair mean concentrations (Hyslop and White, 2008). Such plots for each of the three classes of PM_{10} species are shown: elements (arsenic, Figure 5-2); ions (sulfate, Figure 5-3); and carbon (organic carbon, Figure 5-4). In each figure, panels (a) and (b) are the scaled arithmetic differences and panels (c) and (d) are the scaled relative differences using a linear and logarithmic concentration axis, respectively, to emphasize the high- and low-ends of the concentration ranges. As concentration values approached the detection limit, the relative concentration differences increased but remained within ~20% for all species (including those not shown). At concentrations much greater than the MDL (over 3-to-5 times the MDL), the relative differences decreased with increasing concentration values and for many species approached an asymptotic value. These patterns indicate that a concentration dependent error structure is justified. For the remainder of the analysis, the dataset was screened to exclude all sample pairs with one-or-both concentration values below the detection limit for estimation of the error structures applicable over the entire detectable concentration range.

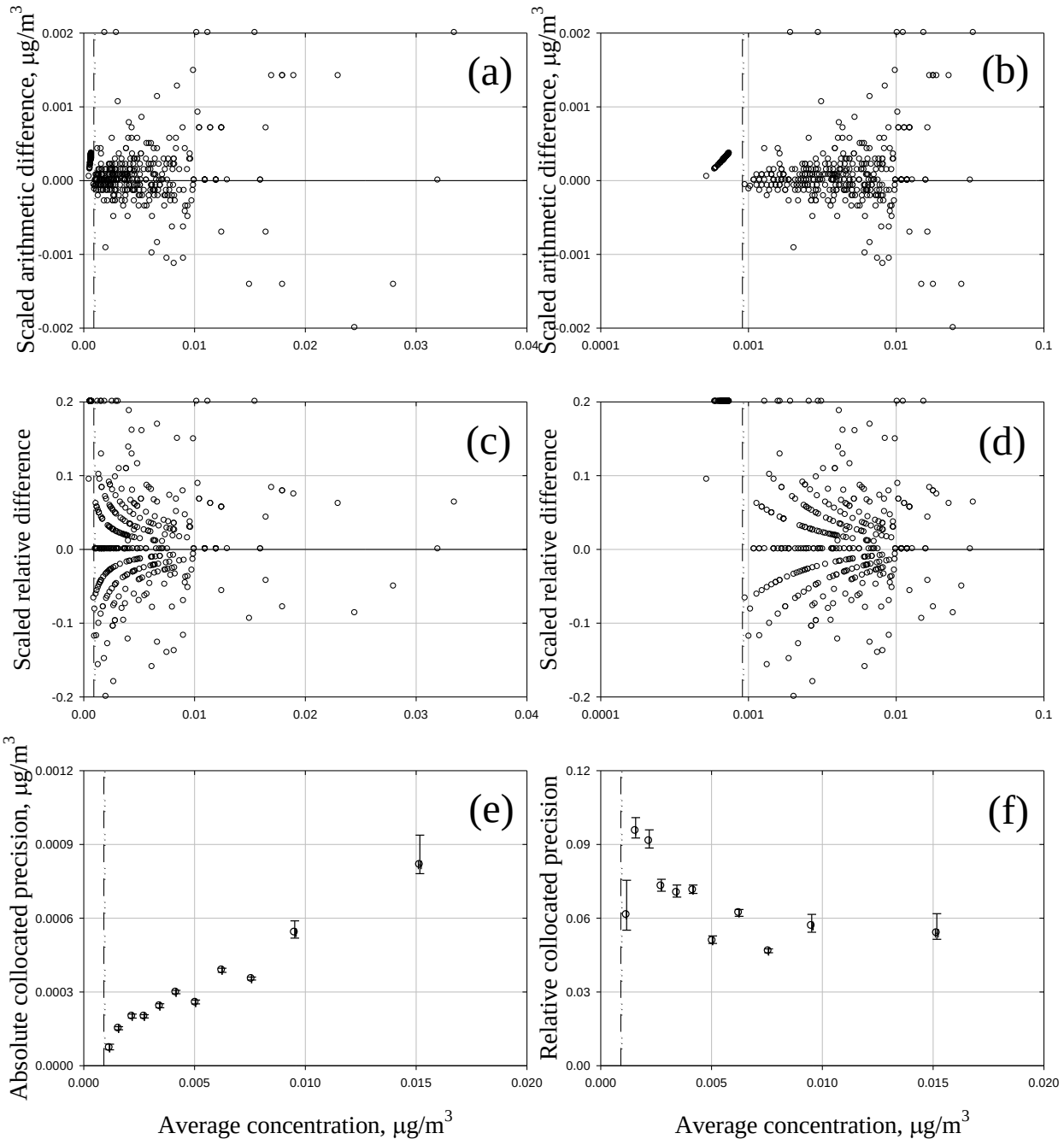


Figure 5-2. SSP collocated PM_{10} data for arsenic: scaled arithmetic difference, (a) linear and (b) logarithmic concentration scale; scaled relative difference, (c) linear and (d) logarithmic concentration scale; (e) binned absolute collocated precision; and (f) binned relative collocated precision. The x-axis of all graphs is the average concentrations and the dashed vertical line is the MDL. Extreme values lying beyond the y-axis range are placed on the border.

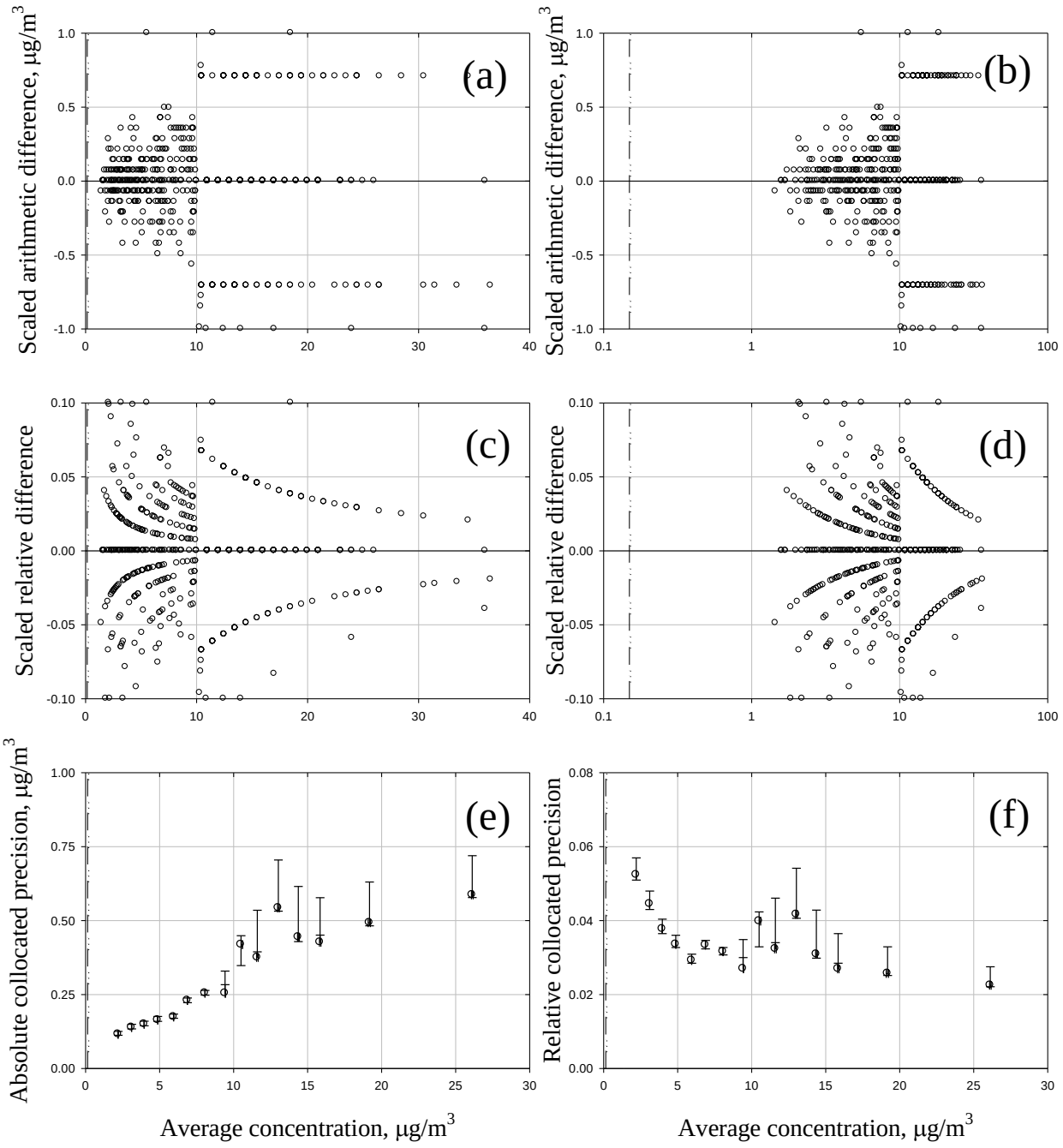


Figure 5-3. SSP collocated PM_{10} data for sulfate: scaled arithmetic difference, (a) linear and (b) logarithmic concentration scale; scaled relative difference, (c) linear and (d) logarithmic concentration scale; (e) binned absolute collocated precision; and (f) binned relative collocated precision. The x-axis of all graphs is the average concentrations and the dashed vertical line is the MDL. Extreme values lying beyond the y-axis range are placed on the border.

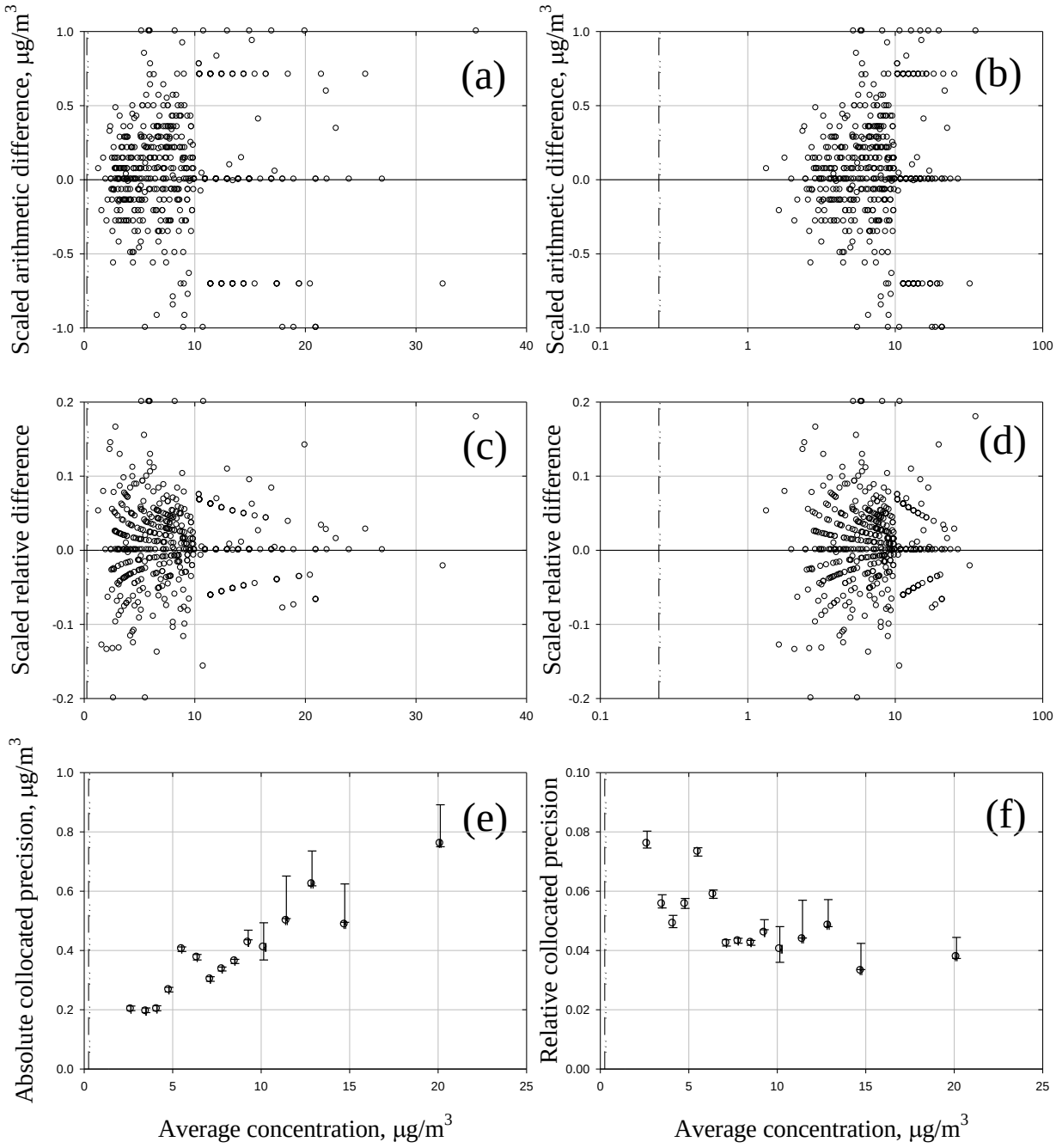


Figure 5-4. SSP collocated PM_{10} data for organic carbon: scaled arithmetic difference, (a) linear and (b) logarithmic concentration scale; scaled relative difference, (c) linear and (d) logarithmic concentration scale; (e) binned absolute collocated precision; and (f) binned relative collocated precision. The x-axis of all graphs is the average concentrations and the dashed vertical line is the MDL. Extreme values lying beyond the y-axis range are placed on the border.

5.4.2. Estimating error structures

The concentration dependence of collocated precision was quantified using the methodology of Wade et al. (2008). For each species *j*, the sample pairs were ordered by decreasing sample pair mean concentrations and grouped into bins (15 bins; each with nominally 34 data pairs). The collocated precision was calculated for each bin of concentration values using equation (5-1) where σ_{kj} is the absolute collocated precision for the N_{kj} sample pairs in bin *k*, and C_{1ij} and C_{2ij} are the primary and secondary sampler concentrations, respectively, for sample *i*. $\hat{C}_{kj}$ in equation (5-2) is the arithmetic mean concentration for the N_{kj} samples pairs in bin *k*. For each bin, the relative collocated precision was calculated as the ratio of the absolute collocated precision to the average concentration (i.e. $\sigma_{kj}/\hat{C}_{kj}$).

$$\sigma_{kj} = \sqrt{\frac{1}{2\times N_{kj}}\sum_{i=1}^{N_{kj}}\left(C_{1ij}-C_{2ij}\right)^2} \quad (5\text{-}1)$$

$$\hat{C}_{kj} = \frac{1}{N_{kj}}\sum_{i=1}^{N_{kj}}\left(\frac{C_{1ij}+C_{2ij}}{2}\right) \quad (5\text{-}2)$$

Based on an inspection of binned collocated precision plots, a linear error structure was determined to be representative for most species and thus was used for the remainder of the analysis. Equation (5-3) shows the assumed error structure where U_{ij} = uncertainty for species *j* in sample *i*; a_j = additive (constant) error term for species *j*; b_j = multiplicative error term for species *j*; and C_{ij} = concentration value for species *j* in sample *i*. The multiplicative term governs the uncertainty for high concentration values while the additive term dominates the uncertainty for low concentration values.

$$U_{ij} = a_j + b_j \times C_{ij} \quad (5\text{-}3)$$

Several approaches were used to estimate the additive and multiplicative error terms in equation (5-3) and four cases are presented which differ in the data conditioning and error term estimation strategy. The error terms were estimated using an approach inspired by Polissar *et al.* (1998) (*Polissar*) and also by weighted linear least squares regression (*Weighted*). The *Polissar* approach sets the uncertainty (U_{ij}) to (5/6)×MDL for concentration values below the MDL and sets the additive error term (a_j) to MDL/3 for concentration values above the MDL (Polissar *et al.*, 1998). The multiplicative error term (b_j) was calculated using the mean of the binned relative collocated precision for the six bins with highest concentrations; this approach is generally consistent with the recommendation of Hyslop and White (2008) to estimate the relative precision using concentrations much greater than the MDL, such as 3-to-5 times MDL, to avoid being influenced by the additive term's uncertainty contribution at low concentration values. The *Weighted* approach used a weighted linear regression where the weight for each bin, w_{kj}, was calculated as the inverse squared absolute collocated precision (σ_{kj}) for the sample pairs in bin *k* as shown in equation (5-4). The weighted error structure coefficients were then calculated using equations (5-5) and (5-6) (Taylor, 1982). To measure the goodness of regression fits, the mean square error, MSE (Taylor, 1982), of the residual between the binned relative collocated precision and the estimated precision from the fitted error structures were evaluated for each species.

$$w_{kj} = 1/\sigma_{kj}^2 \tag{5-4}$$

$$a_j = \frac{(\sum w_{kj}{C_{kj}}^2)(\sum w_{kj}\sigma_{kj}) - (\sum w_{kj}C_{kj})(\sum w_{kj}C_{kj}\sigma_{kj})}{(\sum w_{kj})(\sum w_{kj}{C_{kj}}^2) - (\sum w_{kj}C_{kj})^2} \tag{5-5}$$

$$b_j = \frac{(\sum w_{kj})(\sum w_{kj}C_{kj}\sigma_{kj}) - (\sum w_{kj}C_{kj})(\sum w_{kj}\sigma_{kj})}{(\sum w_{kj})(\sum w_{kj}{C_{kj}}^2) - (\sum w_{kj}C_{kj})^2} \tag{5-6}$$

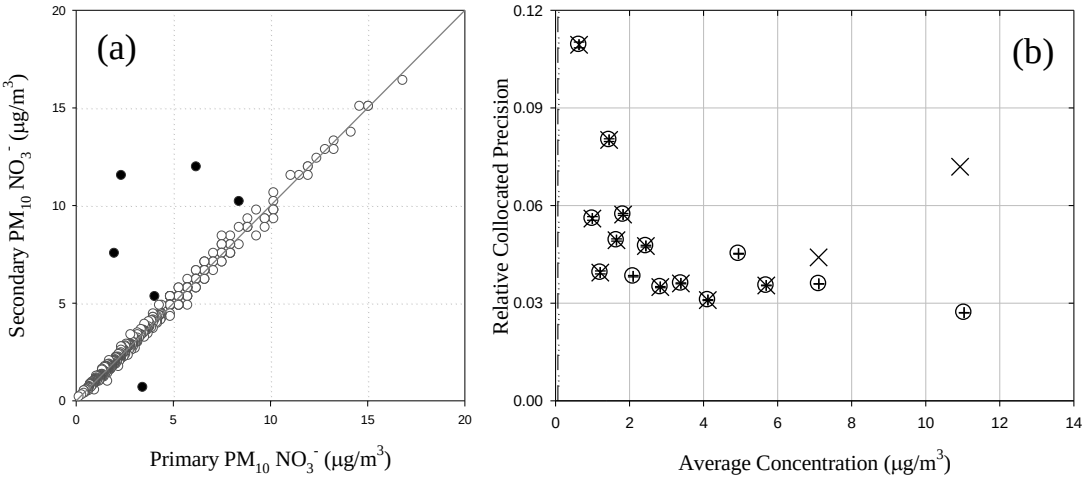


Figure 5-5. SSP collocated PM_{10} data for nitrate: (a) scatter plot with conditioned data used in precision estimation (open circles) and rejected outliers (closed circles); and (b) binned relative collocated precision with all data (cross) and outliers rejected (circles) along with error bars representing 5^{th} and 95^{th} percentile values from the Monte Carlo-type simulations adding an addition, randomly-generated significant figure to each concentration value. The dashed vertical line is the MDL.

The *Polissar* and *Weighted* linear regression approaches were first applied to data sets with all sample pair concentrations above the detection limit (unconditioned) and then to data sets screened for sample pairs deemed as outliers (conditioned). A scattergram of primary and secondary sampler concentrations, such as PM_{10} nitrate shown in Figure 5-5 (a), demonstrates the presence of such sample pairs. The outliers were iteratively removed through sequential application of the Grubb's Test (Grubbs, 1969) with the goal of minimizing the MSE with least number of sample pairs removed for each species. The number of sample pairs rejected as outliers are tabulated in Table 5-3. Less than 2% of the sample pairs were rejected as outliers from the PM_{10} species dataset. MSE values for each of the four cases and for all species are tabulated in Table 5-3. For the unconditioned data, the quality of fit is not consistently better for one estimation approach compared to the other (i.e. *Polissar* versus *Weighted*) because the outliers destabilize the multiplicative term estimates. For the *Polissar* method, conditioning the

data reduces the MSE for most but not all species. In contrast, for the *Weighted* method the MSE decreases for all species upon conditioning the data and in most cases the improvement is dramatic. Overall, the *Conditioned Weighted* case provides the best fit when using the MSE as the figure of merit; the error terms for this case are tabulated in Table 5-3. The absolute and relative precisions for the conditioned binned data few species are also shown in Figure 5-2-4 panels (e) and (f), respectively. The trends are not strictly monotonic and the factors influencing such behavior are examined in Section 5.4.4.

Table 5-3. Error structure coefficients (a_j, b_j) derived by weighted linear regression of refined SSP site collocated PM_{10} dataset i.e., after removing outliers and values below MDL. Mean square errors (MSE) for regression fits derived for the four estimation approaches described in the text. 'Rejected N (%)' is number of data pairs deemed as outliers that are excluded from estimation of error structure coefficients.

Species	Rejected N (%)	Mean Square Error (MSE)				Regression Coefficient	
		Polissar	*Weighted*	*Conditioned Polissar*	*Conditioned Weighted*	a_j (μg/m^3)	b_j
Al	9 (2)	2.1 E-4	3.1 E-4	0.9 E-4	0.2 E-4	0.0046	0.071
As	5 (1)	10.1 E-8	1.0 E-8	10.4 E-8	0.2 E-8	0.0001	0.050
Ca	2 (0)	11.4 E-3	17.0 E-3	0.5 E-3	0.1 E-3	0.0217	0.043
Cd	3 (1)	26.7 E-9	0.8 E-9	25.2 E-9	0.2 E-9	0.00002	0.058
Fe	7 (1)	11.1 E-4	16.2 E-4	0.5 E-4	0.2 E-4	0.0044	0.030
Mg	4 (1)	20.0 E-5	19.7 E-5	6.5 E-5	0.3 E-5	0.0049	0.020
Mn	2 (0)	11.8 E-6	1.23 E-6	11.2 E-6	0.0 E-6	0.0002	0.029
Ni	1 (0)	162.0 E-8	2.6 E-8	156.0 E-8	0.9 E-8	0.0003	0.050
Pb	3 (1)	9.4 E-6	6.5 E-6	9.8 E-6	0.5 E-6	0.0002	0.035
V	3 (1)	14.1 E-6	0.5 E-6	14.4 E-6	0.2 E-6	0.00001	0.047
Zn	8 (2)	24.0 E-5	14.4 E-5	5.3 E-5	0.2 E-5	0.0016	0.032
Cl^-	3 (1)	3.2 E-3	2.9 E-3	0.9 E-3	0.4 E-3	0.0141	0.047
K^+	3 (1)	156.0 E-4	13.6 E-4	133.0 E-4	0.5 E-4	0.0118	0.028
Na^+	4 (1)	348.0 E-3	479.0 E-3	11.5 E-3	0.3 E-3	0.0264	0.018
NH_4^+	3 (1)	3.0 E-3	2.1 E-3	0.5 E-3	0.4 E-3	0.0169	0.028
NO_3^-	6 (1)	48.8 E-3	96.1 E-3	1.5 E-3	0.6 E-3	0.0429	0.026
SO_4^{2-}	4 (1)	5.6 E-2	4.0 E-2	1.1 E-2	0.7 E-2	0.0597	0.025
EC	9 (2)	14.6 E-3	12.4 E-3	6.0 E-3	0.7 E-3	0.1148	0.023
OC	4 (1)	5.8 E-2	6.9 E-2	1.3 E-2	0.8 E-2	0.1117	0.032

5.4.3. Evaluation of error structures using collocated TW dataset

Collocated PM_{10} data from the TW site, spanning 1999-2008 (517 most recent sample pairs for consistency with SSP dataset) were used as an independent check of the modeled error structures. Reported MDL values provided (or estimated) and the *effective* MDL, estimated as three times the additive error coefficient (a_j) (Polissar et al., 1998) based on the *Conditioned Weighted* regression analysis of the collocated SSP data, are compared in Table 5-4.

Table 5-4. Metrics for the TW site collocated PM_{10} dataset with 517 total number of sample pairs. *Effective MDL* refers to the MDL values estimated from analysis of SSP site collocated PM_{10} dataset. N < MDL (%) is the number of sample pairs with one-or-both values below the laboratory-reported MDL. Rejected N (%) is the number of sample pairs deemed as outliers and excluded from collocated precision calculations. Mean square error (MSE) for Unconditioned (excludes values below MDL) and Conditioned (excludes outliers and values below MDL) TW collocated dataset regressed on the estimated error structures derived for SSP collocated dataset are also presented.

Species	*Effective MDL* (μg/m^3)	Reported MDL (μg/m^3)	N < MDL (%)	Rejected N (%)	Mean Square Error	
					Unconditioned	Conditioned
Al	0.0137	0.0190	3 (1)	2 (0)	1.5 E-4	0.5 E-4
As	0.0002	0.0009	141 (27)	0 (0)	0.2 E-7	0.2 E-7
Ca	0.0651	0.0449	1 (0)	1 (0)	0.2 E-3	0.1 E-3
Cd	0.0001	0.0004	153 (30)	0 (0)	0.3 E-8	0.3 E-8
Fe	0.0133	0.0225	0 (0)	7 (1)	2.0 E-4	0.3 E-4
Mg	0.0146	0.0225	1 (0)	5 (1)	1.8 E-5	0.4 E-5
Mn	0.0007	0.0092	166 (32)	1 (0)	3.0 E-6	0.1 E-6
Ni	0.0008	0.0031	184 (36)	0 (0)	0.2 E-7	0.2 E-7
Pb	0.0006	0.0090	115 (22)	1 (0)	4.5 E-6	0.4 E-6
V	0.00002	0.0090	241 (47)	0 (0)	0.1 E-6	0.1 E-6
Zn	0.0048	0.0225	18 (3)	2 (0)	19.8 E-4	0.0 E-4
Cl^-	0.0424	0.0276	31 (6)	0 (0)	0.1 E-3	0.1 E-3
K^+	0.0354	0.3143	219 (42)	0 (0)	0.1 E-3	0.1 E-3
Na^+	0.0792	0.2991	34 (7)	5 (1)	0.9 E-3	0.2 E-3
NH_4^+	0.0508	0.0553	3 (1)	1 (0)	1.5 E-3	0.2 E-3
NO_3^-	0.1286	0.0662	0 (0)	1 (0)	1.4 E-3	0.6 E-3
SO_4^{2-}	0.1792	0.1496	0 (0)	4 (1)	5.0 E-2	0.9 E-2
EC	0.3443	0.2500	0 (0)	9 (2)	1.5 E-2	0.2 E-2
OC	0.3352	0.2500	0 (0)	14 (3)	9.6 E-1	0.2 E-1

The effective MDL values are typically lower than the reported MDLs, suggesting that the reported MDL values are generally conservative for generating error structures for the source apportionment modeling. The collocated precision estimated using TW dataset is generally fitted well by the error structure estimated using SSP dataset, such as As shown in Figure 5-6 (a). Some species had one or two concentration bins that were not well characterized by the estimated error structure. In few such cases, the regression fit for the TW PM_{10} species improved with the exclusion of few sample pairs, such as four sample pairs for SO_4^{2-} (Figure 5-6 (d)), one sample pair for Pb (Figure 5-6 (b)) and 3% each for EC (Figure 5-6 (e)) and OC (Figure 5-6 (f)). The characteristics of TW dataset and the MSEs for collocated precision calculated using unconditioned (all reported samples above MDL) and conditioned (dataset screened for outliers) collocated TW dataset gauged against the *Conditioned Weighted* error structure derived from collocated SSP dataset are also summarized in Table 5-4. Overall, favorable comparisons across these datasets add confidence towards using the *Conditioned Weighted* error structures from Table 5-3 for source apportionment modeling of the Hong Kong speciation dataset.

PM_{10} collocated species data from both sites were also examined for covariance in measurement errors across the species. Measurement error in concentration values should ideally be random and independent of errors measured across different species, but Hyslop and White (2011) observed correlation in measurement error for a few species in IMPROVE and STN datasets. To check for such relationships in Hong Kong collocated PM_{10} dataset, a scatterplot matrix of the observed species concentration differences, defined as $\Delta_{ji} = \ln(C_{2i}/C_{1i})$ for species j, were plotted for all species. For both SSP and TW collocated datasets, the inter-species correlation coefficients for measurement differences were typically below 0.15 indicating no significant

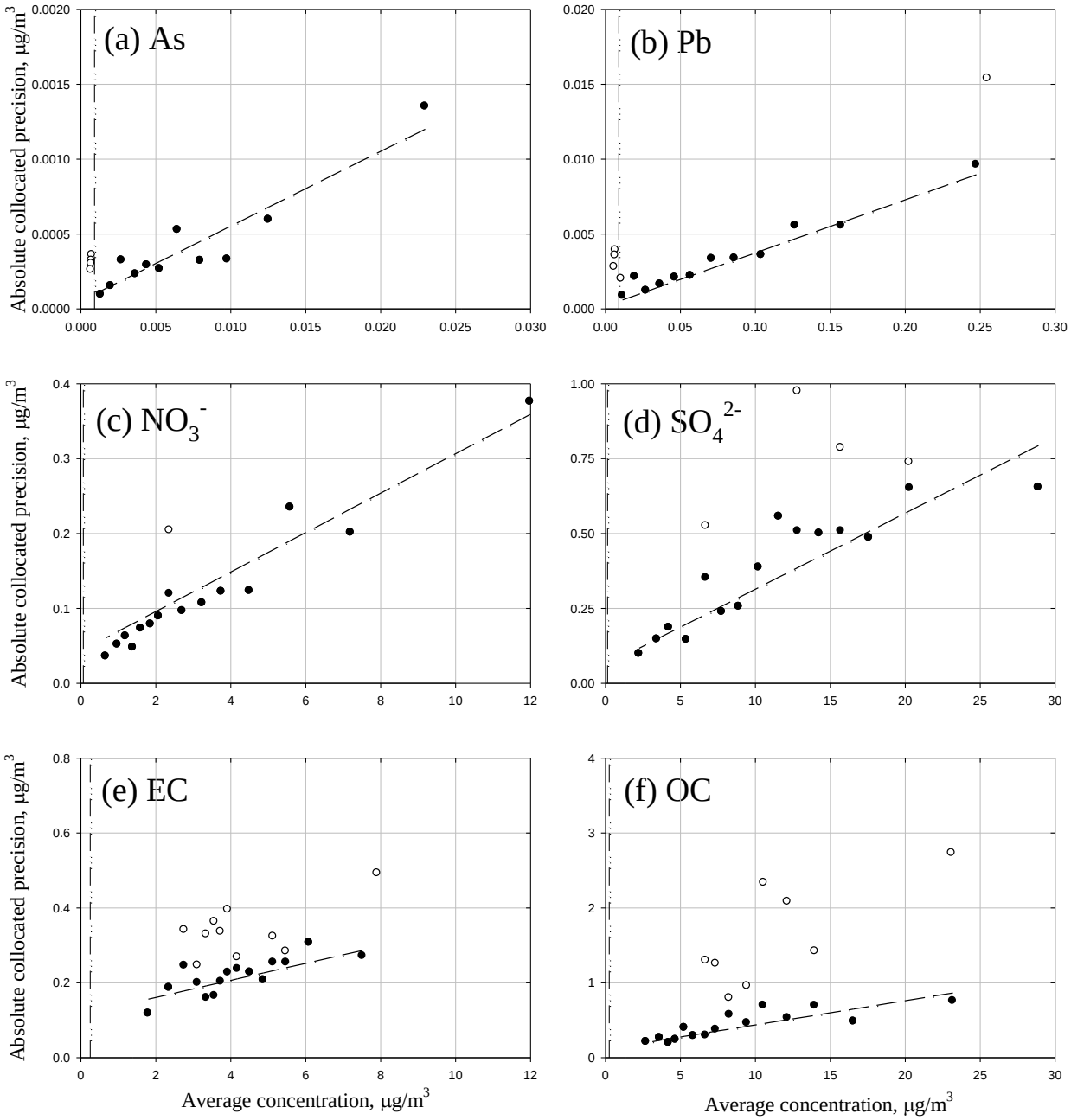


Figure 5-6. Binned absolute collocated precision calculated using unconditioned (open circles) and conditioned (shaded circles) TW collocated PM_{10} dataset along with the error structures estimated derived from SSP collocated data (dashed lines). The dot-do-dash vertical lines are the MDL values.

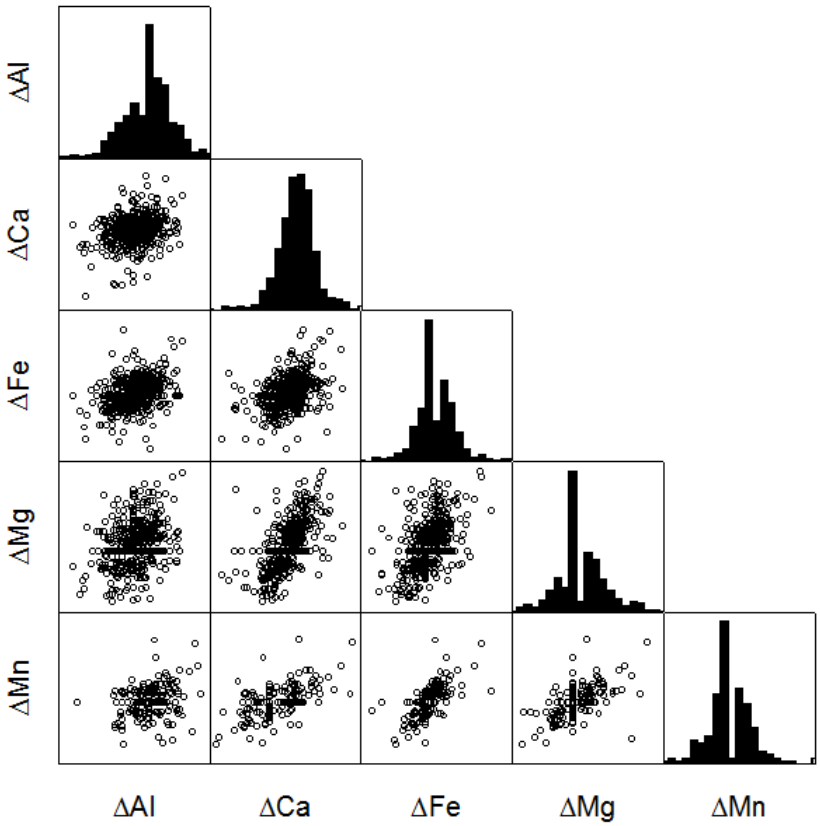


Figure 5-7. Scatterplot matrix of observed differences, $\Delta_{ji} = \ln(C_{2i}/C_{1i})$, where j= Al, Ca, Fe, Mg and Mn for collocated PM_{10} data from TW. Off the diagonal are scatterplots of the observed differences in one species verses the other. Along the diagonal are histograms of the observed differences for each species.

covariance in measurement errors. However, stronger associations in measurement errors were observed across species typically associated with soil/dust profile. For example, Figure 5-7 shows the scatterplot matrix of measurement differences for collocated TW species when correlation coefficients across Al, Ca, Fe, Mg and Mn ranged between 0.25 to 0.64, with highest correlation coefficients for Fe-Mn (0.64), Ca-Mg (0.56) and Mg-Mn (0.52). The implications of covariance in species measurement errors on source apportionment modeling is beyond the scope of this study, however this analysis furthers the concerns regarding data quality of ambient air quality monitoring networks presented in Hyslop and White (2011).

5.4.4. Influence of precision in data reporting and outliers

Figure 5-2 - 5-4 also demonstrates stratification or banding of scaled arithmetic and relative differences across certain concentration ranges. For example, in panel (b) of Figure 5-3, sulfate concentrations greater than 10 $\mu g/m^3$ were stratified into three bands, at 0 and $\pm$ 0.707 $\mu g/m^3$ corresponding to 0 and $\pm$ 1 $\mu g/m^3$ differences in sample pair concentrations respectively. Similarly, between 1 to 10 $\mu g/m^3$ the bands corresponded to concentration differences of 0.0, $\pm$ 0.1, $\pm$ 0.2 $\mu g/m^3$ and so on. This behavior arises from the number of significant figures retained in reporting the measured species concentration. For most species, the laboratory reported the data as two significant digits regardless of the range of the concentration values i.e., $X.X \times 10^Y$ $\mu g/m^3$ (or 0.0XX, 0.XX, X.X, XX.). In the case of NO_3^- and NH_4^+, data for the secondary sampler were reported with several significant digits (XX.XXXX $\mu g/m^3$) while data from the primary sampler were reported up to two significant digits after the decimal point i.e., XX.XX $\mu g/m^3$ (or 0.XX, X.XX, XX.XX). For EC and OC, the samples were analyzed by different laboratories during the study period and resulted in 90% of the data being reported as two significant digits ($X.X \times 10^Y$ $\mu g/m^3$) and the remained reported with several significant digits (XX.XXXX $\mu g/m^3$). The differences in number of significant digits retained for specie-specific reported concentration values enables contrasting the influence of precision in data reporting towards the generation of uncertainty estimates. Implications of stratifications in concentration differences were examined by performing Monte Carlo-type stimulations i.e., by gauging the variability in collocated precisions estimated for a distribution of pseudo collocated datasets generated by assigning additional significant digits to the reported concentration values.

For conditioned SSP species concentrations reported as two significant digits (X.X × 10^{Y} μg/m^3), additional significant digits were assigned such that concentrations greater than 10 μg/m^3, which were originally reported as integers (XX.), were replaced by a random real number in the interval [XX. - 0.5, XX. + 0.5]. Similarly, concentrations between 1 to 10 μg/m^3, reported as one digit after the decimal point (X.X), were replaced with a random real number in the interval [X.X – 0.05, X.X + 0.05] and so on, for concentrations between 0.1 to 1 μg/m^3, 0.01 to 0.1 μg/m^3 etc. In accordance with the precision in reported concentration values, concentrations at the transition points reported as 10 μg/m^3 was assigned a random real number in the interval [9.95, 10.5];1 μg/m^3 reported concentration was assigned a random real number in the interval [0.995, 1.05]; and so on. The influence of one such randomization on collocated PM_{10} sulfate concentrations at the 10 μg/m^3 transition point is shown in Figure 5-8. Concentrations greater than 10 μg/m^3 in the original dataset were reported in the increments of ± 1 μg/m^3, while concentrations between 1

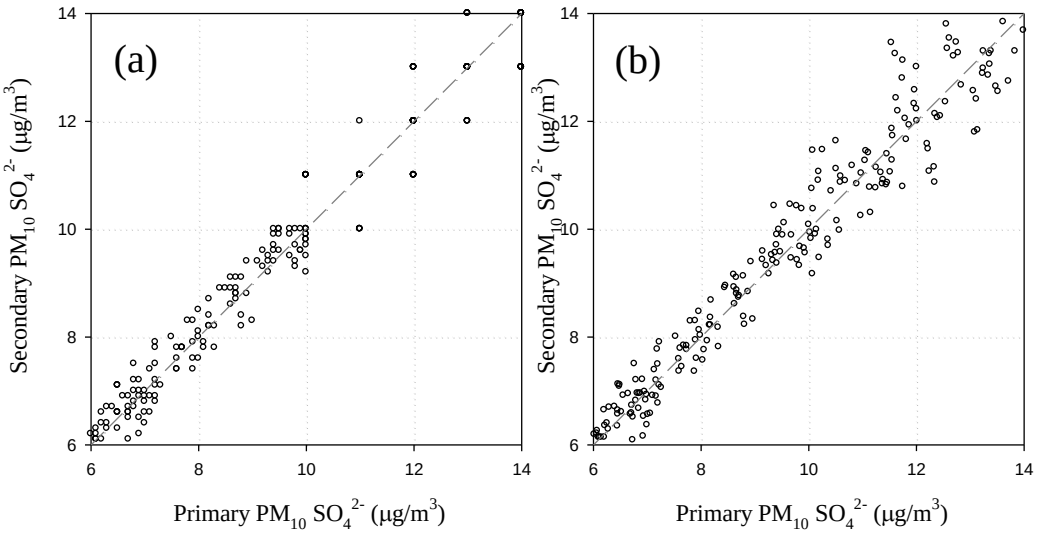


Figure 5-8. Scatter plot of conditioned SSP collocated PM_{10} sulfate data: (a) original data; and (b) one example of randomized concentration values at the transition point of 10 μg/m^3. Dashed line is the 1-to-1 line.

and 10 $\mu g/m^3$ were reported in ± 0.1 $\mu g/m^3$ increments. The randomization breaks the discretized reported values (Figure 5-8 (a)) to generate a pseudo dataset representative of ambient concentrations (Figure 5-8 (b)). For each species, this randomization was repeated fifty times and the binned collocated precision were recalculated for each pseudo dataset generated to obtain a distribution of precision estimates.

The 5^{th} and the 95^{th} percentile of these fifty precision estimates are also plotted in panels (e) and (f) of Figure 5-2 - 5-4, as error bars representing the range of variation in estimated binned precision. At the points of transition, such as 10 $\mu g/m^3$ in Figure 5-3 (f) for sulfate, the error bars are broader for bins corresponding to average concentrations > 10 $\mu g/m^3$ than for bins corresponding to average concentrations < 10 $\mu g/m^3$ because of the difference in precision of reporting concentrations across 10 $\mu g/m^3$. Further, the collocated precision estimated using the reported concentrations are generally positioned below the 10^{th} percentile of the precision range estimated from the pseudo datasets. This indicates that the rounding of concentrations results in the measurement error to be underestimated. Reporting concentrations > 10 $\mu g/m^3$ as values rounded to two digits (XX. $\mu g/m^3$) results in nearly one-third of the data within the corresponding bins to have identical concentrations for both samplers. Thus, the value of binned collocated precision (absolute/relative) calculated as shown in equation (5-1) are reduced and result in false improvement in measurement error than would be expected if the concentrations were not reported as two significant digits. In contrast to sulfate and carbon fractions, the reported nitrate and ammonium ion concentrations for the secondary sampler were not rounded and the primary sampler concentrations were rounded up to two digits after the decimal point (XX.XX $\mu g/m^3$). Thus, the pseudo datasets for these two species were generated by replacing the concentrations from only the primary sampler with a random real number between [XX.XX –

0.005, XX.XX + 0.005]. The error bars in Figure 5-5 (b), representing the 5^{th} and 95^{th} percentiles of the collocated precision distribution estimated from the fifty pseudo PM_{10} nitrate datasets, are indistinguishable from the collocated precision calculated from the reported concentrations. The nitrate concentrations ranged between 1.08 to 6.64 $\mu g/m^3$ (10^{th} and 90^{th} percentile, respectively). Reporting high-end concentrations as decimal numbers with up to two significant digits (XX.XX $\mu g/m^3$) allows the concentrations to be deterministic and results in robust collocated precision, while reporting such concentrations as only two significant digits ($X.X \times 10^Y$ $\mu g/m^3$) can result in less precise high-end concentration values. In contrast, such differences in reporting concentrations at the low-end of the concentration range have insignificant influence on the collocated precision as the concentration differences at the low-end are governed by the detectability rather than the absolute concentration values.

Banding of the concentration values alone does not explain the deviations in monotonic behavior of the binned precision as a function of concentration. Outliers also have significant influence on the concentration dependence of precision. For example, Figure 5-5 (a) shows the SSP PM_{10} nitrate sample pairs rejected as outliers from the collocated precision estimation and Figure 5-5 (b) shows the binned relative collocated precision before and after removing the outliers. In the presence of outliers and/or bins with a high number of samples deviating from the 1-to-1 line in the scattergram, the binned precision containing the outliers deviates from the monotonic trend observed over the entire concentration range. Examination of rejected sample pairs across different species and analysis methods can provide further insight into the characteristics of the outliers. For example, rejection of the same sample pair irrespective of the species and analytical method (such as SSP sample pair collected on 02/11/1999) suggests contamination or a sampling artifact. Rejection of a sample pair for several species that share the same analytical method

(e.g. species measured by IC for SSP sample pair collected on 03/01/2000) indicates an artifact in analytical measurement. Such errors from contamination or mishandling are isolated cases that are not characterized by the central tendency of measurement uncertainty. While such cases typically cannot be identified from the concentration time series of a single sampler, collocated datasets provide information about the frequency of such outliers that should be taken into consideration when interpreting single sampler datasets from the same network.

5.5. $PM_{2.5}$ Speciation Dataset Uncertainty

A preliminary estimation of error structures for the $PM_{2.5}$ dataset was also generated for source apportionment modeling by following the methodology used for the PM_{10} dataset. Collocated data collected at all four sites in the $PM_{2.5}$ monitoring network, comprising one remote, one roadside and two urban/neighborhood stations, was combined to obtain 40 ion and carbon fraction each and 47 metal sample pairs. No significant bias was observed between data collected by primary and secondary samplers for all species, with an exception of $PM_{2.5}$ EC (RMA metrics for data from secondary sampler against the primary sampler: slope 0.880 ± 0.038, intercept 0.236 ± 0.246 and $R^2 = 0.984$). Site-specific EC concentrations, typically contributed by vehicle exhaust, vary over distinct concentration ranges depending on the site characteristics. The bias resulting from high EC concentrations at roadside station cannot be corrected for sampler-to-sampler bias due to limited site-specific dataset. The site-specific or multi-site aggregated collocated EC concentrations should be examined further when a larger dataset is available. Collocated precision was calculated by grouping the paired concentrations into five to six bins with nominally eight sample pairs per bin. Due to the limited dataset, the $PM_{2.5}$ dataset was conditioned to remove only the outliers but not the values below the MDL as all concentrations values were as reported by the analytical laboratory. For few species up to 4

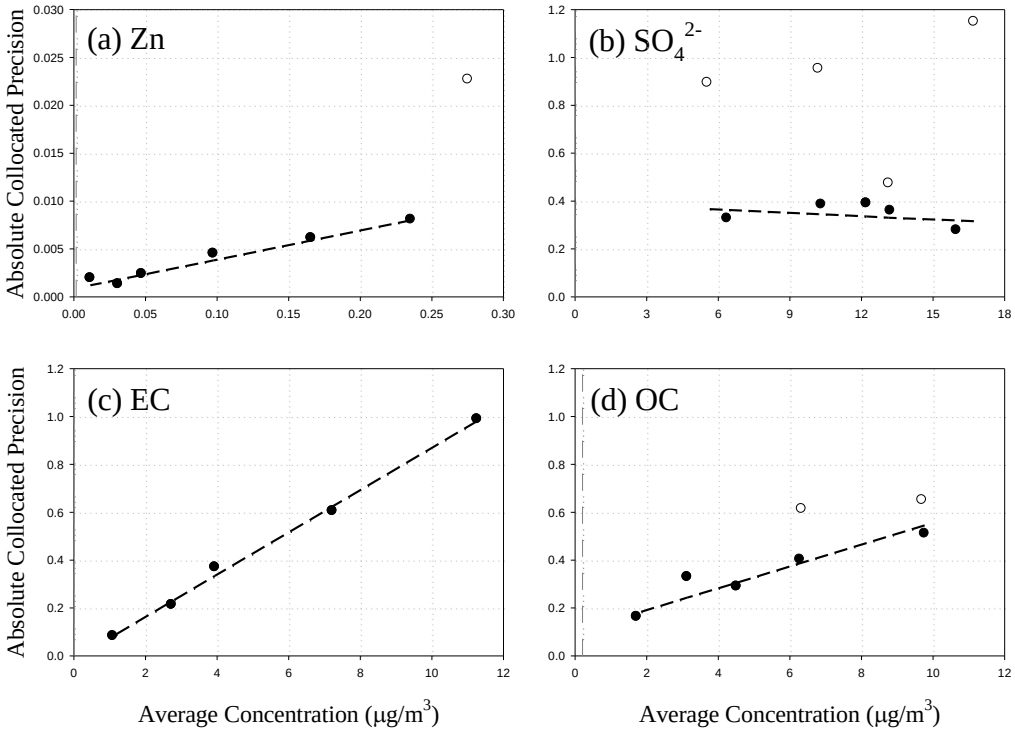


Figure 5-9. Examples of the binned absolute collocated precision for the $PM_{2.5}$ data set using unconditioned data (open circles) and conditioned data with outliers excluded (shaded circles); and the error structures estimated using *Conditioned Weighted* (dashed lines) regressions. The dash-dot-dash vertical lines are the MDL values.

sample pairs were excluded as outliers with an exception of 8 sample pairs for sulfate. Among the four approaches for estimating the error structure utilized for PM_{10} dataset, the *Conditioned Weighted* linear regression provided the least MSEs and hence, the best regression fit. The coefficients for error structure (a_j, b_j) are tabulated in Table 5-2 and the estimated error structures are plotted against the binned absolute collocated precision for few species are shown in Figure 5-9. Concentration dependence of collocated precision of sulfate (Figure 5-9 (b)) is confounded by the lack of sufficient data and/or issues with analytical measurements. In contrast, collocated precisions for other species, even with concentrations below the MDL, are well captured by the estimated error structure. The relationship between the additive error terms (a_j) derived from the collocated data and the MDL and limit of quantification (LOQ) values provided by analytical

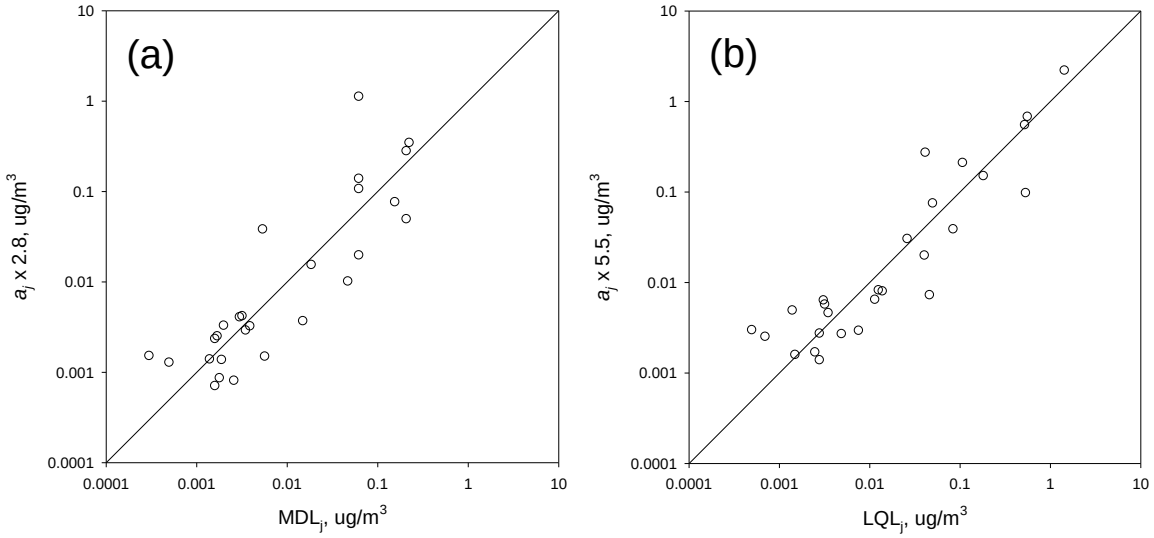


Figure 5-10. Additive term (a_j) calculated from the weighted regression of binned collocated precision on conditioned $PM_{2.5}$ concentrations versus: (a) reported MDL; and (b) LQL.

laboratory are examined in Figure 5-10. In each case a_j, was multiplied by a single-valued parameter that was adjusted to minimize the sum-of-squares difference between the datasets after log-transforming the data. For the MDL, the best-fit relationship was a_j = MDL/2.8, which is in excellent agreement with the value of 3 assumed by Polissar et al. (1998). High correlation was also observed between a_j and the LQL with best-fit relationship a_j = LOQ/5.5. For a more robust assessment of error structures for the HK $PM_{2.5}$ dataset, a larger collocated $PM_{2.5}$ dataset, when available, should be examined.

5.6. Conclusion

PM_{10} collocated species data from the Hong Kong air quality network was used to assess measurement precision and generate error structures for source apportionment modeling. The error structure was determined to be concentration dependent. Error structures representative of the entire concentration range were generated using weighted linear regression of the binned collocated precision for concentrations above the MDL. Outliers resulting from occasional

artifacts in sampling or analytical procedures are encountered in routine ambient sampling, but are not characterized by the measurement uncertainty of the general dataset. Removal of a few sample pairs that are deemed outliers resulted in well-behaved trends in precision over the entire concentration range. The scaled and relative difference plots also highlight banding on concentration differences that result from precision in reporting concentration values. To gauge the impact of banding of concentration differences resulting from precision in data reporting on precision estimation, Monte-Carlo-type simulations were performed by systematically randomizing the discreet concentration values. Rounding of concentrations for data reporting can result in less precise values at the high end of the concentration range. Air quality data encompasses many species, each varying over different concentration ranges. One plausible approach is to report each concentration value to the same degree of confidence, i.e. number of significant digits, as the method detection limit which is, the smallest acceptable measured concentration value for the species, assuming it is appropriately characterized.

5.7. Disclaimer

The content of this chapter does not necessarily reflect the views and policies of the Hong Kong Special Administrative Region (HKSAR) Government, nor does mention of trade names or commercial products constitute an endorsement or recommendation of their use.

5.8. Acknowledgments

We acknowledge the support from HKEPD (Tender Ref AS09-056) and assistance of Damgy H.L. Chan at HKEPD. This work was also sponsored by the Research Grant Council of Hong Kong, China (615406), the University Grant Committee (GMGS07/08.EG03) and the Hong

Kong University of Science and Technology, Fok Ying Tung Graduate School

(NRC06/07.SC01).

5.9. References

Barwick, V.J.; Ellison, S.L.R; Fairman, B.; Estimation of Uncertainties in ICP-MS Analysis: A Practical Methodology, *Analytica Chimica Acta*, **1999**, 394, 281-291.

Bohonak, A.J.; RMA: Software for Reduced Major Axis Regression; Version 1.17, San Diego State University; http://www.bio.sdsu.edu/pub/andy/rma.html (accessed Nov. 22, 2013).

Chueinta, W.; Hopke, P.K.; Paatero, P.; Investigation of Sources of Atmospheric Aerosol at Urban and Suburban Residential Areas of Thailand by Positive Matrix Factorization, *Atmospheric Environment*, **2000**, 34, 3319-3329.

Grubbs, F.; Procedures for Detecting Outlying Observations in Samples, *Technometrics*, **1969**, 11(1), pp. 1-21.

Hyslop, N.P.; White, W.H.; An Evaluation of Interagency Monitoring of Protected Visual Environments (IMPROVE) Collocated Precision and Uncertainty Estimates, *Atmospheric Environment*, **2008**, 42, 2691-2705.

Hyslop, N.P.; White, W.H.; Estimating Precision Using Duplicate Measurements, *Journal of Air & Waste Management Association*, **2009**, 59, 1032-1039.

Hyslop, N.P.; White, W.H.; Identifying Sources of Uncertainty from the Inter-Species Covariance of Measurement Errors, *Environmental Science & Technology*, **2011**, 45, 4030-4037.

International Vocabulary of Basic and General Terms in Metrology (VIM); DGUIDE 99999; International Standardization Organization: Geneva, Switzerland, **2004**.

Ito, K.; Xue, N.; Thurston, G.; Spatial Variation of $PM_{2.5}$ Chemical Species and Source-Apportioned Mass Concentrations in New York City, *Atmospheric Environment*, **2004**, 38, 5269-5282.

Karakas, S.Y.; Uncertainty Assessment Including Sampling for Determination of Metals in Lichen Samples using ICP-OES, *Journal of Chemical Metrology*, **2007**, 1:1,10-24

Kim, E.; Hopke, P.K.; Improving Source Identification of Fine Particles in a Rural Northeastern U.S. Area Utilizing Temperature-Resolved Carbon Fractions, *Journal of Geophysical Research*, **2004**, 109, DOI 10.1029/2003JD004199.

Kim, E.; Hopke, P.K.; Qin, Y.; Estimation of Organic Carbon Black Values and Error Structures of the Speciation Trends Network Data for Source Apportionment, *Journal of Air & Waste Management Association*, **2005**, 55, 1190-1199.

Lau, A.K.H.; Yuan Z.; Bian, E.Q.; Trend and Apportionment Analysis of Air Pollution Sources over Hong Kong, Project Number #AS06-435, prepared by the Hong Kong University of Science and Technology for the Hong Kong Environmental Protection Department, November **2010**.

Maykut, N.N; Lewtas, J.; Kim, E.; Larson, T.V.; Source Apportionment of PM2.5 at an Urban IMPROVE Site in Seattle, Washington, *Environmental Science & Technology,* **2003**, 37, 5135-5142.

Tylor, B.N.; Kuyatt, C.E.; Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, National Institute of Standards and Technology, Technical Notes 1297, **1994** (http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf)

Polissar, A.V; Hopke, P.K.; Paatero, P.; Malm, W.C.; Sisler, J.F.; Atmospheric Aerosol over Alaska 2. Elemental Composition and Sources, *Journal of Geophysical Research*, **1998**, 103, 19045-19057.

Polissar, A.V; Hopke, P.K.; Poirot, R.I.; Atmospheric Aerosol over Vermont: Elemental Composition and Sources, *Environmental Science & Technology*, **2001**, 35, 4604-4621.

Ramadan, Z.; Song, X.H.; Hopke, P.K.; Identification of Sources of Phoneix Aerosol by Positive Matrix Factorization, *Journal of Air & Waste Management Association*, **2000**, 50, 1308-1320.

Reff, A.; Eberly, S.I.; Bhave, P.V.; Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods, *Journal of Air & Waste Management Association*, **2007**, 57, 146-154.

Taylor, J..R.; An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, **1982**, University Science Books.

Wade, K.S.; Brown, S.G.; Turner, J.R.; Garlock, J.L.; Concentration Value Uncertainty Estimates for Source Apportionment Modeling of Chemical Speciation Network Data, *Proceedings of the Air & Waste Management Association 2008 Annual Conference*, **2008**, Paper #632.

Yadav, V.; Yuan, Z.; Lau, A.K.H.; Louie, P.K.K.; Turner, J.R.; Particulate matter impact to Hong Kong region by weight-of-evidence approach from multi-site network data; ***To be submitted***.

Yuan, Z.; Yadav, V.; Turner, J.R.; Louie, P.K.K.; Lau, A.K.H.; Long-term Trends of Ambient Particulate Matter Emission source contributions and the Accountability of Control Strategies – A Case Study in Hong Kong over a Decade (1998-2008), *Atmospheric Environment,* **2013**, 76, 21-31.

Chapter 6 : Summary

As part of this dissertation work, data analysis strategies were developed and evaluated for interpreting intraurban spatiotemporal variability in concentrations measured across a network of monitoring sites. The application of the proposed strategies to networks with varying characteristics provided insights into the unifying aspects of the approach and also highlighted those aspects that needed to be customized to accommodate the distinctive features of the specific datasets. In this chapter, the advantages, limitations and recommendations for future work on these interconnected approaches are discussed.

6.1. Measurement error

Data quality can profoundly impact the variability metrics and their interpretation. An analysis to generate error structures for the Hong Kong PM_{10} speciation network using collocated data revealed structural issues in the data reporting that led to false representation of measurement precision. This finding also has implications to variability assessments. For example, a data reporting convention based on fixed digit scientific notation (e.g., $X.X*10^{Y}$) could lead to concentration differences of 0.1 being reported as 1.4 versus 1.5 in one decadal range but in the next higher decadal range values that are 11.4 and 11.5 would be reported as 11 and 12, respectively. These concentration differences could strongly influence the variability metrics. One possible way to avoid such misrepresentation of variability is to report all concentrations to the same degree of accuracy as the detection limit of the species and provide appropriate uncertainty estimates.

Another issue encountered in several of the data analyses was that concentrations near or below the detection limit can dramatically influence the variability metrics. In particular, metrics that

are based on relative differences, such as the COD, can appear to exhibit heterogeneity which actually stems from the relatively increased noise for concentrations near or below the MDL. The common approach of imputing concentration values below the MDL with ½ the MDL can lead to a false increase in homogeneity. While the data can be censored to remove such cases, the resulting variability metrics no longer represent the behavior of the entire dataset. Furthermore, outliers in measured concentrations can also skew the interpretation of variability and can be very challenging to isolate. While there are no clear solutions to these issues, it is important to identify cases where such issues arise and clearly document how they are handled.

Analytical uncertainty of concentrations typically provided by analytical laboratories only evaluates the precision in the estimated sample concentration. The factors contributing to the occurrence of covariance in measurement error across species and its influences on the representativeness of measurement uncertainty estimates needs to be further evaluated. Collocated measurements, if available, aid in evaluating measurement precision for the combined sampling and analysis protocol. Further, COD for collocated data is a measure of precision and it could be used, alone or as a *PCC-COD plot*, to provide a context for interpreting inter-site COD values. However, on-site collocated measurements are resource-intensive and may or may not be conducted, depending on the project resources and objectives. In the absence of such datasets, the variability in measured concentrations may be evaluated using similar datasets that were collected and analyzed by other agencies as part of routine monitoring.

6.2. Interpretation of statistical metrics for gauging variability

The interpretation of PCC and COD, in a relative sense across sites in a monitoring network, is graphically aided by the *PCC-COD plot*, which merges the temporal and spatial tracking ability

of the individual metrics. In the *PCC-COD plot* using measured concentrations, a clustering of site pairs towards the homogeneity end of the *variability line* indicates general uniformity in species concentrations across the monitoring network. Plots constructed using site-specific measured and baseline concentrations can have several advantages. First, the sites that track the baseline are clustered near the homogeneity end of the *variability line* and thus the plot aids in better understanding of both the site-specific and baseline time series. Second, there is a reduction in dimensionality because each site appears only once in the plot rather than all the possible site pair combinations. Third, by comparing the site-specific measured concentrations to a common baseline, which can be a significant portion of the measured concentrations, the inter-site variability is amplified and sites displaying greater spatiotemporal variability can be better categorized. The cautions that are applied to the interpretation of the individual metrics extend to this plot as well. As previously stated, the presence of concentrations near/below the MDL or outliers in the dataset can drastically influence these metrics.

In many but not all cases the sites tend to cluster along the *variability line* that spans the (COD, PCC) range from absolute homogeneity (0, 1) to absolute heterogeneity (1, 0). Further work is needed to explain why this pattern is often observed as well as to improve the interpretation of a site-pair/sites' position on the plot. One possible approach is to perform dispersion modeling for emission sources(s) to create modeled concentration fields and examine the influence of receptor locations on PCC and COD. While such exercises may provide insights on the various regimes of the *PCC-COD plot*, interpreting the variability in observed ambient concentrations based on modeled concentrations can be challenging. Ambient concentrations in each network are dependent on various factors such as the spatial scale of the network, zone of influence of the emission sources, emission source profiles, meteorological conditions etc. Despite such issues

regarding the interpretation of the *variability line*, it does provide a reference to gauge the relative spatiotemporal variability across the sites.

6.3. Baseline-excess apportionment

The most significant advantage, and the limitation, of the baseline-excess apportionment is that the baseline can be subjectively defined by the researcher depending on the objectives of the study and in this regard is similar to other conventional tools typically used such as receptor modeling tools where the user subjectively decides on the number of factors (sources) representative for the dataset. The baseline can be any reference concentration time series, such as concentrations from a background site or other site with well understood concentration trends. In the absence of such a reference, the variability inherent to the concentrations measured across the network can be utilized to estimate a baseline concentration time series. For monitoring networks with just a few sites, the selection of the baseline was restricted to the minimum concentration measured across the network. However, networks with a large number of sites provide greater flexibility to choose the baseline to establish the general characteristics of the network-wide uniform contributions. The foremost requirement for meaningful interpretation of variability with respect to the defined baseline is to understand the baseline and the factors contributing to it. The *PCC-COD plot* (constructed using site-specific measured and baseline concentrations) along with examination of summary statistics of the baseline and site-specific excess concentrations aid in identifying the sites contributing to the baseline as well as establishing the representativeness of the background site, if available.

One of the key assumptions of this approach relies on network-wide uniform concentration burdens from emission sources located far away from the network. Interpreting variability will

be challenging for networks with sources located outside the network yet close enough to result in spatially differential impacts across the network. The applicability of this approach is also questionable for species that have the potential to chemically react over the spatial domain of the network. The baseline-excess apportionment does not provide the absolute local- (for the particular network) or large-scale source emission contributions but it does utilize the inherent concentration variability across the network to establish semi-quantitative estimates of network-scale and larger-scale emission source contributions. Thus, this approach can be particularly useful for monitoring networks with significant and temporally varying larger-scale contributions because it can isolate deviations resulting from local source impacts. For networks with dominant local-scale contributions, this approach still provides insights for the sites that display relatively high variability. In either case, the sites with higher variability are distinguished and can be examined to identify the underlying drivers, whether anthropogenic, natural (geographic barriers, meteorology etc.) or both. With a clear understanding of such constraints, the baseline-excess apportionment, when applicable, can significantly improve evaluation of spatiotemporal variability.

6.4. Estimating emission source zones

Extending the application of the baseline-excess apportionment to pollution rose-type graphical tools, such as the CPF and NWR plots, also aids in refined estimation of the directionality of local emission sources. Using concentration differences (either with respect to the defined baseline or between sites) focuses on the bearings that correspond to the observed excess concentrations. However, such concentration difference time series are prone to higher propagated uncertainty than the measured concentrations. Collocated data, if available, can be used to evaluate if the concentration differences are merely noise or truly represent ambient

variability. More work is needed to evaluate the efficacy of estimating source bearings through exercises such as peak-fitting of the NWR plots. In particular, uncertainty estimates for the resolved bearings are needed. Furthermore, the measured ambient concentrations at the receptor sites are affected by various attributes of the contributing sources, such as the number of sources, their spatial and temporal scales of influence, fluctuations in emission rates, etc. In the absence of refined uncertainties for the resolved bearings, large conservative uncertainties, such as the optimized smoothing parameter of the NWR plot, can be used for the triangulation of emission source zones.

Source bearings resolved for ambient concentrations at each site may triangulate to multiple emission zones. Such ambiguities in estimation of emission source locations can result from factors such as the presence of closely sited sources or sources in corridors, intermittent source emissions, and from errors in measurements or modeling. The deduction approach requires that multiple receptor sites be impacted by concentrations from the same source which may not occur if the data is sparse for certain wind directions or if the impact at a receptor site is caused by an exceptional event that occurs during periods when the wind direction causes only one site to be affected. Furthermore, the efficacy of meteorological parameters at evaluating concentration-wind direction relationships also depends on having sufficient frequency of concentration values with varying wind directions. Despite such limitations, the approach narrows down the spatial extent of the emission zones which contribute to elevated receptor concentrations. The sources identified in such triangulated zones should be evaluated using emission inventories, if available. Their impact at the receptor sites can be further validated using other tools such as dispersion modeling (AERMOD, CALPUFF etc.).

6.5. Concluding remarks

The data analysis strategies developed and evaluated as part of this dissertation contribute to various aspects of air quality data assessment. While monitoring networks have distinctive features and are designed for different objectives, the data-driven approach exploits the unifying characteristics of the network data structure to aid in understanding the factors contributing to variability across a variety of monitoring networks. The most significant contribution of these strategies is the opportunity to combine their results with other, conventional tools into a weight-of-evidence towards gauging spatiotemporal variability for effective air quality management.

Appendix A: Summary of tasks performed for this dissertation.

St. Louis Community Air Toxics Project

- Operated a PM_{10}HiVol sampler at Washington University in St. Louis during 2008.
- Optimized a sample digestion and ICP-MS analysis protocol for arsenic and certain other air toxics metals. Extensive QA/QC characterization of blank levels, interferences, recoveries (using NIST standard reference materials), etc.
- Performed filter digestion by hot acid digestion using a ModBlock, elemental analysis by ICP-MS and archived more than five hundred PM samples from the four site network in Missouri and additional samples provided by Illinois EPA.
- Performed preliminary intraurban variability analysis of air toxics metals in St. Louis.
- Developed approach to interpret scattergram of Pearson's correlation coefficient (PCC) and coefficient of divergence (COD) (i.e., *PCC-COD plot*) to combine temporal tracking ability of PCC with COD's ability to gauge spatial variability.
- Developed the *baseline-excess approach* to utilize the variability inherent to the data structure for gauging intraurban variability.
- Refined triangulation approach to identify putative emission sources using Conditional Probability Function (CPF) plots of site-specific excess concentrations and validated resolved emission zones with the National Emission Inventory, 2008.

Dearborn Midwest Rail Study

- Developed and implemented a comprehensive dispersion modeling framework for PM emissions generating activities at the Rougemere Rail Yard. (Emission inventory prepared by Sierra Research, Inc.)
- Developed modeling strategies and input files for dispersion modeling (AERMOD) and interpreted the modeling results.
- Performed spatial variability analyses of hourly carbon measurements from a three month field-campaign at a three-site monitoring network in Dearborn, MI.
- Extended the *PCC-COD plot* and *baseline-excess approach* to gauge variability and identify the background sites for this field campaign.

- Performed Non-parametric wind regression (NWR) and Peak-fitting (using IGOR-PRO^{TM} software) of inter-site concentration differences to examine concentration-wind direction relationship.
- Refined emission source zones estimation by systematic triangulation of resolved NWR peaks.

Variability of Sulfur Di-oxide over Southeast Michigan

- Extended the *baseline-excess approach* to gauge variability in sulfur di-oxide measurements at a three-site monitoring network in Detroit, MI.
- Identified the industrial area around Zug Island, MI as the sulfur dioxide emission zone causing exceedances, leading the Michigan Department of Environmental Quality to develop a detailed emission inventory for the area.
- This analysis was included in the State of Michigan's attainment designation request as part of their weight-of-evidence for where the nonattainment area boundary should be drawn

Hong Kong Air Quality Study

- Developed sample-specific uncertainty estimates using collocated precision for source apportionment studies.
- Identified issues with precision in data reporting and performed sensitivity tests to examine the impact of rounding of reported concentrations and outliers on uncertainty estimation.
- Performed source apportionment of PM_{10} and $PM_{2.5}$ speciation data collected at ten sites using EPA Positive Matrix Factorization (PMF), EPA UNMIX and Absolute Principle Component Analysis (APCA).
- Compared and contrasted source apportionment models through sensitivity studies.
- Examined long-term temporal and seasonal variability in derived source contributions.
- Extended the *PCC-COD plot* and *baseline-excess approach* to gauge variability in hourly mass measured over 1998-2008 by Tapered Element Oscillating Microbalance (TEOM) monitors at fourteen sites.
- Identified meteorological and geographic factors modulating spatial variability of PM

Coarse PM Speciation Pilot Study

- Managed a year-long field campaign during 2010-2011 at the East St. Louis site.
- Responsible for all routine operation, maintenance and extensive troubleshooting of two sequential FRM samplers, two sequential dichotomous samplers, and a MOUDI cascade impactor.

Evaluation of Cooper Environmental's Xact 620 semi-continuous field XRF monitor

- Performed digestion and elemental analysis by ICP-MS of HiVol filter samples collected concurrently with Xact 620.
- Evaluated Xact 620 bias against filter-based sampling and precision using collocated data.

Mongolia Air Quality Project

- Performed filter digestion by hot acid digestion using ModBlock and elemental analysis by ICP-MS for ~160 PM_{10} samples collected in Mongolia.

Federal Equivalent Method (FEM) testing

- Performed Federal Equivalent Method (FEM) testing of Teledyne-Advanced Pollution Instrumentations' Model 602 Beta Particle Measurement System during two-month long field campaign in 2011-2012.
- Responsible for daily routine operations and maintenance of three Model 602 Beta and six BGI PM_{10} samplers.

Appendix B: Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies – A case study in Hong Kong over a decade (1998-2008)[*].

B.1. Abstract

Despite extensive emission control measures targeting motor vehicles and to a lesser extent other sources, annual-average PM_{10} mass concentrations in Hong Kong have remained relatively constant for the past several years and for some air quality metrics, such as the frequency of poor visibility days, conditions have degraded. The underlying drivers for these long-term trends were examined by performing source apportionment on eleven years (1998-2008) of data for seven monitoring sites in the Hong Kong PM_{10} chemical speciation network. Nine factors were resolved using Positive Matrix Factorization. These factors were assigned to emission source categories that were classified as local (operationally defined as within the Hong Kong Special Administrative Region) or non-local based on temporal and spatial patterns in the source contribution estimates. This data-driven analysis provides strong evidence that local controls on motor vehicle emissions have been effective in reducing motor vehicle-related ambient PM_{10} burdens with annual-average contributions at neighborhood- and larger-scale monitoring stations decreasing by ~6 $\mu g/m^3$ over the eleven year period. However, this improvement has been offset by an increase in annual-average contributions from non-local contributions, especially secondary sulfate and nitrate, of ~8 $\mu g/m^3$ over the same time period. As a result, non-local

[*] Published paper: "Yuan, Z.; Yadav, V.; Turner, J.R.; Louie, P.K.K.; Lau, A.K.H.; Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies – A case study in Hong Kong over a decade (1998-2008), *Atmospheric Environment*, **2013**, 76, 21-31."

source contributions to urban-scale PM_{10} have increased from 58% in 1998 to 70% in 2008. Most of the motor vehicle-related decrease and non-local source driven increase occurred over the period 1998-2004 with more modest changes thereafter. Non-local contributions increased most dramatically for the secondary sulfate and secondary nitrate factors and thus combustion-related control strategies, including but not limited to power plants, are needed for sources located in the Pearl River Delta and more distant regions to improve air quality conditions in Hong Kong. PMF-resolved source contribution estimates were also used to examine differential contributions of emission source categories during high PM episodes compared to study-average behavior. While contributions from all source categories increased to some extent on high PM days, the increases were disproportionately high for the non-local sources. Thus, controls on emission sources located outside the Hong Kong Special Administrative Region will be needed to effectively decrease the frequency and severity of high PM episodes.

B.2. Introduction

During the past twenty years, the environmental authorities in Hong Kong have undertaken a series of air pollution control measures including but not limited to requirements for low-sulfur diesel fuel, lead-free gasoline, three-way catalytic converters, conversion of diesel taxis and minibuses to liquefied petroleum gas, and capping the VOC content of some solvents and consumer products. Despite these measures, the long-term trends for various ambient air quality metrics show either no improvement or a worsening of conditions. For example, the open circles in Figure B-1 show the annual-average PM_{10} mass concentration composited over six monitoring stations that collectively represent urban-scale conditions (these sites are described later in this chapter). Hourly data from the PM_{10} monitoring network were used to construct the annual

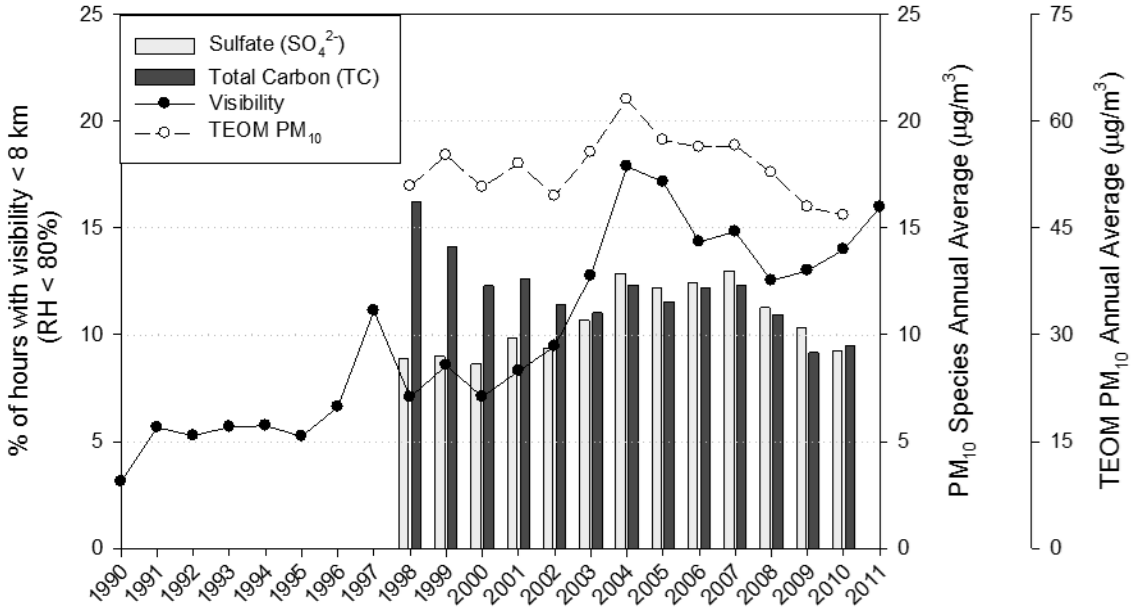


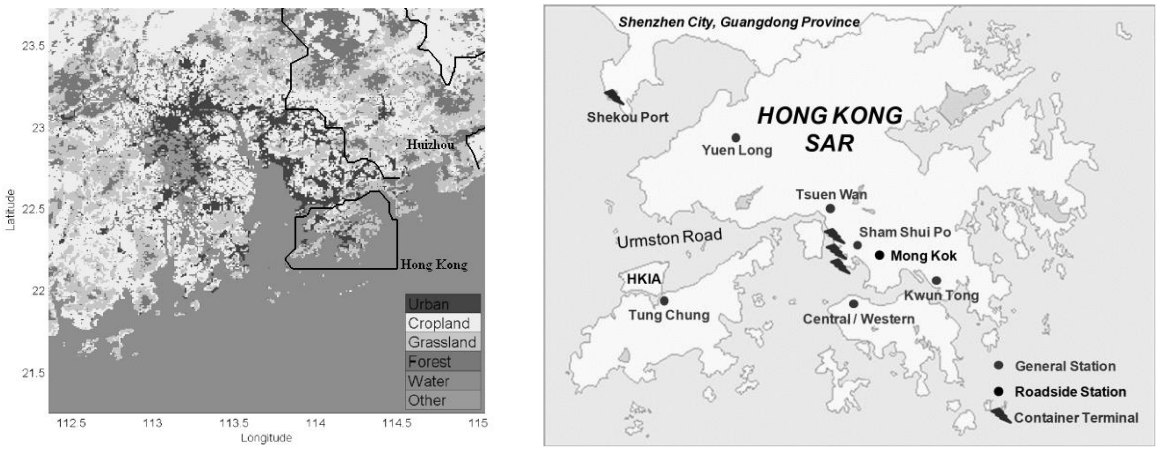
Figure B-1. Historical trends for visibility impairment and PM_{10} mass, sulfate and total carbon. Data sources and calculation methodology are described in the text.

averages. Annual-average PM_{10} mass was relatively constant over the period from 1998-2008 with a 9% increase in mass for the five years starting in 2004 compared to the five year period starting in 1998. The solid circles in Figure B-1 show the percentage of hours with poor visibility (visibility less than 8 km with relative humidity less than or equal to 80%) measured at the Hong Kong Observatory (HKO, 2012). The frequency of poor visibility days increased throughout the 1990s with more dramatic increases during first half of the 2000s. A local maximum was observed in 2004 (18% of hours) with the frequency of poor visibility hours decreasing through 2008 and then increasing again through 2011. The frequency of hours with poor visibility increased by 47% for the five years starting in 2004 compared to the five year period starting in 1998. The substantially greater worsening of visibility compared to the more modest increase in PM_{10} mass has been driven by changes in chemical composition of the PM_{10} aerosol. Figure B-1 also shows that over the period from 1998 to about 2004, for the same six sites used to construct the PM_{10} TEOM mass trend, the annual average total carbon dramatically

decreased while annual average sulfate significantly increased. Thus, while there was little change in PM_{10} mass during this period, the composition shifted towards an aerosol with higher light extinction efficiency that degraded the visibility more efficiently (Andreae et al., 2008).

The two most pervasive air pollution issues in Hong Kong are street-level (especially near roadway) air quality and regional-scale air quality that affects not only Hong Kong but also the entire Pearl River Delta (PRD). The emission inventory reveals gasoline and diesel vehicles to be the main sources of street-level pollution (Zheng et al., 2009a). Regional-scale air pollution, however, is caused by the cumulative impacts from numerous emission source categories such as motor vehicles, marine vessel emissions, industry and power plants located in Hong Kong, the greater PRD region, and beyond. As shown in Figure B-2, the PRD region of Guangdong Province, which historically was Hong Kong's hinterland, has become highly urbanized and indeed has developed as the largest workshop in the world. Over the past fifteen years, pollutant emissions within the PRD region have increased dramatically as a result of booming industrial activities and relatively modest pollution control regulations. Therefore, the pollutants transported from the greater PRD region to Hong Kong are expected to have correspondingly increased.

Effective air quality management necessarily requires setting priorities and therefore it is important to identify the emission source categories leading to deterioration of air quality in Hong Kong and quantify the long-term trends in the contributions from such sources. To the extent there are large contributions from within the Hong Kong Special Administrative Region (HKSAR) the local government can implement additional controls on. However, if the dominant contributions are from the greater PRD area and more distant regions, a regional planning process that involves cooperation between Hong Kong, Guangdong Province and perhaps other



(a) (b)

Figure B-2. (a) Land-use map of the Pearl River Delta in 2003, and (b) geographical distribution of the Air Quality Monitoring Stations in the Hong Kong Special Administrate Region.

provinces in Mainland China will be necessary to efficiently manage air quality in Hong Kong.

This study was commissioned to extend our previous work to apportion ambient PM_{10} levels in Hong Kong (Yuan et al., 2006; Huang et al., 2009) with emphasis on further identifying and quantifying the drivers for the long terms trends in Hong Kong air quality such as those features demonstrated in Figure B-1. Insights were sought into whether air quality benefits from prior control measures could be elucidated from routine monitoring data. This information could be used to frame and prioritize future control policies. Following the methodology of the previous studies, receptor modeling was performed on the PM_{10} speciation data.

B.3. Data collection and analysis

For more than twenty years, the Hong Kong Environmental Protection Department (HKEPD) has operated a PM_{10} chemical speciation network (network) that has included several monitoring stations throughout the HKSAR. 24-hour integrated (midnight-to-midnight) samples are collected on a 1-in-6 day frequency using High Volume PM_{10} samplers with quartz filters. The filters are analyzed for gravimetric mass, elements by inductively coupled plasma atomic

emission spectroscopy and ions by ion chromatography. Since January 1998, a thermal/optical transmittance method following the NIOSH 5040 protocol (Chow et al., 2001) has been used to obtain organic carbon (OC) and elemental carbon (EC). Table B-1 summarizes the valid sampling periods and environmental characteristics for each of the seven stations that have operated nearly continuously over the 1998-2008 period and were used in this study. The first six stations (YL through KT) are neighborhood-scale sites and collectively are considered to represent urban-scale PM_{10} levels albeit with some underlying spatial variability. For the purpose of this study they are termed "general stations" and this descriptor is used when analyses are presented as the average over these six sites. General stations are typically located on rooftops of four- to six-story buildings with unobstructed airflow from most directions. The remaining site is the MK station that is located only a few meters above the ground in a heavily trafficked area to gauge roadside exposures by pedestrians. Towards the end of 2000, the MK station was relocated to its current location and started operation on 2 January 2001. This relocation caused a discontinuity in the time series for PM_{10} mass and some species. Therefore, data for MK is included only for the period from 2001 onward. Detailed information about the network can be found in Yuan et al. (2006) and Section BS.1.

Source apportionment modeling was performed using Positive Matrix Factorization (PMF). PMF is a factor analytic method that provides a parameterized convergence scheme and constrains all the elements in the factor score and factor loading matrices to be positive (within a narrow tolerance). These features are attractive for the modeling of environmental data sets and PMF has been widely used for the source apportionment of different ambient pollutants, including particulate matter (e.g. Lee et al., 1999; Lee et al., 2003), volatile organic compounds

Table B-1. Monitoring station characteristics

Station		Count	Start Date	End Date	Major Data Gap	Station Characteristics
Yuen Long	YL	654	Jan. 2, 1998	Dec. 26, 2008		Urban / Residential
Tung Chung	TC	575	Apr. 3, 1999	Dec. 28, 2008		New Town / Residential
Tsuen Wan	TW	631	Jan. 3, 1998	Dec. 26, 2008	Feb. 6, 2003 – Apr. 23, 2003	Urban / Residential / Commercial
Sham Shui Po	SSP	586	Jan. 11, 1998	Dec. 26, 2008	Nov. 29, 2003 – Mar. 16, 2004; May 12, 2007 – Jun. 5, 2008	Urban/ Commercial
Central / Western	CW	641	Jan. 2, 1998	Dec. 26, 2008	Nov. 9, 2001 – Mar. 27, 2002	Urban / Residential
Kwun Tong	KT	598	Jan. 4, 1998	Dec. 31, 2008	Oct. 31, 1998 – Jan. 11, 1999; Apr. 13, 2002 – Feb. 1, 2003	Commercial / Residential / Near ferry pier
Mong Kok	MK	482	Jan. 2, 2001	Dec. 27, 2008		Roadside, relocated at the end of 2000

(e.g. Song et al., 2007; Lau et al., 2010), polycyclic aromatic hydrocarbons (e.g. Larsen et al., 2003), and organic aerosols (e.g. Lanz et al., 2007). In this study, PMF modeling was performed using EPA PMF Version 3.0 (USEPA, 2008).

The goal of PMF is to identify a number of factors *p*, the species profile *f* of each source, and the amount of mass *g* contributed by each factor to each individual sample, as shown in Equation (B-1) (USEPA, 2008).

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \qquad \text{(B-1)}$$

where x_{ij} is the element in the data matrix with *n* number of samples and *m* chemical species measured, and e_{ij} is the residual for the element corresponding to the i^{th} sample and the j^{th} species. The PMF solution minimizes the object function *Q* (Equation (B-2)), based upon the data point-specific uncertainties u_{ij}.

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{u_{ij}} \right] \qquad \text{(B-2)}$$

It is therefore important to provide plausible uncertainty estimates that can reflect the variability and artifact in the processes of sampling, chemical analysis and transformation of pollutant profiles in the air.

In this study, concentration dependent error structures were assumed with the form $U_{ij} = a_j + b_j \times C_{ij}$ where U_{ij} is the uncertainty for species *j* in sample *i*, C_{ij} is the mass concentration of species *j* in sample *i*, and a_j and b_j are the species-specific additive and multiplicative contributions, respectively, to the uncertainty. For each species the a_j and b_j coefficients were derived from an analysis of collocated samples routinely collected at the SSP station and evaluated using

collocated samples also routinely collected at the TW station. The methodology to develop the error structures from collocated data is summarized in Section BS.2. Table B-2 lists the derived a_j and b_j values, effective MDL and percentage of samples below the effective MDL for each species.

Data from all sites were combined into a single dataset of 4167 samples to generate a consistent set of source profiles across the sites. The quartz fiber filter gravimetric mass (hereafter denoted as QTM) was used as the total mass variable. The rationale for using QTM as the total mass variable is discussed in Section BS.3. 10% extra modeling uncertainty was imposed to account for possible temporal changes in the source profiles and other sources of variability. Twenty base runs were performed and the run with the minimum *Q* value was selected as the base run solution. 100 bootstrap runs were conducted with minimum correlation R-value of 0.6 to examine the stability and to estimate the uncertainty of the base run solution. All of the bootstrapped factors were uniquely mapped to a factor from the base solution, and no factors were unmapped. For the *Q* values derived from the twenty base runs, the standard deviation was only ~0.004% of the mean, indicating a very stable solution. For this modeling, an *Fpeak* value of zero, indicating no rotation performed, resulted in the most meaningful results. Modeling was performed for four to ten factors and the nine factor solution was deemed to be most representative, as discussed in detail in Section BS.4.

Putative emission source categories were mapped onto the factors by identifying the tracer(s) for each source which are typically the species that exclusively or largely reside in a particular source. Table B-3 lists the emission source categories and corresponding tracers used for matching PMF-resolved factors to the sources. It is noted that the bootstrap method accounts for

Table B-2. Additive (a_j) and multiplicative (b_j) error terms derived using collocated data, effective minimum detection limit (defined in the text), and the percentage of samples below the effective detection limit.

Species j	a_j (μg/m^3)	b_j	Effective MDL (μg/m^3)	% of samples below Effective MDL
Al	0.046	0.071	0.014	1%
As	0.0001	0.050	0.0002	25%
Ca	0.022	0.043	0.065	0%
Cd	0.00002	0.058	0.0001	28%
Cl^-	0.014	0.047	0.042	0%
Fe	0.004	0.030	0.013	0%
K^+	0.012	0.028	0.035	0%
Mg	0.005	0.020	0.015	0%
Mn	0.0002	0.029	0.001	27%
Na^+	0.026	0.018	0.079	0%
NH_4^+	0.017	0.028	0.051	0%
Ni	0.0003	0.050	0.0008	36%
NO_3^-	0.043	0.026	0.129	0%
Pb	0.0002	0.035	0.001	20%
SO_4^{2-}	0.060	0.025	0.179	0%
V	0.00001	0.047	0.00002	48%
Zn	0.002	0.032	0.005	6%
EC	0.115	0.023	0.344	0%
OC	0.112	0.032	0.335	0%

Table B-3. Tracers applied for source identification and assignment of factors as primarily of local (within the HKSAR) or non-local origin.

Source	Tracers	Dominant Spatial Scale
Vehicle Exhaust	EC, OC	local
Residual Oil	Ni, V	local
Fresh Sea Salt	Na^+, Mg^{2+}, Cl^-	local
Aged Sea Salt	Na^+, Mg^{2+}, NO_3^-, SO_4^{2-}	local
Crustal Soil / Dust	Al, Ca, Mg, Fe	non-local
Secondary Sulfate	NH_4^+, SO_4^{2-}	non-local
Secondary Nitrate	NH_4^+, NO_3^-	non-local
Trace Metals	Zn	non-local
Coal Combustion / Biomass Burning	As, Cd, Pb, K^+, OC	non-local

some but not all sources of uncertainty and thus is likely a lower bound on the true uncertainties in the modeling results. The PM_{10} emission source categories assigned to the nine factors are vehicular exhaust, residual oil combustion, fresh sea salt, aged sea salt, crustal material/soil, secondary sulfate, secondary nitrate, trace metals from sources such as zinc smelting, and coal combustion admixed with biomass burning, Detailed information on source identification can be found in Yuan et al. (2006) and Section BS.4.

To examine the robustness of the PMF source apportionment results, additional receptor models including principal components analysis with absolute principal components scores (APCA, Thurston and Spengler, 1985) and Unmix (USEPA, 2007) were also applied. Generally, the three models yielded similar estimations for sources of vehicle exhaust, residual oil, crustal soil, secondary sulfate, and coal combustion / biomass burning. Considering that APCA and Unmix cannot resolve distinct aged sea salt and secondary nitrate factors and significant negative loadings onto a few factors were observed, the PMF results were considered the most reasonable and used for the remainder of the study. APCA and Unmix results were provided in Section BS.5 for reference.

B.4. Results and Discussion

B.4.1. Source contributions and their seasonal and spatial variations

As previously discussed, one objective of the source apportionment modeling was to determine local (again, operationally identified as sources within the HKSAR) and non-local emission source categories and their respective contributions because the control of local sources is fully within Hong Kong's jurisdiction while controlling non-local sources requires regional

cooperation. The approach largely focused on assigning sources as local or non-local based on seasonal and spatial patterns in the PMF-modeled source contribution estimates (SCE).

Hong Kong is situated in the subtropics along the southeast coast of Eurasia Continent. The seasonal evolution of weather in Hong Kong is closely related to and controlled by seasonal evolution of the East Asian Monsoon system. In summer, a large surface low-pressure system develops over Asia in response to rising surface temperature over the continent. Consequently, cyclonic winds flow into the Asian landmass in the lower levels, leading to southerly or southwesterly prevailing winds in Hong Kong bringing clean oceanic air mass from the sea. In winter the temperatures inland become lower than that of the ocean, leading to anti-cyclonic flow over the Asian landmass. This wintertime pattern drives mainly northerly or northeasterly surface and synoptic winds that transport particulate matter to Hong Kong from the PRD and regions further inland. In light of these characteristics, seasonal and spatial patterns for the receptor modeled SCEs were examined from the perspective that local source SCEs typically should have relatively larger spatial variation over the network and smaller seasonal variation, while non-local sources should exhibit smaller spatial variation and larger seasonal variation. Based on the prevailing synoptic systems, for this analysis the seasons in Hong Kong were defined as summer from 16 May to 15 September; winter from 16 November to 15 March of following year; and a transitional season, including both spring and fall, from 16 March to 15 May and from 16 September to 15 November.

Table B-4 lists the annual-, wintertime-, summertime-, and transition season- averaged source contributions at general stations and roadside station during the 11-year period from 1998 to 2008. Based on the general station SCEs, the secondary sulfate factor is the largest contributor to PM_{10} mass and accounts for 22% of the period-average ambient PM_{10} in Hong Kong. SCEs

Table B-4.Annual-, wintertime-, summertime-, and transition season- averaged source contribution estimates (unit: μg/m^3) and contribution percentages for general stations from 1998 to 2008 and for the roadside station from 2001 to 2008 (unit: μg/m^3)

	General stations (YL, TC, TW, SSP, CW & KT)					Roadside station (MK)				
	Annual	Winter	Summer	Transitional	Win/Sum	Annual	Winter	Summer	Transitional	Win/Sum
Vehicle Exhaust	8.5 (15%)	8.2 (11%)	8.9 (25%)	8.4 (15%)	0.92	21.3 (30%)	19.0 (22%)	23.3 (47%)	21.4 (28%)	0.82
Residual Oil	0.2 (0.4%)	0.2 (0.3%)	0.3 (1%)	0.2 (0.4%)	0.74	0.3 (0.4%)	0.2 (0.2%)	0.3 (1%)	0.3 (0.4%)	0.57
Fresh Sea Salt	1.8 (3%)	2.1 (3%)	1.8 (5%)	1.4 (2%)	1.2	2.3 (3%)	2.5 (3%)	2.8 (6%)	1.7 (2%)	0.90
Aged Sea Salt	7.1 (13%)	6.2 (8%)	6.5 (19%)	8.6 (15%)	0.95	7.2 (10%)	6.5 (7%)	6.5 (13%)	8.5 (11%)	1.0
Crustal Soil / Dust	6.7 (12%)	9.7 (13%)	3.2 (9%)	7.3 (13%)	3.0	7.5 (11%)	10.3 (12%)	4.0 (8%)	8.4 (11%)	2.6
Secondary Sulfate	12.3 (22%)	16.1 (22%)	6.9 (20%)	14.1 (25%)	2.3	13.4 (19%)	16.4 (19%)	6.9 (14%)	17.1 (22%)	2.4
Secondary Nitrate	8.3 (15%)	14.9 (20%)	3.1 (9%)	7.2 (13%)	4.8	11.0 (16%)	18.2 (21%)	4.5 (9%)	10.6 (14%)	4.0
Trace Metals	2.6 (5%)	4.1 (5%)	1.0 (3%)	2.9 (5%)	4.0	2.8 (4%)	4.1 (5%)	0.9 (2%)	3.4 (4%)	4.4
Coal Comb. / Biomass Burning	7.3 (13%)	12.7 (17%)	2.6 (7%)	7.1 (12%)	5.0	6.2 (9%)	10.4 (12%)	1.7 (3%)	6.7 (9%)	6.1
Local sources	17.6 (32%)	16.7 (22%)	17.4 (50%)	18.6 (32%)	1.0	31.0 (44%)	28.2 (33%)	32.9 (66%)	32.0 (41%)	0.86
Non-local sources	37.2 (67%)	57.6 (77%)	16.8 (48%)	38.6 (67%)	3.4	41.0 (58%)	59.3 (69%)	18.1 (36%)	46.3 (60%)	3.3
Residual*	0.5 (1%)	0.6 (1%)	0.8 (2%)	0.2 (0.3%)		-1.1 (-2%)	-0.9 (-1%)	-1.2 (-3%)	-1.1 (-1%)	
Total	55.3	74.8	35.1	57.4		70.9	86.6	49.7	77.1	

* Residual refers to particle mass that is not apportioned to the PMF-resolved factors, i.e. the total mass in the residual matrix.

for vehicle exhaust, aged sea salt, crustal soil, secondary nitrate, and coal combustion / biomass burning are comparable with each emission source category responsible for ~12-15% of the PM_{10} mass. Contributions from residual oil, fresh sea salt, crustal soil, and trace metals are generally below 5%.

Seasonal and spatial variability in the SCEs are presented in Figure B-3. The results in both Table B-4 and Figure B-3 show that for sources such as vehicle exhaust, residual oil, fresh sea salt, and aged sea salt, the summertime contributions are comparable to or even slightly higher than the wintertime contributions. The weak seasonal dependence and the general lack of emission sources in the summertime upwind direction (i.e. SE/SW) leads to these sources being characterized as local. On the other hand, general stations SCEs for the crustal/soil, secondary sulfate, secondary nitrate, trace metals and coal combustion / biomass burning, emission source categories were 3-5 times higher in winter than in summer. With prevailing surface winds from NW/NE in winter bringing air masses from the PRD and regions further inland to Hong Kong, these sources are presumed to be non-local (i.e. from the PRD and larger regional scales). The assignment of emission source categories as being predominantly local or non-local is summarized in Table B-3. Sources that were assigned as local based on seasonal behavior also show relatively large spatial variation across the network. In particular, the vehicle exhaust SCE at MK is nearly 2.5 times that at the general stations. Non-local sources, such as crustal soil, secondary sulfate, secondary nitrate, and trace metals, and coal combustion / biomass burning, show relatively consistent SCEs across the network. These spatial patterns are another indicator for the geographical nature of each emission source category. However, when the general stations are ordered from northwest to southeast (Figure B-3), a spatial gradient in the SCEs is

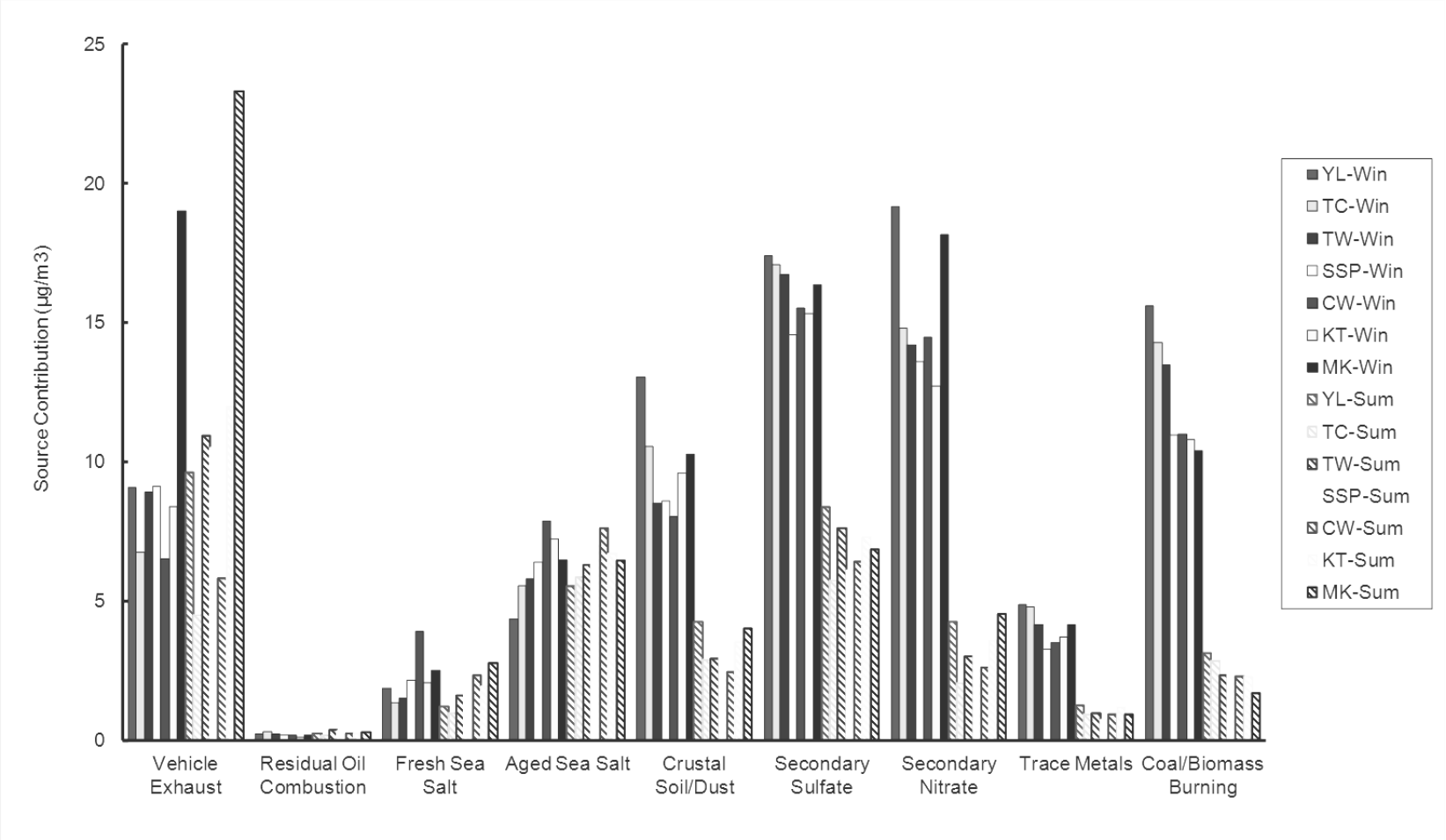


Figure B-3. Site-specific PMF-resolved summer (XX-Sum) and winter (XX-Win) source contribution estimates (µg/m^3) across the network, with general stations (YL to KT) from northwest to southeast (i.e., in the order of increasing distance from PRD region).

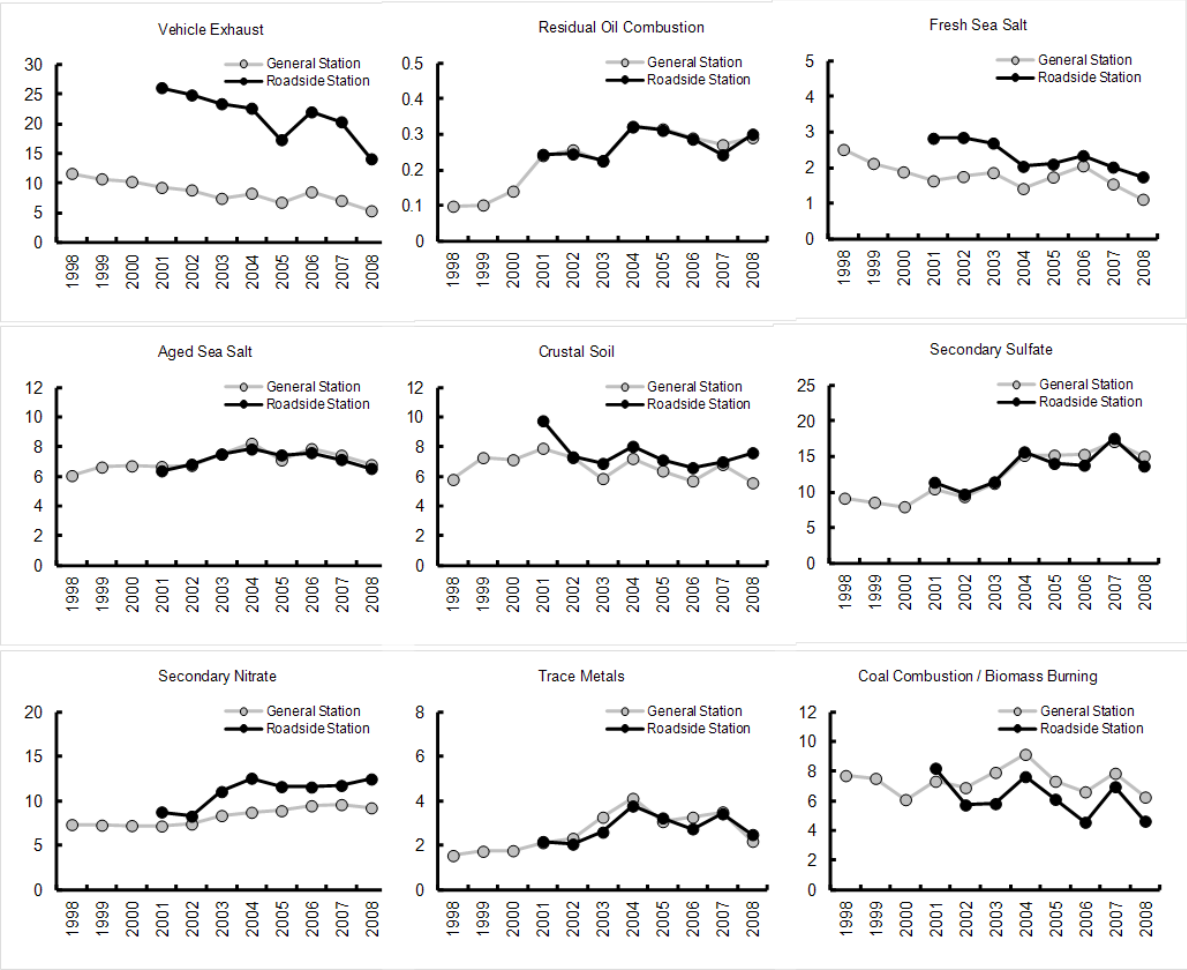


Figure B-4. PMF-resolved annual average source contribution estimates ($\mu g/m^3$) at general stations and roadside station for the period 1998-2008.

discernible for most of the non-local sources with SCEs decreasing with increasing distance from the PRD. This pattern suggests that some of the non-local sources may be sufficiently close to Hong Kong to result in a spatial gradient in source contributions across the network. In contrast, such spatial gradients were not observed for the local sources.

Having classified the geographic nature of each emission source category, the relative contributions of local and non-local sources to ambient PM_{10} levels observed in Hong Kong were calculated. For the general stations, non-local sources contributed about 67% and local sources contributed about 32% to the grand-average PM_{10} during the period from 1998 to 2008. Non-

local contributions exhibited relatively higher contributions in winter (77%) compared to summer (48%). At the roadside station, non-local sources contributed about 58% and local sources contributed about 44% to the PM_{10} burdens from 2001 to 2008 (there was a 2% over prediction in the estimates). Non-local contributions were 69% in winter and 36% in summer.

B.4.2. Long-term (eleven-year) trends in annual source contribution estimates

Inter-annual trends in SCEs over the eleven-year study period were examined for the general stations and the roadside station. Figure B-4 shows the trends in annual average SCEs for each of the nine resolved factors. Vehicle exhaust annual-average SCEs monotonically decreased over the six year period from 1998 to 2003 at the general stations with similar behavior at the roadside station for 2001 to 2003. Despite small rebounds in 2004 and 2006, the vehicle exhaust annual average SCEs at the general stations continued to decrease through 2008. This decreasing trend demonstrates the efficacy of vehicular emission control measures adopted in Hong Kong. For the general stations, the annual average contributions from vehicles were ~5$\mu g/m^3$ in 2008 which corresponds to a 54% reduction from the contribution of nearly 12 $\mu g/m^3$ in 1998. Reasons for the modest SCE increases in 2004 and 2006 are not known; they might arise from year-to-year variations in atmospheric stability that affect pollutant dispersion or from increases in emissions in response to economic changes or other activity-based drivers. For example, over the period 1998-2007 the Hong Kong Gross Domestic Product was lowest in 2003 (HKEPD, 2008a) and thus 2004 was a rebound year in economic activity. Also, the HKSAR PM emission inventory shows a monotonic decrease in annual PM emission road transport emissions from 1998 to 2007 with annual reductions averaging ~300 tonnes/year (HKEPD, 2008b). However, the smallest emission reduction was in 2006 (50 tonnes/year) which coincides with one of the years with a modest increase in the motor vehicle exhaust SCE.

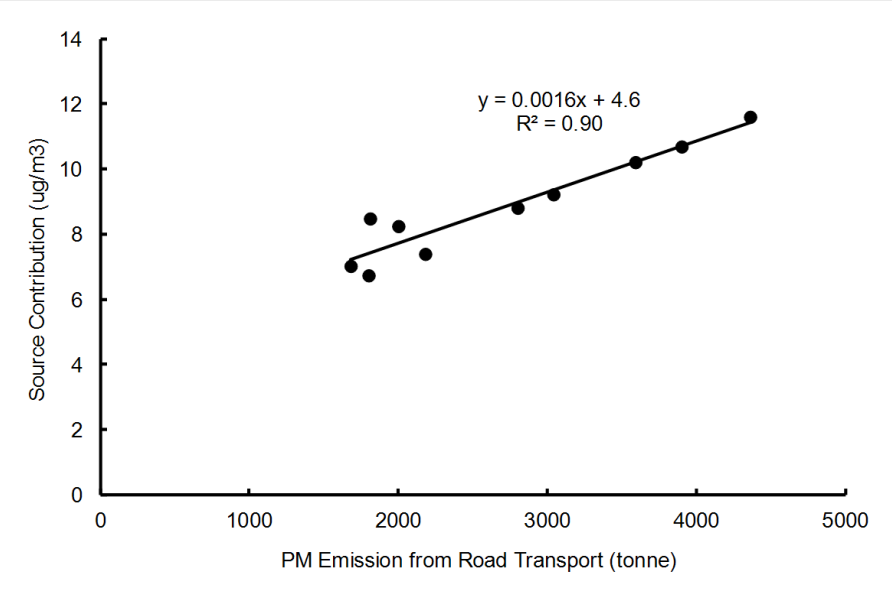


Figure B-5. Correlation between the PMF-resolved vehicle exhaust source contribution estimates at general stations and estimation of PM emissions from road transport. The emission inventory data are from HKEPD (2008b). The data were fitted with an ordinary least squares regression.

Figure B-5 shows that for the years from 1998 to 2007, there is a high correlation between PMF-modeled vehicle SCEs at the general stations and the annual PM emissions from road transport estimated by HKEPD (2008b). The correlation coefficient of 0.90 increases to 0.99 if data from 2004 and 2006 are excluded (the emission inventory was not available for 2008). This high correlation adds confidence to the assignment of the motor vehicle factor and its long term trend. When the PM emissions from road transport decreased by 1000 tonnes per year, the annual average urban-scale ambient PM_{10} burden in Hong Kong decreased by about 1.6 $\mu g/m^3$. The correlation coefficient at roadside MK station (0.70) is lower, likely because the emission variations near the single roadside station respond differently to the emission control programs compared to the greater Hong Kong region. The large intercept for the ordinary least squares regression (4.6 $\mu g/m^3$) may result from an underestimation of Hong Kong road transport emission inventory, non-local motor vehicle contributions to the resolved factor, or admixing of contributions from other largely carbonaceous source categories into the factor assigned to

vehicle exhaust. These different scenarios do not alter the interpretation of the long-term trend but do affect the total PM_{10} SCE assigned to local vehicular exhaust and thus additional work is needed to further refine the motor vehicle contribution estimates.

In contrast to the decreasing trend for vehicle exhaust, annual average SCEs for residual oil increased from 1998 to 2004 and has been generally constant thereafter. In terms of relative changes the residual oil factor increased more than any other factor over the eleven year period. For the general stations, the residual oil factor annual average SCE was 0.3 $\mu g/m^3$ in 2008 which is threefold higher than the contribution of 0.1 $\mu g/m^3$ in 1998. The relatively high loading of sulfate onto this factor (Table BS-2) is not surprising. In Hong Kong, residual oil is mainly used as marine vessel fuel and oceangoing cargo vessels are legally allowed to use fuels with sulfur content up to 4.5%. In contrast, the sulfur content of motor diesel and petrol/gasoline dispensed in Hong Kong is currently capped at 0.005% (Gall and Van Rafelghem, 2006). While there are numerous marine vessel lanes throughout the navigable waters near Hong Kong and the greater PRD, the highest summertime residual oil combustion SCEs are observed at TW followed by SSP and MK (Figure B-3). This spatial pattern is consistent with significant emissions from the high density of container terminal operations located south and southwest of these three stations (Figure B-2).

Sea salt factors were classified as fresh or aged with the fresh being rich in chloride and the aged being depleted in chloride and enriched in sulfate and nitrate. Figure B-4 demonstrates that over the eleven-year period, fresh sea salt contributions to PM_{10} modestly decreased (2.0 $\mu g/m^3$ for the five year period starting 1998 and 1.6 $\mu g/m^3$ for the five year period starting 2004) while aged sea salt contributions increased (6.6 $\mu g/m^3$ for the five year period starting 1998 and 7.4

$\mu g/m^3$ for the five year period starting 2004). The annual average sodium ion concentrations for the general stations data were quite stable over the eleven-year period with a maximum deviation from the grand mean being 11%; this local maximum, observed in 2006, is reflected in the aged sea salt SCE time series. Similarly, sodium ion apportioned to the combined sea salt factors was stable over the study period with a maximum annual deviation from the grand mean being 12%. Thus, the sea salt SCE patterns indicate negligible change in the sea salt emissions but an increase over time in the atmospheric processing of sea salt, perhaps in response to increased transport of inorganic PM from non-local sources as described below. This behavior drives an increase in overall sea salt SCEs because sulfate and nitrate have higher molecular weight than the chloride they are replacing.

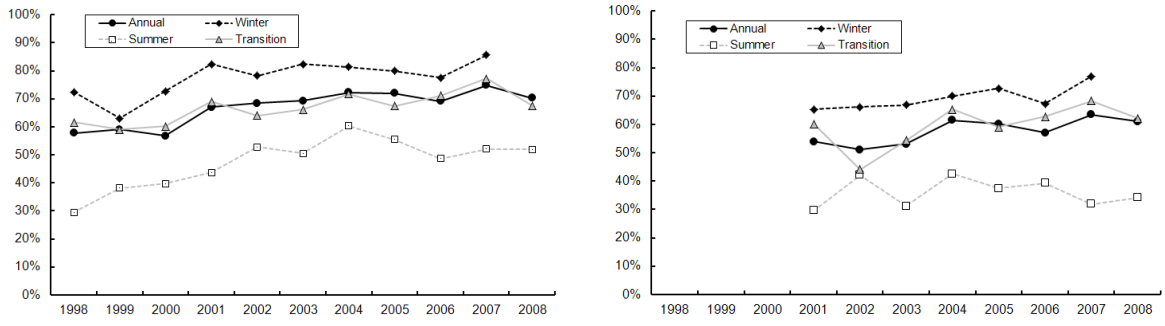
Annual-average SCEs for the crustal soil / dust factor were relatively constant over the study period. Both spatial and temporal patterns suggest this factor has non-local origins. In addition to re-suspended material, this factor might have contributions from industrial sources. For example, a 2006 emission inventory for the PRD assigned more than half of the primary PM_{10} emissions to industrial sources with the dominant sub-categories being nonmetallic mineral products (He et al. 2011). This material might have composition similar to soil/dust.

Both secondary sulfate and secondary nitrate SCEs increased from 1998 to 2004 and have been relatively constant thereafter through 2008. Secondary sulfate SCEs for the general stations increased from 9 $\mu g/m^3$ in 1998 to 17 $\mu g/m^3$ in 2007, corresponding to a ~90% increase. Meanwhile, secondary nitrate SCEs at the general stations increased from 7 $\mu g/m^3$ in 1998 to more than 9 $\mu g/m^3$ in 2007, corresponding to a ~30% increase. Figure B-3 demonstrates that the spatial variability in sulfate is relatively small across the network with a modest gradient of decreasing SCEs from the north to the south. These spatial patterns and the large seasonal

differences strongly implicate non-local sources as the dominant contributors to secondary sulfate in Hong Kong. Patterns for nitrate are generally similar to sulfate for the general stations; the behavior of nitrate at the MK roadside station is discussed in section B.3.3.

Contributions from the trace metals factor increased from 1998 to 2004 and have been relatively constant thereafter (Figure B-4). The local maxima in 2004 and 2007 track the secondary sulfate factor SCE and thus might result from inter-annual variability in synoptic transport patterns. This factor could represent a broad range of industrial emission sources. A 2006 emission inventory for the PRD assigned 55% of the primary PM_{10} emissions to industrial sources with the dominant sub-categories being nonmetallic mineral products followed by pulp and paper industry and light manufacturing (He et al. 2011). It is possible that the PM_{10} mass loading onto the trace metals factor has substantial contributions from other industrial sources, such as those mentioned, that are located in the same region as the metals-emitting sources and are admixed into the factor.

There was no discernible trend for the SCEs of coal combustion / biomass burning factor over the period 1998-2008 with annual average contributions of 6.6 $\mu g/m^3$ for the five year period starting 1998 and 6.5 $\mu g/m^3$ for the five year period starting 2004. Similar to the secondary sulfate and trace metals factors, there were local maxima in 2004 and 2007 that might reflect differences in synoptic transport patterns. PRD biomass burning $PM_{2.5}$ emissions, which are dominated by domestic biofuel burning and field burning of rice straw, decreased by ~20% over the period 2003-2007 (He et al. 2011). However, a 2006 emission inventory assigned only 2% of the PRD primary PM_{10} emissions to biomass burning (Zheng et al., 2009b) and this change might be too small to detect in the ambient data. The same study allocated 21% of the PRD



(a) (b)

Figure B-6. Long-term trends of relative contributions from non-local sources in winter, summer, and transitional seasons, and for all sampling days: (a) general stations; and (b) at roadside station.

primary PM_{10} emissions to power plants. Emissions from industrial combustion sources could also be admixed into this factor.

Figure B-6 shows the long-term trend of relative contributions from non-local sources for winter, summer, transition seasons, and all sampling days. The contribution of non-local sources at the general stations increased from 58% in 1998 to 72% in 2004 (~7 μg/m^3 increase) and has been relatively constant thereafter albeit with some year-to-year variability. This trend arises from the net effect of the decreasing contributions from vehicle emissions, driven by emission controls, being more than offset by the increase in contributions from non-local sources such as secondary sulfate, secondary nitrate, and trace metals. Percentage contributions from non-local sources are highest during the wintertime and increased from 72% in the winter of 1998-1999 to 86% in the winter of 2007-2008. In the latter case, contributions from local sources are estimated to be only 1/7 of the wintertime ambient PM_{10} and this demonstrates the increased importance of controlling sources from the PRD and beyond to improve air quality conditions in Hong Kong.

B.4.3. General station *vs.* Roadside station

As expected, contributions from vehicle exhaust are much greater at the roadside stations compared to the general stations with grand average (1998-2008) vehicle exhaust SCEs of 8.5 $\mu g/m^3$ for the general stations and 21.3 $\mu g/m^3$ at the MK roadside station (Table B-4). Figure B-4 and Table B-4 demonstrate relatively similar SCEs between the general stations and the roadside station for all remaining emission source categories except fresh sea salt, secondary nitrate, and coal combustion / biomass burning. Elevated SCEs for fresh seas salt at MK results from its geographical location in the center of Kowloon Peninsula; the station is relatively close to the sea compared with those stations in the New Territories. Indeed, Figure B-3 shows that the fresh sea salt SCE at MK is quite similar to the SCEs at the nearby SSP and KT stations. Secondary nitrate is also elevated at the roadside station compared to the composited general stations, indicating that apart from non-local contribution, local contribution cannot be neglected at roadside. The tall building and narrow roads in urban Hong Kong prevent the vehicle-emitting pollutants from dispersing, leading to a "street canyon effect" that makes nitrogen oxides have sufficient time to accumulate and oxidize to form secondary nitrate (Yim et al., 2009). The coal combustion / biomass burning SCE is lower at the roadside station compared to the general stations. In this case, there are significant SCE gradients across the network with higher SCEs to the north and lower SCEs to the south; the roadside station SCE is comparable to SCEs for the nearby SSP and KT stations. Thus, with the exception of vehicle exhaust and secondary nitrate, both local and non-local sources impact the general stations and roadside stations in similar ways despite differences in the sampling environment including the sampling height.

B.4.4. Source characteristics during high PM days

Synoptic weather conditions are an important driver in the occurrence of pollution episodes. They either stabilize the boundary layer that degrade air ventilation, or favor pollutant transport from high emission areas. Adverse health effects can arise from both chronic and acute exposures to PM (Wang et al., 2002). Thus, there is motivation to not only reduce average PM levels but also to "shave the peaks" off the highest PM concentration days. For example, in the United States these objectives are operationalized through having both annual and 24-hour standards for $PM_{2.5}$. This issue motivates the evaluation of whether certain emission source categories have disproportionately high contributions on high PM days compared to average behavior.

In this study, we set the threshold value for high PM to be 100 $\mu g/m^3$, the value of World Health Organization (WHO)'s interim target-2 and the proposed air quality objective / index for 24-hour PM_{10} in Hong Kong. The threshold criterion of 100 $\mu g/m^3$ yields a total of 408 high PM sample-days at the seven sites from 1998 to 2008. The number of days each year exceeding the PM_{10} mass measured at each station is listed in Table B-5. For reference, a 1-in-6 day sampling schedule would have 61 sampling days per year for 100% data completeness. While the annual average PM_{10} mass has remained relatively constant over the eleven year study period (Figure B-1), the frequency of high PM days at the general stations nearly doubled for the five year period starting 2004 compared to the five year period starting 1998.

Figure B-7 shows the frequencies of high PM days (numerical values provided at the bottom of Table B-5) and the average source contributions by the nine identified sources at individual stations on high PM days during 1998-2008. Apart from the roadside station, the northern most

Table B-5. Year-specific number of high PM days (nominally 61 sample days per year). Asterisks denote years with significant data completeness gaps (Table B-1).

No. high PM days	YL	TC	TW	SSP	CW	KT	MK
1998	11		4	4	3	2*	
1999	5	0*	7	2	3	3	
2000	7	5	3	5	2	1	
2001	8	3	2	1	3*	2*	10
2002	4	2	2	2	0*	4*	5
2003	8	8	4	3*	4	2	8
2004	15	5	8	1*	2	3	15
2005	15	3	7	7	9	6	11
2006	11	4	7	7	6	5	9
2007	10	10	7	3	6	8	17
2008	10	7	7	1	4	2	8
High PM Days	104	47	58	36	42	38	83
Total Days	654	575	631	586	641	598	482
% High PM Days	15.9%	8.2%	9.2%	6.1%	6.6%	6.4%	17.2%

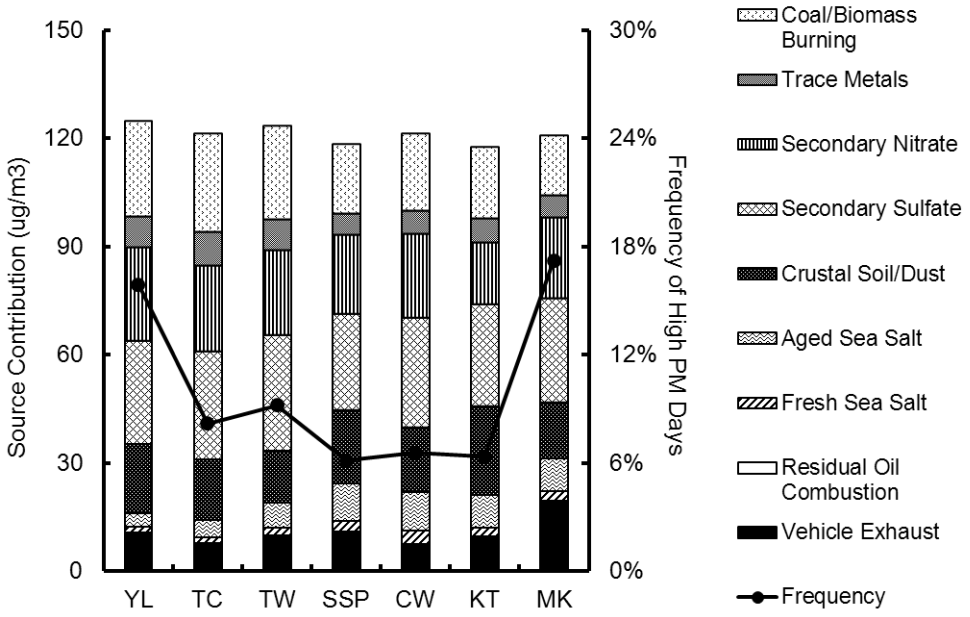


Figure B-7. Frequencies of the high PM days by site (i.e., the number high PM samples divided by the total number of collected samples at each station) and the PMF-resolved source contribution estimates averaged over the high PM days.

general station in the network, YL had the highest frequency of high PM days (16%). SSP, KT and CW, located in the southern portion of the HKSAR, had the lowest frequency of high PM days (6-7%). Overall, the frequency of high PM days in Hong Kong was 10% during 1998-2008.

The average mass concentrations on high PM days were similar among all stations (Figure B-7), with a variation of only 2%, despite the stations being distinct in their local environments (i.e., residential, commercial, industrial, roadside, or near container terminal). The highest average mass concentration (125 $\mu g/m^3$) occurred at YL and the lowest average mass concentration (118 $\mu g/m^3$) occurred at KT. Both the highest frequency and the highest mass concentration occurred at YL in the northwest, while both the lowest frequencies and lowest mass concentrations occurred at CW, TC and KT in the southeast. The narrow range of average concentrations on high PM days and the concentration gradient from northwest to southeast suggests that the non-local PM sources, especially in areas to the north of Hong Kong including the PRD, are significant contributors on high PM days.

The source contributions on high PM days were compared with the average values over the 1998-2008 study period to identify the sources that had elevated contributions on the high PM days. Figure B-8 plots the ratios of average source contributions on high PM days to that for the entire study period at individual stations. Mass contributions from the three local emissions source categories - vehicle exhaust, fresh and aged sea salt - did not increase significantly on high PM days. While the residual oil combustion (the remaining local source) contribution on high PM days was at most twice the study average, the absolute value itself is less than 1 $\mu g/m^3$ and thus is not a major contributor to the total mass on high PM days. In contrast, contributions from crustal soil, secondary sulfate, secondary nitrate, trace metals and coal combustion /

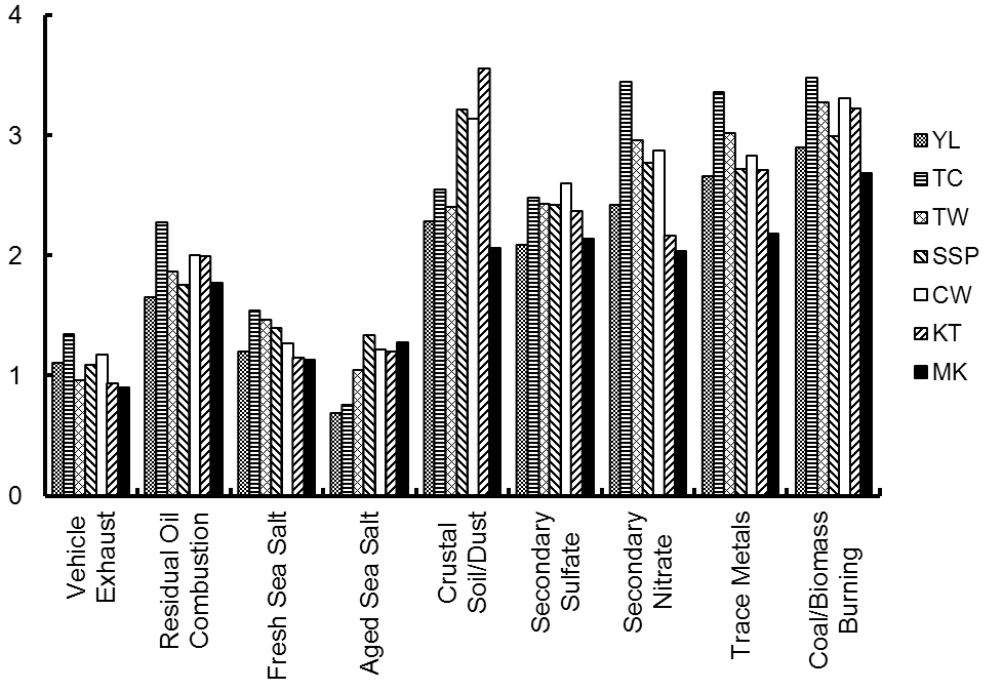


Figure B-8. Ratios of average source contribution estimates on high PM days to the source contribution estimates averaged over all samples.

biomass burning were on average two to three times higher on high PM days. These observations indicate that pollutant transport from outside Hong Kong is disproportionately responsible for high PM days rather than accumulation of local pollutants due to unfavorable atmospheric dispersion over the region.

B.5. Conclusions and policy implications

PM_{10} chemical speciation network data for Hong Kong collected from 1998 to 2008 was used to apportion PM_{10} mass to emission source categories. Spatial and seasonal patterns in the PMF-resolved source contribution estimates were used to classify sources as local (emissions largely from within the HKSAR) or non-local (emissions largely from the PRD region and more distant sources). The time series of annual-average SCEs were examined for each factor. There is strong evidence for the effectiveness of local control measures on vehicle exhaust with

contributions at the general stations decreasing more than 50% between 1998 and 2008. Most of this decrease occurred prior to 2004 which is consistent with the timeline for implementation of measures to reduce vehicle emissions (HKEPD, 2011). During this period, the PMF-modeled annual average source contribution estimates for vehicle exhaust decreased linearly with the HKEPD road transport PM emission inventory. Vehicle-related emission control strategies implemented over the previous fifteen years has resulted in a substantial positive impact by decreasing ambient burdens of – and thus exposures to – vehicle exhaust. In contrast to the late 1990s, vehicle exhaust is no longer the dominant emission source category contributing to PM_{10} burdens at the general stations, with its contribution decreased from 22% in 1998 to 10% in 2008.

The PM_{10} air quality improvements from vehicle emissions controls have been offset by increased contributions from non-local sources which increased from being 58% of the PM_{10} burden in 1998 to 70% in 2008. In wintertime, the season with persistently high levels of PM air pollution, the relative contribution from non-local sources is even higher, increasing from 72% in the winter of 1998-1999 to 86% in the winter of 2007-2008. If current air quality conditions are maintained then on average the existing monitoring stations in Hong Kong will violate the proposed 24-hour PM_{10} AQO of 100 $\mu g/m^3$ on about 10% of days. Future air quality conditions in Hong Kong will depend upon cooperation from provinces in mainland China. Such cooperation is already underway with Guangdong Province and perhaps a coordinated, regional-scale air quality management plan will best serve both the Hong Kong and neighboring provinces.

Source apportionment modeling suggests that PM from residual oil combustion has increased over the study period. These emissions are attributed to marine vessels and this increase is

generally consistent with increased marine activity in the greater PRD region over the past decade. While the absolute PM mass impacts from this source category appear to be relatively small, residual oil combustion emissions do warrant further attention because these emissions tend to be rich in air toxics metals such as nickel (Lau et al., 2003). From a public health standpoint there is substantial motivation to better characterize marine-related activities/emissions and consider the implementation of air quality management plans targeting this sector such as near-shore fuel switching or more stringent fuel quality specifications. Such programs will require cooperation with Guangdong to be most effective.

Residual oil combustion is not the only source of air toxics metals in Hong Kong. For example, As, Cd, and Pb loaded onto the coal combustion / biomass burning factor and these same elements plus Mn loaded onto the trace metals factor. These metals can have several adverse effects such as acting as biochemical catalysts to cause severe molecular damage and induction of biochemical synthesis pathways (Halliwell and Gutteridge 1999). Ambient levels of air toxics metals should be evaluated from a risk perspective to assess whether targeted controls are warranted.

The prominent role that emissions from sources outside the HKSAR exert on Hong Kong PM air quality highlights the need for regional air quality planning and management. Hong Kong and Guangdong have already taken a pioneering role in establishing China's first Regional Air Quality Monitoring Network. The monitoring results are released on the web as a Regional Air Quality Index (GDEPB, 2012). Since March 2012, Hong Kong and Guangdong has started releasing the hourly pollutant concentration data, including $PM_{2.5}$, to the public. This is the first city cluster in China to regularly release hourly pollutant concentration data. In recent years, a series of control measures were adopted in Guangdong to reduce pollutant emissions. Measures

included requiring all large-scale thermal power plants to install and operate flue gas desulphurization units and to install continuous emissions monitoring systems with real-time on-line access by local authorities, closing down small power plants and other major polluting industries (e.g., cement plants and iron and steel plants with low production capacity), restricting the growth of the motorcycle populations in key cities, requiring newly registered motor vehicles in the PRD to comply with the National III standards (on a par with the Euro III standards), supplying the National III standard motor fuels, increasing use of nuclear power and natural gas, etc. These actions are laudable, yet additional measures will be needed to reverse the trend of deteriorating air quality in Hong Kong and the greater PRD region.

Although emissions from non-local sources have increased in relative importance, vehicle emissions remain among the most dominant source categories at the roadside station and further vehicle emission reductions within Hong Kong would be beneficial. For example, Lau et al. (2007) used SO_2 as a surrogate to classify daily air pollution burdens in 2006 as being dominated by local or regional sources. They estimated that Hong Kong was predominantly affected mainly by regional sources on 132 days (or 36% of the time) and by local sources on 192 days (53% of the time). The findings were significant, as it showed that people in Hong Kong were exposed to local pollution for longer periods of time over a year, hence measures that target local emissions would also greatly reduce public's exposure to air pollutants and the health risks associated with them. In addition, advances in the performance of internal combustion engines often results in more ultrafine particles which may have significant health impacts (Delfino et al., 2005). The European Union is now looking into the prospect of regulating ultrafine particles starting in 2013 (ECFA, 2011). These issues should be considered when planning the next round

of vehicle emissions control and may motivate interest to convert the existing vehicle fleet to low- or no-emission models to comprehensively address the vehicle emissions problem.

B.6. Acknowledgements

We acknowledge the support from Hong Kong Environmental Protection Department (Tender Ref AS09-056). This work was also sponsored by the Research Grant Council of Hong Kong, China (615406), the University Grant Committee (GMGS07/08.EG03) and the Fok Ying Tung Graduate School (NRC06/07.SC01).

B.7.Disclaimer

The content of this chapter does not necessarily reflect the views and policies of the Hong Kong Special Administrative Region (HKSAR) Government, nor does mention of trade names or commercial products constitute an endorsement or recommendation of their use.

B.8 References

Andreae, M.O., Schmid, O., Yang, H., Chand, D., Yu, J.Z., Zeng, L.M. and Zhang, Y.H.; Optical properties and chemical composition of the atmospheric aerosol in urban Guangzhou, China, *Atmospheric Environment*, **2008**, 42, 6335-6342.

Chow, J.C., Watson, J.G., Crow, D., Lowenthal, D.H. and Merrifield, T.; Comparison of IMPROVE and NIOSH carbon measurements, *Aerosol Science and Technology*, **2001**, 34, 23-34.

Delfino, R.J., Sioutas, C. and Malik, S.; Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health, *Environmental Health Perspective*, **2005**, 113, 934-946.

European Federation of Clean Air and Environmental Protection Associations (ECFA), **2011**. ECFA Newsletter, 12, available at http://efca.net/uploads/file/EFCA%20Newsletter%20nr%2012.pdf

Gall, C. and Van Rafelghem, M.; **2006**. Marine emission reduction options for Hong Kong and the Pearl River Delta (available at: http://www.civic-exchange.org/publications/2006/marineemission-e.pdf)

Guangdong Environmental Protection Bureau (GDEPB), **2012**. Pearl River Delta Regional Air Quality Index (http://www-app.gdepb.gov.cn/raqi3/RAQI_en.htm, accessed March 31, 2012).

Halliwell, B. and Gutteridge, J.M.C.. Free radicals in Biology and Medicine, **1999**, Oxford University Press, Oxford, United Kingdom.

He, M., Zheng, J.Y., Yin, S.S. and Zhang, L.J.; Trends, temporal and spatial characteristics, and uncertainties in biomass burning emissions in the Pearl River Delta, China, *Atmospheric Environment*, **2011** 45, 4051-5059.

Ho, K.F., Lee, S.C., Chan, C.K., Yu, J.C., Chow, J.C. and Yao, X.H.; Characterization of chemical species in $PM_{2.5}$ and PM_{10} aerosols in Hong Kong, *Atmospheric Environment*, **2003**, 37, 31-39.

Hong Kong Environmental Protection Department (HKEPD), **2008a**. Changes in Emissions Relative to Economic Growth, available at http://www.epd.gov.hk/epd/english/environmentinhk/air/data/emission_inve.html#1, accessed March 31, 2012.

Hong Kong Environmental Protection Department (HKEPD), **2008b**. Hong Kong Air Pollutant Emission Inventory, available at http://www.epd.gov.hk/epd/english/environmentinhk/air/data/emission_inve.html#1, accessed March 31, 2012.

Hong Kong Environmental Protection Department (HKEPD), **2011**. Clearing the Air at Street Level, available at (http://www.epd.gov.hk/epd/english/environmentinhk/air/prob_solutions/cleaning_air_atroad.html, accessed March 31, 2012)

Hong Kong Observatory (HKO); Number of Hours of Reduced Visibility, **2012**, available at http://www.hko.gov.hk/cis/statistic/hko_redvis_statistic_e.htm accessed Dec 8, 2013)

Huang, X.F., Yu, J.Z., Yuan, Z.B., Lau, A.K.H. and Louie, P.K.K.; Source analysis of high particulate days in Hong Kong, *Atmospheric Environment*, **2009**, 43, 1196-1203.

Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C. and Prevot, A.S.H.; Source apportionment of submicron organic aerosols at an urban site by factor analytical modeling of aerosol mass spectra, *Atmospheric Chemistry & Physics*, **2007**, 7, 1503-1522.

Larson, R.K. and Baker, J.E.; Source apportionment of polycyclic aromatic hydrocabons in the urban atmosphere: A comparison of three methods, *Environmental Science & Technology*, **2003**, 37, 1873-1881.

Lau, A., Lo, A., Gray, J., Yuan, Z.B. and Loh, C.;. Relative significance of local vs. regional sources: Hong Kong's air pollution, **2007**, available at (http://www.civic-exchange.org/eng/upload/files/200703_HKAirPollution.pdf accessed Dec 8 2013).

Lau, A.K.H., Yu, J.Z., Wong, T.W., Yu, I.T.S. and Poore, M.W.; Assessment of toxic air pollutant measurement in Hong Kong, **2003**, available at (http://www.epd.gov.hk/epd/english/environmentinhk/air/studyrpts/assessment_of_tap_measurements.html accessed Dec 8, 2013)

Lau, A.K.H., Yuan, Z.B., Yu, J.Z. and Louie, P.K.K.; Source apportionment of ambient volatile organic compounds in Hong Kong, *Science of the Total Environment*, **2010**, 408, 4138-4149.

Lee, E., Chan, C.K. and Paatero, P.; Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, *Atmospheric Environment*, **1999**, 33, 3201-3212.

Lee, P.K.H., Brook, J.R., Dabek-Zlotorzynsk, E. and Mabury, S.A.; Identification of the major sources contributing to $PM_{2.5}$ observed in Toronto, *Environmental Science & Technology*, **2003**, 37, 4831-4840.

Polissar, A.V., Hopke, P.K., Paatero, P., Malm, W.C. and Sisler, J.F.; Atmospheric aerosol over Alaska 2. Elemental composition and sources, *Journal of Geophysical Research – Atmospheres*, **1998**, 103, 19045-19057.

Song, Y., Shao, M., Liu, Y., Lu, S.H., Kuster, W., Goldan, P. and Xie, S.D.; Source apportionment of ambient volatile organic compounds in Beijing, *Environmental Science & Technology*, **2007**, 41, 4348-4353.

Thurston, G.D. and Spengler, J.D.; A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment*, **1985**, 19, 9-25.

U.S. Environmental Protection Agency (USEPA); EPA Unmix 6.0 Fundamentals and User Guide, **2007**, available at http://www.epa.gov/heasd/products/unmix/unmix.html, accessed March 31, 2012.

U.S. Environmental Protection Agency (USEPA); EPA PMF 3.0 Fundamentals and User Guide, **2008**, available at http://www.epa.gov/heasd/products/pmf/pmf.html, accessed March 31, 2012.

Wang, T.W., Tam, W.S., Yu, T.S. and Wong, A.H.S.; Associations between daily mortalities from respiratory and cardiovascular diseases and air pollution in Hong Kong, China, *Occupational Environmental Medicine*, **2002**, 59, 30-35.

World Health Organization (WHO); WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide, summary of risk assessment, **2006**, available at (http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_eng.pdf accessed on Dec 8, 2013)

Yim, S.H.L., Fung, J.C.H., Lau, A.K.H and Kot, S.C.; Air ventilation impacts of the "wall effect" resulting from the alignment of high-rise building, *Atmospheric Environment*, **2009**, 43, 4982-4994.

Yuan, Z.B., Lau, A.K.H., Zhang, H.Y., Yu, J.Z., Louie, P.K.K. and Fung, J.C.H.; Identification and spatiotemporal variations of dominant PM_{10} sources over Hong Kong, *Atmospheric Environment*, **2006**, 40, 1803-1815.

Zheng, J.Y., Shao, M., Che, W.W., Zhang, L.J., Zhong, L.J., Zhang, Y.H. and Streets, D.; Speciated VOC emission inventory and spatial patterns of ozone formation potential in the Pearl River Delta, China, *Environmental Science and Technology*, **2009a**, 43, 8580-8586.

Zheng, J.Y., Zhang, L.J., Che, W.W., Zheng, Z.Y. and Yin, S.S.; A highly resolved temporal and spatial air pollutant emission inventory for the Pearl River Delta region, China and its uncertainty assessment, *Atmospheric Environment*, **2009b**, 43, 5112-5122.

B.S. Supplemental Information for Appendix B:" Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies – A case study in Hong Kong over a decade (1998-2008)".

B.S.1. PM_{10} monitoring and chemical speciation in Hong Kong

In this study, PM_{10} samples were collected for 24 hours onto 8×10 in^2 quartz fiber filters (without any pre-treatment) using High Volume PM_{10} samplers operating at the flow rate of 1.13 m^3/min. Chemical analyses were performed on sections taken from the quartz fiber filter. PM carbon was measured using thermal / optical transmittance to obtain organic carbon (OC) and elemental carbon (EC) mass loadings. The analysis protocol conformed to NIOSH Method 5040 (Birch et al., 1996) and the detailed analytical procedure was described in Sin et al. (2002). The OC sampling artifact issue was discussed in Yu et al. (2004). PM elements were quantified using USEPA Method IO-3 in conjunction with two analytical methods (USEPA, 1999). Inductively coupled plasma atomic emission spectroscopy (ICP-AES) was used for Al, Ba, Be, Ca, Cd, Cr, Cu, Fe, Mg, Mn, Ni, Pb, V and Zn, while Flow Injection Analysis – Atomic Absorption (FIA-AA) was used for As, Hg and Se. PM ions including SO_4^{2-}, NO_3^-, Cl^-, Br^-, NH_4^+, Na^+ and K^+, were analyzed by extracting portions of the filters with waterand analyzing the extracts by ion chromatography (IC). It is noted that the nitrate data might have large measurement uncertainties because the sampling technique - only a quartz filter and with no denuder - could lead to negative sampling artifacts from evaporative loss of aerosol nitrate or positive artifacts from gaseous nitric acid retained on the filter through adsorption or reaction with alkaline aerosol constituents. The PM_{10} analyses were subjected to the strict in-house QA/QC procedures of the

Hong Kong Environmental Protection Department (HKEPD, 2003). Typically, the control limits for the PM_{10} measurements are ±10% for both accuracy and precision.

27 chemical components, including As, Be, Cd, Ni, Pb, Cr, Al, Mn, Fe, Ca, Mg, V, Zn, Ba, Cu, Hg, Se, Na^+, K^+, Cl^-, Br^-, SO_4^{2-}, BAP, NH_4^+, NO_3^-, OC, and EC, together with sample mass, were quantified by the aforementioned analytical methods. Following the methodology of Huang et al. (1999), species with more than 50% of samples below detection limit were excluded from the data analyses. Otherwise, values below detection limit were replaced by half of the detection limit. Cu was also excluded because of apparent contamination of a number of samples. All the species were reported in $\mu g/m^3$, except for OC and EC, which were reported in $\mu gC/m^3$. After data filtering, 19 species, including As, Cd, Ni, Pb, Al, Mn, Fe, Ca, Mg, V, Zn, Na^+, K^+, Cl^-, SO_4^{2-}, NH_4^+, NO_3^-, OC, and EC, were retained for the source apportionment analysis.

In order to provide consistent source profiles for the data collected at different stations in the chemical speciation network, the data from all stations were combined to construct a single dataset with 4167 with samples. This approach is reasonable because Hong Kong is a small territory with linear dimension only about 40km by 30km; within the 24-h sampling period, pollutants from any local or remote sources have a fair chance to impact on receptors anywhere within the territory. Hence, measurements from all of the monitoring stations can be considered as independent samples for the same set of sources and be combined in our receptor modeling analyses. This is not the same as combing data sets across distinct airsheds (e.g. data from Hong Kong and Beijing), which should not be done in receptor model analyses. The relative strength of impact on an individual receptor depends on its proximity to the source, the profile for the same source at the different receptors should remain similar if the species are stable over the time scales representative of transport between the stations.

B.S.2. Assessment of measurement error for the PM_{10} speciation dataset

Collocated PM_{10} sampling is routinely performed at two sites in the Hong Kong network – Sham Shui Po (SSP) and Tsuen Wan (TW). Although analysis of the collocated data can capture many, but not all, sources of measurement error, the analysis does provide valuable insights into the data quality by quantifying the aggregate errors from field sampling operations and analytical sensitivity. The analysis adopted the methodology of Hyslop and White (2008a, b) with some deviations to account for differences in the information available. Hyslop and White used field blanks data to estimate "critical limits" for the concentration values and compared these estimates to the method detection limits (MDL) reported for the IMPROVE network. Subsequently, they examined the collocated precision as a function of concentration. From these two analyses, key derived data quality metrics were the critical limits and the collocated precision for the subset of data with concentration values at least three times greater than the method detection limit. In our work, field blanks data were not available. Thus, the method detection limit values provided by HKEPD could not be independently examined. Well-behaved trends between concentrations above MDL and the estimated precision were observed to break down for concentrations below the MDL, with generally poor and fluctuating precision for concentrations below MDL. Hence, sampling days with one or both concentrations below the MDL were excluded from the error structure estimation (additive and multiplicative error terms of uncertainty). SSP collocated data for 1998 to 2008 was used to generate error structures which were then evaluated using the corresponding TW collocated dataset for 1999 to 2008.

$$U_{ij} = a_j + b_j \times C_{ij} \qquad \text{(BS-1)}$$

Following Polissar et al. (1998), the error structure was assumed to have the form shown in equation (BS-1), where U_{ij} = uncertainty for species j in sample i; a_j = additive (constant) error

term for species *j*; b_j = multiplicative error term for species *j*; and C_{ij} = concentration value for species *j* in sample *i*. The first error estimation strategy used the approach of Polissar et al. (1998) which sets the additive error term to MDL/3. The multiplicative error term was first estimated by taking the mean of the binned relative collocated precision for the six bins with highest concentration values. The additive and multiplicative errors for this approach are presented in the "Unweighted" columns of Table BS-1. The regression fits using this approach were poor for most species and thus another approach was taken.

Table BS-1. Additive and multiplicative error terms derived using the SSP collocated data set using three estimation approaches. Coefficient a_j has units μg/m^3.

Species	Unweighted		Weighted		Refined Weighted (Outliers Removed)	
	a_j	b_j	a_j	b_j	a_j	b_j
Al	0.0063	0.095	0.0076	0.022	0.0046	0.071
As	0.0003	0.074	0.0002	0.003	0.0001	0.050
Ca	0.0150	0.092	0.0349	0.002	0.0217	0.043
Cd	0.0001	0.066	0.0001	-0.006	0.00002	0.058
Cl^-	0.0092	0.101	0.0125	0.017	0.0141	0.047
Fe	0.0075	0.054	0.0132	-0.003	0.0044	0.030
K^+	0.1048	0.084	0.0568	-0.037	0.0118	0.028
Mg	0.0075	0.071	0.0076	0.001	0.0049	0.020
Mn	0.0031	0.253	0.0013	-0.025	0.0002	0.029
Na^+	0.0997	0.144	0.0452	0.014	0.0264	0.018
NH_4^+	0.0184	0.037	0.0219	0.014	0.0169	0.028
Ni	0.0010	0.100	0.0007	-0.018	0.0003	0.050
NO_3^-	0.0221	0.057	0.0242	0.034	0.0429	0.026
Pb	0.0030	0.039	0.0020	0.005	0.0002	0.035
SO_4^{2-}	0.0499	0.040	0.1261	0.009	0.0597	0.025
V	0.0030	0.113	0.0012	-0.018	0.00001	0.047
Zn	0.0075	0.072	0.0043	0.002	0.0016	0.032
EC	0.0833	0.052	0.2087	-0.010	0.1148	0.023
OC	0.0833	0.055	0.2634	-0.003	0.1117	0.032

The second approach used a weighted linear regression to estimate both the additive and multiplicative error terms using all fourteen bins. The weight w_{kj} for species *j* and bin *k* was taken as the inverse of square of the absolute collocated precision (σ_k) for bin *k* (dropping the *j* subscripts on each parameter) as shown in equation (BS-2), with the coefficients for the error structure calculated using equations (BS-3) and (BS-4).

$$w_k = 1/\sigma_k^2 \quad \text{(BS-2)}$$

$$a_j = \frac{(\sum w_k {C_k}^2)(\sum w_k \sigma_k) - (\sum w_k C_k)(\sum w_k C_k \sigma_k)}{(\sum w_k)(\sum w_k {C_k}^2) - (\sum w_k C_k)^2} \quad \text{(BS-3)}$$

$$b_j = \frac{(\sum w_k)(\sum w_k C_k \sigma_k) - (\sum w_k C_k)(\sum w_k \sigma_k)}{(\sum w_k)(\sum w_k {C_k}^2) - (\sum w_k C_k)^2} \quad \text{(BS-4)}$$

For most species the weighted error estimates, summarized in the "Weighted" columns of Table BS-1, provided better fit than the unweighted error estimates. The weighted error estimates were further refined by a third approach which included removing individual sample pairs that were subjectively deemed outliers and significantly influenced the collocated precision for a given bin. These results are presented in the "Refined Weighted" columns of Table BS-1 and these values are recommended for the source apportionment modeling.

Collocated data from the TW site were used as an independent check on the modeled error structures. Figure BS-1 shows for several species the binned collocated precision for the TW data (open markers for raw data and shaded markers for data after removing outliers) and the error structures generated from the "Refined Weighted" regression coefficients for the SSP collocated data (dashed lines). In general, the concentration-dependent collocated precision for the TW data is well captured by the SSP-generated error structures. A few species had one or two concentration bins that were not explained by this error structure, for instance the highest

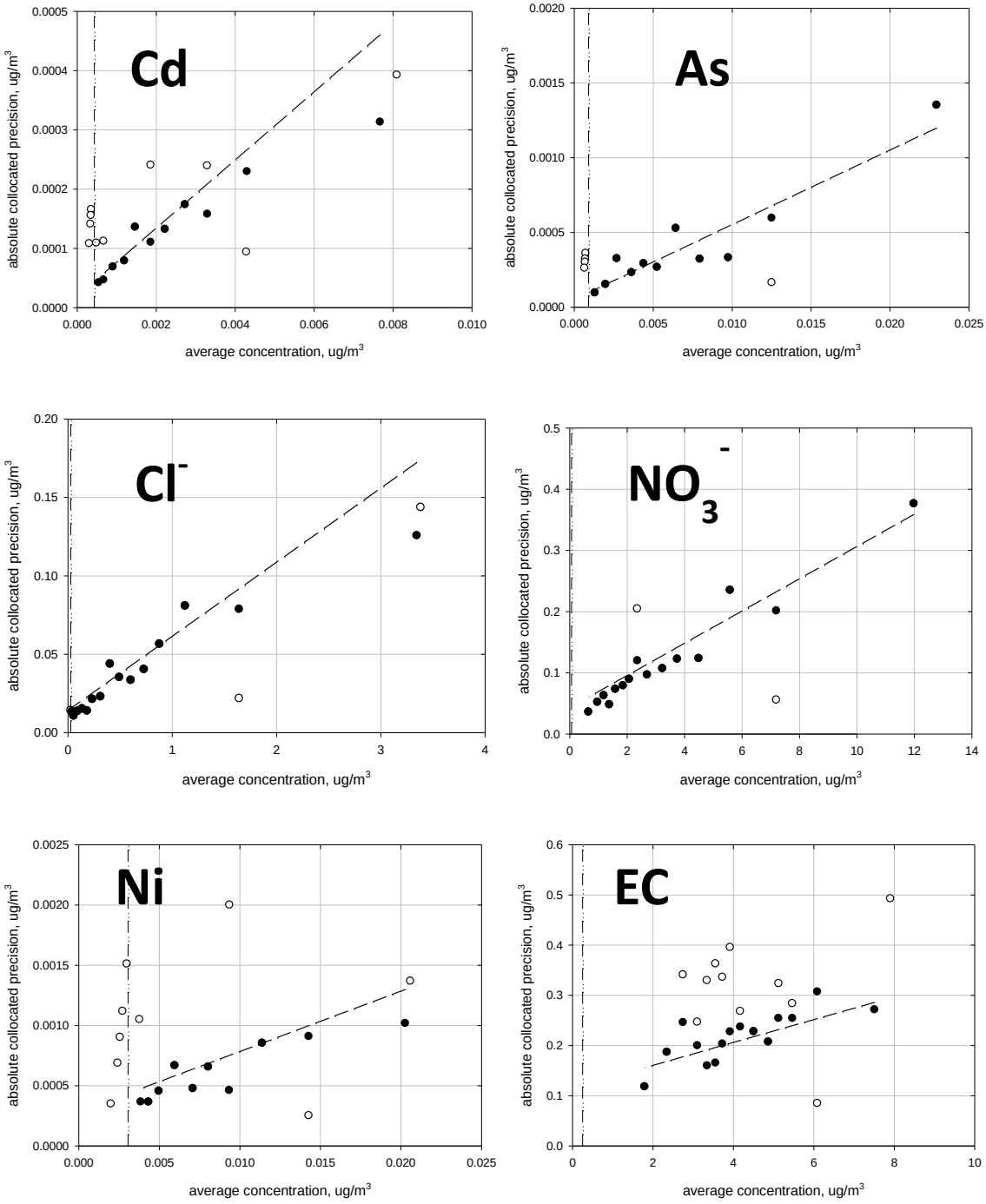


Figure BS-1. Six examples of the binned absolute collocated precision for the TW data set using raw data (open circles) and outliers removed (shaded circles); and the error structures estimated regressions of the SSP collocated data (dashed lines). The dot-do-dash vertical lines are the MDL values.

concentration bins for As (Figure BS-1) were not matched well by the error structure. Overall, these favorable comparisons add confidence to using the "Refined Weighted" error structures of Table BS-1 for source apportionment modeling.

B.S.3. Justification of using quartz filter gravimetric mass (QTM) as the sample total mass

Traditionally, concerns have been raised about the quality of QTM data such as bias arising from the hygroscopic nature of quartz fiber and the possibility of losing fibers during filter handling between pre-weighing and after-weighing. This measurement artifact could lead to the reported gravimetric mass being biased low. However, for particle mass loadings representative of conditions in Hong Kong, this artifact might be relatively small. Given that the sum-of-species approach on average accounts for only 62% of the QTM, a significant portion of the particulate matter mass is not being apportioned and this missing mass could bias the source contribution estimates. Therefore, we conducted an analysis to determine whether it would be more appropriate to apportion the sum-of-species mass or the QTM mass.

In Hong Kong the particulate matter Air Quality Objectives (AQOs) are assessed using data from the network of PM_{10} TEOM monitors operated at 50°C. Thus, from an air quality planning and management perspective, the ideal case would be source apportionment of the aerosol mass as measured by the TEOM. The remainder of this section focuses on comparisons of QTM mass to TEOM mass to evaluate the appropriateness of apportioning QTM as a proxy for the TEOM mass.

For these comparisons the TEOM hourly PM_{10} mass data were conditioned by imputing concentration values for cases with missing data for only one or two consecutive hours. These data gaps were filled by linear interpolation of the concentration values bounding the data gaps.

Subsequently, only those days with 24 hours of TEOM data (after imputing) were included for the comparison to the QTM data.

For reasons noted below, data from the Yuen Long (YL) station were chosen for analysis. Figure BS-2a shows a scattergram of the daily-average TEOM PM_{10} mass versus the daily-integrated QTM PM_{10} mass measured at YL. Overall, the agreement is quite good and there is no clear evidence of systematic bias in the QTM mass from fibers flaking off the filter. Excluding the four highest data points which are truly extreme values, a reduced major axis regression of TEOM on QTM yields slope 0.906 ± 0.015 (95% C.L.) and intercept 2.8 ± 1.0 $\mu g/m^3$. The grand mean TEOM and QTM concentrations were 59 and 62 $\mu g/m^3$, respectively, resulting in a TEOM-to-QTM ratio of means of 0.95. These trends suggest that, at least in an average sense, QTM mass is a very good proxy for TEOM mass. Figure BS-2b shows the CUSUM time series for the PM_{10} mass difference (TEOM minus QTM) at YL. The gross feature is a monotonic decay which means the QTM mass is greater than the TEOM mass, consistent with the above results. CUSUM plots for most other stations exhibited notable slope changes along the time series which suggest changes in the QTM-TEOM mass relationship such as might arise from even small changes in TEOM performance. The stability of the QTM-TEOM mass relationship, as evidenced by the CUSUM plot, was the main criterion for choosing YL for this analysis. The local maxima and minima along the CUSUM time series correspond to seasonal patterns in the TEOM-to-QTM relationship. Figure BS-2c is a time series of the sample day specific difference (TEOM minus QTM). In the summer the TEOM concentrations are modestly put persistently higher than the QTM concentrations. In contrast, in the winter the QTM concentrations are often significantly higher than the TEOM concentrations. The net effect is a 3 $\mu g/m^3$ difference between the two measurements but there are larger season-specific

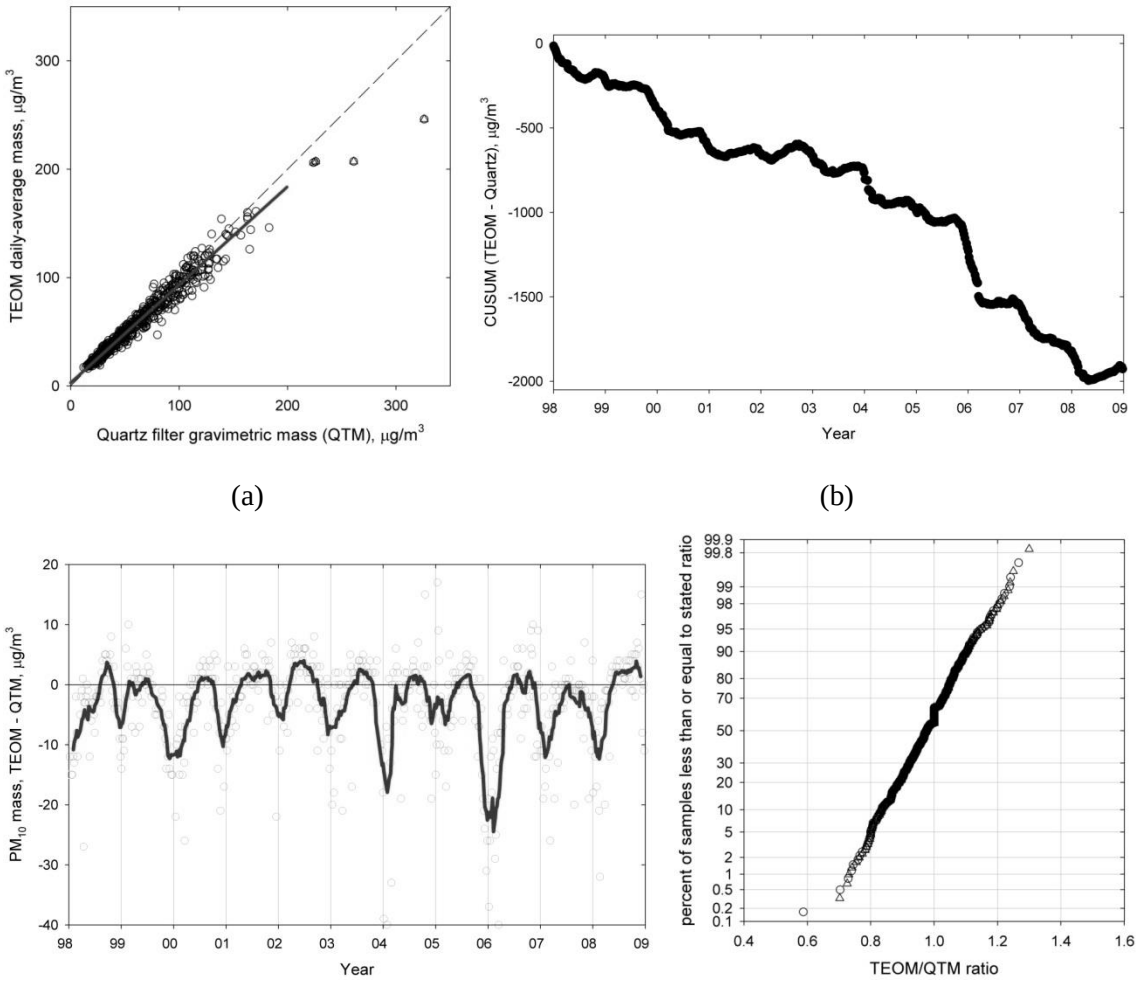


Figure BS-2. Relationships between PM_{10} 24-average TEOM mass and 24-hour integrated QTM mass at Yuen Long (YL): (a) scatter plot of TEOM on QTM; (b) cumulative summation (CUSUM) time series plot for TEOM minus QTM; (c) time series of TEOM minus QTM (markers) including a centered 11-sample arithmetic mean smoother (line); and (d) cumulative distribution of the TEOM/QTM mass ratio.

differences that cancel to some extent when calculating the grand averages. The reason for the summertime pattern is unknown. The wintertime patterns likely arises from semivolatile compounds (ammonium nitrate and OC) being retained by the quartz filter to varying degrees depending on environmental conditions whereas the TEOM when operated at 50°C is measuring nonvolatile mass.

Overall, the average difference between the daily TEOM-to-QTM mass is only 3%. Figure S2d shows the cumulative distribution of the TEOM/QTM ratio. The cumulative distribution is a straight line on a probability plot and thus is well represented by a normal distribution. The wintertime pattern of TEOM mass being modestly higher than QTM mass coincides with relatively low PM_{10} mass concentrations, whereas the summertime pattern of QTM mass significantly greater than TEOM mass coincides with relatively high PM_{10} mass concentrations. Thus, the absolute value of the *relative* difference is nearly the same for both seasons. While there is clearly some seasonal dependence in the relationships between TEOM mass and QTM mass, the QTM data are a reasonably good proxy for the TEOM mass which supports its use in the source apportionment modeling. Furthermore, the QTM mass represents the mass on the filter used for the speciation analysis and therefore is internally consistent as a total variable.

The advantages of apportioning QTM include: (i) both the gravimetric mass and all of the species (elements, ions, and carbon) are measured from the same filter so there is consistency between the gravimetric mass measurement artifacts and the species-specific artifacts; and (ii) all mass is being apportioned, not just the species that were measured. Furthermore, mass closure for the factor loadings can be used as a quality check on the representativeness of the unidentified (unmeasured) mass that is assigned to each factor.

B.S.4. PMF source identification

PMF derived source profiles and their uncertainties are tabulated in Table BS-2, and the percentage explained variations (EV) are plotted in Figure BS-3. In the PMF analysis, the number of factors was set as 4 and increase 1 at a time until the resolved source profiles are not fully explainable. When the number of factor was set to 8, factors of trace metals and coal

Table BS-2. Source profiles and uncertainties derived by PMF. Source profiles are expressed in mass percentages.

	Veh. Exh.	Residual Oil	Fresh S.S.	Aged S.S.	Crustal Soil	Sec. Sulfate	Sec. Nitrate	Trace Metal	Coal Comb.
Al	0%	2±1%	0%	0%	2%	0%	0%	1%	0%
As	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ca	1%	7±4%	1%	1%	6%	0%	0%	4±1%	1%
Cd	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cl^-	0%	0%	44±3%	0%	0%	0%	1%	0%	0%
Fe	1%	2±1%	0%	0%	3%	0%	0%	4±1%	0%
K^+	0%	3±2%	0%	1%	0%	0%	0%	0%	6%
Mg	0%	0%	2%	2%	1%	0%	0%	0%	0%
Mn	0%	0%	0%	0%	0%	0%	0%	0%	0%
Na^+	0%	0%	19±1%	16%	0%	0%	0%	0%	0%
NH_4^+	0%	0±3%	2%	0%	0%	18±1%	9%	0±1%	0±1%
Ni	0%	1±1%	0%	0%	0%	0%	0%	0%	0%
NO_3^-	0%	0±2%	0±2%	6±1%	4±1%	0%	34±1%	0±1%	0±1%
Pb	0%	0%	0%	0%	0%	0%	0%	1%	0%
SO_4^{2-}	4%	11±11%	0%	39±1%	5±1%	51±2%	0%	11±4%	10±2%
V	0%	4±3%	0%	0%	0%	0%	0%	0%	0%
Zn	0%	1±1%	0%	0%	0%	0%	0%	5±1%	0%
EC	34±1%	0±9%	7±1%	4±1%	0±1%	0±1%	0±1%	0±1%	6±2%
OC	32±1%	31±17%	4±2%	0%	0±1%	5±1%	23±1%	5±5%	34±2%
Unidentified	28%	38%	20%	30%	78%	25%	32%	68%	42%

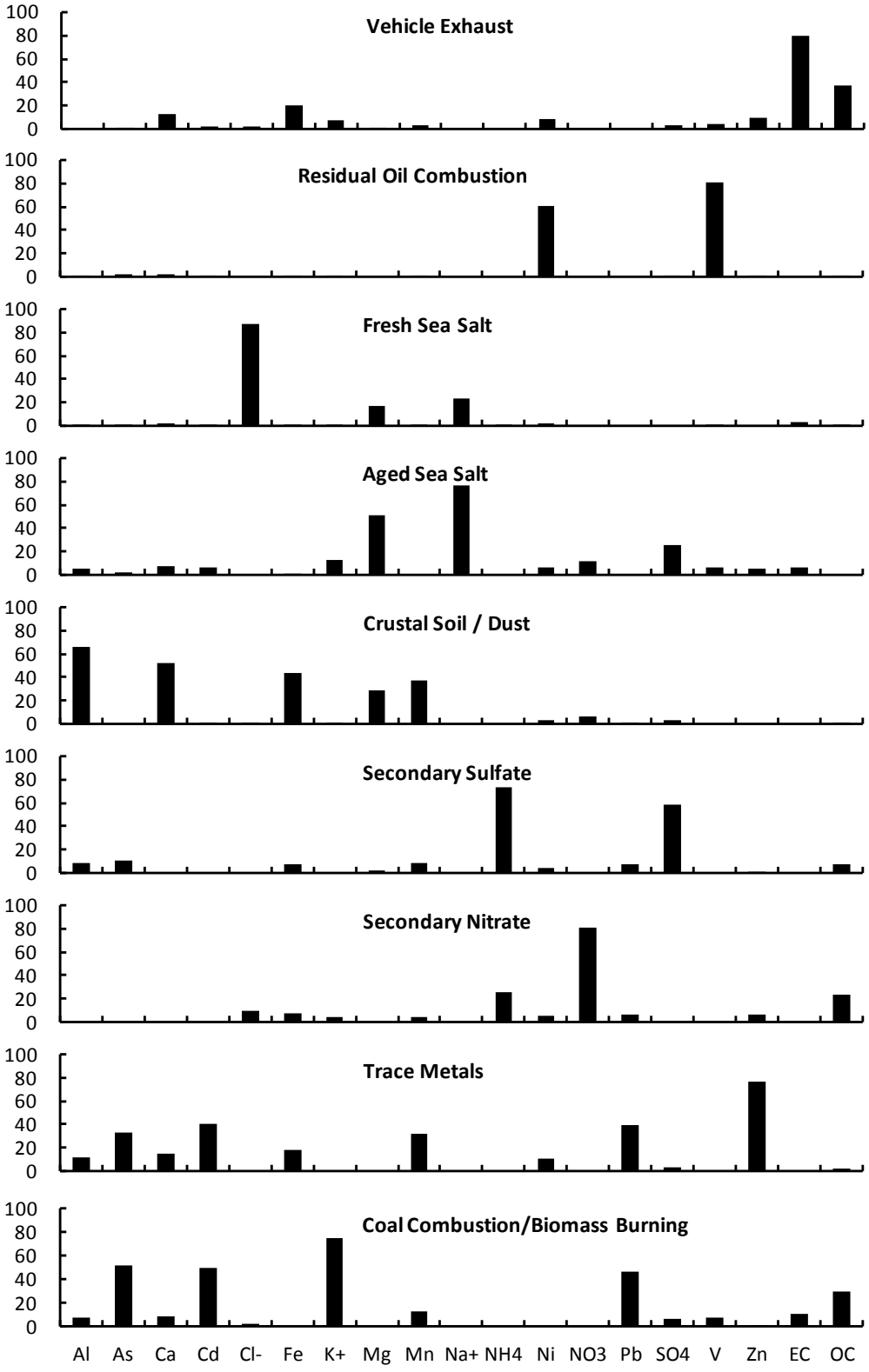


Figure BS-3. Percentage Explained Variation (EV) for the source profiles derived by PMF.

combustion / biomass burning merged together. When the number of factor was set to 10, the coal combustion / biomass burning factor was dissolved into two, as shown in Table BS-3 of the Supplemental Information. Some key species of this factor, e.g. Ca, K^+ and SO_4^{2-}, were smeared over the two factors. OC was entirely apportioned into 'Unresolved Factor 1' while EC was entirely apportioned into 'Unresolved Factor 2'. Both factors were not representative of emission profiles from any known sources. Therefore, we believe the 9-factor model generated the most reasonable results.

The first source in Figure BS-3 was identified as vehicle exhaust based upon the abundances of EC, OC and certain amount of Fe. Both diesel- and gasoline-powered vehicles generate large amount of carbonaceous compounds but the ratios of OC to EC in the emission profile can be quite different. Therefore the EC/OC ratio in this derived source profile represents the weighted average of emissions from all vehicles running in Hong Kong. As discussed in the main body of the manuscript, there might be other carbonaceous aerosol sources admixed into this factor but there is strong evidence that its temporal trend is governed by motor vehicle exhaust. The second source identified was distinguished by large EV values for Ni and V, which are good indicators for residual oil combustion (Chow and Watson, 2002). Although the relative contribution to the total PM_{10} mass is small, this factor is exceptionally stable. The third source was characterized by large loadings of Cl^-, Na^+ and Mg which collectively are a signature for fresh sea salt. The three species accounted for 44%, 19% and 2% of the total mass of this aerosol source, respectively, which is in close agreement with their corresponding percentage composition in seawater of 55%, 30% and 4% (Goldberg, 1963). The presence of small amount of OC (4%) and EC (7%) may be related to marine emissions in or around Hong Kong waters. Na^+ and Mg were also found in the fourth factor, along with significant loadings of nitrate and

Table BS-3. Profiles of relevant factors when the number of factors was set as 8, 9 and 10 in the PMF analysis. Source profiles are expressed in mass percentages.

Number of sources set in the PMF analysis	8	9		10		
Source	Trace Metals + Coal Comb. / Bio. Burning	Trace Metals	Coal Comb. / Bio. Burning	Trace Metals	Unresolved Factor 1	Unresolved Factor 2
Al	1%	1%	0%	1%	0%	0%
As	0%	0%	0%	0%	0%	0%
Ca	2%	4%	1%	3%	1%	2%
Cd	0%	0%	0%	0%	0%	0%
Cl^-	0%	0%	0%	1%	0%	0%
Fe	1%	4%	0%	3%	0%	0%
K^+	5%	0%	6%	0%	6%	6%
Mg	0%	0%	0%	0%	0%	0%
Mn	0%	0%	0%	0%	0%	0%
Na^+	0%	0%	0%	0%	0%	0%
NH_4^+	0%	0%	0%	0%	0%	0%
Ni	0%	0%	0%	0%	0%	0%
NO_3^-	0%	0%	0%	0%	0%	0%
Pb	1%	1%	0%	0%	0%	2%
SO_4^{2-}	11%	11%	10%	7%	7%	23%
V	0%	0%	0%	0%	0%	0%
Zn	2%	5%	0%	5%	0%	1%
EC	0%	0%	6%	0%	0%	15%
OC	25%	5%	34%	17%	42%	0%
Unidentified	53%	68%	42%	62%	44%	50%

sulfate. We assigned this factor to aged sea-salt because a lack of chloride and abundance of nitrate and sulfate suggests this is sea salt from which the chloride has been displaced by nitrate and sulfate. The fifth source has large contributions from Al, Ca, Fe and Mg, suggesting crustal or dust aerosols. This is mostly likely related to exposed soil, unpaved roads and construction activities the greater PRD region, nonmetallic mineral industries, and as well as other crustal materials in the background continental airmass. The sixth and seventh sources were associated with significant amount of ammonium and sulfate, and ammonium and nitrate, respectively, indicating they are secondary sulfate and nitrate aerosols. The eighth source was distinguished with large presence of Zn, suggesting it was mostly from trace metals processing. The ninth source was associated with large EV for As, Cd, Fe, K^+, Pb, and OC, probably a mixture of biomass burning and coal combustion from power plants and industrial combustion sources in the PRD.

B.S.5. Source apportionment by APCA and Unmix

The results presented in this study were derived from the PMF modeling and the PMF-resolved source profiles are presented in Table BS-2. Source apportionment was also conducted using APCA and Unmix. Table BS-4 shows the source contribution estimates for 4-to-9 factor solutions from APCA, mapped onto the PMF-resolved source categories where appropriate. Results are presented for both constrained (Table BS-4a) and unconstrained (Table BS-4b) regressions to reconstruct source contribution estimates from the principle components analysis. For the unconstrained regression, the constant term should tend to zero for a good APCA solution. Table BS-4b shows that a large constant was obtained for the four-factor solution and thus it is rejected; this is consistent with the scree plot demonstrating that at least five factors should be retained. For the 5-to-9 factor solutions the constant was smallest for six and nine

Table BS-4. APCA-modeled source contribution estimates for 4-to-9 factor solutions, mapped onto the PMF-resolved source categories. All concentration values in $\mu g/m^3$.

(a) Constrained multivariable regression (constant term not included)

No. of Factors	Coal Comb.	Sea Salt	Residual oil	Crustal Soil	Veh. Exh.	Sec. Sulfate	Sec. Nitrate	OC Factor	Chloride Factor
9	7.84	4.15	3.24	5.07	8.45	8.30	-0.49	4.83	-0.36
8	8.22	3.83	3.07	5.26	10.64	9.20	1.59	-1.03	
7	8.99	3.78	3.10	5.14	10.26	8.41	1.16		
6	7.08	4.01	2.91	5.27	10.43	11.26			
5	17.09	6.53	4.63	4.18	8.23				
4	14.62	6.50	10.45	8.25					

(b) Unconstrained multivariable regression (constant term included)

No. of Factors	Coal Comb.	Sea Salt	Residual oil	Crustal Soil	Veh. Exh.	Sec. Sulfate	Sec. Nitrate	OC Factor	Chloride Factor	Const.
9	7.81	4.01	3.21	5.04	8.30	8.26	-0.49	4.80	-0.38	0.52
8	8.08	3.27	2.94	5.11	9.78	8.96	1.57	-1.09		2.45
7	8.88	3.36	3.00	5.00	9.60	8.17	1.16			1.92
6	7.05	3.84	2.86	5.19	10.17	11.14				0.82
5	16.70	5.75	4.36	3.99	7.53					2.75
4	13.76	4.52	8.51	7.08						7.21

factors. Unlike PMF it is possible to obtain large negative values for the factor loadings (source profiles) and factor scores (source contribution estimates). The presence of negative values is mathematically possible but not physically realistic, and this becomes an important criterion for accepting or rejecting certain solutions. The eight- and nine-factor solutions exhibit unacceptably large negative source contribution estimates for one or more factors and thus they are rejected. The seven factor solution includes Residual Oil, Crustal/Soil, Vehicular Exhaust, Secondary Nitrate and Secondary Sulfate distinct factors; Coal Combustion and Biomass Burning source categories lumped into a single factor; and Aged Sea Salt and Fresh Sea Salt also lumped into a single factor. Within the Secondary Sulfate factor, sulfate ion accounts for 37% of the mass assigned to that factor and chloride ion has a negative loading corresponding to 40% of the total observed chloride. The Secondary Nitrate factor contains 38% nitrate ion by mass along with other trace metals including Fe, K, Pb, Cl, etc. With six factors, nitrate ion is smeared over the resolved six factors. The Secondary Sulfate factor still has negative chloride loading corresponding to 30% of the observed chloride, but now contains ~45% sulfate. With five factor solution, most of the sulfate is apportioned to the Coal Combustion and Biomass Burning factor. Further, the Crustal/Soil factor is smeared with Biomass Burning. Note that even with higher number of factors (7-9), the Coal Combustion and Biomass Burning source categories could not be separately resolved. Similarly, the Aged Sea Salt and Fresh Sea Salt factors could not be separately resolved without allowing high negative mass for chloride in the Aged Sea Salt factor. The six- and seven-factor solutions seem most plausible and ultimately the six factor solution was selected for additional analysis.

Six factors were also deemed as the most reasonable solution for Unmix analysis. Tables BS-5 - BS-6 present the source profiles derived by APCA and Unmix, respectively, expressed as mass

Table BS-5. Source profiles derived from APCA analysis for a six factor solution. The values are expressed as mass percentages.

Species	Sec.Sulfate	Veh.Exh.	Coal Comb.	Crustal Soil	Sea Salt	Residual Oil
Al	0.21	-0.09	0.64	3.26	0.32	0.25
As	0.01	0.00	0.03	0.02	-0.01	0.02
Ca	0.14	0.94	1.85	7.99	1.51	0.75
Cd	0.00	0.00	0.01	0.01	0.00	0.00
Cl^-	-1.97	0.56	0.30	-0.50	29.69	-1.48
Fe	0.58	1.18	1.26	4.71	0.11	0.73
K^+	1.03	0.64	3.49	2.56	0.26	-0.05
Mg	0.25	-0.03	-0.06	1.87	3.46	-0.04
Mn	0.03	0.01	0.07	0.16	-0.01	0.02
Na^+	2.22	0.18	-1.45	1.96	28.65	0.83
NH_4^+	16.22	-0.20	10.59	3.63	-2.24	13.02
Ni	0.01	0.01	0.01	0.01	0.00	0.11
NO_3^-	10.14	3.04	8.32	8.63	24.09	7.62
Pb	0.11	0.04	0.41	0.28	-0.10	0.10
SO_4^{2-}	46.89	1.54	26.73	26.66	12.74	34.65
V	0.01	0.02	0.01	0.00	-0.01	0.26
EC	0.50	34.66	0.90	4.68	3.55	10.08
OM*	23.61	57.51	46.88	34.08	-2.01	33.12

(*) OM = organic matter, assumed to be 1.6×OC.

percentages. The explained mass profiles are quite similar to the PMF-derived profiles presented in Figure BS-3. The key differences are: (1) nitrate ion is loaded primarily onto the Secondary Sulfate factor and the Sea Salt factor in APCA and Unmix whereas it was resolved as a separate factor using PMF. This issue might arise from the artifacts inherent to nitrate measurement using High Volume samplers and quartz fiber filters. Assuming significant nitrate ion is present as ammonium nitrate, its volatility and the artifact behavior could affect its covariance with other species and lead to the observed smearing across factors. However, this artifact is influenced by gaseous nitric acid and ammonia mixing ratios, temperature and humidity and is likely reproducible on a day-specific basis leading to high collocated precision as reflected in the small

Table BS-6. Source profiles derived from Unmix analysis for a six factor solution. The values are expressed as mass percentages.

Species	Sec.Sulfate	Veh.Exh.	Coal Comb.	Crustal Soil	Sea Salt	Residual Oil
Al	0.19	-0.09	0.42	4.75	0.24	0.50
As	0.00	-0.01	0.05	0.00	0.00	0.01
Ca	0.17	1.08	1.39	11.24	2.39	1.38
Cd	0.00	0.00	0.01	0.00	0.00	0.00
Cl^-	-0.77	0.22	0.11	-0.55	53.17	-0.51
Fe	0.62	1.16	0.99	6.57	-0.12	1.14
K^+	0.39	0.81	4.40	1.03	0.96	-1.38
Mg	0.51	0.17	-0.43	3.16	4.50	0.37
Mn	0.02	0.01	0.08	0.19	-0.01	0.01
Na^+	4.00	1.81	-3.24	5.67	38.07	2.87
NH_4^+	13.97	-0.80	11.85	-1.71	-11.88	10.08
Ni	0.01	0.01	0.00	0.01	0.00	0.12
NO_3^-	14.34	1.08	5.22	12.89	23.01	5.98
Pb	0.01	0.05	0.54	0.05	-0.01	-0.05
SO_4^{2-}	39.57	6.50	28.74	21.80	-15.75	31.30
V	0.01	0.01	-0.01	0.01	0.00	0.31
EC	3.40	33.23	-0.19	9.11	9.59	18.37
OM	23.58	54.78	50.08	25.77	-4.16	29.48

(*) OM = organic matter, assumed to be 1.6×OC.

multiplicative term (0.026) in the nitrate error structure (Table B-2). PMF picks up on the combination of high measurement precision and a lack of covariance with other species to assign nitrate as a distinct factor; (2) Sea Salt was resolved as a single factor in APCA and Unmix but as two factors – Fresh Sea Salt and Aged Sea Salt – using PMF; and (3) a significant negative mass concentration of chloride ion (relative to the study-average observed chloride average concentration) was apportioned to the Secondary Sulfate and Residual Oil factor in both APCA and Unmix. The anomalous behavior for modeled chloride ion is not surprising since it is a reactive species and can be displaced from sea salt by forming hydrochloric acid. The ion displacement likely has high day-to-day variability which would add significant noise to the source profile representing true fresh sea salt emissions.

Qualitatively, the source profiles of Tables BS-5 (APCA) and BS-6 (Unmix) are similar tothe profiles presented in Table BS-2 after taking into consideration the lumping of factors in APCA and Unmix that were separately resolved using PMF.

B.S.6. References

Birch, M.E. and Cary, R.A.; Elemental carbon-based method for monitoring occupational exposures to particulate diesel exhaust, *Aerosol Science & Technology*, **1996**, 25, 221-241.

Chow, J.C. and Watson, J.G.; Review of $PM_{2.5}$ and PM_{10} apportionment for fossil fuel combustion and other sources by chemical mass balance receptor model, *Energy and Fuels*, **2002**, 16, 222-260.

Goldberg, E.D.; Chemistry- the oceans as a chemical systems. In: Hill,H.M.(E.d.), Composition of Sea Water, Comparative and Descriptive Oceanography, Vol.2 of the sea, **1963**, Wiley-Interscience, New York, 3-25.

Hong Kong Environmental Protection Department (HKEPD), **2003**. HKSAR Air Quality in Hong Kong 2002, available at http://www.epd.gov.hk/epd/english/environmentinhk/air/air_quality/files/aqr02e.pdf

Huang, S., Rahn, K.A. and Arimoto, R.; Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island, *Atmospheric Environment*, **1999**, 33, 2169-2185.

Hyslop, N.P. and White, W.H.; An empirical approach to estimating detection limits using collocated data, *Environmental Science & Technology*, **2008a**, 42, 5235-5240.

Hyslop, N.P. and White, W.H.; An evaluation of interagency monitoring of protected visual environment (IMPROVE) collocated prevision and uncertainty estimates, *Atmospheric Environment*, **2008b**, 42, 2691-2705.

Sin, D.W.M., Fung, W.H. and Lam, C.H.; Measurement of carbonaceous aerosols: validation and comparison of a solvent extraction – gas chromatographic method and a thermal optical transmittance method, *Analyst*, **2002**, 127, 614-622.

United States Environmental Protection Agency (USEPA), **1999**. Office of Research and Development, Compendium Method IO-3: Chemical Species Analysis of Filter-Collected Suspended Particulate Matter (SPM), available at http://www.epa.gov/ttn/amtic/inorg.html

Yu, J.Z., Tung, J.W.T., Wu, A.W.M., Lau, A.K.H, Louie, P.K.K. and Fung, J.C.H.; Abundance and seasonal characteristics of elemental and organic carbon in Hong Kong PM_{10}, *Atmospheric Environment*, **2004**, 38, 1511-1521.

Appendix C: Supplementary Information for Chapter 4.

C.1. Particulate matter datasets

Table 4-1 (in Section 4.3 of the Chapter 4) summarizes the environmental characteristics of various monitoring sites and the valid sampling periods for ambient particulate matter datasets used in this study. 'General station' refers to monitoring sites located in urban areas (commercial/residential), usually on top of public buildings (20-80 m off the ground) to gauge neighborhood-level exposure. 'Roadside' monitoring stations are usually located few meters above ground at busy intersections to gauge vehicular traffic exposure; while 'Remote' station are located in regions isolated from urban development to measure regional background conditions.

PM_{10} speciation data was collected from 1998 to 2008 at 1-in-6 day frequency on quartz filters and analyzed for gravimetric mass, elements by Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AMS), ions by Ion Chromatography (IC) and organic carbon (OC) and elemental carbon (EC) by a thermal/optical transmittance method following the NIOSH 5040 protocol (Chow et al., 2001). PM_{10} speciation sampling at various sites was conducted on different day of the week such that on any particular day, at most three to four different sites had sampling conducted on the same day.

Three one-year 24-hour integrated $PM_{2.5}$ speciation datasets were collected during 2000/01, 2004/05 and 2008/09 at 1-in-6 day frequency. Sampling in 2000-01 was conducted at three sites; Mong Kok (MK), Tsuen Wan (TW), and Hok Tsui (HT) which represent roadside, urban/general, and remote conditions, respectively. During the latter two study periods, $PM_{2.5}$

sampling was also conducted at a second urban site; Yuen Long (YL). Unlike PM_{10} sampling, $PM_{2.5}$ samples were collected on the same day across the network using Rupprecht & Patashnick (now Thermo) Partisol samplers on Teflon filters for gravimetric analysis and elemental analysis by X-Ray Fluorescence (XRF). Samples were also collected on quartz filters for gravimetric analysis and anion analysis by Ion Chromatography (IC), ammonium ion by Automated Colorimetry, sodium and potassium ions by Atomic Absorption Spectrometry (AAS), carbon analysis by the IMPROVE Thermo-Optical Reflectance (TOR) method.

While there were differences in the field sampling hardware and analytical methods for $PM_{2.5}$ and PM_{10} speciation datasets, comparisons between corresponding species concentrations provided insights into potential data quality issues and, on which species dominated the fine ($PM_{2.5}$) or coarse fraction ($PM_{10-2.5}$). Figure C-1 shows the Reduced Major Axis (RMA) regression slope as proxy for $PM_{2.5}/PM_{10}$ ratio for total PM mass and measured species

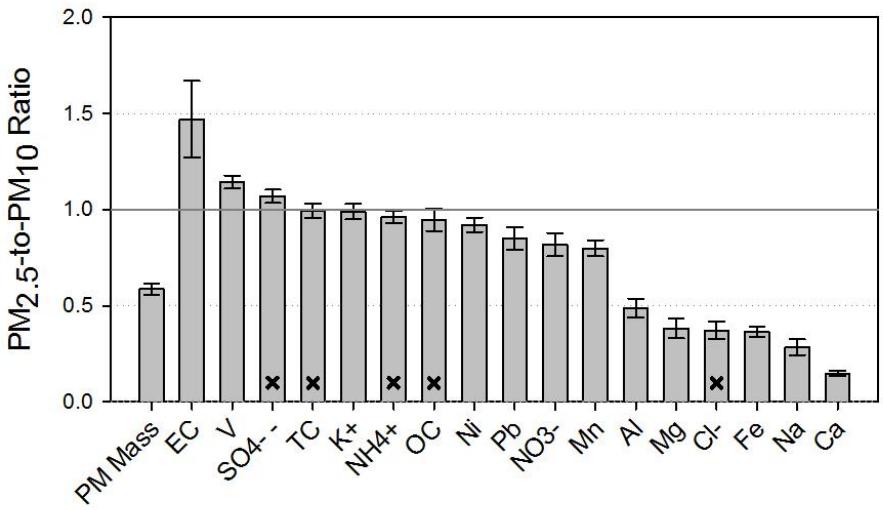


Figure C-1. Reduced major axis (RMA) regression slope as proxy for $PM_{2.5}$ to PM_{10} ratio for total mass and species concentrations. The marked (x) species have regression constant distinct from zero over its 95% confidence interval.

concentration utilized in this study. Other species reported were rejected from the remainder of the analysis due to lack of sufficient data above the detection limit and/or due to concerns over the species data quality. In general, $PM_{2.5}$ mass accounted for 59% of the PM_{10} mass in Hong Kong, indicating significant fraction of particles to be in fine fraction. Good agreement between $PM_{2.5}$ and PM_{10} total carbon (TC) measurements demonstrates predominance of carbonaceous matter in fine fraction. Some deviations, including $PM_{2.5}$ TC greater than PM_{10} TC, could arise from measurement artifacts due to different flow rates of the samplers and/or differences in analytical measurement methods for carbon fractions. $PM_{2.5}$ OC values were less than or equal to the PM_{10} values while considerable $PM_{2.5}$ EC values were greater than or equal to the PM_{10} values. The EC trend is not realistic for a consistent measurement methodology but in this case both the OC and EC trends could arise from differences in the thermo-optical analysis protocols – NIOSH thermo-optical transmittance (TOT) for the PM_{10} samples and IMPROVE TOR for the $PM_{2.5}$ samples. High $PM_{2.5}/PM_{10}$ ratio for SO_4^{2-} and NH_4^+ indicates these species to be present in the fine size range with excellent agreement between the $PM_{2.5}$ and PM_{10} data. In contrast, PM_{10} NO_3^- concentrations were often much higher than $PM_{2.5}$ NO_3^-. The NO_3^- measurements may suffer from differential measurement artifacts; such as uncertainty in measurement due to volatility of NO_3^- (Chow et al., 2002). Al, Ca, Fe and Mn all showed significant coarse particle loadings. Ca and Fe concentrations in the $PM_{2.5}$ and PM_{10} samples were very highly correlated which suggests the same sources contribute to both fine and coarse size modes. Cl^- was also overwhelmingly in the coarse fraction. Mn, Ni, Pb and V were predominantly in the fine fraction; with some Mn and Pb in the coarse fraction. V was substantially higher in $PM_{2.5}$ samples than PM_{10} samples, suggesting measurement artifacts with at least one of the methods.

On the other hand elemental $PM_{2.5}$ K; measured by XRF, showed very good agreement with PM_{10} K^+ measured by IC.

Hourly PM_{10} mass collected at fourteen sites from 1998-2008 by TEOM monitors are also utilized in this study. Missing values for single or two consecutive hours were imputed by linear interpolation of adjacent reported hourly values and the insignificant impact on the calculated daily averages was confirmed by performing Monte Carlo simulations using a jackknife method. Depending on the site, this resulted in 84-95% daily data completeness for calculating the daily average TEOM mass concentrations for days with all 24 hourly measured mass.

C.2. Source apportionment of $PM_{2.5}$ speciation dataset

Positive Matrix Factorization (PMF) has been successfully applied to access particulate source contributions in various regions/urban environments; such as Alaska (Polissar et al., 1998), Phoenix (Ramadan et al., 2000), St. Louis (Wang et al., 2011) and northeastern US cities (Song et al., 2001), as well as for various ambient pollutants, such as fine particulate matter (Lee et al., 2006), polycyclic aromatic hydrocarbons (Larsen and Baker, 2003) and volatile organic compounds (Lau et al., 2010). General receptor modeling methodology can be stated in terms of the contribution from *p* sources to all of the species in a given sample (eq. CS-1), where, x_{ij} is the *j*th species concentration measured in the *i*th sample, g_{ik} is the particulate mass concentration from the *k*th source contributing to the *i*th sample, f_{kj} is the *j*th species mass fraction from the *k*th source, e_{ij} is the residual associated with the *j*th species concentration measured in the *i*th sample, and *p* is the total number of sources. The PMF solution utilizes sample-specific uncertainties for each species (u_{ij}) to minimize the objective function *Q(E)* (eq. CS-2).

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \qquad \text{(CS-1)}$$

$$Q(E) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{u_{ij}} \right]^2 \quad \text{(CS-2)}$$

Record-specific uncertainty estimates (u_{ij}) utilized in PMF were generated by analyzing the collocated data collected during the study period using the methodology presented in Supplementary Information of Yuan et al., 2013. The concentration-dependent error structure for each species was calculated as: $u_{ij} = a_j + b_j \times C_{ij}$ (Polissar et al., 1998), where u_{ij} is the uncertainty of species *j* and sample *i*, C_{ij} is the mass concentration, and a_j and b_j are the species-specific additive and multiplicative coefficients respectively derived from the analysis of the collocated data.

Collocated $PM_{2.5}$ gravimetric mass measurement on quartz-fiber filter and Teflon-membrane filter showed good agreement (RMA regression slope = 1.02, intercept = 2.04 μg/m^3 and R^2 = 0.99; with Teflon filter mass on x-axis). The positive intercept could be attributed to adsorption of organic vapors onto quartz-fiber filter (Chow et al., 1996). For consistency with PM_{10} source apportionment, quartz gravimetric mass is used as 'Total Variable'. To account for various sources of variability, such as temporal variation in source profiles, modeling was performed with 10% extra modeling uncertainty. $PM_{2.5}$ data from all four sites and three sampling years was combined into a single dataset. Sensitivity studies performed by executing PMF modeling on subsets of the $PM_{2.5}$ data include: one run for each site including all years (four runs); one run for each year including all sites (three runs); runs for each site and year combination (eleven runs). Results from these exercises are tabulated in Table C-1 and, were generally consistent with the PMF analysis using the single dataset obtained by combining all sites and years. Thus, the results discussed are based on the single combined dataset. Each modeling run included 20 base runs and the base run with minimum *Q* value was retained as the solution. Solutions for six

Table C-1. $PM_{2.5}$ mass contributions for the eight PMF-resolved factors stratified by runs based on site and sampling periods. All concentrations are reported in $\mu g/m^3$ (percentages in parenthesis).

Site & Period	Secondary Sulfate		Vehicle Exhaust		Secondary Nitrate		Biomass Burning		Trace Metals		Residual Oil		Soil/Dust		Fresh Sea Salt		Total
All Sites	12.7	(31.5)	10.4	(25.9)	5.0	(12.5)	4.3	(10.7)	2.7	(6.6)	2.2	(5.4)	1.6	(4.0)	1.4	(3.4)	40.2
2000-01	10.2	(24.2)	16.4	(38.9)	2.3	(5.5)	6.5	(15.4)	1.4	(3.3)	1.8	(4.4)	1.9	(4.5)	1.6	(3.8)	42.1
2004-05	15.4	(34.7)	10.2	(23.1)	5.0	(11.4)	4.6	(10.3)	3.5	(8.0)	2.5	(5.6)	1.6	(3.6)	1.4	(3.2)	44.3
2008-09	11.7	(33.4)	6.8	(19.3)	6.7	(19.0)	2.6	(7.4)	2.6	(7.5)	2.1	(5.9)	1.4	(4.0)	1.2	(3.4)	35.1
TW	13.2	(34.9)	7.6	(20.2)	4.6	(12.1)	4.8	(12.8)	2.7	(7.0)	2.4	(6.4)	1.6	(4.1)	0.9	(2.5)	37.9
2000-01	11.0	(29.3)	9.6	(25.6)	3.0	(8.0)	7.3	(19.4)	1.8	(4.7)	2.0	(5.2)	1.8	(4.9)	1.0	(2.8)	37.3
2004-05	16.0	(37.2)	8.6	(19.9)	4.3	(9.9)	4.8	(11.2)	3.6	(8.4)	3.1	(7.1)	1.7	(3.9)	1.0	(2.4)	43.1
2008-09	12.5	(37.4)	5.1	(15.3)	6.2	(18.5)	2.8	(8.4)	2.5	(7.5)	2.3	(6.8)	1.2	(3.7)	0.8	(2.4)	33.6
MK	11.3	(20.6)	25.4	(46.4)	5.9	(10.8)	4.0	(7.4)	2.6	(4.8)	2.1	(3.9)	1.8	(3.3)	1.5	(2.8)	54.7
2000-01	9.1	(14.4)	37.9	(59.8)	2.9	(4.6)	6.9	(10.9)	1.3	(2.0)	1.9	(3.0)	1.8	(2.9)	1.6	(2.5)	63.4
2004-05	14.2	(25.0)	23.1	(40.7)	6.2	(10.9)	3.9	(6.8)	3.7	(6.6)	2.4	(4.1)	1.8	(3.1)	1.6	(2.8)	56.8
2008-09	10.3	(22.8)	16.8	(37.1)	8.3	(18.3)	1.8	(3.9)	2.8	(6.2)	2.1	(4.6)	1.8	(4.0)	1.4	(3.1)	45.3
YL	14.6	(35.3)	5.9	(14.2)	8.2	(19.8)	4.0	(9.6)	3.5	(8.4)	2.1	(5.0)	1.5	(3.6)	1.7	(4.0)	41.3
2000-01	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2004-05	16.7	(35.3)	7.3	(15.4)	8.1	(17.1)	4.9	(10.5)	4.0	(8.5)	2.5	(5.2)	1.6	(3.4)	2.2	(4.6)	47.2
2008-09	12.7	(35.3)	4.6	(12.8)	8.3	(23.0)	3.1	(8.7)	3.0	(8.4)	1.7	(4.8)	1.4	(3.8)	1.2	(3.3)	36.0
HT	12.1	(45.1)	1.3	(4.9)	2.2	(8.1)	4.2	(15.6)	2.1	(7.8)	2.0	(7.5)	1.5	(5.5)	1.5	(5.5)	26.9
2000-01	10.5	(40.9)	1.6	(6.4)	1.0	(4.0)	5.4	(21.0)	1.2	(4.7)	1.7	(6.5)	2.0	(7.8)	2.2	(8.7)	25.5
2004-05	14.5	(48.4)	2.0	(6.8)	1.6	(5.4)	4.6	(15.5)	2.8	(9.4)	2.1	(7.1)	1.3	(4.2)	1.0	(3.2)	30.0
2008-09	11.3	(45.0)	0.3	(1.3)	3.8	(15.1)	2.7	(10.8)	2.2	(8.7)	2.2	(9.0)	1.2	(4.8)	1.3	(5.4)	25.1

to ten factors were examined and the eight factor solution was deemed optimal. The factors were assigned to the corresponding source categories based on the tracer species (Yuan et al., 2013); and in the order of decreasing source contribution estimates over the three study periods were: Secondary Sulfate, Vehicular Exhaust, Secondary Nitrate, Biomass Burning, Trace Metals, Residual Oil Combustion, Soil/Dust and Chlorine/Fresh Sea Salt. With more than eight factors, various trace metals apportioned to different factors and could not be related to any specific source categories. When the number of factors was reduced below eight, various identifiable factors admixed. 100 bootstrap runs with minimum correlation R-value of 0.6 on the eight factor PMF base solution indicated stable solution with all the bootstrapped factors uniquely mapped to a factor from the base solution.

Source apportionment through additional receptor models including, principal components analysis with absolute principal component scores (APCA, Thurston and Spengler, 1985) and US-EPA Unmix 6.0 (USEPA, 2007) were also investigated with the same set of species to examine the robustness of PMF modeling results. In both APCA and Unmix models, Na, Mg and Zn dominated separate factors with large negative values assigned to other species in the profile, until reducing the number of factors forced these species to merge with other factors. Hence, these species were iteratively rejected from the apportionment. In APCA model, NO_3^- was smeared across various factors and nearly 50% of daily factor contribution estimates associated with secondary nitrate, sea salt and biomass burning were assigned negative values, leading to near-zero annual average factor contributions. Hence, $PM_{2.5}$ sources identified by APCA apportionment were deemed impractical. Unmix model resulted in a seven factor solution with secondary sulfate, vehicle exhaust, biomass burning, secondary nitrate, residual oil combustion, soil/dust and sea salt identified as the main factors contributing to the PM mass.

The trace metal species were smeared across different factors. Unmix factor contribution estimates showed good agreement with PMF estimated factor contributions. Grand average contribution estimates between PMF and Unmix for the two dominant factors, secondary sulfate and vehicle exhaust; differed by less than 5% between the two models, despite the relatively high scatter in day-specific secondary sulfate factor apportionment. Residual oil combustion also showed modest agreement between the models with 15% difference. For the remaining factors, both the absolute and relative differences between the models were relatively larger. The added value of utilizing sample-specific uncertainties and non-negativity constraint incorporated in PMF was subjectively preferred over the edge detection capabilities of Unmix for this study.

Figure C-2 shows the eight PMF-resolved $PM_{2.5}$ source profiles expressed as mass fraction. With 32% of the total mass apportioned to secondary sulfate factor, it was the dominant contributor to $PM_{2.5}$ mass followed by vehicular exhaust and secondary nitrate at 26% and 13% respectively. The sulfate factor was dominated by sulfate ion (58%) and ammonium ion (18%), with ammonium-to-sulfate molar ratio of 1.6, which is between the stoichiometric ratio for ammonium bisulfate (1.0) and ammonium sulfate (2.0). The molar ammonium-to-nitrate ratio of 1.5 was considerably higher than the stoichiometric ratio of 1:1 for ammonium nitrate. Such discrepancies in apportionment of $SO_4^{2-}/NO_3^-/NH_4^+$ could arise from multiple scenarios; such as measurement/analysis artifact in SO_4^{2-} values indicated by better correlation of collocated sulfur measured by XRF over IC measurement of SO_4^{2-}, volatility of NO_3^- and/or SO_4^{2-}/NO_3^- displacement of chloride in sea salt (Russell et al., 1994; Saul et al., 2006). Further, $SO_4^{2-}/NO_3^-/NH_4^+$ were smeared across various factors. Sea salt factor was dominated by sulfate (14%), OC (11%), nitrate (8%) and EC (4%), with only 6% of the factor mass accounted by chloride. However, 90% of the XRF measured Cl was apportioned to the sea salt factor identified by

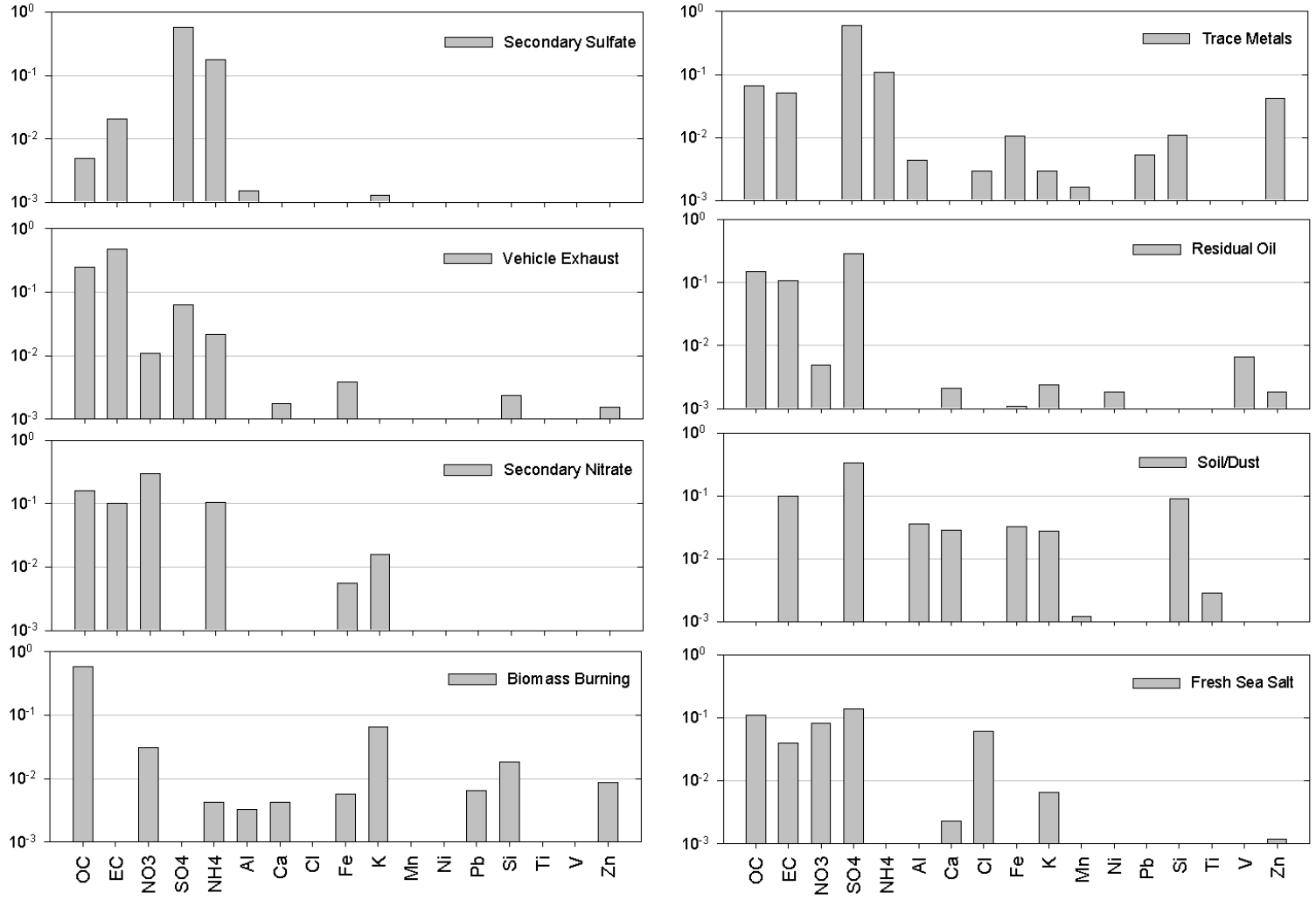


Figure C-2. $PM_{2.5}$ source profiles, expressed as mass fractions for the eight PMF-resolved factors.

chloride alone due to lack of quality $PM_{2.5}$ sodium data. Biomass burning factor contributing 11% to the $PM_{2.5}$ mass was dominated by OC (59%) and accounted for 63% of the measured potassium. 58% of Pb apportioned to this factor suggests similar source region for biomass burning and trace metals (or industrial emissions) factors. Presence of various elements associated with crustal material (Al/Ca/Fe/Si) also indicates admixing of resuspended soil with biomass burning activities. The relative distribution of major crustal elements (Al, Ca, Fe and Si) in the ambient samples and resolved soil/dust factor indicates the PMF-resolved soil profile to be a composite of county park soil, urban soil, and paved road dust profiles published in Ho et al., 2003. PMF-resolved soil/dust factor was based on the modeling data aggregated across four sites each with potentially distinct soil type impacts. Ni and V, key indicators of residual oil combustion, were present at low mass fractions in the corresponding factor, with factor mass dominated by sulfate (28%). V/Ni ratio of 3.6 was high but plausible given $PM_{2.5}$ V concentrations were biased high compared to PM_{10} V concentration. Vehicular exhaust factor dominated by EC and OC had high EC/OC ratio of 1.9 in contrast to the EC/OC ratio of 1.3 for PM_{10} vehicular exhaust factor, which is expected of samples analyzed by IMPROVE TOR method in contrast to NIOSH TOT method used for PM_{10} samples. Despite the various caveats in apportionment of individual species, the PMF-resolved factors provided an optimal representation of source categories also independently identified by source apportionment modeling of PM_{10} speciation dataset.

Site specific $PM_{2.5}$ annual-average contribution estimates are plotted in Figure C-3. Vehicle exhaust contribution was highest at MK roadside station and very low at HT remote station, which is consistent with respective site characteristics. Further, vehicle exhausts contributions decreased by ~58%; from 38 $\mu g/m^3$ in 2001 to 17 $\mu g/m^3$ in 2009 at roadside station, as well as by

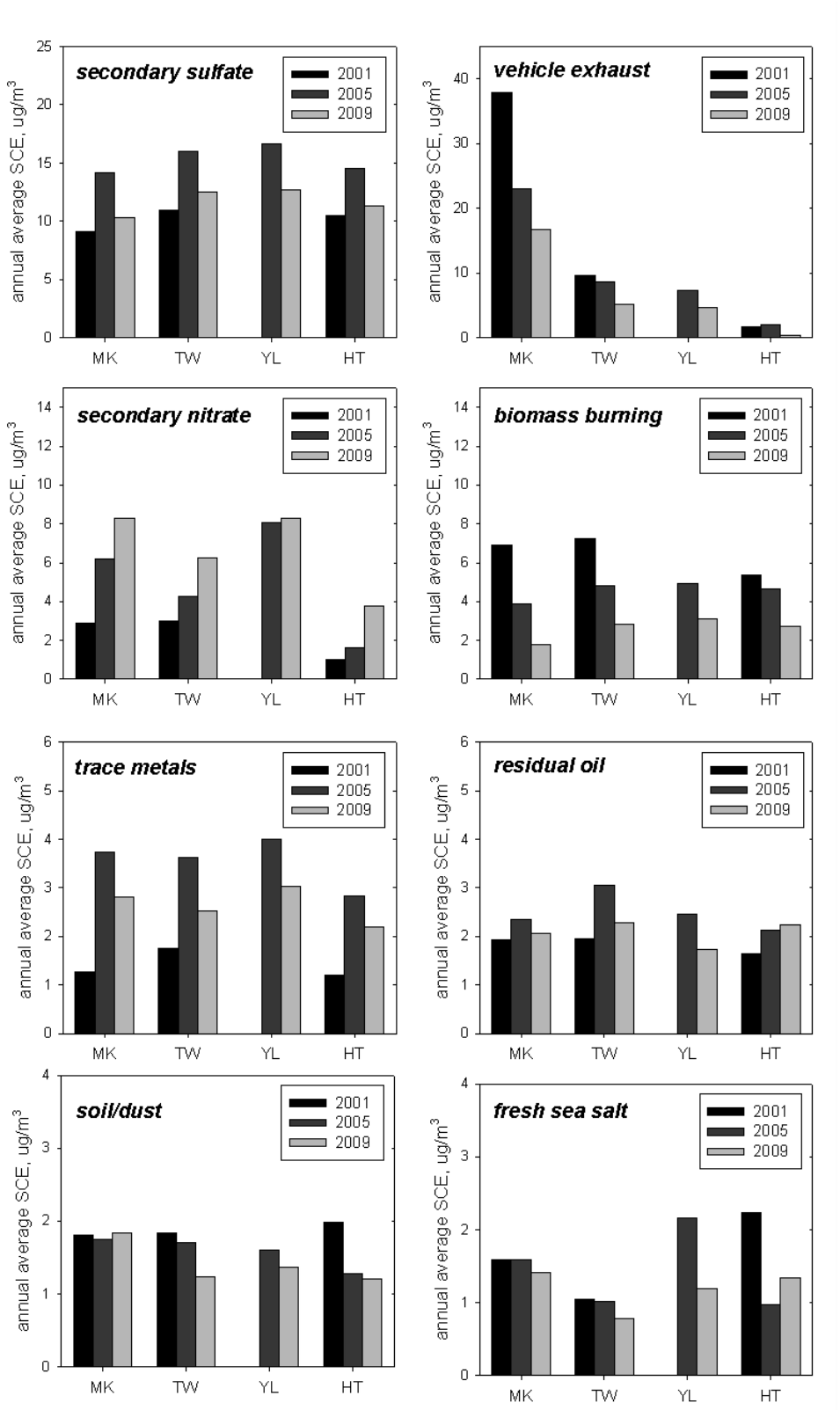


Figure C-3. Distribution of annual average source contribution estimates by site and year for each of the eight PMF-resolved $PM_{2.5}$ factors.

~50% at TW general station, consistent with PM_{10} PMF results. Secondary sulfate factor had consistent patterns across the four sites with factor contribution to $PM_{2.5}$ mass increasing from 24% in 2001 to 33% in 2009, with highest contributions in 2005 irrespective of the sites characteristic. Similarly, irrespective of the site characteristic, mass apportioned to secondary nitrate factor increased at all sites from ~10% in 2001 to ~20% in 2009. However the nitrate contributions were lowest at the HT remote station and highest at the MK roadside station and YL general station. Despite concerns over measurement and apportionment of NO_3^- in this study, NO_x emissions from vehicular and marine emissions (Kim et al., 2012) can contribute to spatial gradient based on site characteristics. Biomass burning contributions monotonically decreased over the study period independent of sites characteristic while trace metals contributions modestly increased. However, the mass apportioned to these two factors combined together consistently contributed to ~18% of the $PM_{2.5}$ mass over the study period. The remaining factors; crustal/soil, sea salt, and residual oil combustion showed neither significant site-specific variation nor clear temporal change.

C.3. Estimating local- and larger-scale contributions from TEOM dataset

Hourly PM_{10} TEOM mass data collected daily at up to fourteen sites during 1998 to 2008 is utilized by developing day-specific semi-quantitative estimates of local-versus larger-scale contributions. This approach utilizes intraurban variability in PM arising from emission sources acting over various scales as conceptualized in Figure C-4. Non-local PM impacts in this study are collectively labeled as synoptic-scale even though synoptic-scale (i.e. spatial domain > 1000 km) PM impacts may be admixed with regional-scale (~50-1000 km) PM burdens from Pearl River Delta source emissions. Depending on the monitoring station locations, the sites

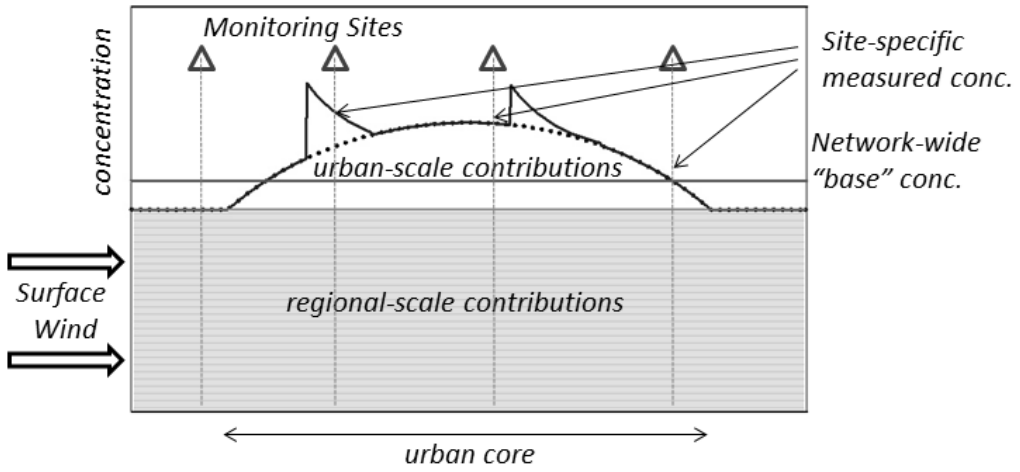


Figure C-4. Conceptualization of the influence of emissions on observed concentrations at monitoring sites over an urban area.

experience differential PM impacts from urban- and finer-scale (i.e. local-scale or < 50 km) sources in addition to the presumed homogeneous synoptic-scale impacts. The PM mass data measured across the network can now be utilized to construct a time series for a network-wide baseline to differentiate the network-wide, uniform behavior from the site-specific PM impacts. Despite the limitations posed by this simplified representation of a complex system, this approach dampens the fluctuations in synoptic-scale impacts arising from year-to-year variability in synoptic weather patterns and thereby enables a refined interpretation of the spatiotemporal patterns in site-specific PM.

The daily average TEOM mass measured at each station over 1998-2008 was apportioned into a network-wide daily baseline mass and a site-specific daily excess mass to broadly capture the PM trends representative of the general stations. Days with 24-hour average TEOM mass concentrations for at least twelve stations were defined as valid network days. The metric used to estimate daily baseline concentration is subjective and can be chosen as the minimum or x^{th}-

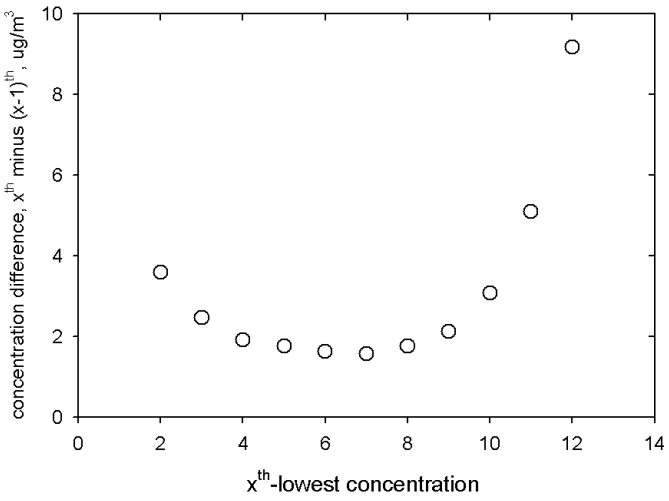


Figure C-5. Change in the grand average base concentration upon increased the metric from the $(x-1)^{th}$ lowest value to the x^{th}-lowest value.

ranked lowest measured concentration across the network of monitoring stations. Figure C-5 shows change in the grand average baseline concentration as the metric is varied from one rank to next (e.g. from the 4^{th}-lowest to the 5^{th}-lowest). As the metric was changed from the 4^{th}-lowest to the 8^{th}-lowest value, the baseline mass concentration increased by ~2 $\mu g/m^3$ per step. The general uniformity in PM_{10} concentrations across the general stations in the network allows any of these metrics to represent the base concentration with no impact on the qualitative patterns but up to a 10 $\mu g/m^3$ difference in the actual value assigned to the base. Scatter plots of the daily 5^{th}-minimum value against the daily 4^{th}- and 6^{th}-minimum values were clustered tightly about the 1:1 line, demonstrating stable concentration distribution across all general stations on a daily basis. Hence, the 5^{th}-lowest daily average TEOM mass measured across the network was chosen as the daily baseline mass concentration to represent the PM captured by the general stations. The site-specific excess mass thus refers to the daily average mass measured at each site in excess of the baseline even if the excess mass with respect to the defined baseline is negative at some sites.

C.4. Clustering of air mass trajectories

The trajectories were clustered based on refinements to the methodology of Dorling et al., 1992 as shown in Figure C-6. A subset of *N* trajectories was randomly chosen as seeds. For each trajectory, the 2D Great Circle distance separating the trajectory from the same time stamp in each seed was calculated. The separation distances were summed along the trajectory history and the trajectory was assigned to the seed which minimizes the separation distance accumulated along the trajectory path. After assigning all trajectories to a cluster, new mean trajectories were calculated to represent the clusters and replace the seeds. The entire process of assigning trajectories to clusters was repeated until there was no change in the trajectory-cluster matching with each update to the cluster-mean trajectories.

The two spatially closest clusters were identified and the cluster with the fewer trajectories was removed. The trajectory assignment and cluster pruning process was repeated to obtain results for sets of $(N\text{-}1)\rightarrow 2$ clusters. For each set of clusters, a global separation distance was calculated based on the separation of each trajectory from the mean trajectory of the respective cluster,

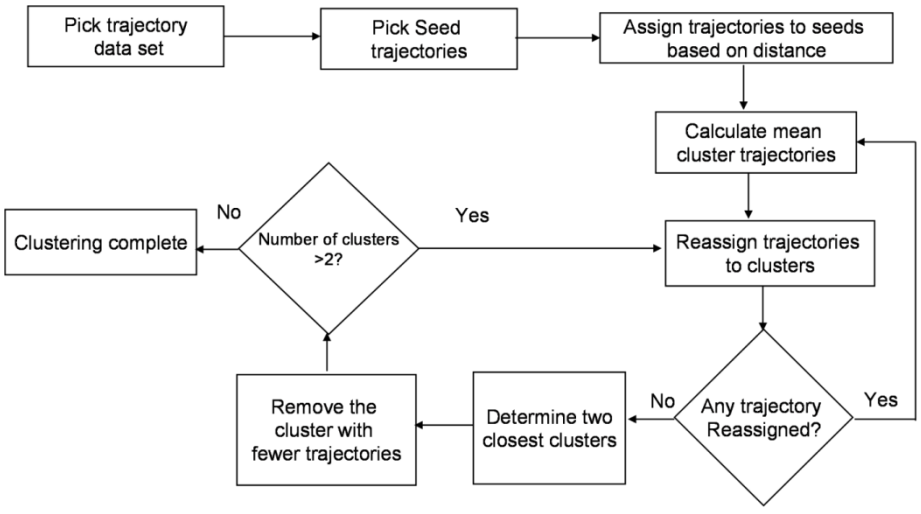


Figure C-6. Flow chart for the cluster analysis methodology used in this study.

summed over all trajectories in the dataset. The cluster set for which an increase in number of clusters did not significantly reduce the global separation distance was chosen as the final set of clusters. The results from this clustering analysis included the assignment of each trajectory to a cluster, a measure of its separation distance from the mean trajectory of the respective cluster, and a flag indicating whether the trajectory was complete (seven days) or truncated. In this study, a ten trajectory subset (N = 10 seeds) was used to eventually obtain a five trajectory cluster solution. As a sensitivity check, clustering of only 2009 trajectories with sixteen seeds was performed and resulted in a nearly identical five-cluster solution. Trajectory time stamps were converted from UTC to Hong Kong local time.

Five air mass transport patterns were resolved with the following characteristics: (1) Slow ECC – relatively slow moving air masses transported from north with the centroid trajectory of the cluster located along the eastern coast of mainland China (ECC = East Coast of China); (2) Fast ECC – relatively fast moving air masses transported from north with the centroid trajectory of the cluster located along the eastern coast of China; (3) Stagnant/circulating – stagnant/circulating air masses that resided over the greater HKSAR area for much of the seven day period; (4) S/SW – South/southwesterly flow into Hong Kong; and (5) East – Easterly flow from the Pacific Ocean into Hong Kong.

Monthly distribution of air mass transport patterns over 2000-2009 are shown in Figure C-7(a). During the winter months (September to March) of the study period, the synoptic conditions were dominated by northerly air masses nominally from the eastern portion of mainland China. Slow ECC most frequently occurred during the month of October (65%) while December was the month of most frequent occurrence of the Fast ECC (40%). These two clusters, transecting the industrial/urbanized eastern region of mainland China represent the air mass trajectories

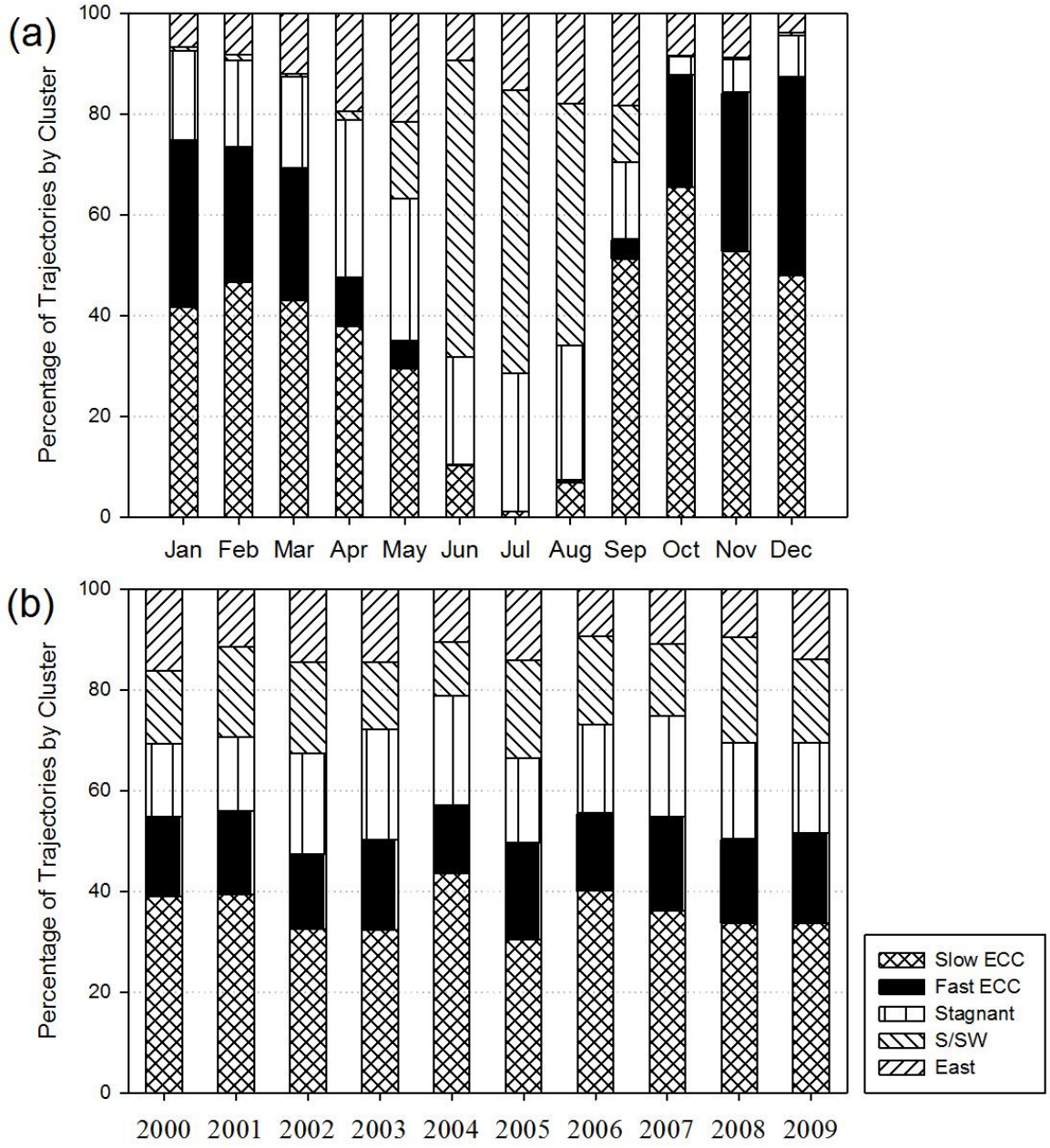


Figure C-7. Distributions of the air mass back trajectory cluster assignments by (a) month and (b) year, 2000-2009. Each day included four air mass back trajectories.

associated with high pollutant levels. Beginning in January the Stagnant/circulating air masses became more prevalent with relatively high frequency of occurrence in April (32%) and May (30%). These months represent the seasonal transition from the wintertime north/northeasterly

air masses to the summertime south/southwesterly air masses. S/SW air mass patterns prevailed during the monsoon/summer season (June (58%), July (54%) and August (50%)), with only limited occurrence of Slow and Fast ECC during June (10%), July (1%) and August (7%). The summer months also experienced higher occurrences of East air masses than other months. The occurrences of S/SW and East air masses during October through March were limited with the highest occurrence of ~13%. These oceanic air masses represent air mass trajectories associated with low pollutant levels. September was also a transition month leading into the fall/winter season with prevailing northerly air mass patterns. Figure C-7(b) shows the temporal distribution of air mass transport patterns over 2000 to 2009. The frequency of Slow ECC ranged between 30%-40% for each year with the highest frequency of occurrence in 2004 (44%). The occurrence frequency of Stagnant/circulating air mass was generally below 20%, but reached 22% in 2003 and 2004. The frequency of S/SW and East air masses combined was ~30% for each year during 2000-2009. Distinct seasonal and consistent temporal patterns in the trajectory clusters demonstrate that the seasons in Hong Kong are good proxies for corresponding air mass transport patterns.

C.5. Relationships between PM burdens and synoptic air mass patterns

Meteorological pathways leading to the observed PM burdens over the decade can be meaningfully analyzed using clustered air mass trajectories for well represented mass datasets. Daily average PM_{10} TEOM mass collected during 2000-2008 at fourteen sites allowed for site-specific PM impact assessment based on the synoptic patterns. PM_{10} speciation samples collected every year across ten sites on different days of the week; with 2-4 general stations concurrently sampling on any given day provided a large dataset with well-understood emission source linkage. The daily PMF-modeled PM_{10} contributions for each factor were averaged over

the concurrently sampling general stations for generating a daily network-wide factor contribution to exploit the uniformity in measured PM mass at general stations for maximum data completeness. This approach was also consistent with the 1° resolution of the meteorological data set used to generate the trajectories for the cluster analysis. Such spatial averaging exploits the sampling network design to yield a daily time series with 67% completeness compared to a maximum completeness of 17% for data from a single station. $PM_{2.5}$ speciation dataset collected over three one-year campaigns on the same day of the week at only 3-4 sites across the network accounts for only 5% of days overlapping with valid cluster assignment during 2000-2008; and, can introduce significant uncertainty in the results. Hence, only PMF-modeled network-average PM_{10} factor contributions and TEOM measured PM_{10} mass were analyzed using the clustered trajectories.

Air mass trajectory cluster associations of network-average PM_{10} factors are examined in Figure C-8 by plotting the PMF-resolved factor contributions; scaled such that the average contribution over all samples is unity for each factor to facilitate comparison across different source categories. PMF-resolved factors associated with non-local PM impacts, such as secondary sulfate, secondary nitrate and biomass burning exhibited similar trajectory association, with highest scaled factor contributions associated to Slow and Fast ECC clusters and lowest for S/SW. In contrast, the distribution of scaled vehicle exhaust contributions was independent of the synoptic cluster class and is consistent with characteristics of a local source. Residual oil combustion and aged sea salt factors also do not show any significant trajectory cluster association. High association with Fast ECC for crustal/soil factor suggests transport of crustal matter from the continent and/or resuspension of local soil due to fast flowing air masses impacting Hong Kong. To quantify synoptic cluster-specific variations in PM impacts for each

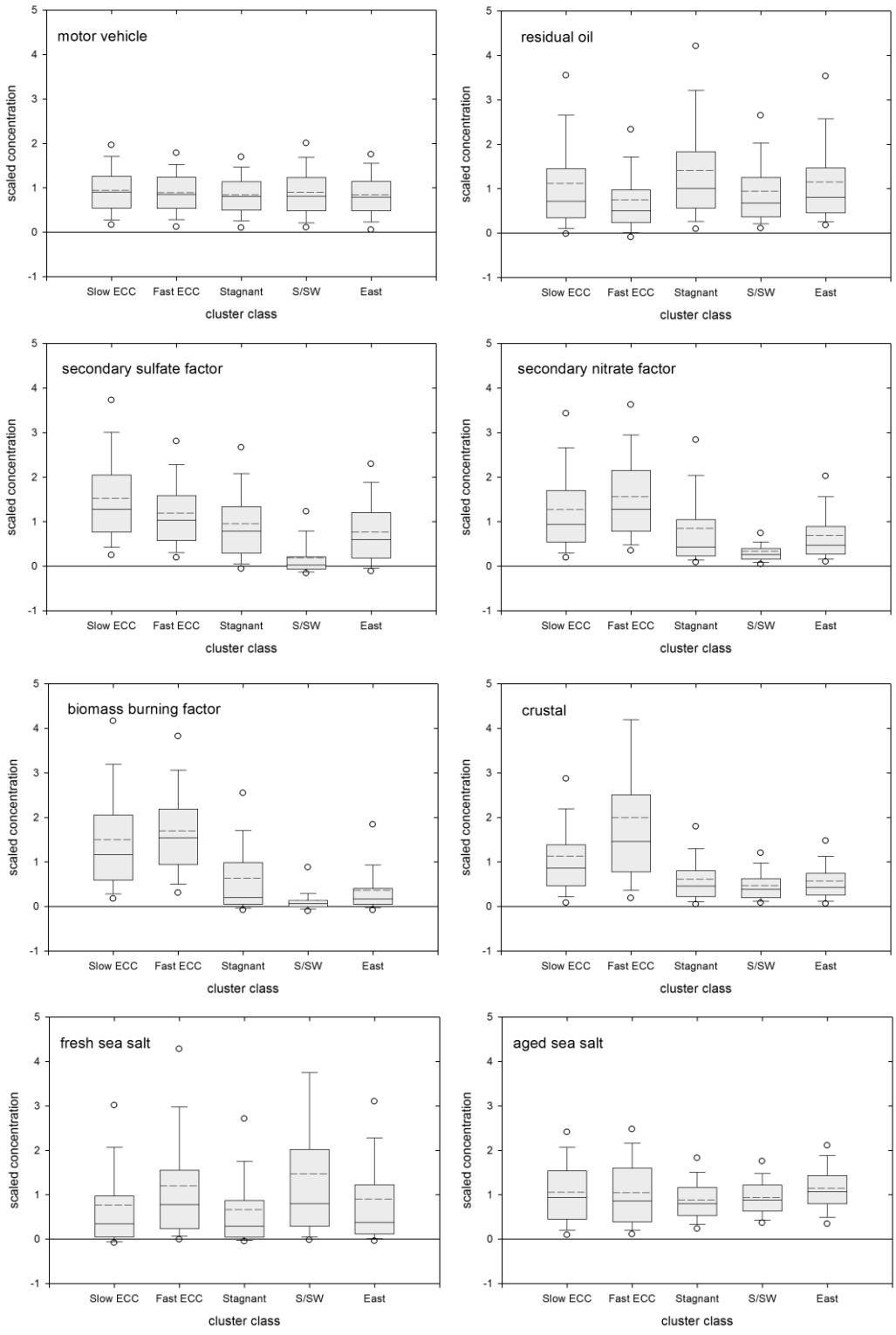


Figure C-8. Distribution of PMF-modeled PM_{10} scaled source contributions estimates corresponding to each synoptic class pattern. The interior solid line is the median, the dashed line is the arithmetic mean, and the circles are 5th and 95th percentiles.

PMF-resolved factor over 2000-2008, the annual scaled median PM_{10} contributions were regressed on year using a linear-least squares regression for each air mass cluster. The regression slopes, tabulated in Table 2 of Chapter 4, represent the linearized rate of change in scaled annual median contribution by factor and synoptic pattern. Minor changes were observed for crustal/soil and smelting factors, but no statistically significant changes were observed for residual oil combustion, fresh sea salt, aged sea salt, secondary nitrate and biomass burning factors associated with any of the trajectory clusters. On the other hand, the two largest PM contributors, vehicle exhaust and secondary sulfate factors exhibited statistically significant, complimentary changes over the decade with very distinct air mass cluster associations and are discussed in Chapter 4.

A similar analysis was also performed on the site-specific PM_{10} TEOM mass measurements to examine the changes in PM patterns on a station-by-station basis. For each station, the daily average PM_{10} TEOM mass concentrations were stratified by air mass cluster assignments to calculate the annual median mass for each cluster. For each synoptic pattern, the site-specific annual concentrations were regressed on year using a linear-least squares regression. The regression slope, representing a linearized rate of change in PM_{10} mass per year over 2000-2008 are tabulated in Table C-2 for changes larger than its 95% confidence interval.

Table C-2. Change in PM_{10} TEOM mass ($\mu g/m^3$/year) from a linear least-squares regression of annual median concentration on year for each air mass transport cluster. Values for only statistically different change from zero at the 95% confidence level are shown.

Air Mass	YL	SSP	KT	EN	TM	TC	MK	CL	CB
Slow ECC	2.6 ± 2.1				2.7 ± 1.8	2.6 ± 1.8			
Fast ECC	1.9 ± 1.8				2.3 ± 1.3				
Stagnant		-1.4 ± 1.1	-2.5 ± 1.9						-4.0 ± 2.3
S/SW					0.9 ± 0.6		-1.5 ± 0.8		-3.3 ± 1.4
East		-1.6 ± 1.5	-2.2 ± 1.3	-1.4 ± 1.0			-3.1 ± 1.6	-2.0 ± 1.1	-4.5 ± 3.5

For example, for Slow ECC air masses, PM_{10} mass at the CB roadside station decreased at a linearized rate of 1.2 μg/m^3/year but this change was statistically indistinguishable from zero at the 95% confidence level (2.6 μg/m^3/year). In contrast, the increase in PM_{10} mass during 2000-2008 by 2.6-2.7 μg/m^3/year at TC, YL and TM is statistically significant for the Slow ECC cluster. For oceanic air masses (S/SW and East), the change in PM_{10} mass was either statistically indistinguishable from zero or decreased irrespective of the site characteristic (roadside or general). The increase in PM_{10} mass at TM station, located to the northeastern edge of Hong Kong, for S/SW air masses could result from the air masses transecting the urban core of Hong Kong.

C.6. References

Chow, J.C.; Watson, J.G.; Lu, Z.; Lowenthal, D.H.; Frazier, C.A.; Solomon, P.A., Thuillier, R.H.; Magliano, K.; Descriptive analysis of $PM_{2.5}$ and PM_{10} at regionally representative locations during SJVAQS/AUSPEX. *Atmospheric Environment*, **1996**, 30(12), 2079– 2112.

Chow, J.C.; Watson, J.G.; Crow, D.A.; Lowenthal, D.H.; Merrifield, T.; Comparison of IMPROVE and NIOSH carbon measurements. *Aerosol Science & Technology*, **2001**, 34(1), 1–12.

Chow, J.C.; Watson, J.G.; Review of $PM_{2.5}$ and PM_{10} apportionment for fossil fuel combustion and other sources by chemical mass balance receptor model. *Energy and Fuels*, **2002**, 16, 222‑260.

Dorling, S.R.; Davies, T.D.; Pierce, C.E.; Cluster analysis: A technique for estimating the synoptic meteorological controls on air and pollution chemistry. *Atmospheric Environment*, **1992**, 26A, 2575-2581.

Ho, K.F.; Lee, S.C.; Chan, C.K.; Yu, J.C.; Chow, J.C.; Yao, X.H.; Characterization of chemical species in PM2.5 and PM10 aerosols in Hong Kong. *Atmospheric Environment*, **2003**, 37, 31–39.

Kim, M.J.; Park, R.J.; Kim, J.J.; Urban air quality modeling with full O_3-NO_x-VOC chemistry: Implications for O_3 and PM air quality in a street canyon. *Atmospheric Environment*, **2012**, 47, 330-340

Larsen, R.K.; Baker, J.E.; Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. *Environmental Science & Technology*, **2003**, 37, 1873-1881.

Lau, A.K.H.; Yuan, Z.; Yu, J.Z.; Louie, P.K.K.; Source apportionment of ambient volatile organic compounds in Hong Kong. *Atmospheric Environment*, **2010**, 408, 4138-4149.

Lee, J. H.; Hopke, P.K.; Turner, J.R.; Source identification of airborne $PM_{2.5}$ at the St. Louis-Midwest Supersite. *Journal of Geophysical Research: Atmospheres,* **2006**, 111, D10S10, doi:10.1029/2005JD006329.

Polissar, A.V.; Hopke, P.K.; Paatero, P.; Malm, W.C.; Sisler, J.F.; Atmospheric aerosol over Alaska 2. Elemental composition and sources. *Journal of Geophysical Research,* **1998**, 103, 19045-19057.

Ramadan, Z.; Song, X.H.; Hopke, P.K.; Identification of source of Phoenix aerosol by positive matrix factorization. *Journal of Air & Waste Management Association*, **2000**, 50: 1308–1320.

Russell, L.M.; Pandis, S.N.; Seinfeld J.H.; Aerosol production and growth in the marine boundary layer. *Journal of Geophysical Research*, **1994**, 99, 20989-21003.

Saul, T.D.; Tolocka, M.P.; Johnston, M.V.; Reactive uptake of nitric acid onto sodium chloride aerosols across a wide range of relative humidities. *The Journal of Physical Chemistry A*, **2006**, 110, 7614-7620.

Song, X.H.; Polissar, A.V.; Hopke, P.K.; Source of fine particle composition in the northeastern U.S. *Atmospheric Environment,* **2001**, 35, 5277-5286.

Thurston, G.D.; Spengler, J.D.; A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment*, **1985**, 19, 9-25.

U.S. Environmental Protection Agency (USEPA), **2007**. EPA Unmix 6.0 - Fundamentals and user guide. http://www.epa.gov/heasd/products/unmix/unmix.html (accessed 21.03.13)

Wang, G.; Hopke, P.K.; Turner, J.R.; Using highly time resolved fine particulate compositions to find particle sources in St. Louis, MO *Atmospheric Pollution Research* 2, **2011**, 219-230

Yadav, V.; Turner, J.R.; Yuan, Z.; Lau, A.K.H.; Louie, P.K.K.; Assessment of measurement error for the Hong Kong PM_{10} and $PM_{2.5}$ speciation datasets. *To Be Submitted in* **2013**.

Yuan, Z.; Yadav, V.; Turner, J.R.; Louie, P.K.K.; Lau, A.K.H., Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies in Hong Kong over 1998-2008. *Atmospheric Environment*, **2013**, 76, 21-31.

Curriculum Vita

Varun Yadav

Doctor of Philosophy (Ph.D.) in Energy, Environment and Chemical Engineering, 2013.
Department of Energy, Environment and Chemical Engineering,
Washington University in St. Louis, St. Louis, MO, USA

Bachelor of Technology (B.Tech.) in Civil Engineering, 2006.
Department of Civil Engineering,
Indian Institute of Technology (IIT) Bombay, Mumbai, India

Courses taken at Washington University in St Louis:

- Transport Phenomena in Energy, Environmental and Chemical Engineering
- Mathematical Methods in Engineering
- Aerosol Science and Technology
- Kinetics and Reaction Engineering Principles
- Sustainable Air Quality
- Remote Sensing
- Mathematics of Modern Engineering I
- Environmental Chemistry
- Partial Differential Equations
- Measurement Techniques for Particle Characterization

Publications

Yadav, V.; Turner, J.; Downs, J.; Rowles, T.; "Gauging intraurban variability of ambient particulate matter from network of monitoring sites"
(Submitted to Atmospheric Environment on 18 October 2013)

Yadav, V.; Yuan, Z.; Louie, P.K.K.; Lau, A.K.H.; Turner, J.R.; "A weight of evidence approach using network datasets to characterize drivers of ambient particulate matter air quality in Hong Kong"
(Under review with co-authors)

Yadav, V.; Yuan, Z.; Louie, P.K.K.; Lau, A.K.H.; Turner, J.R.; "Assessing measurement error using collocated ambient particulate matter speciation datasets from Hong Kong"
(Under review with co-authors)

Yadav, V.; Turner, J.; "Estimating local emission source zones using high-time resolution carbon measurements across a monitoring network in a multi-source industrial area."
(In preparation)

Yuan, Z.; Yadav, V.; Turner, J.R.; Louie, P.K.K.; Lau, A.K.H.; "Long-term trends of ambient particulate matter emission source contributions and the accountability of control strategies in Hong Kong over 1998-2008", *Atmospheric Environment*, **2013**, 76, 21-31.

Feinberg, S.N.; Yadav, V.; Heiken, J.G.; Turner, J.R., "Midwest Rail Study: Modeled Near-Field Impacts of $PM_{2.5}$ Emissions from Railyard Activities", *Transportation Research Record*, 106-114, **2012.**

Conference Proceedings

Feinberg, S.N.; Yadav, V.; Turner, J.R., "Weight of Evidence Analysis for Attributing Monitored Sulfur Dioxide Exceedances to Emission Source Zones". Paper #13326. *Air and Waste Management Association 106th Annual Conference*, Chicago, IL, June 26-28, **2013**.

Yadav, V., Turner, J.; Yuan, Z.; Lau, A.; Louie, P.; "Assessment of Measurement Error for the Hong Kong PM_{10} and $PM_{2.5}$ Speciation". Paper # 66. *A&WMA Air Quality Measurement Methods and Technology Conference*, Durham, NC, April 24-26, **2012**.

Turner, J.; Yadav, V.; Feinberg, S.N.; Du, L; "Ambient Particulate Matter: Intraurban Variability". *Gateway Society - Hazardous Material Managers (GSHMM) Regulatory Update Seminar*, St. Louis, MO, April 21, **2012**. [Platform Presentation]

Yadav, V., Turner, J.; Yuan, Z.; Lau, A.; Louie, P.; "A Conceptual Model for Ambient Particulate Matter over Hong Kong". *American Association for Aerosol Research (AAAR) 30th Annual Conference*, Orlando, FL, October 3-7, **2011**. [Platform Presentation]

Feinberg, S.N.; Yadav, V.; Heiken, J.G.; Turner, J.R.; "Midwest Rail Study: Modeled Near-Field Impacts of $PM_{2.5}$ Emissions from Railyard Activities" *Transportation Research Record (TRB) 90th Annual Meeting*, Washington D.C., January 23-27, **2011**. [Poster]

Yadav, V., Turner, J.; Petterson, K.; Cooper, J.; "Field Performance Evaluation of the Cooper Environmental Services Ambient Metals Monitor (Xact 620) for near-Real Time PM_{10} Metals Monitoring". *American Association for Aerosol Research (AAAR) 29th Annual Conference*, Portland, OR, October 25-29, **2010**. [Platform Presentation]

Robinson, A.; Heindorf, M.A.; Hodges, C.; Koerber, M.; Turner, J.; Yadav, V.; "Neighborhood-Scale Spatial and Temporal Patterns for Ambient Carbonaceous Particulate Matter in Detroit, Michigan". Paper No. 2009-A-314-AWMA. *A&WMA 102nd Annual Conference*, Detroit, MI, June 16-19, **2009**.

Yadav, V.; Turner, J.; Downs, J.; Rowles, T.; "Intraurban Variability of Ambient Particulate Matter Arsenic and Other Air Toxics Metals in St. Louis". Paper No. 2009-A-370-AWMA. *Air and Waste Management Association 102nd Annual Conference*, Detroit, MI, June 16-19, **2009**. [Platform Presentation]

Yadav, V.; Turner, J.; Downs, J.; Rowles, T.; "High Time Resolution PM_{10} Metals by the Cooper Environmental Services Ambient Metals monitor (Xact 620): Field Performance Evaluation and Data trends for St. Louis". Paper No. 2009-A-146-AWMA. *Air and Waste Management Association 102nd Annual Conference*, Detroit, MI, June 16-19, **2009**.

Wetherell, W., Downs, J.; Rowles, T.; Turner, J.; Yadav,V.; "Operating Experience with a Near-Real-Time Ambient Airborne Metals Instrument System (Cooper Environmental Systems Xact 620)". *National Air Monitoring Conference*, Nashville, TN, Nov 2-5, **2009**.

Turner, J., Yadav, V.; Downs, J.; Wetherell, W.; Rowles, T.; "Measurements and Data Analysis for Source Attribution of Ambient Particulate Matter Arsenic and Other Air Toxics Metals in St. Louis". *National Air Monitoring Conference*, Nashville, TN, Nov 2-5, **2009**.