SOURCE APPORTIONMENT OF POLYCYCLIC AROMATIC HYDROCARBONS IN DIVERSE ENVIRONMENTAL MEDIA USING RECEPTOR AND PHYSICO-CHEMICAL MODELS.

By

Uwayemi Modesoyin Sofowote, B.Sc.

A Thesis Submitted to the School of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

McMaster University

Doctor of Philosophy (2010)	McMaster University
(Chemistry)	Hamilton, Ontario

TITLE:	Source apportionment of polycyclic aromatic hydrocarbons in diverse environmental media using receptor and physico-chemical models.
AUTHOR:	Uwayemi Modesoyin Sofowote, B.Sc. (University of Ibadan)
SUPERVISOR:	Professor B.E. McCarry
NUMBER OF PAGES:	xv, 166

Abstract

This thesis focused on the interpretation of environmental data obtained from chemical analyses of complex and divergent environmental media. Environmental monitoring of pollutants yields large datasets which may be difficult to process using simplistic univariate approaches. Polycyclic aromatic hydrocarbons (PAH) are environmental organic contaminants that have important health implications and are ubiquitous in the environment. These compounds are regularly monitored in all environmental compartments but delineating their impacting sources from analytical data is usually challenging and thus not routinely attempted.

The work contained in this thesis discusses how source information can be obtained by the application of (a) qualitative relationships, (b) physico-chemical models dealing with phase change and partitioning, as well as (c) receptor models using factor analysis to PAH data collected in samples from the atmosphere and in suspended sediments.

There are unique concepts and new applications embodied in this thesis. They include: (a) The development and definition of a new diagnostic relationship that discriminates between industrial and urban sources of PAH, (b) the integration of qualitative diagnostic correlations with measured concentration data to improve the diagnostic potential of these simple non-multivariate relationships, (c) the integration of local meteorological data with receptor modeling using factor analysis to explicitly identify impacting PAH source regions, and finally (d) the first known instance of the use of a hybridized receptor modeling methodology to explicitly identify, quantify and spatially map the potential source regions of PAH on a hemispheric scale.

An additional effect of the work contained herein is that it shows how much information which could be useful to environmentalists and policy-makers in designing effective abatement and control strategies if available is usually left un-extracted from environmental monitoring activities.

Preface

This publication-based thesis contains results from five years of research at McMaster University, NWRI and NLET by the author. The focus of the thesis is the application and integration of many source apportionment techniques for the comprehensive assessment of sources of polycyclic aromatic hydrocarbons in diverse environmental media. The work herein is derived from three research projects: The first project was in collaboration with Dr. Chris Marvin of the National Water Research Institute, Environment Canada. The second was a re-processing and re-interpretation of archival data originally sourced by Dr. Laurie Allan while the third project was in conjunction with Dr. Hayley Hung of the Air Quality Research Division, Environment Canada.

Acknowledgements

Firstly, I am indebted to my supervisor Dr. Brian E. McCarry, who has allowed me to do meaningful research in his laboratory for five years. His understanding of my abilities coupled with the constant intellectually-stimulating discussions fostered in me the capacity for independent and critical thinking. These life-changing potentials can only be realized in the proximity of the very best mentors.

I am lucky to have interacted with and learned from Dr. Philip Britz-McKibbin who introduced me to multivariate analyses and fundamental chemometrics. I am also grateful to have met and worked with Dr. Chris Marvin and Dr. Hayley Hung at various periods of my graduate research. Thank you for the opportunities you afforded me to learn and develop.

My lab colleagues, past and present and people I met at McMaster University who are now my good friends: David Dam, Sujan Fernando, Ruzni Hanifa, Dr. Libia Saborido, Roger Luckham, Catherine Amoateng, Ken Chalcraft, Tarlika Persaud and Elna Deglint. Of course, huge portions of my research would have been totally impossible without timely help and assistance from Ankit K. Rastogi.

To all my numerous collaborators- Johnny Westgate, Ed Sverko, Ivy D'Sa, Dr. Yushan Su, Julia Jia, Dr. Laurie Allan, NLET Ultra-trace Laboratory staff and the many others, I am grateful for all the help and consideration when it mattered most.

Finally, I would like to thank my family who have been nothing but supportive all these years: My sister – Lorade Meseko, my two brothers – Wande Sofowote and Bibi Sofowote, my aunt – Mrs Modupe Orebiyi, and my Dad – Segun Sofowote, thank you so much. To my Mom – Motunlayo Adefunke Sofowote, I must give special thanks for the continuous encouragement and unflinching affection.

Table of Contents

List of Figures

Chapter One:

Chapter Two:

Chapter Three:

Chapter Four:

Chapter Five:

Chapter Six:

List of Tables

Chapter One:

Chapter Two:

Chapter Three:

Chapter Four:

Chapter Five:

Chapter Six:

List of Abbreviations

AOC: Area of concern
ASE: Accelerated solvent extraction
CMC: Canadian Meteorological Centre
DW: Downwind
EI: Electron ionization
EPA: United States Environmental Protection Agency
FA: Factor analysis
GC/MS: Gas chromatography coupled to mass spectrometry
GFF: Glass fibre filter
Hi-Vol: High volume
HYSPLT: Hybrid single-particle Lagrangian integrated trajectory
INCATPA: Intercontinental atmospheric transport of anthropogenic pollutants to the Arctic
IPY: International Polar Year
LFL: Little Fox Lake
MDL: Method detection limit
ME: Multilinear engine
MLR: Multilinear regression
MODIS: Moderate resolution imaging spectroradiometer
NASA: National Aeronautics and Space Administration
NIST: National Institute of Science and Technology
NLET: National Laboratory for Environmental Testing
PAC: Polycyclic aromatic compounds
PAH: Polycyclic aromatic hydrocarbon
PASH: Polycyclic aromatic sulphur heterocycle
PBDE: Polybrominated diphenyl ether
PCA: Principal component analysis
PCB: Polychlorinated biphenyl
PM$_{2.5}$: Particulate matter with 2.5 micron diameter size
PM$_{10}$: Particulate matter with 10 micron diameter size
PMF: Positive matrix factorization
POP: Persistent organic pollutant
PSCF: Potential source contribution function
PUF: Polyurethane foam
S.D.: Standard deviation
SRM: Standard reference material
TPAH: Total polycyclic aromatic hydrocarbons

TSP: Total suspended particulate
UTC: Coordinated universal time
UW: Upwind

Chapter One

Introduction

The study of contaminants in the environment can be categorized into broad groups dealing with their sources and fates in multiple environmental compartments. The fate of a contaminant in the environment is studied to determine its longevity or persistence and its potential adverse impacts on the ecosystem (Halsall 2007) while the study of the nature and significance of sources of the contaminants is required to develop effective control and abatement strategies. Thus, source apportionment studies apply principles of chemistry and multivariate statistics to delineate underlying sources of environmental contaminants.

The application of modern instrumentation and principles of analytical chemistry will allow environmental scientists to determine the qualitative and quantitative characteristics of individual contaminant species in a given environmental matrix with high sensitivity and selectivity. One of the challenges for the environmental scientist is to determine the significance of the results from a source apportionment perspective.

The objective of my graduate studies was to integrate qualitative and quantitative source apportionment models for comprehensive interpretation of pollution data. Polycyclic aromatic hydrocarbons (PAH) and related compounds were the subject of the source apportionment studies carried out in different environmental matrices throughout the five year period of my doctorate research. PAH are a class of semi-volatile organic contaminants composed of multiply fused benzene rings formed from incomplete combustion of fuels and organic matter in general. Some PAH are known to have mutagenic and carcinogenic properties (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans and International Agency for Research on Cancer 2010; Legzdins et al. 1995; Metcalfe et al. 1990). Unlike other organic pollutants, PAH are difficult to apportion to any one source due primarily to their ubiquitous presence in the environment and their generation from multiple pollution source types, anthropogenic or natural. They are found in all environmental compartments comprising air, water, sediments and soils. Since these compounds are ubiquitous and sometimes harmful, it may be necessary to delineate their sources and determine their relative contributions so as to develop targeted control measures.

1.1. What is source apportionment?

Source apportionment of pollutants in the environment is an integrated discipline which requires the knowledge of analytical chemistry and the use of mathematical approaches (that range in complexity from univariate /marker speciation techniques to multivariate analysis) for the identification and possible quantification of the impacts of pollutional sources.

It is usually important to determine the source(s) of a contaminant with elevated levels in the environment to formulate effective strategies and policies for its control. This is relatively straightforward when a point source is not far from the sampling (receptor) site. In the absence of a point source in close proximity to the sampling site, simple source-receptor relationships are not straightforward, i.e., it may not be easy to delineate the sources responsible for an increased level of a pollutant, thus, the development of proper control strategies of the pollution may be impossible if source apportionment is not accomplished. Source apportionment requires a keen knowledge of the underlying physico-chemical and mathematical principles governing pollutant behaviour in the environment.

Source apportionment techniques can either be simplified or multivariate. An excellent example of the power of simplified source apportionment using well-established molecular markers is the work of researchers who traced tar balls that washed up the coasts of Malaysia to crude oils from the Middle East and South East Asia even though no point source was proximate to the site (Pauzi Zakaria, Okuda and Takada 2001). The relatively easy approach they employed was based on the knowledge of the relative distribution of time-invariant (non-degradable) aliphatic species in the collected samples on comparison to known potential sources (in this case, crude oils). Such unique species are usually called molecular markers or biomarkers in environmental studies. The successful use of environmental markers for source apportionment will yield qualitative information regarding the impacting sources. This level of information may however still be insufficient for the development of more aggressive control measures. Thus, simplified source apportionment approaches suffer from the problems below that limit their wider applicability.

1. *Only very few true 'marker' compounds exist*: Invariably, all chemical species will eventually undergo some form of physical and/or chemical transformation in all environmental compartments. These transformations range from minor to severe and may alter the relative distributions of markers. Noted

examples of useful biomarkers are hopanes and steranes which are homocyclic aliphatic hydrocarbons arising from the bacterial processes involved in the formation of crude oil. An excellent example that illustrates the suitability of these compounds for source tracing work due to their time-invariance is the work of researchers who were able to identify the source of a twenty-two year old crude oil sample based on the relative distributions of these compounds (Wang, Fingas and Sergy 1994). Individual PAH compounds are not usually considered source markers due their generation from multiple sources. Retene, is a special case, produced in a specific synthetic pathway from abietic acid found in conifers (Ramdahl 1983), and has been used to trace softwood combustion emissions (Gogou et al. 1996).

2. *Not all sources of pollution have distinct markers*: Pollution sources related to combustion and natural emissions all generate a wide range of pollutants which overlap significantly from one source type to the other. An example is the generation of ubiquitous polycyclic aromatic hydrocarbons (PAH) from all known combustion sources leading to a lack of specificity or uniqueness. Specificity can still be potentially achieved if different sources yield different but *constant* distributions of the pollutant compounds. The use of the relative distributions/ratios of isomers or homologues of ubiquitous species such as PAH has been attempted over the years for source apportionment e.g. (Yunker et al. 2002; Bence, Kvenvolden and Kennicutt II 1996). These approaches are however simplistic estimators of sources for the following reasons: - *1.* Many source types do not generate pollutant species in any consistent manner thus ratio thresholds used for source identifications lack specificity. Galerneau recently reviewed the specificity of many source types and showed that the ranges employed for unique diagnosis overlapped significantly among all source types (Galarneau 2008). *2.* Even PAH ratios measured in samples impacted solely by a single source can change significantly over time. Kim et al. recently reviewed the effect of photolysis on commonly used diagnostic ratios for PAH sorbed to representative environmental substrates and found significant changes over a 500-hour period (Kim et al. 2009). There are no generalized ways to correct for these changes which may lead to misidentifications of sources (especially when the source

identifications are based on literature-reported ratio thresholds) if no other models are employed for source apportionment. Thus, the use of these ratios may only be feasible if the potential sources are geographically close to the sampling site.

3. *Quantitative source apportionment using simplified approaches is not generally feasible*: The identification of a unique marker may be able to give information regarding the impacting source, but the contribution (relative or absolute) of that impact is not easily extractable from the measurement of the abundance of the unique species. Other non-marker species which may be more important from a health perspective may also be present in the source of pollution, thus a study of the co-variation of species in pollutant source types may be useful. This multi-variable approach not only better accounts for more species in the source thus increasing the reliability of the source apportionment outcomes but gives additional insight to the origin of the source of pollution. In other words, quantitative source apportionment requires multivariate approaches (Larsen and Baker 2003; Sofowote, McCarry and Marvin 2008)

Multivariate approaches to source apportionment can be readily understood in the context of receptor modeling. They are more complex to understand and use, but they give a more rounded picture of the impacting sources of pollution. Simply explained, if a receptor site is impacted by multiple sources of pollution which contribute varying amounts of a given pollutant, then the total amount of the pollutant measured or sampled is the sum of the product of the pollutant abundance in the sources *and* the fractional contribution of the sources to the sample (the figure below summarizes the definition above and a detailed discussion on receptor modeling is given elsewhere in this introduction). The import of this statement is profound; what the analytical chemist perceives as a simple value of concentration of a given pollutant in a data matrix can be mathematically de-convoluted into different sources and contributions. Methods to delineate these sources (or more correctly the source profiles) and their contributions include chemical mass balance and factor analysis. As implied from the definition above, the quantitative apportionment obtained with receptor modeling is an intrinsic characteristic of the model formulation. The complexity of these receptor models arises from the requirement to resolve each concentration value in the data matrix into sources and contributions.

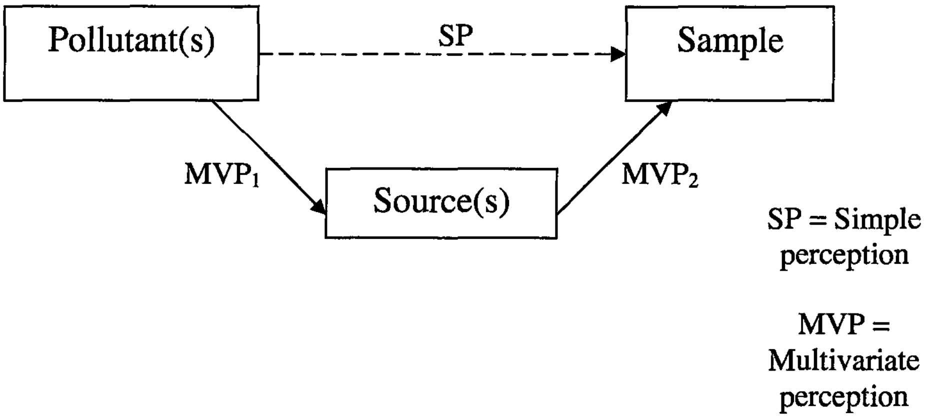


Figure 1: Diagram summarizing the goal of multivariate analysis in source apportionment using receptor modeling. Dashed line (SP) represents simple perception of concentration values (pollutant in sample) in a data matrix. Solid lines indicate the multivariate steps it actually takes to yield the values in the data matrix. MVP_1 = Pollutant(s) in source(s) values or source profiles. MVP_2 = Source(s) in sample values or source contributions.

Multivariate receptor modeling may also be affected by chemical transformation/depletion processes which may lead to spurious identifications when dealing with species which are relatively susceptible to transformations (such as PAH) but because source characterizations are done using entire factor profiles and not a few compounds, the effect of these transformations on source identifications is lesser. In all cases, knowledge of the dominant transformation processes affecting the physical state (vapour phase or substrate-bound) of the pollutant species measured in their given environmental compartment will help determine the actual feasibility of the source apportionment outcomes. Thus, for relatively susceptible pollutants (e.g. PAH), source apportionment using multivariate models should yield more reliable outcomes when these compounds are examined in the bound-phase (i.e., atmospheric particulate, bottom sediment and soil samples) where they are less available for chemical transformation processes.

1.2. Source apportionment methods used in environmental studies

The discussion below touches on the types of source apportionment methods used in environmental contaminant studies with special focus on PAH.

These methods can be regarded as either qualitative (e.g. the use of unique molecular markers and diagnostic ratios) or quantitative methods (e.g. multivariate statistical analyses). In general, qualitative methods are usually uni-variate or bi-variate in nature and may not be representative of the entire pollution data set. Thus, they are relatively easier to understand and use but are not as reliable as multivariate models which also offer the possibility for quantitative apportionment.

1.2.1. Unique molecular markers

A source of pollution may be characterized qualitatively by the distinct presence of a given 'marker' compound, for instance, in the past, enhanced levels of lead in the atmosphere have been used as a unique tracer for gasoline emissions (Greenberg et al. 1981) . The use of unique tracers for source apportionment is problematic as enumerated above. As an example, the use of lead as an anti-knocking agent has been discontinued in gasoline. Organic compounds such as PAH have also been reported as tracers of pollution but their use is unreliable since they are generated from multiple source types, thus true PAH source tracers do not exist. Retene is an exception and has been extensively used as a resinous wood combustion marker since it is a specific by-product of the thermal cracking of abietic acid (Ramdahl 1983).

1.2.2. Diagnostic ratios

Diagnostic ratios are particularly suited to organic species such as alkanes and PAH although some ratios of metals have been used to account for specific sources of pollution (Cheng et al. 1993). Diagnostic ratios of PAH (Yunker et al. 2002; Bence, Kvenvolden and Kennicutt II 1996; Fang et al. 2006; Gogou et al. 1996) can be employed in bi-variate plots to offer more qualitative source discrimination of PAH. Bi-variate plots are more reliable for source apportionment (i.e., the outcomes are more representative of the impacting source types) than the use of single marker compounds since at least four compounds would have been used to construct the plot. This approach gives useful results primarily when there are only one or two primary sources. A small number of chemically distinct impacting sources will ensure that ratio thresholds are fairly unique. The diagnostic ratio approach is not limited to homocyclic PAH; ratios of 3- and 4-ringed polyaromatic sulfur heterocycles (PASH or thia-arenes) have been used to distinguish between coal tar pollution and diesel exhaust emissions (McCarry et al. 1996) in air particulate and sediments (Marvin et al. 2000). PAH are however known to undergo chemical and photolytically-initiated oxidation depending on type of PAH and physico-chemical properties of the environmental compartment in which they are found. Thus, a diagnostic ratio based on compounds which may have undergone environmental degradation is questionable and any source apportionment based solely on this is incomplete. Also, thresholds of PAH ratios are not as diagnostic or unique as once believed. In a recent review on source specificity of airborne PAH, Galarneau (2008) showed that the diagnostic power of PAH ratios commonly used for source apportionment is diminished since the huge intra-variability in the source signatures reduces the specificity of the ratios.

1.2.3. Receptor Modeling

In complex urban-industrialized environments, sources of pollution need to be identified to develop meaningful control and abatement strategies for the contaminants they generate. Some contaminant species found in a given matrix are not derived from a single source, for example, PAH arise from multiple sources. Thus, quantitative source apportionment of ubiquitous, well mixed contaminants will require multivariate statistics. The use of multivariate models in environmental applications is referred to as receptor modeling, a term that has been used primarily in air quality monitoring studies to describe a set of mass apportionment models to account for sources of pollution and their relative contributions at a receptor (sampling) site (Hopke 1985). The two main approaches to receptor modeling are factor analyses and chemical mass balance; the latter being the method of choice when the impacting sources and their profiles have been pre-determined (Bi et al. 2007). Factor analyses, on the other hand, identify and quantify factors contributing to pollution data sets without prior knowledge of impacting source profiles or source characteristics (Guo, Wang and Louie 2004). This approach has been applied in studies involving atmospheric pollutants, both organic (Larsen and Baker 2003) and inorganic (Almeida et al. 2006), to sediments (Christensen and Bzdusek 2005) and soils (Wang et al. 2007). 'Factor' and 'source' in many literature texts are used interchangeably.

The table below (Table 1) details some important studies dealing with many aspects and techniques of source apportionment of PAH in a chronological order.

Table 1: Summary of some selected studies dealing with the development of source apportionment of PAH from early qualitative comparisons to advanced factor analytic receptor modeling.

Study	Year	Purpose of Study	PAH Source Apportionment Significance
(Grimmer, Naujack and Schneider 1982)	1982	Differentiation of gasoline emissions- impacted and coal-impacted source profiles	This study showed that PAH profiles were different for these two important source types
(Rogge et al. 1993a; Rogge et al. 1993b)	1993	Generation of chemical profiles of organic contaminants associated with fine aerosols	The source profiles (including PAH) generated from these studies can be used in CMB or FA for source identifications
(Wild and Jones 1995)	1995	Generation of multi-compartment inventories of PAH in an urban environment	Creation of UK source inventories which can be used for identifications of factors are detailed
(Harrison, Smith and Luhana 1996; Kavouras et al. 2001)	1996, 2001	Application of factor analyses to contaminant data including PAH from urban particulate to extract generic and specific pollutional sources	These studies showed that PAH can be combined with other inorganic or aliphatic source tracers for identification of pollutional sources by FA-regression methods
(Schauer et al. 1999)	1999	Profiling organic contaminants in emissions from mobile sources	Factor profiles of organic contaminants including PAH were generated for mobile diesel emissions
(Allan and McCarry 1999)	1999	Use of thia-arenes (PASH) for source apportionment	Reviewed the development of many PAH ratio-based source apportionments
(Yunker et al. 2002)	2002	Source apportionment of PAH using diagnostic ratios	Compilation of studies describing and applying common PAH ratios for source apportionment
(Larsen and	2003	Source apportionment of	First study to apply PMF

Baker 2003)		PAH from urban air	to stand-alone PAH data for identification and quantification of factors
(Bzdusek et al. 2004)	2004	Source apportionment of PAH from water sediments	First study to apply FA with non-negativity constraints to PAH from sediments
(Galarneau 2008)	2008	Examination of the diagnostic qualities and applicability of commonly used PAH ratios for source apportionment	Details the pitfalls of using PAH ratios for source apportionment. Suggestions for improvements of source apportionments were also made
(Brandli et al. 2008)	2008	Factor analytical approach for source apportionment of PAH collected in soils from Europe	Compilation of PAH profiles (supplementary information) of many source types

1.3. Receptor Models used in this thesis

Two factor analysis methods were extensively used in my research, they are: principal component analysis (PCA) and positive matrix factorization (PMF). PCA is the commonest and easiest factor analytical tool used for receptor modeling (Almeida et al. 2006); for mass apportionment applications, multiple linear regression (MLR) of the standard deviate of the measured mass or concentration is applied to principal component scores (Larsen and Baker 2003; Harrison and Yin 2000). PCA often yields unrealistic (negative) source profiles and contributions, thus reducing the accuracy of the factor analysis. Other receptor models have been developed to overcome the negativity shortcomings of PCA. Positive matrix factorization (PMF), a receptor modeling tool developed in the early 1990s by Paatero and Tapper (1993), is one such model that has been applied to source apportionment of aerosols in large cities,(Lee et al. 2003), in bulk wet deposition (Anttila et al. 1995) and in characterization of contaminated soils (Vaccaro et al. 2007) and sediments (Bzdusek and Christensen 2006; Bzdusek, Lu and Christensen 2006).

While factor analyses may yield profiles/loadings of all included variables and their relative contributions/scores to samples collected at a receptor site, they do not explicitly reveal a potential geographical source region of the identified factors impacting the site. The spatial mapping of sources of pollution is particularly critical in studies assessing long-range impacts. Meteorological information in the form of air mass back trajectories can be intelligently employed to provide more information about the atmospheric transport pathway of contaminant species collected at a receptor site and by extension, their potential source regions. Potential source contribution function PSCF (Ashbaugh, Malm and Sadeh 1985) is one form of residence time analysis of back trajectory endpoints within a fixed (grid cell) space. PSCF constitutes the conditional probability that a given grid cell is a source region of the contaminant species of interest if air mass trajectories passing through the cell on their way to the receptor site contain measured levels of the contaminant above a subjectively pre-defined threshold. Many studies have employed PSCF to give a spatial rendering of contaminant species including inorganic components of aerosols (Zeng and Hopke 1989; Ara Begum et al. 2005). In the 1990s, researchers studied aerosols collected in the Canadian High Arctic (Alert, Nunavut) during the winter months using multi-layered back trajectories for their PSCF computations and identified geographically disperse source regions in Asia and Europe (Cheng et al. 1993) for non-marine sulphate and some inorganic constituents of air particulates. Source regions of organic contaminants have also been spatially mapped using the PSCF technique. In continental North America, researchers have explored the source regions of gas-phase PAH, polychlorinated biphenyls (PCB) and pesticides sampled at receptor sites around the Laurentian Great Lakes (Hafner and Hites 2003). Hsu et al. (Hsu, Holsen and Hopke 2003) explored the source regions of PCB measured at sampling sites in Chicago.

PSCF has also been combined with factor analytic models such as principal component analysis (Zeng and Hopke 1989) and more recently, PMF, for spatial mapping of source profiles identified by the factor analyses. The PMF-PSCF methodology has been employed by the research group of Philip Hopke (*e.g., see* Hopke, Zhou and Poirot 2005; Poirot et al. 2001; Pekney et al. 2006; Polissar, Hopke and Poirot 2001; Polissar et al. 1999; Gao et al. 2006; Kim et al. 2004; Kim et al. 2005; Heo, Hopke and Yi 2009; Zhou, Hopke and Zhao 2009) for assessing the long-range transport potential of constituents of aerosols from source regions. Du and Rodenburg (2007) identified and mapped four PCB factors from their PMF-PSCF analyses on samples collected in Camden, Ohio. Liu et al. (2003) used the PMF-PSCF methodology to identify and map out six factors in Stockton, New York and seven factors in Potsdam, New York using metals, inorganic species and PAH as the contaminant species of interest.

The receptor models described above vary in complexity and are applicable with varying degrees of accuracy to pollutional data from any environmental matrix since mass transfer of chemical species occurs between environmental compartments and physico-chemical parameters such as chemical reactivity, partition coefficients and solubility govern the distribution of these compounds in the environment. Thus the right environmental compartment(s) must be studied to obtain useful information about the impacting sources. The table below (Table 2) summarizes four key characteristics of all the PAH source apportionment techniques discussed in this thesis.

Table 2: Comparisons of the properties of common PAH source apportionment techniques

	Property			
Technique	**Multivariate**	**Quantitative**	**Output Quality**	**Ease of use**
Biomarkers	No	No	Low	High
Diagnostic ratios	No	No	Low-Average	High
Receptor models	Yes	Yes	High	Low

1.4. Objectives of the Thesis

PAH pollution data from sediment and air samples were used as the contaminant species of interest to which qualitative and quantitative source apportionment models were applied. Sample sets used in this research may have been impacted by both short- and long- range pollution sources. These samples included suspended sediment samples from Hamilton Harbour collected between 2002 and 2006, air particulate samples from Hamilton collected in 1995, and air samples (gas- and particle-phase) collected at a sub-Arctic site in the Yukon Territory, Canada between 2007 and 2009.

The research conducted in this thesis had four main goals:

1. The application of a variety of source apportionment models to complex environmental data sets originating from hemispherical (long-range impacts) and local (short-range impacts) pollution studies to reveal hidden source information.
2. The integration of common qualitative and quantitative models for more robust data interpretation.
3. The integration of meteorological data and common source apportionment models for assessing sources of pollution over relatively short and long ranges.
4. Application of a series of physico-chemical models to understand the temporal variations of partitioning behaviour of semi-volatile organic contaminants.

1.5. Thesis Layout

This thesis is a publication-based thesis and Chapters 2 – 6 are research manuscripts that are either currently under review for publication or have been previously published. Each of these chapters will start with a preface detailing the contributions of each of the authors. Chapter 2 focuses on the partitioning of PAH between the gas- and particle-phases in samples collected from a sub-Arctic site in the Yukon. This work revealed that the potential sources (which could be local or from relatively long distances) of impacting PAH collected at the sampling site affected their partitioning between both phases (Sofowote et al. 2010b).

Chapter 3 assesses the long-range transport of PAH to a relatively pristine sub-Arctic site by integrating a factor analytic model (PMF) with a back trajectory residence time analysis (PSCF). This work is the first work to apply the PMF-PSCF methodology for organic contaminants on a hemispherical scale (*submitted* (Sofowote et al. 2010a).

Chapter 4 discusses a site-dependent determination of thresholds for common diagnostic PAH ratios in air pollution data from an urban-industrial environment. This new approach is more practical than relying solely on thresholds reported in the literature. A new diagnostic ratio was proposed and then combined with PAH atmospheric concentrations and wind direction data for source apportionment more quantitative than with diagnostic ratios alone (Sofowote, Allan and McCarry 2010b).

Chapter 5 discusses the application and comparison of two factor analytic models (PMF and PCA) on the same PAH data used in Chapter 4. PMF results were combined with wind direction to reveal hidden trends of industrial impacts (Sofowote, Allan and McCarry 2010a).

Chapter 6 focuses on the application of PCA and PMF to PAH pollution data from suspended sediments collected in Hamilton Harbour. The impacts of legacy and continuous sources within the harbour were quantified using a PMF-MLR hybrid. This work was one of the first works to show that PMF could be applied successfully to environmental data arising from sediments (Sofowote, McCarry and Marvin 2008).

Chapter 7 contains a general discussion of the results and provides a summary of the most important findings of the preceding chapters in this thesis. The Appendix sections contain the detailed sample preparation methods used for air samples collected from the Yukon Territory, summary of the mathematical principles of the receptor models used as well as common structures of some polycyclic aromatic compounds encountered in this research. The attached compact disc contains supplementary information for all the manuscripts included in this thesis as well as other useful data such as concentration data sets used for the statistical analyses.

References

Allan, L.M. and McCarry, B.E., 1999. Thia-arenes as pollution source tracers in urban air particulate. Ph.D., McMaster University.

Almeida, S.M., Pio, C.A., Freitas, M.C., Reis, M.A., Trancoso, M.A., 2006. Approaching PM(2.5) and PM(2.5-10) source apportionment by mass balance analysis, principal component analysis and particle size distribution. Science of the Total Environment 368, 663-74.

Anttila, P., Paatero, P., Tapper, U., Jarvinen, O., 1995. Source identification of bulk wet depostion in Finland by positive matrix factorization. Atmospheric Environment 29, 1705-18.

Ara Begum, B., Kim, E., Jeong, C., Lee, D., Hopke, P.K., 2005. Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. Atmospheric Environment 39, 3719-3724.

Ashbaugh, L.L., Malm, W.C., Sadeh, W.Z., 1985. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. Atmospheric Environment (1967) 19, 1263-1270.

Bence, A.E., Kvenvolden, K.A., Kennicutt II, M.C., 1996. Organic geochemistry applied to environmental assessments of Prince William sound, Alaska, after the Exxon Valdez oil spill--a review. Organic Geochemistry 24, 7-42.

Bi, X., Feng, Y., Wu, J., Wang, Y., Zhu, T., 2007. Source apportionment of PM10 in six cities of northern China. Atmospheric Environment 21, 903-12.

Brandli, R.C., Bucheli, T.D., Ammann, S., Desaules, A., Keller, A., Blum, F., Stahel, W.A., 2008. Critical evaluation of PAH source apportionment tools using data from the Swiss soil monitoring network. Journal of Environmental Monitoring : JEM 10, 1278-1286.

Bzdusek, P.A. and Christensen, E.R., 2006. Comparison of a new variant of PMF with other receptor modeling methods using artificial and real sediment PCB data sets. Environmetrics 17, 387-403.

Bzdusek, P.A., Lu, J., Christensen, E.R., 2006. PCB congeners and dechlorination in sediments of Sheboygan River, Wisconsin, determined by matrix factorization. Environmental Science & Technology 40, 120-9.

Bzdusek, P.A., Christensen, E.R., Li, A., Zou, Q., 2004. Source apportionment of sediment PAHs in Lake Calumet, Chicago: Application of factor analysis with nonnegative constraints. Environmental Science & Technology 38, 97-103.

Cheng, M.D., Hopke, P.K., Barrie, L., Rippe, A., Olson, M., Landsberger, S., 1993. Qualitative determination of source regions of aerosol in Canadian High Arctic. Environmental Science & Technology 27, 2063-2071.

Christensen, E.R. and Bzdusek, P.A., 2005. PAHs in sediments of the Black River and the Ashtabula River, Ohio: Source apportionment by factor analysis. Water Research 39, 511-24.

Du, S. and Rodenburg, L.A., 2007. Source identification of atmospheric PCBs in Philadelphia/Camden using positive matrix factorization followed by the potential source contribution function. Atmospheric Environment 41, 8596-8608.

Fang, G.C., Wu, Y.S., Chang, C.N., Ho, T.T., 2006. A study of polycyclic aromatic hydrocarbons concentrations and source identifications by methods of diagnostic ratio and principal component analysis at Taichung Chemical Harbor near Taiwan Strait. Chemosphere 64, 1233-42.

Galarneau, E., 2008. Source specificity and atmospheric processing of airborne PAHs: Implications for source apportionment. Atmospheric Environment 42, 8139-8149.

Gao, N., Gildemeister, A.E., Krumhansl, K., Lafferty, K., Hopke, P.K., Kim, E., Poirot, R.L., 2006. Sources of fine particulate species in ambient air over Lake Champlain Basin, VT. Journal of the Air & Waste Management Association (1995) 56, 1607-1620.

Gogou, A.I., Stratigakis, N., Kanakidou, M., Stephanou, E.G., 1996. Organic aerosols in eastern Mediterranean: Components source reconciliation by using molecular markers and atmospheric back trajectories. Organic Geochemistry 25, 79-96.

Greenberg, A., Bozzelli, J.W., Cannova, F., Forstner, E., Giorgio, P., Stout, D., Yokoyama, R., 1981. Correlations between lead and coronene concentrations at urban, suburban, and industrial sites in New Jersey. Environmental Science & Technology 15, 566-570.

Grimmer, G., Naujack, K., Schneider, D., 1982. Profile analysis of polycyclic aromatic hydrocarbons by glass capillary gas chromatography in atmospheric suspended particulate matter in the nanogram range collecting 10 m^3 of air. Fresenius' Journal of Analytical Chemistry 311, 475-484.

Guo, H., Wang, T., Louie, P.K., 2004. Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model. Environmental Pollution 129, 489-98.

Hafner, W.D. and Hites, R.A., 2003. Potential sources of pesticides, PCBs, and PAHs to the atmosphere of the Great Lakes. Environmental Science & Technology 37, 3764-3773.

Halsall, C.J. 2007. Environmental Organic Chemistry. In: Harrison, R.M. (Ed.), Principles of Environmental Chemistry, RSC Publishing, Cambridge, 279-313.

Harrison, R.M. and Yin, J., 2000. Particulate matter in the atmosphere: Which particle properties are important for its effects on health? The Science of The Total Environment 249, 85-101.

Harrison, R.M., Smith, D.J.T., Luhana, L., 1996. Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, U.K. Environmental Science & Technology 30, 825-832.

Heo, J.-., Hopke, P.K., Yi, S.-., 2009. Source apportionment of PM2.5 in Seoul, Korea. Atmospheric Chemistry and Physics 9, 4957-4971.

Hopke, P., 1985. Receptor Modeling in Environmental Chemistry. Wiley-Interscience, NY,

Hopke, P.K., Zhou, L., Poirot, R.L., 2005. Reconciling trajectory ensemble receptor model results with emissions. Environmental Science & Technology 39, 7980-7983.

Hsu, Y., Holsen, T.M., Hopke, P.K., 2003. Locating and quantifying PCB sources in Chicago: Receptor modeling and field sampling. Environmental Science & Technology 37, 681-690.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans and International Agency for Research on Cancer, 2010. Some Non-Heterocyclic Polycyclic Aromatic Hydrocarbons and some Related Exposures. World Health Organization, International Agency for Research on Cancer ; Geneva, Switzerland : Distributed by WHO Press, Lyon, France; Geneva, Switzerland, 853.

Kavouras, I.G., Koutrakis, P., Tsapakis, M., Lagoudaki, E., Stephanou, E.G., Von Baer, D., Oyola, P., 2001. Source apportionment of urban particulate aliphatic and polynuclear aromatic hydrocarbons (PAHs) using multivariate methods. Environmental Science & Technology 35, 2288-94.

Kim, D., Kumfer, B.M., Anastasio, C., Kennedy, I.M., Young, T.M., 2009. Environmental aging of polycyclic aromatic hydrocarbons on soot and its effect on source identification. Chemosphere 76, 1075-1081.

Kim, E., Hopke, P.K., Kenski, D.M., Koerber, M., 2005. Sources of fine particles in a rural midwestern U.S. area. Environmental Science & Technology 39, 4953-4960.

Kim, E., Hopke, P., Larson, T., Maykut, N., Lewtas, J., 2004. Factor analysis of Seattle fine particles. Aerosol Science & Technology 38, 724-738.

Larsen, R.K.,3rd and Baker, J.E., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. Environmental Science & Technology 37, 1873-81.

Lee, P.K., Brook, J.R., Dabek-Zlotorzynska, E., Mabury, S.A., 2003. Identification of the major sources contributing to PM2.5 observed in Toronto. Environmental Science & Technology 37, 4831-40.

Legzdins, A.E., McCarry, B.E., Marvin, C.H., Bryant, D.W., 1995. Methodology for bioassay-directed fractionation studies of air particulate material and other complex environmental matrices. International Journal of Environmental Analytical Chemistry 60, 79.

Liu, W., Hopke, P.K., Han, Y., Yi, S., Holsen, T.M., Cybart, S., Kozlowski, K., Milligan, M., 2003. Application of receptor modeling to atmospheric constituents at Potsdam and Stockton, NY. Atmospheric Environment 37, 4997-5007.

Marvin, C.H., McCarry, B.E., Villella, J., Allan, L.M., Bryant, D.W., 2000. Chemical and biological profiles of sediments as indicators of sources of genotoxic contamination in Hamilton Harbour. part I: Analysis of polycyclic aromatic hydrocarbons and thia-arene compounds. Chemosphere 41, 979-988.

McCarry, B.E., Allan, L.M., Legzdins, A.E., Lundrigan, J.A., Marvin, C.H., Bryant, D.W., 1996. Thia-arenes as pollution source tracers in urban air particulate. Polycyclic Aromatic Compounds 11, 75-82; 75.

Metcalfe, C.D., Balch, G.C., Cairns, V.W., Fitzsimons, J.D., Dunn, B.P., 1990. Carcinogenic and genotoxic activity of extracts from contaminated sediments in western Lake Ontario. Science of the total environment 94, 125-41.

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and Intelligent Laboratory Systems 18, 183-194.

Pauzi Zakaria, M., Okuda, T., Takada, H., 2001. Polycyclic aromatic hydrocarbon (PAHs) and hopanes in stranded tar-balls on the coasts of peninsular Malaysia: Applications of biomarkers for identifying sources of oil pollution. Marine pollution bulletin 42, 1357-1366.

Pekney, N.J., Davidson, C.I., Zhou, L., Hopke, P.K., 2006. Application of PSCF and CPF to PMF-modeled sources of $PM_{2.5}$ in Pittsburgh. Aerosol Science and Technology 40, 952.

Poirot, R.L., Wishinski, P.R., Hopke, P.K., Polissar, A.V., 2001. Comparative application of multiple receptor methods to identify aerosol sources in Northern Vermont. Environmental Science & Technology 35, 4622-4636.

Polissar, A.V., Hopke, P.K., Paatero, P., Kaufmann, Y.J., Hall, D.K., Bodhaine, B.A., Dutton, E.G., Harris, J.M., 1999. The aerosol at Barrow, Alaska: Long-term trends and source locations. Atmospheric Environment 33, 2441-2458.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001. Atmospheric aerosol over Vermont: chemical composition and sources. Environmental Science & Technology 35, 4604-4621.

Ramdahl, T., 1983. Retene - a molecular marker of wood combustion in ambient air. Nature 306, 580-582.

Rogge, W.F., Hildemann, L.M., Mazurek, M.A., Cass, G.R., Simoneit, B.R.T., 1993a. Sources of fine organic aerosol. 2. Noncatalyst and catalyst-equipped automobiles and heavy-duty diesel trucks. Environmental Science & Technology 27, 636-651.

Rogge, W.F., Hildemann, L.M., Mazurek, M.A., Cass, G.R., Simoneit, B.R.T., 1993b. Sources of fine organic aerosol. 5. Natural gas home appliances. Environmental Science & Technology 27, 2736-2744.

Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, B.R.T., 1999. Measurement of emissions from air pollution sources. 2. C1 through C30 organic compounds from medium duty diesel trucks. Environmental Science & Technology 33, 1578-1587.

Sofowote, U.M., Allan, L.M., McCarry, B.E., 2010a. A comparative study of two factor analytic models applied to PAH data from inhalable air particulate collected in an urban-industrial environment. Journal of Environmental Monitoring 12, 425-433.

Sofowote, U.M., Allan, L.M., McCarry, B.E., 2010b. Evaluation of PAH diagnostic ratios as source apportionment tools for air particulates collected in an urban-industrial environment. Journal of Environmental Monitoring 12, 417-424.

Sofowote, U.M., Hung, H., Rastogi, A.K., Westgate, J.N., Su, Y., DeLuca, P., McCarry, B.E., 2010a. Assessing the long-range transport of PAH to a sub-Arctic site using positive matrix factorization and potential source contribution function (ATMENV-D-10-00565) *submitted.* Atmospheric Environment

Sofowote, U.M., Hung, H., Westgate, J.N., Su, Y., Sverko, E., D'Sa, I., Roach, P., Fellin, P., McCarry, B.E., 2010b. The Gas/Particle partitioning of polycyclic aromatic hydrocarbons collected at a sub-Arctic site in Canada. Atmospheric Environment 44, 4919-4926.

Sofowote, U.M., McCarry, B.E., Marvin, C.H., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. Environmental Science & Technology 42, 6007-6014.

Vaccaro, S., Sobiecka, E., Contini, S., Locoro, G., Free, G., Gawlik, B.M., 2007. The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. Chemosphere 69, 1055-63.

Wang, Z., Chen, J., Qiao, X., Yang, P., Tian, F., Huang, L., 2007. Distribution and sources of polycyclic aromatic hydrocarbons from urban to rural soils: A case study in Dalian, China. Chemosphere 68, 965-71.

Wang, Z., Fingas, M., Sergy, G., 1994. Study of 22-year-old arrow oil samples using biomarker compounds by GC/MS. Environmental Science & Technology 28, 1733-1746.

Wild, S.R. and Jones, K.C., 1995. Polynuclear aromatic hydrocarbons in the United Kingdom environment: A preliminary source inventory and budget. Environmental Pollution 88, 91-108.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette, D., Sylvestre, S., 2002. PAHs in the Fraser River Basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry 33, 489-515.

Zeng, Y. and Hopke, P.K., 1989. A study of the sources of acid precipitation in Ontario, canada. Atmospheric Environment (1967) 23, 1499-1509.

Zhou, L., Hopke, P.K., Zhao, W., 2009. Source apportionment of airborne particulate matter for the speciation trends network site in Cleveland, OH. Journal of the Air & Waste Management Association (1995) 59, 321-331.

Chapter Two

The Gas/Particle Partitioning of Polycyclic Aromatic Hydrocarbons collected at a sub-Arctic Site in Canada

The work in this chapter has been published in an article under the title: Uwayemi M. Sofowote[1], Hayley Hung[2], Ankit K. Rastogi[1], John N. Westgate[3], Yushan Su[2], Ed Sverko[4], Ivy D'Sa[4], Patrick Roach[5], Phil Fellin[6], and Brian E. McCarry[1] 2010, Atmospheric Environment 44, 4919-4926 and is reproduced here with permission.

Author Contributions:

Uwayemi Sofowote: Performed some of the sample collection and all the sample cleanups jointly with Johnny Westgate; carried out all the PAH quantification, data mining, interpretation, statistical analyses, and wrote the first draft of the manuscript.

Dr. Hayley Hung: Initiated this project (IPY-INCATPA) and was responsible for logistics. Wrote and edited several parts of the manuscript

Ankit Rastogi: Created and ran a script for merging and visualizing the back trajectories and generated the geo-referenced map in Figure 4

John Westgate: Performed some of the sample collection and sample cleanups jointly with Uwayemi Sofowote.

Dr. Yushan Su: Supplied the final pressure-corrected PAH concentrations used in the article; was responsible for ensuring concentration data integrity. Wrote and edited several parts of the manuscript.

Ed Sverko: Responsible for sample cleanups at NLET, Burlington.

Ivy D'Sa: Supervised many aspects of sample cleanups and instrumental analyses. Responsible for quality assurance and control of the raw PAH data.

Patrick Roach: Supervised all sample logistics in the Yukon.

Phil Fellin: Generated all raw sample extracts and designed many aspects of the sampling protocol.

Dr. Brian McCarry: Supervised many aspects of the data interpretation. Edited the final manuscript.

Supplementary Information:

Supplementary tables and figures for this manuscript can be found in the attached compact disc.

Author Affiliations:

[1]Department of Chemistry and Chemical Biology
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[2]Air Quality Research Division
Environment Canada
4905 Dufferin Street, Toronto, Ontario, M3H 5T4, Canada

[3]Department of Physical and Environmental Sciences
University of Toronto Scarborough
1265 Military Trail, Toronto, Ontario, M1C 1A4, Canada

[4]National Laboratory for Environmental Testing
Environment Canada
867 Lakeshore Road, Burlington, Ontario L7R 4A6, Canada

[5]Indian and Northern Affairs Canada (Yukon)
415c - 300 Main St., Whitehorse, Yukon, Y1A 2B5, Canada

[6]AirZOne One Ltd.
222 Matheson Boulevard East, Mississauga, Ontario, L4Z 1X1, Canada

Corresponding author email address: mccarry@mcmaster.ca; telephone: +1 (905) 525-9140

The Gas/Particle Partitioning of Polycyclic Aromatic Hydrocarbons collected at a sub-Arctic Site in Canada

Uwayemi M. Sofowote, Hayley Hung, Ankit K. Rastogi, John N. Westgate, Yushan Su, Ed Sverko, Ivy D'Sa, Pat Roach, Phil Fellin, Brian E. McCarry

ABSTRACT

Polycyclic aromatic hydrocarbons (PAH) were measured in air samples at a remote air monitoring site established in the Yukon Territory, Canada as part of a global project (International Polar Year; IPY) to study the potential for atmospheric long range transport of anthropogenic pollutants to the Arctic. Gas- and particle-phase PAH were collected in polyurethane foam plugs and on glass fibre filters respectively from August 2007 to October 2009. PAH concentrations were found to be highest in the winter months and lowest in summer. The gas/particle partitioning coefficients of 3-5 ringed PAH were computed and seasonal averages were compared. In the summertime, lower molecular mass PAH exhibited relatively higher partitioning into the particle-phase. This particle-phase partitioning led to the shallowest slopes being recorded during summer for the log-log correlation plots between the PAH partition coefficients and their sub-cooled vapour pressures. Air mass back trajectories suggest that local impacts may be more important during the summer time which is marked by increased camping activities at camping sites in the proximity of the sampling station. In conclusion, both summer and wintertime variations in PAH concentrations and gas/particle partitioning are considered to be source- and phototransformation-dependent rather than dependent on temperature-driven shifts in equilibrium partitioning.

1. Introduction

Polycyclic aromatic hydrocarbons (PAH) are a class of ubiquitous organic contaminants that are made up of multiple fused benzene rings. They are products of incomplete combustion from both anthropogenic and natural sources. Some PAH are known to be mutagenic and potentially carcinogenic, thus, they could pose adverse health effects to higher organisms (Metcalfe et al. 1990; International Agency for Research on Cancer 2010). Many studies have shown significant correlations between anthropogenic pollution associated with particulate-bound pollutants, increased mortality rates and increased hospitalization rates for respiratory ailments and cardiovascular issues (Pope III et al. 2004; Pope III et al. 2002; Jerrett et al. 2005). While there is still debate over whether these health effects are caused by particulate matter alone, the compounds entrained in them or both (Harrison and Yin 2000), a lot of knowledge has been garnered regarding the gas/particle partitioning behaviour of some of these pollutants.

PAH have been found in the Arctic and atmospheric long range transport is believed to be a significant entry process of these contaminants (Halsall 2004; Halsall et al. 1997). As part of many research activities initiated under the International Polar Year program (IPY), a Canadian project (IPY-INCATPA; INterContinental Atmospheric Transport of Anthropogenic Pollutants to the Arctic) was created to study the potential for long range atmospheric transport of a wide range of organic pollutants originating from lower latitudes to the sparsely populated high Arctic (Environment Canada 2009).

The Canadian INCATPA air monitoring site is located in a mountainous region near Little Fox Lake (LFL) in Yukon Territory, Canada (61° 21' N, 135° 38' W; 1138m a.s.l.) and is equipped with a high volume (Hi-Vol) air sampler for the collection of gas- and particle- phase contaminants including PAH. Levels of PAH in high altitude regions may be controlled by the surrounding topography (Choi et al. 2009). In the Arctic, the highest levels of PAH are usually observed during the winter months (Halsall et al. 1997). Halsall et al. reported that the haze period (November to April) is usually typified by high levels of anthropogenic aerosols, thus it is particularly important to study the changes that occur in PAH partitioning from summer to winter.

This work focuses on the temporal variation in the partitioning coefficients of common PAH collected in a sub-Arctic environment.

2. Experimental Section

2.1. Air Sampling: Weekly air samples were collected at the Little Fox Lake, Yukon site (LFL; see Figure S1; Supporting Information for a map of all INCATPA sites) from August 2007 to October 2009. Over the course of the project, two daily sampling events were initiated between August 26 - September 08, 2007 and September 15 – September 28, 2008. Most of the weekly sampling was continuous; however, in 2009, three months of sampling were missing due to maintenance issues with the sampler. This study focuses on all samples collected from August 2007 to December 2008. Gas- and particle- phase samples were collected on two inline polyurethane foam (PUF) plugs and on glass fibre filters

(GFF) respectively housed in a high volume (Hi-Vol) air sampler. This Hi-Vol air sampler has been previously described (Halsall et al. 1997). The Hi-Vol sampler was fitted with a 10μm diameter particle size selective inlet and the flow rate was calibrated so that approximately 12000 m^3 of air was collected over a seven-day period. To minimize sample handling and the potential for human error at this fairly remote site, total suspended particulates (TSP) were not sampled or determined at LFL during the course of this project. $PM_{2.5}$ were however collected at a National Air Pollution Surveillance Program (NAPS) sampling station (NAPS House, Whitehorse, Yukon) situated relatively close to LFL.

2.2. Sample Extraction and Clean-up: All samples were subjected to soxhlet extraction in hexane for a period of 24 h. All extractions were carried out by Airzone One Ltd. (Mississauga, ON, Canada). After extraction, a complete sample set comprised raw extracts from one glass fibre filter (GFF) and two polyurethane foam plugs (PUF). Previous studies with the same Hi-Vol sampler model under similar conditions had shown that the potential for breakthrough from the front to the back PUF was very low (Halsall et al. 1997), thus, the PUF raw extracts were combined before further analysis.

One half of each raw extract was further cleaned-up at the National Laboratory for Environmental Testing (NLET, Environment Canada, Burlington, ON, Canada) using NLET protocols for pre-existing Arctic sites monitoring programs. Briefly, raw extracts were evaporated to 1 mL under nitrogen and transferred onto a silica gel column (8g, activated overnight at 160°C) wet packed with top and bottom anhydrous sodium sulphate plugs (1 cm, fired at 600°C). Elution with hexane (70 - 90 mL) afforded Fraction A while elution with hexane/dichloromethane (1:1, v/v 130 mL) afforded Fraction B. Fractions A and B were evaporated, reduced under nitrogen, and solvent-exchanged into iso-octane (1 mL).

2.3. GC/MS Analysis and Quality Control: All GC/MS analyses were performed at NLET and all quality control parameters applied to this study met or exceeded the basic criteria for acceptable data. For PAH analysis, a 100 μL aliquot each sample was obtained by combining 45 μL each of Fractions A and B and adding 10 μL of internal standard solution containing three deuterated PAH. Relative response factors for each PAH were obtained by running six calibration standards containing the target compounds in concentrations ranging from 10 – 2000 pg μL^{-1} on an Agilent 6890 gas chromatograph/5973 mass selective detector system using electron impact ionization (EI^+) in selected ion monitoring (SIM) mode and equipped with a J&W DB-5ms column (30 m x 0.25 mm internal diameter, 0.25 μm film). Twenty-one PAH and 4 other polycyclic aromatic compounds were routinely quantified by this protocol where possible. Four deuterated PAH surrogates were spiked into the raw extracts and method blanks before silica gel clean-up and recoveries (n > 200 samples)were: naphthalene-d_8, 93.3±11.4%; fluorene-d_{10}, 101±11.7%; pyrene-d_{10}, 104±10.9%; perylene-d_{12}, 94.2±10.5%. In all a total of thirteen compounds (indene, naphthalene, 1- & 2-methylnaphthalene, acenapthylene, acenapthene, tetrahydronaphthalene, β-chloronaphthalene, dibenzothiophene, perylene, indeno[1,2,3-*cd*]pyrene, benzo[*a*]pyrene, dibenz[*a, h*]anthracene) were excluded from further data interpretation either because their concentrations were consistently below their method detection limits in one or both phases or they had poor sampling efficiencies. Field blanks comprising pairs of GFF and PUF were collected once every 4 weeks for the weekly samples and once a week for the daily samples over the course of a year and were subjected to the same sample preparation and analytical schemes as the real samples. Method detection limits in pg m^{-3} (determined as yearly mean concentration of PAH for all field blanks + 3S.D) for PAH of interest can be found in Table 1.

2.4. Data Analysis: Data used in this work have been blank corrected. The 12 PAH remaining after initial data screening were used for this work even though benzo[*ghi*]perylene was almost exclusively found in the particle-phase making its partition coefficient incalculable in >95% of the samples. Parameters for computing temperature-dependent constants $\log P_L^o$ were readily available for all compounds except retene and were sourced from the recent reports of Ma et al. (Ma et al. 2010) and elsewhere when necessary (e.g. parameters for chrysene, benzo[*b*]fluoranthene, & benzo[*k*]fluoranthene were from (Odabasi, Cetin and Sofuoglu 2006) and benzo[*e*]pyrene from (Hinckley et al. 1990)) for log-log relationships with the partition coefficients. All data analyses were performed in a spreadsheet (Microsoft Excel, Microsoft Office 2003).

Five-day air mass back trajectories were obtained from the Canadian Meteorological Centre (CMC) for LFL and were plotted using a GIS mapping program (ArcGIS 9.3).

The partitioning between the gas- and particle phase for many freely exchangeable organic contaminants may be correlated with their sub-cooled vapour pressure (Pankow 1987; Lohmann and Lammel 2004). At equilibrium, in the absence of sampling artefacts, and assuming that the organic compound is completely available for gas-particle exchange from the sorbent material, its partitioning coefficient may correlate fairly well with its vapour pressure by the linear equation below for both adsorption and absorption phenomena (Pankow 1998).

$$\log K_p = m \log P_L^o + b \qquad (1)$$

As noted in Section 2.1 above, TSP were not collected at LFL. Freely available data for $PM_{2.5}$ collected at a NAPS sampling station situated relatively close (60° 42' N, 135° 3' W, 669m a.s.l.; http://www.etc-cte.ec.gc.ca/napsdata/main.aspx?dir=D%3a%5cexternal%5cclients%5cRiver%5cnapsdata%5cDATA) to LFL were used as TSP surrogates for the computation of log K_p in relevant portions of this study. This assumption was made because the bulk of PAH are known to partition to particulate matter with effective diameter <2.5μm (Chrysikou and Samara 2009; Kaupp and McLachlan 1999; Kaupp and McLachlan 2000; Liu et al. 2006; Venkataraman, Lyons and Friedlander 1994; Venkataraman and Friedlander 1994). Furthermore, data from the work of Kaupp and McLachan at a temperate sampling site show that $PM_{2.5}$ accounts for ≥70% of all particulate matter collected in February (winter) and August (summer) (Kaupp and McLachlan 1999).

3. Results and Discussion

3.1. Temporal Variation of PAH in the Gas- and Particle- Phase: The plot in Figure 1 is a time series of the sum of the concentrations of 12 PAH ($TPAH_{12}$, see Table 1 for the PAH) in the gas- and particle- phase. As noted above, the other PAH were not included because their levels were below zero after blank correction or due to poor extraction recoveries stated above. The PAH concentrations at this site are higher in fall-winter than in spring-summer. The gas phase PAH dominate the entire time series at this sub-Arctic site but the colder period (fall-winter) shows an enrichment of the particle phase PAH by a factor of about 2 when compared to the spring-summer of 2008. This is in agreement with the observed trend for a monitoring site at Tagish, Yukon (60° 20' N, 134° 12' W) reported by Halsall et al. (Halsall et al. 1997); this location is proximate to the Little Fox Lake site but is now decommissioned).

The plots in Figure S2 reveal that PAH with molecular mass 228-276 Da are the ones getting enriched in the particle-phase in the fall-winter months and are thus responsible for the increase in particle-phase $TPAH_{12}$ observed during this period. The middle plot (Figure 1b) also reveals the seasonality of $PM_{2.5}$ with greatest levels in the winter of 2008, though other 2008 maxima can be found in the spring and summer times. Interestingly, the winter maximum for $PM_{2.5}$ does not exactly coincide with the 2008 winter maximum for most particle-bound PAH. This may indicate that the sample with highest PAH levels recorded in winter of 2008 has different source origins from the sample/period with the highest $PM_{2.5}$.

Table 1: Arithmetic mean, standard deviation (S.D.) and method detection limits (MDL) for 12 PAH (*in* pg m-^{3}) monitored in gas- (PUF) and particle- (Filter) phase samples from Little Fox Lake, Yukon, Canada between August 2007 and December 2008. PAH in boldface are used for gas/particle partition modeling

		PUF				**Filter**			
PAH	Abbreviation	Geometric Mean	Arithmetic Mean	S.D	Mean MDL	Geometric Mean	Arithmetic Mean	S.D	Mean MDL
fluorene	FLUOR	23	50	61	5.8	0.99	2.5	3.8	1.8
phenanthrene	PHEN	67	92	85	43	5.9	20	31	16
anthracene	ANTH	2.2	5.1	11	3.6	0.41	1.8	3.2	0.39
fluoranthene	FLUORT	14	22	23	12	3.4	9.1	15	2.6
pyrene	PYR	14	25	31	31	5.7	17	32	7.8
Retene	RET	34	57	80	13	5.5	18	33	1.1
benz[*a*]anthracene	BAA	0.59	1.2	1.9	0.22	0.71	2.4	5.8	0.27
chrysene	CHRY	1.5	3.3	4.4	0.84	1.2	3.7	6.6	0.30
benzo[*b*]fluoranthene	BBF	0.50	0.80	0.68	0.24	1.2	3.8	7.2	0.080
benzo[*k*]fluoranthene	BKF	0.24	0.52	0.56	0.24	0.87	2.8	4.7	0.080
benzo[*e*]pyrene	BEP	0.26	0.41	0.37	0.21	0.83	2.5	4.2	0.070
benzo[*ghi*]perylene	B_GHI_P	0.040	0.10	0.090	0.71	0.80	2.0	3.0	0.14

This seasonality was investigated further by the use of a classical phase change relationship at equilibrium. The phase change from liquid or solid to gas for most organic compounds including persistent organic pollutants (POP) can be predicted by their enthalpies of vaporization or sublimation using the thermodynamic relationship in the Clausius-Clapeyron equation (Atkins 1994) which yields a negative slope (i.e., concentration/partial pressure decreases with temperature) when the natural logarithm of the pressure lnP is plotted against the inverse of the absolute temperature $1/T$ (see examples of PCB in (Hillery et al. 1997)). The negative slope of the plot is equal to – ($\Delta_{vap}H$/R) and the enthalpy of the phase change is a positive number calculable by multiplying the slope by R, the gas constant. Using the ideal gas equation, concentrations of the most abundant gas phase PAH were converted to partial pressures in atmospheres and the plots in Figure S3 show the temperature dependence of their lnP. Only two PAH (phenanthrene and pyrene) yielded negative slopes albeit insignificant at the 95% confidence level for pyrene. The others had statistically significant positive slopes which is anomalous for phase changes predictable due to enthalpies alone. Gas-phase PAH concentrations at this site generally increase with a decrease in average temperature. Clearly, some other mechanism(s) is at work with regards to PAH which cannot be captured by phase change thermodynamics alone. Cortes et al. observed similar phenomena with gas-phase PAH collected near the Great Lakes (Cortes et al. 2000) and suggested that other seasonal sources (e.g. winter-time residential heating) which were not correlated with the ambient temperature may be responsible.

3.2. Seasonal variation of gas/particle partitioning behaviour of PAH at LFL:

The partitioning of PAH between the gas- and particle-phases also exhibits seasonality for some PAH. Statistical evaluation of the seasonal log K_p averages using the Welch *t*-test for comparing arithmetic means with unequal sample sizes and variances (Welch 1947) shows that the heavier PAH (especially benzo[*k*]fluoranthene and benzo[*e*]pyrene) had partitioning coefficients that were usually not statistically different in the consecutive progression from one season to the next. The lightest PAH fluorene, however, had partitioning coefficients that were statistically different from Summer 2007 to Fall 2007, Winter 2008 to Spring 2008, and Summer 2008 to Fall 2008. Other lower molecular mass PAH (phenanthrene, anthracene, fluoranthene and pyrene) had log K_p averages with varying statistically significant differences from Summer 2007 to Fall 2008. In general winter log K_p were usually statistically different from the preceding or succeeding season. A summary of the side-by-side comparisons of all the seasons and the summer *vs.* winter periods can be found in the supplementary information section and Figure 2 below shows the seasonal variations in the log K_p for 11 PAH in box-and-whisker plots.

A plot of the seasonal averages of log K_p against the temperature-dependent $\log P_L^o$ using ten PAH (benzo[*ghi*]perylene partitioned almost exclusively into the particle-phase, thus log K_p could not be computed; the parameters for computing the temperature-dependent sub-cooled vapour pressure for retene are not readily available) is shown below in Figure 3 for representative purposes only while Table S3 in the supplementary information section contains all the individual *m* and *b* parameters for each of the weekly and daily samples. In general, the regression lines have steeper slopes in the fall/winter period in comparison with the spring/summer period.

During the spring/summer period, it appears that the lower right end of the plot where lower molecular mass PAH are found experiences a boost in partitioning to the particle-phase. This may arise from filter sorption artefacts in which gas-phase PAH are co-collected on the filters. While filter sorption artefacts were not explicitly investigated in this study, it is not expected to be the main reason for the spring/summer time increase in particle-phase sorption of lower molecular mass PAH because the increasingly warmer temperatures during this time of the year is expected to preferentially favour their partitioning to the gas-phase. Thus, we speculate that these spring/summer deviations may be a case in which the particle-bound lower molecular mass PAH are not totally available for gas-particle exchange but are still largely extractable during the extraction step of chemical analysis. Arp et al. reported similar deviations for PAH collected in Europe and suggested a "shell-like" entrapment of PAH to the elemental carbon fraction of particulate matter, thus, leading to slow/non-exchangeability (Arp, Schwarzenbach and Goss 2008). Indeed, from theoretical arguments, Pankow and Bidleman had shown that non-exchangeability will affect the partitioning of lower molecular mass PAH more than the heavier ones (Pankow and Bidleman 1991). These results suggest that seasonal differences in impacting sources may be responsible for the seasonal variation in the gas/particle partitioning of lower molecular mass PAH.

We speculate that summer samples collected at LFL were generally more affected by fresh local emissions containing slowly exchanging or non-exchangeable low molecular mass PAH; summer camping activities at camping grounds relatively close to the monitoring site may lead to an increase in PAH from vehicular transportation, fuel and wood combustion. To investigate this possibility further, we computed the particle fraction (φ) for PAH in the fourteen daily samples (August 25, 2007 – September 08, 2007) collected in the summer of 2007 (Table 2) and examined their 5-day air mass back trajectories collected at 0000, 0600, 1200 and 1800 h UTC and obtained for a starting height of 1138m a.s.l. (the Hi-Vol sampler elevation) at LFL (see Figure 4).

Table 2: Particle fraction φ for 11 PAH collected from August 25, 2007 to September 08, 2007. Particle fraction values greater than 0.1 have been highlighted for lower molecular mass PAH (166 – 202 Da). Missing values indicate a mathematically impossible solution.

SAMPLE COLLECTION END DATE	FLUOR	PHEN	ANTH	FLUORT	PYR	RETENE	BAA	CHRY	B_B_F	B_K_F	B_E_P
26/08/2007	**0.22**	0.060	**0.13**	**0.22**		0.0		0.19	0.15	0.0	0.0
27/08/2007	**0.18**	**0.21**	**0.10**	**0.25**		0.040	0.0	0.17	0.39	0.45	0.45
28/08/2007	**0.11**	**0.17**	**0.35**	**0.25**		0.17	1.0	0.35	0.60	0.92	0.95
29/08/2007	**0.60**	**0.52**	**0.32**	**0.55**	**1.0**	0.28	0.64	0.29	0.31	0.63	0.27
30/08/2007	**0.80**	**0.76**	**0.33**	**0.76**	**1.0**	0.50	1.0	0.42	0.53	1.0	1.0
31/08/2007	**0.28**	**0.40**	**0.28**	**0.37**	**1.0**	0.070	1.0	0.21	0.40	0.43	1.0
01/09/2007	**0.29**	**0.47**	**0.40**	**0.61**	**1.0**	0.070	0.33	0.44	0.52	0.89	0.12
02/09/2007	**0.59**	**0.38**	**0.19**	**0.68**	**1.0**	0.12	0.61	0.31	0.79	0.95	1.0
03/09/2007	**0.40**	**0.54**	**0.57**	**0.73**	**1.0**	0.24	1.0	0.40	0.53	1.0	1.0
04/09/2007	**0.21**	**0.40**	**0.50**	**0.61**	**1.0**	0.17	0.47	0.43	0.49	0.92	1.0
05/09/2007	**0.21**	**0.42**	**0.25**	**0.47**		0.050	0.0	0.31	0.58	0.47	0.93
06/09/2007	**1.0**	**0.80**	**0.42**	**1.0**	**1.0**	0.43	0.49	0.52	0.12	1.0	
07/09/2007	**0.29**	**0.31**	**0.24**	**0.55**		0.21	0.68	0.50	0.53	0.62	1.0
08/09/2007	**0.91**	**0.73**	**0.42**	**0.97**		0.26	0.72	0.67	0.52	0.97	1.0

Clearly, the endpoints for the back trajectories for all 14 days have their densest population in the proximity of the receptor site and in general, the farthest reaches of these 5-day trajectories are either – 1. still in the Yukon, 2. arriving from the Northwest Territories, 3. arriving from the northern portions of British Columbia, or, 4. arriving from within the Pacific Ocean with straying endpoints from Alaska. In other words, the back trajectories seem to originate in regions that are relatively close to the sampling site even at their farthest. This would suggest that the sources of PAH at LFL during this period in the summer are most likely related to local activities. The ellipse in Figure 1b also indicates that the $PM_{2.5}$ collected around the LFL receptor site were relatively high during this 14-day period. Thus, we believe that summer activities at camping grounds close to the sampler may yield sooty particles which cause the slow/non-exchangeability of lower molecular mass PAH and result in the shallower slopes for the summer samples in Figure 3. Stabilization of PAH by sooty particles has been reported by previous researchers (Behymer and Hites 1988; Gustafsson et al. 1997) and it may result in PAH being unavailable for true equilibrium gas/particle partitioning.

PAH may also be more susceptible to photolytically-initiated transformation reactions in the atmosphere in the summer compared to wintertime due to increased sunlight. The UV-photolysis of ozone yields hydroxyl radicals and homogeneous atmospheric reactions of organic compounds with the OH radical is considered the most significant atmospheric loss process for these organic compounds including PAH (Atkinson and Arey 1994; Cooper 1996; Brubaker and Hites 1998). Low molecular mass PAH typically found in the gas-phase will be more affected by these atmospheric transformation reactions. Kim et al. have recently investigated and compiled the half-lives obtained upon direct irradiation of PAH sorbed to representative environmental substrates and their study shows that low molecular mass PAH are the least stable on these environmental surfaces (Kim et al. 2009). The combined effect of these processes may cause a greater depletion in gas-phase concentrations of lighter PAH leading to their relatively higher summer log K_p values and may also be a contributory factor in the shallower slopes observed in Figure 3. We note that all these are speculative assumptions at this point and suggest that one of the goals for future studies at this sub-Arctic site should be the proper determination of the actual causal factors of the summertime increase in log K_p (i.e., source- and/or photolytic transformation-dependent) of low molecular mass PAH by conducting carefully designed field and control experiments. This would help determine the seasonal and overall importance of local sources in outcomes of long-range studies.

4. Conclusion

At Little Fox Lake which is a sub-Arctic site, the colder periods showed the highest levels in PAH concentrations and $PM_{2.5}$. An enrichment of particle-associated PAH (see Figure 1) was also observed during the winter season while the warmer periods were marked by low PAH concentrations and lower levels of heavier molecular mass PAH normally associated with the particle-phase. This is consistent with studies by other researchers who have observed that the highest concentrations of anthropogenic particulates in the Arctic which arrive via long-range atmospheric transport occur during winter and the lowest concentrations occur during summer (Halsall et al. 1997; Polissar et al. 1999). The wintertime increase in gas- and particle-phase PAH concentrations observed at this sparsely populated sub-Arctic site may be generally reflective of the increased residential heating requirements in the northern hemisphere occurring during this time.

In the summertime, the partitioning of lower molecular mass PAH seems to exhibit some non-exchangeability though the potentially significant effect of photolytically-initiated atmospheric transformations on gas-phase PAH during this period could not be ruled out. Air mass trajectories suggest that local sources may be more important than long-range source regions during this time. Local summer sources that could give rise to this phenomenon may include summer camping activities at camping grounds close to the receptor site. Our accompanying article explores factor analysis and air mass back trajectories to identify and quantify the underlying sources of PAH impacting this site (Sofowote et al. 2010).

Acknowledgement

Funding for this project from the Government of Canada Program for International Polar Year (IPY) is hereby acknowledged. The authors would like to acknowledge the National Air Pollution Surveillance Program (NAPS) for providing the $PM_{2.5}$ data for the NAPS House station in Whitehorse, Yukon.

Appendix A. Supplementary Data

Supplementary data including partitioning coefficients for samples collected at LFL during the course of this work can be found in Supp. Data_Gas_Part_Final.doc.

REFERENCES

Arp, H.P.H., Schwarzenbach, R.P., Goss, K., 2008. Ambient Gas/Particle partitioning. 2: The influence of particle source and temperature on sorption to dry terrestrial aerosols. Environmental Science & Technology 42, 5951-5957.

Atkins, P.W., 1994. Physical Chemistry. W.H. Freeman, New York,

Atkinson, R. and Arey, J., 1994. Atmospheric chemistry of gas-phase polycyclic aromatic hydrocarbons: Formation of atmospheric mutagens. Environmental Health Perspectives 102 Suppl 4, 117-126.

Behymer, T.D. and Hites, R.A., 1988. Photolysis of polycyclic aromatic hydrocarbons absorbed on fly ash. Environmental Science & Technology 22, 1311-1319.

Brubaker, W.W. and Hites, R.A., 1998. OH reaction kinetics of polycyclic aromatic hydrocarbons and polychlorinated dibenzo-p-dioxins and dibenzofurans. The Journal of Physical Chemistry A 102, 915-921.

Choi, S., Shunthirasingham, C., Daly, G.L., Xiao, H., Lei, Y.D., Wania, F., 2009. Levels of polycyclic aromatic hydrocarbons in Canadian mountain air and soil are controlled by proximity to roads. Environmental Pollution 157, 3199-3206.

Chrysikou, L.P. and Samara, C.A., 2009. Seasonal variation of the size distribution of urban particulate matter and associated organic pollutants in the ambient air. Atmospheric Environment 43, 4557-4569.

Cooper, D.J., 1996. Estimation of hydroxyl radical concentrations in the marine atmospheric boundary layer using a reactive atmospheric tracer. Journal of Atmospheric Chemistry 25, 97-113.

Cortes, D.R., Basu, I., Sweet, C.W., Hites, R.A., 2000. Temporal trends in and influence of wind on PAH concentrations measured near the Great Lakes. Environmental Science & Technology 34, 356-360.

Environment Canada, 2009. Intercontinental atmospheric transport of anthropogenic pollutants to the Arctic: Http://www.ec.gc.ca.libaccess.lib.mcmaster.ca/api-ipy/default.asp?lang=En&n=100FA705-1. 2010.

Gustafsson, O., Haghseta, F., Chan, C., MacFarlane, J., Gschwend, P.M., 1997. Quantification of the dilute sedimentary soot phase: Implications for PAH speciation and bioavailability. Environmental Science & Technology 31, 203-9.

Halsall, C.J., Barrie, L.A., Fellin, P., Muir, D.C.G., Billeck, B.N., Lockhart, L., Rovinsky, F.Y., Kononov, E.Y., Pastukhov, B., 1997. Spatial and temporal variation of polycyclic aromatic hydrocarbons in the Arctic atmosphere. Environmental Science & Technology 31, 3593-3599.

Halsall, C.J., 2004. Investigating the occurrence of persistent organic pollutants (POPs) in the Arctic: Their atmospheric behaviour and interaction with the seasonal snow pack. Environmental Pollution 128, 163-175.

Harrison, R.M. and Yin, J., 2000. Particulate matter in the atmosphere: Which particle properties are important for its effects on health? The Science of The Total Environment 249, 85-101.

Hillery, B.R., Basu, I., Sweet, C.W., Hites, R.A., 1997. Temporal and spatial trends in a long-term study of gas-phase PCB concentrations near the Great Lakes. Environmental Science & Technology 31, 1811-1816.

Hinckley, D.A., Bidleman, T.F., Foreman, W.T., Tuschall, J.R., 1990. Determination of vapor pressures for nonpolar and semipolar organic compounds from gas chromatograhic retention data. Journal of Chemical & Engineering Data 35, 232-237.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans and International Agency for Research on Cancer, 2010. Some Non-Heterocyclic Polycyclic Aromatic Hydrocarbons and some Related Exposures. World Health Organization, International Agency for Research on Cancer ; Geneva, Switzerland : Distributed by WHO Press, Lyon, France; Geneva, Switzerland, 853.

Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A.I.I.I., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J., 2005. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16, 727-736.

Kaupp, H. and McLachlan, M.S., 2000. Distribution of polychlorinated dibenzo-P-dioxins and dibenzofurans (PCDD/Fs) and polycyclic aromatic hydrocarbons (PAHs) within the full size range of atmospheric particles. Atmospheric Environment 34, 73-83.

Kaupp, H. and McLachlan, M.S., 1999. Atmospheric particle size distributions of polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) and polycyclic aromatic hydrocarbons (PAHs) and their implications for wet and dry deposition. Atmospheric Environment 33, 85-95.

Kim, D., Kumfer, B.M., Anastasio, C., Kennedy, I.M., Young, T.M., 2009. Environmental aging of polycyclic aromatic hydrocarbons on soot and its effect on source identification. Chemosphere 76, 1075-1081.

Liu, Y., Liu, L., Lin, J., Tang, N., Hayakawa, K., 2006. Distribution and characterization of polycyclic aromatic hydrocarbon compounds in airborne particulates of East Asia. China Particuology 4, 283-292.

Lohmann, R. and Lammel, G., 2004. Adsorptive and absorptive contributions to the gas-particle partitioning of polycyclic aromatic hydrocarbons: State of knowledge and recommended parametrization for modeling. Environmental Science & Technology 38, 3793-3803.

Ma, Y., Lei, Y.D., Xiao, H., Wania, F., Wang, W., 2010. Critical review and recommended values for the physical-chemical property data of 15 polycyclic aromatic hydrocarbons at 25 °C. Journal of Chemical & Engineering Data 55, 819-825.

Metcalfe, C.D., Balch, G.C., Cairns, V.W., Fitzsimons, J.D., Dunn, B.P., 1990. Carcinogenic and genotoxic activity of extracts from contaminated sediments in western lake ontario. Science of the total environment 94, 125-41.

Odabasi, M., Cetin, E., Sofuoglu, A., 2006. Determination of octanol–air partition coefficients and supercooled liquid vapor pressures of PAHs as a function of temperature: Application to gas–particle partitioning in an urban atmosphere. Atmospheric Environment 40, 6615-6625.

Pankow, J.F. and Bidleman, T.F., 1991. Effects of temperature, TSP and per cent non-exchangeable material in determining the gas-particle partitioning of organic compounds. Atmospheric Environment Part A, General Topics 25, 2241-2249.

Pankow, J.F., 1987. Review and comparative analysis of the theories on partitioning between the gas and aerosol particulate phases in the atmosphere. Atmospheric Environment (1967) 21, 2275-2283.

Pankow, J.F., 1998. Further discussion of the octanol/air partition coefficient koa as a correlating parameter for gas/particle partitioning coefficients. Atmospheric Environment 32, 1493-1497.

Polissar, A.V., Hopke, P.K., Paatero, P., Kaufmann, Y.J., Hall, D.K., Bodhaine, B.A., Dutton, E.G., Harris, J.M., 1999. The aerosol at Barrow, Alaska: Long-term trends and source locations. Atmospheric Environment 33, 2441-2458.

Pope III, C.A., Burnett, R.T., Thurston, G.D., Thun, M.J., Calle, E.E., Krewski, D., Godleski, J.J., 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. Circulation 109, 71-77.

Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132-1141.

Sofowote, U.M., Hung, H., Rastogi, A.K., Westgate, J.N., Su, Y., DeLuca, P., McCarry, B.E., 2010. Assessing the long-range transport of PAH to a sub-Arctic site using positive matrix factorization and potential source contribution function (ATMENV-D-10-00565) *submitted.* Atmospheric Environment

Venkataraman, C., Lyons, J.M., Friedlander, S.K., 1994. Size distributions of polycyclic aromatic haydrocarbons and elemental carbon. 1. Sampling, measurement methods and source characterization. Environmental Science & Technology 28, 555-62.

Venkataraman, C. and Friedlander, S.K., 1994. Size distributions of polycyclic aromatic hydrocarbons and elemental carbon. 2. Ambient measurements and effects of atmospheric processes. Environmental Science & Technology 28, 563-572.

Welch, B.L., 1947. The Generalization of 'Student's' Problem when several Different Population Variances are Involved. Biometrika 34, 28-35.

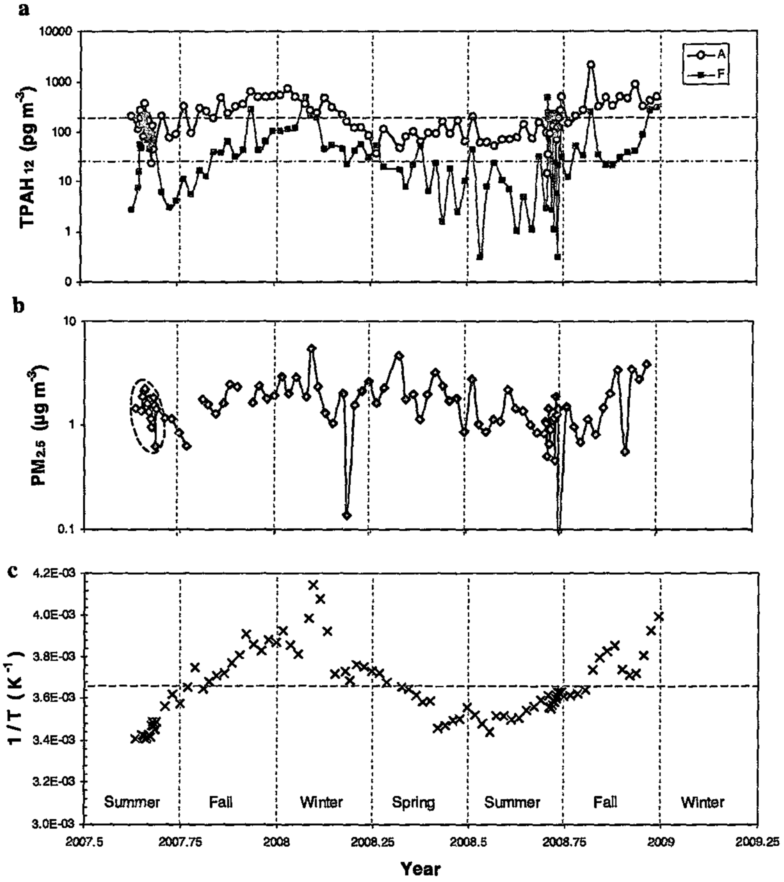


Figure 1: (a) Temporal variation of blank-corrected $TPAH_{12}$ in the gas- and particle- phase at LFL (A; top line with diamonds and F; bottom line with squares respectively). Dashed lines in Figure 1a represent the geometric means of the $TPAH_{12}$ in a given phase. (b) Temporal variation of $PM_{2.5}$ collected at the closest PM sampling site. The ellipse is explained later. (c) Variation of temperature at sampling site. Dashed line represents arithmetic mean temperature for period of study (1.24°C).

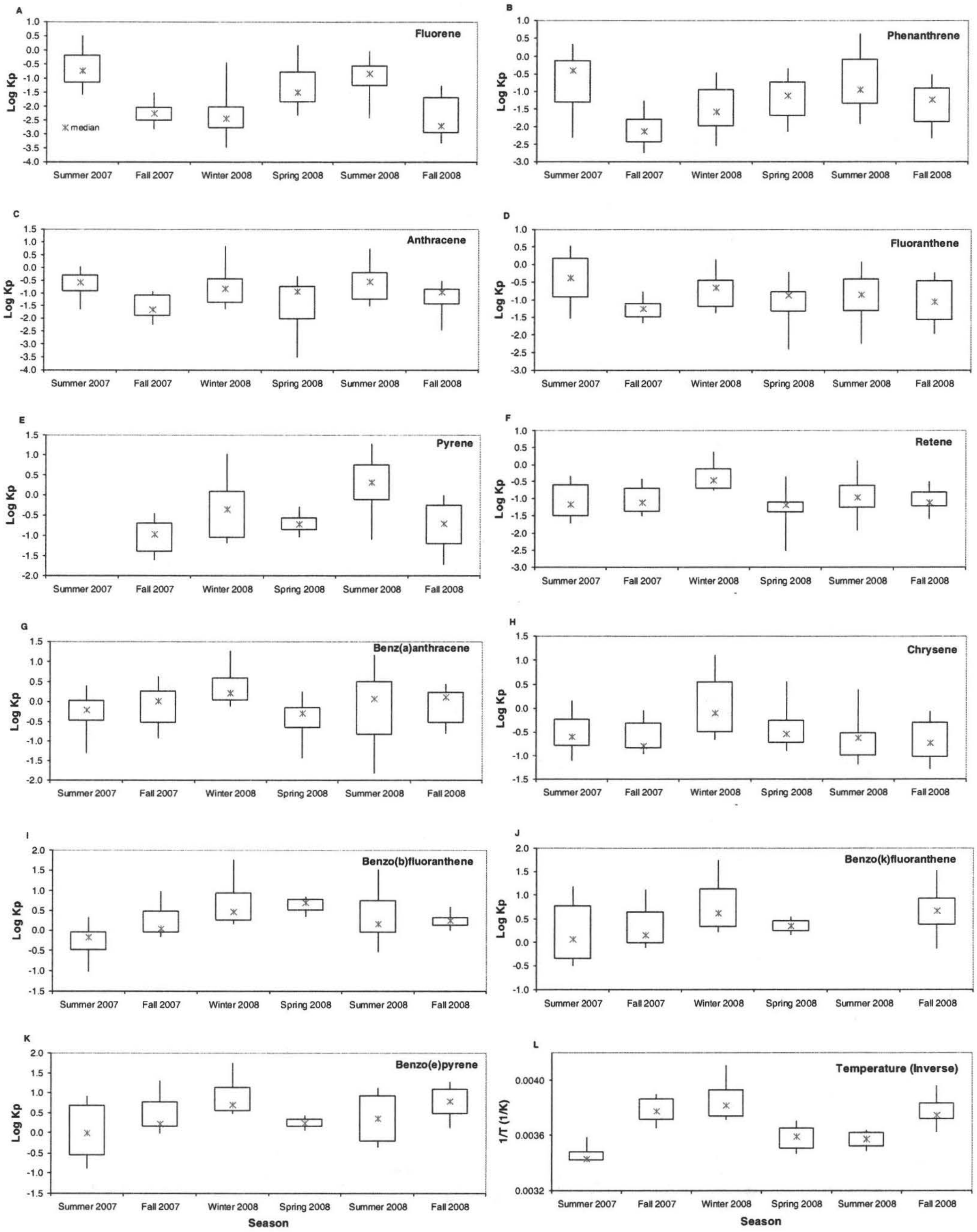


Figure 2: Box and whisker plots of the seasonal of averages log K_p for 11 PAH and temperature measured at LFL between August 2007 and December 2008. The error bars represent the 5th and 95th percentile and the star represents the median value.

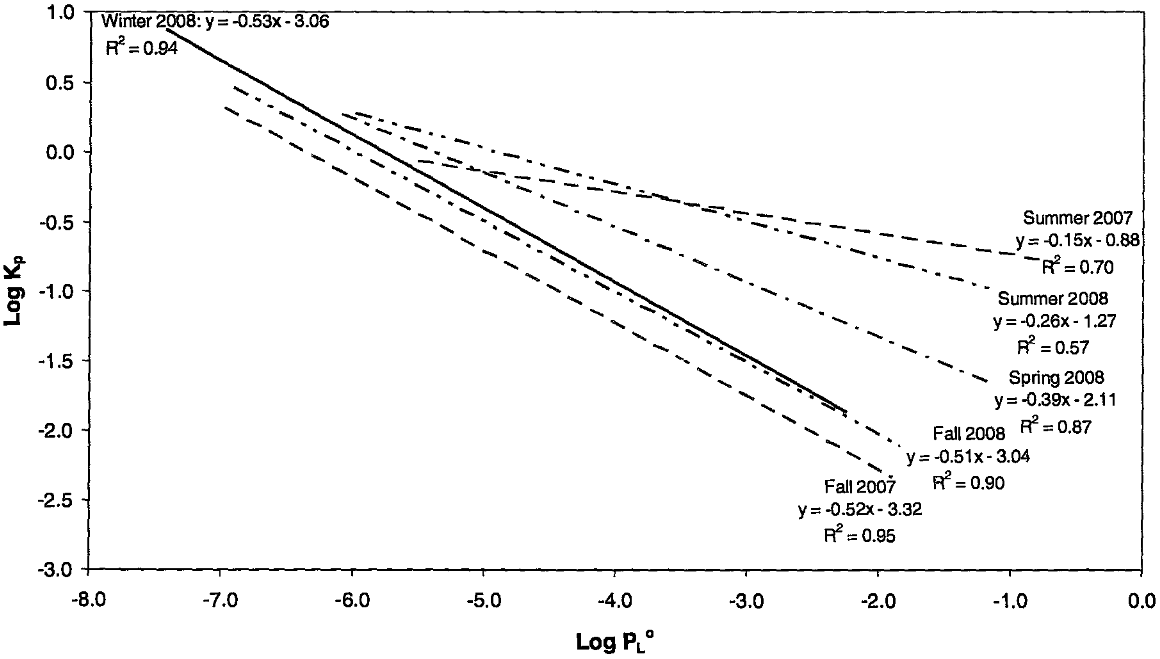


Figure 3: Average log K_p versus logP_L° for 10 PAH (benzo[*a*]pyrene and benzo[*ghi*]perylene partition almost exclusively into the particle-phase when detected) collected at Little Fox Lake classed by seasons between August 2007 and December 2008.

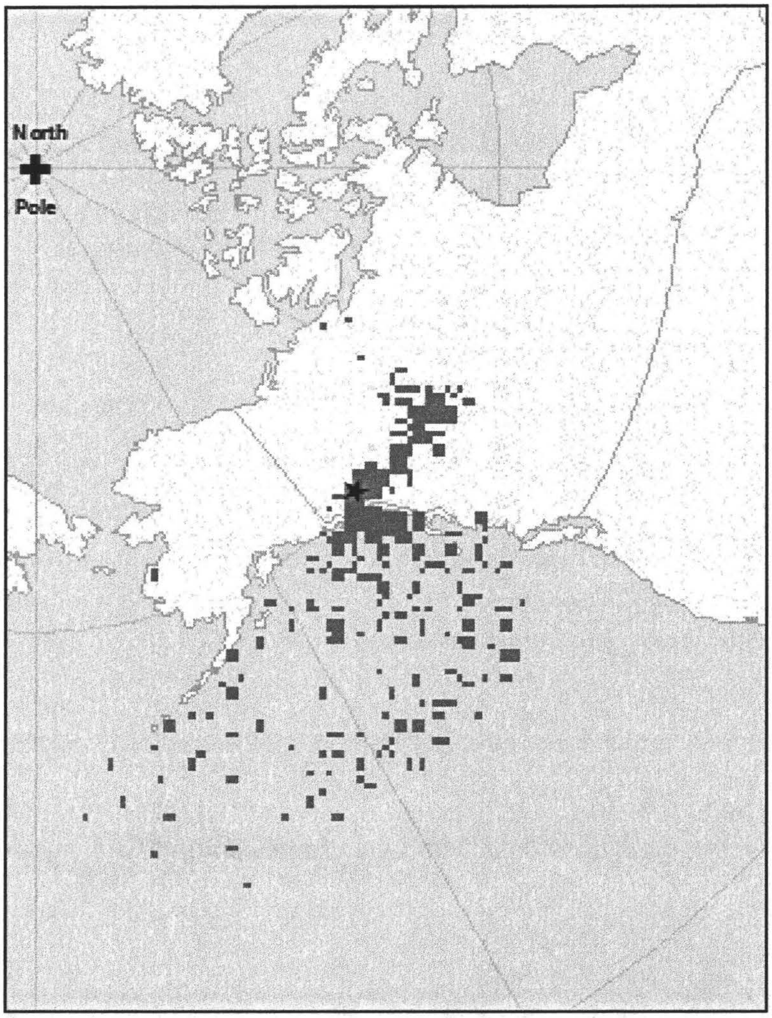


Figure 4: Five-day air mass back trajectories collected from August 25, 2007 to September 08, 2007 for LFL. The densest aggregation of endpoints occurs around the sampling site indicating the importance of local impacts during this time.

Chapter Three

Assessing the Long-range Transport of PAH to a sub-Arctic Site using Positive Matrix Factorization and Potential Source Contribution Function

The work in this chapter has been submitted in an article authored by: Uwayemi M. Sofowote[1], Hayley Hung[2], Ankit K. Rastogi[1], John N. Westgate[3], Patrick F. Deluca[4], Yushan Su[2], and Brian E. McCarry[1] to *Atmospheric Environment* for consideration for publication.

Author Contributions:

Uwayemi Sofowote: Performed some of the sample collection and all the sample cleanups jointly with Johnny Westgate; carried out all the PAH quantification, data mining, interpretation, statistical analyses, and wrote the first draft of the manuscript.

Dr. Hayley Hung: Initiated this project (IPY-INCATPA) and was responsible for logistics.

Ankit Rastogi: Designed a PSCF source code, generated all PSCF input matrices.

John Westgate: Performed some of the sample collection and sample cleanups jointly with Uwayemi Sofowote.

Patrick DeLuca: Supplied the basic input ASCII file for PSCF raster matrices. Designed the final PSCF maps used in the article and edited sections of the manuscript.

Dr. Yushan Su: Supplied the final pressure-corrected PAH concentrations used in the article; was responsible for ensuring concentration data integrity.

Dr. Brian McCarry: Supervised many aspects of the data interpretation. Edited and wrote many portions of the final manuscript.

Supplementary Information:

Supplementary tables and figures for this manuscript can be found in the attached compact disc.

Author Affiliations:

[1]Department of Chemistry and Chemical Biology
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[2]Air Quality Research Division
Environment Canada
4905 Dufferin Street, Toronto, Ontario, M3H 5T4, Canada

[3]Department of Physical and Environmental Sciences
University of Toronto Scarborough
1265 Military Trail, Toronto, Ontario, M1C 1A4, Canada

[4]Centre for Spatial Analysis, School of Geography and Earth Sciences
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4K1

Corresponding author email address: mccarry@mcmaster.ca; telephone: +1 (905) 525-9140

Assessing the Long-range Transport of PAH to a sub-Arctic Site using Positive Matrix Factorization and Potential Source Contribution Function

Uwayemi M. Sofowote, Hayley Hung, Ankit K. Rastogi, John N. Westgate, Patrick F. Deluca, Yushan Su, Brian E. McCarry

Abstract

Gas-phase and particle-phase atmospheric samples collected in a sparsely populated sub-Arctic environment in the Yukon Territory, Canada were analyzed for a wide range of organic pollutants including polycyclic aromatic hydrocarbons (PAH). Receptor modeling using positive matrix factorization (PMF) was applied to a PAH data set from samples collected between August 2007 and December 2008 to afford four factors. These factors were designated as coal combustion emissions, particle-phase wood combustion emissions, gas-phase wood combustion emissions, and unburned petroleum/petrogenic emissions. The multiple linear regression-derived average contributions of these factors to the total PAH concentrations were 14% for coal combustion, 6% for particle-phase wood combustion emissions, 46% for gas-phase wood combustion emissions and 34% for petrogenic emissions. When the total PAH concentrations (defined as the sum of twenty-two PAH) and the PMF-modeled PAH concentrations set were compared, the correlation was excellent (R^2=0.97). Ten-day back trajectories starting at four different heights were used in a potential source contribution function analysis (PSCF) to assess the potential source regions of these PAH factors. Mapping the computed PSCF values for the four PMF factors revealed different source regions in the northern hemisphere for each PMF factor. Atmospheric transport of PAH occurred from both relatively short and long distances with both continental (North American) and trans-oceanic (Asian) sources contributing significantly to the total PAH. This study provides evidence of the transport of coal and wood combustion emissions from Asia, continental North America and northern Europe to sub-Arctic Canada (and by extension to the Canadian Arctic) primarily during cooler (fall-winter) months. This study demonstrates for the first time that the combined PMF-PSCF methodology can be used to identify geographically disperse PAH source contributors on a hemispherical scale.

1. Introduction

Long-range atmospheric transport of contaminant species to seemingly pristine regions has been investigated by air quality researchers over the years

(Bailey et al. 2000; Genualdi et al. 2009; Harner et al. 2005; Jaffe et al. 1999; Noël et al. 2009; Primbs et al. 2008). Trans-oceanic transport has been implicated in several cases. In the context of the Arctic environment, both trans-Pacific (Bailey et al. 2000) and trans-Atlantic (Cheng et al. 1993) transport of particulate species from lower latitudes have been identified as entry routes. The maxima of the influx of anthropogenically-derived particulates to the Arctic have been found to occur in the winter-spring period (Cheng et al. 1993; Polissar et al. 1999). This phenomenon is referred to as Arctic haze and its duration has been the timeframe of study for many inorganic constituents of aerosols discussed in the literature.
PAH, a class of ubiquitous and potentially carcinogenic (Metcalfe et al. 1990; Legzdins et al. 1995; International Agency for Research on Cancer 2010), semi-volatile organic contaminants composed of multiple-fused benzene rings, have been found in the Arctic.

As part of many research activities associated with the International Polar Year program (IPY), a Canadian project (IPY-INCATPA; INterContinental Atmospheric Transport of Anthropogenic Pollutants to the Arctic) was established to study the long-range atmospheric transport of a wide range of organic pollutants including PAH originating from lower latitudes to the Arctic (Environment Canada 2009). Atmospheric long-range transport is believed to be a significant entry process of these contaminants (Halsall et al. 1997; Halsall et al. 2001) but not until recently have potential impacting source regions been spatially mapped for an Arctic location (Wang et al. 2010). Atmospheric samples were collected at various sites including one located in a sub-Arctic environment in the Yukon Territory near Little Fox Lake (LFL; 61° 21’ N, 135° 38’ W) for the IPY-INCATPA project. The goal of our study was to apportion and spatially map the sources of PAH impacting this site, thus, the source apportionment work discussed in this article is concerned with establishing the chemical composition of sources of PAH and quantifying their relative contributions.

Source apportionment of PAH can be attempted using receptor modeling, a set of qualitative and quantitative apportionment models used to account for sources of pollution and their relative contributions at a receptor (sampling) site (Hopke 1985). Specifally, we have attempted receptor modeling by using factor analysis (FA) since FA can aid in identifying and quantifying contributory sources of pollution without prior knowledge of all the sources and their characteristics (Hopke 1985). Positive matrix factorization (PMF; (Paatero and Tapper 1993) has become a factor analytic model of choice for quantitative source apportionment of contaminant species in many air quality monitoring studies (Paterson et al. 1999; Xie et al. 1999a; Xie et al. 1999b; Larsen and Baker 2003; Polissar et al. 1998).

While FA may yield factor (source) profiles and their relative contributions to samples collected at a receptor site, they do not explicitly reveal a potential geographical source region of the identified factors impacting the site. Pollutional source regions may not be in the immediate locality of the sampling site. This is particularly true for pristine environments such as the Polar regions. Chemical species that have been observed to pollute these environments are believed to have arrived at these places via long-range transport. In the case of fine particulate matter, volatile and semi-volatile organic species, the atmosphere is believed to be the most significant environmental compartment for transport. Thus, meteorological information in the form of air mass back trajectories may be needed to provide information about the atmospheric transport pathway of contaminant species collected at a receptor site and by extension, their potential source regions (Bailey et al. 2000, Killin et al. 2004). Potential source contribution function (PSCF; Ashbaugh, Malm and Sadeh 1985) is one form of residence time analysis of back trajectory endpoints within a fixed (grid cell) space. PSCF constitutes the conditional probability that a given grid cell is a source region of the contaminant species of interest if air mass trajectories passing through the cell on their way to the receptor site contain measured levels of the contaminant above a pre-defined threshold. Thus PSCF computations identify potential source regions and the preferred pathways of pollutant species to the receptor site. Many studies have employed PSCF to give a spatial rendering of contaminant species including inorganic components of aerosols (Zeng and Hopke 1989; Ara Begum et al. 2005). In the 1990s, Cheng et al. studied aerosols collected in the Canadian High Arctic (Alert, Nunavut) during the winter months using multi-height back trajectories for their PSCF computations and identified geographically disperse source regions in Asia and Europe (Cheng et al. 1993) for non-marine sulphate and some inorganic constituents of air particulates. Source regions of organic contaminants have also been spatially mapped using the PSCF technique. On a continental scale, Hafner and Hites have explored the source regions of gas-phase PAH, polychlorinated biphenyls (PCB) and pesticides sampled at receptor sites around the Laurentian Great Lakes in North America (Hafner and Hites 2003). Hsu et al. explored the source regions of PCB measured at sampling sites in Chicago (Hsu, Holsen and Hopke 2003).

PSCF has also been combined with factor analytic models such as principal component analysis (Zeng and Hopke 1989) and more recently, PMF, for spatial mapping of source profiles identified by the factor analyses. The PMF-PSCF methodology has been employed by Hopke et al. (Polissar et al. 1999; Hopke, Zhou and Poirot 2005; Poirot et al. 2001; Pekney et al. 2006; Polissar, Hopke and Poirot 2001; Gao et al. 2006; Kim et al. 2004; Kim et al. 2005; Heo, Hopke and Yi 2009; Zhou, Hopke and Zhao 2009) for assessing the long-range transport potential of constituents of aerosols from source regions. Du and Rodenburg (Du and Rodenburg 2007) identified and mapped four PCB factors from their PMF-PSCF analyses on samples collected in Camden, Ohio. Liu et al. used the PMF-PSCF methodology to identify and map out six and seven factors in Stockton and Potsdam, New York respectively using inorganic species and PAH (Liu et al. 2003). A stand-alone PMF-PSCF analysis on PAH has not been carried out until now. This work also is the first work to use the PMF-PSCF methodology to assess the potential for long-range transport of organic contaminants on a hemispherical scale.

2. Descriptions of the Receptor Models Used and their Source Apportionment Characteristics

The following section is a brief summary of the source apportionment principles of the receptor models used in this work. More thorough discussions on the formulations and characteristics of these models should be sought in the literature cited.

2.1. Positive Matrix Factorization (PMF): PMF treats factor analysis as a true least squares problem (Paatero and Tapper 1993) i.e., for a data set denoted by matrix X(n x m), X can be solved for as:

$$X = GF + E \qquad (1)$$

where F, G and E are the factor profile, contribution and residual matrices, respectively.

Then, Q(E), an object function of the residual matrix E(n x m) which is to be minimized, is defined as a constrained and weighted least squares function as follows:

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{e_{ij}}{s_{ij}} \right]^2 \qquad (2)$$

where e_{ij} is the residual value defined as:

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \qquad (3)$$

$\hat{x}_{ij}$ is the factor analytic modeled value of the measured x_{ij} value and s_{ij} is the uncertainty associated with the determination of x_{ij}. The values f_{kj} and g_{ik} are elements of the F and G matrices, respectively. The constraints arise because $F \geq 0$, $G \geq 0$, i.e., are not allowed to assume negative values. In a properly weighted analysis, the Q value should be approximately equal to the optimum theoretical Q of the data set estimated as nm-p(n+m) (United States Environmental Protection Agency 2008).

Appropriate uncertainty estimates s_{ij} are imperative for PMF analyses. PMF programs allow the user to generate the S matrix based on the method detection limit and the uncertainty in the chemical measurement of the contaminant species of interest. PMF treats missing data values differently from values below the detection limit for optimized model fit to the data. For example, EPA PMF 3, based on the multilinear engine (ME) program (Paatero 1999) can replace missing data values with the median value of the species prior to analysis. The use of measurement uncertainties may not represent the total uncertainties in a given data set. Chen et al. inform that the use of measurement uncertainties may be an underestimation of the true s_{ij} (Chen et al. 2007). The EPA PMF 3 program allows the addition of extra modeling uncertainty which may arise from variability in source consistency and other indeterminate factors before analysis (United States Environmental Protection Agency 2008). The robust mode is the default mode with EPA PMF 3. This reduces the effect of very large variables by treating them as outliers so that they do not distort the model's fitting process (Hopke 2001).

Estimation of factors in PMF is subjective and there is no hierarchy of factors as with eigenvalue analyses since orthogonality of factors is not a requirement. For our purposes, we have found that the quickest way to estimate the number of unique factors is to initially extract a relatively large number of factors and to step-wisely reduce this number based on the results obtained from bootstrapping (United States Environmental Protection Agency 2008); a feature supplied with the PMF program. By applying and fine-tuning this bootstrapping feature (which provides uncertainty estimates for the factor profiles based on re-sampling the initial factor solutions in a user-specified number of random runs), one can determine in quantitative terms the percentage of bootstrapped factors mapped to their corresponding base factors and thus the uniqueness of a given factor profile. If a relatively high amount of a given bootstrapped factor is spread over more than one base factor, it may suggest that the factor profile solutions are not unique, thus, the number of factors initially extracted may have to be reduced. More details on bootstrapping factor profiles, factor rotations (Fpeak and G-space) and weighting the strengths of variables and other PMF features can be found elsewhere (United States Environmental Protection Agency 2008; Paatero and Hopke 2003; Paatero et al. 2005).

2.2. Potential Source Contribution Function (PSCF): For a given ij^{th} grid cell,

$$\text{PSCF}_{ij} = \frac{m_{ij}}{n_{ij}} \qquad (4)$$

where n_{ij} is the total number of endpoints of back trajectories that reach the receptor site and m_{ij} for a given pollutant or contaminant species is a subset of n_{ij} whose back trajectory endpoints are associated with concentration values higher than the pre-defined threshold for the pollutant. Detailed discussions on the formulation of PSCF can be found elsewhere (Cheng et al. 1993). The use of back trajectories starting at different heights has been encouraged since they not only better approximate the motion of air parcels over rough terrains (Cheng et al. 1993) but they also increase the number of back trajectory endpoints used in the analysis compared to single layer trajectories. The use of back trajectories in ensembles has been noted to give a better estimation of the position of air parcels since random errors will generally be averaged out (Hopke, Zhou and Poirot 2005).

Weighting factors are usually applied to raw PSCF values to reduce the potential for spurious identification of source regions and 'tailing' effects (Cheng et al. 1993; Cheng and Lin 2001; Zhou, Hopke and Liu 2004). These weighting factors are usually a function of n_{ij} and their effect is to eliminate the possibility that grid cells with relatively small n_{ij} values have a PSCF value approaching 1. Sizes of grid cells may be constructed to depend on the time length of the back trajectories used. The accuracy of back trajectories deteriorates with the length of time of the trajectories. Stohl summarises the absolute and relative errors of many kinds of trajectories and notes that the errors could be in the hundreds to thousands of kilometres for travel times in the 36 – 120 h range (Stohl 1998). Cheng et al. (Cheng et al. 1993) and Polissar et al. (Polissar et al. 1999) have used 5^{o} x 5^{o} (lat-long coordinates) grid cells to capture the uncertainty associated with their 5-day back trajectories. It must be re-iterated that while PSCF computations are helpful in the spatial study of source regions of contaminant species, they cannot replace emissions inventories; as noted by other researchers, they should be viewed as providing qualitative information only (Cheng et al. 1993). Also, PSCF will reveal the locations of potential source regions of contaminant species of interest only if these species are borne by air masses that impact the sampling site after passage through the source region, thus, low PSCF values do not necessarily mean an absence of source regions (Polissar, Hopke and Poirot 2001).

3. Experimental Section

3.1. Air Sampling: The Canadian INCATPA air monitoring site is located about 100km north of Whitehorse (60° 43' N, 135° 03' W; population: ~20000) in a mountainous region near Little Fox Lake in Yukon Territory, Canada (LFL; 61° 21' N, 135° 38' W; 1138m a.s.l.) and is equipped with a high volume (Hi-Vol) air sampler for the collection of gas- and particle- phase contaminants including PAH. Weekly air samples were collected at the Little Fox Lake, Yukon site (LFL; see Figure S1; Supporting Information for a map of all INCATPA sites) from August 2007 to October 2009. This study focuses on samples collected from August 12, 2007 to December 28, 2008. Particle- and gas- phase samples were collected on a glass fibre filters (GFF) and two inline polyurethane foam (PUF) plugs, respectively, housed in a high volume (Hi-Vol) air sampler described previously (Halsall et al. 1997). The Hi-Vol sampler was fitted with a 10 μm diameter particle size selective inlet and the flow rate was calibrated so that approximately 12000 m^3 of air were collected over a seven-day period. Ninety-five samples (comprising 67 weekly samples and 28 daily samples) collected from August 2007 to December 2008 were used in this study.

3.2. Sample Extraction and Clean-up: Each PUF plug and the air particulate filter sample was individually subjected to a 24-h Soxhlet extraction (carried out by AirZone One Ltd., Mississauga, ON, Canada) by hexane and dichloromethane, respectively, to afford a gas-phase extract and a particle-phase extract, respectively, for each sampling period. Previous studies using the same sampling and extraction protocols had shown that PUF breakthrough of PAH with $\geq$3 rings was negligible (Halsall et al. 1997); extracts from the two PUF plugs were combined before further analysis. One half of each raw PUF or GFF extract was cleaned-up according to protocols used at the National Laboratory for Environmental Testing (NLET, Environment Canada, Burlington, ON, Canada). Briefly, extracts were carefully evaporated to 1 mL under nitrogen and placed on a silica gel column (8g, activated overnight at 160°C) that had been wet-packed in hexane; the column was prepared with anhydrous sodium sulphate plugs (1 cm, fired at 600°C) top and bottom. Elution with hexane (70-90 mL) afforded Fraction A; elution with hexane/dichloromethane (1:1, v/v, 130 mL) afforded Fraction B. Fractions A and B were evaporated separately, reduced in volume using nitrogen, and solvent-exchanged into iso-octane (1 mL).

3.3. GC/MS Analysis and Quality Control: All GC/MS analyses were performed at NLET on 5% phenylmethylsilicone columns (30m x 0.25mm x 0.25μm film; J&W DB-5ms). Twenty-one PAH and 4 other polycyclic aromatic compounds (PAC) were quantified by this protocol where possible. Four deuterated PAH surrogates were spiked into the raw extracts and method blanks before silica gel clean-up. Their recoveries (*n* > 200 samples) were: naphthalene-d_8, 93±11%; fluorene-d_{10}, 101±12%; pyrene-d_{10}, 104±11%; perylene-d_{12}, 94±11%. As noted above, breakthrough for PAH with molecular masses in the 128–154 Da range was expected to be significant, thus data for a total of ten compounds (including perylene a 252 Da PAH which was almost never detected) were excluded from the data set leaving the 15 un-substituted PAH in Table 1. The entire sample set was re-run at a later date on a 50% phenylmethylsiloxane column (J&W DB-17ht; 30 m x 0.25 mm internal diameter x 0.15 μm film) for the quantification of the seven alkylated phenanthrenes (Poster et al. 2003) also reported in Table 1. Field blanks comprising pairs of GFFs and PUFs were collected once every 4 weeks. Method detection limits in pg m^{-3} (determined as yearly mean concentrations of PAH for all field blanks + 3 standard deviations) for all 22 PAH used in this study can be found in Table 1.

4. Back Trajectories & Data Analysis

Ten-day back trajectories generated by the Canadian Meteorological Center's (CMC) Trajectory Model (D'Amours and Page 2001) were supplied for the Little Fox Lake, Yukon sampling site. The CMC trajectories are three-dimensional and based on analyzed wind fields of the Numerical Weather Prediction Center (NWP). Noel et al. recently reported that the CMC Trajectory Model gave results similar to the National Oceanic and Atmospheric Administration's Hybrid Single-Particle Lagrangian Intergrated Trajectory model (HYSPLIT) for a study investigating trans-Pacific transport of polybrominated diphenyl ethers (PBDE) (Noël et al. 2009). Relevant trajectory outputs for this study included lat-long coordinates and time segments of endpoints.

In the current work, the CMC back trajectories were computed for four starting heights (1138m, 2000m, 3000m and 4000m a.s.l.) and four start times each day (0000 h, 0600 h, 1200 h and 1800 h UTC). The maximum height for effective down-mixing of air masses for this site is not known. Harner et al. had however shown that the maximum likelihood of trans-Pacific transport of organochlorine pesticides to the Fraser Valley, British Columbia (a site also on the west coast of Canada) occurred in the mid-troposphere (Harner et al. 2005). Thus, we believe that the trajectories with higher starting heights (i.e. between 2000m – 4000m) are more suited to capturing trans-Pacific events. In our PSCF computations, trajectories starting at all four heights were used at equal weighting since like Cheng et al. who discussed the improved reliability potentially obtainable with using trajectories starting at different heights for PSCF computations over rough terrains (Cheng et al. 1993), there was no way for us to determine the relative probability of downward material transfer from each of the four trajectory levels. The trajectories used in our work regressed in time for 10 days at six-hour intervals yielding 41 endpoints per 10-day trajectory. In all, over 3.2×10^5 back trajectory endpoints (508 days x 4 start times x 4 starting heights x 41 endpoints) were used.

An in-house code was developed with Java JDK 6 (Oracle Corporation, USA) and JCreator LE 4.50 (Xinox Software, Netherlands) for the computation of PSCF of PAH and PMF factors in 5^o x 5^o grid cells for the entire northern hemisphere; weekly or daily contaminant concentrations were tagged to the air mass back trajectory endpoints. Visualizations of the PSCF were done using the ArcGIS 9.3 mapping program. PMF analyses were carried out using the EPA PMF 3 software program (http://www.epa.gov/heasd/products/pmf/pmf.htm). Multilinear regression (MLR) analyses were carried out on the PMF G-matrix using the Microsoft Excel (Tools/Data Analysis/Regression, Microsoft Office Professional 2003) regression function.

5. Results and Discussion

At the Little Fox Lake (LFL) sampling site, gas-phase PAH accounted for ~80% of the total PAH ($TPAH_{22}$, defined as the sum of the 22 PAH listed in Table 1) on average. Higher molecular mass PAH (≥252 Da) dominated the particle-phase almost exclusively. Gas-phase and particle-phase concentrations were combined to prevent a partition-driven interpretation (Lee et al. 2004).

5.1. PMF Analysis on the LFL Data Set: The data set used was a 95 × 22 matrix (sample number × number of compounds). Nine variables were downweighted as weak based on their signal-to-noise ratios in the model diagnostics, thus the data was effectively fitted by 13 strong variables. PMF was run on the data with 40 initial random starting points and 10% extra modeling uncertainty. A four-factor solution was found to be ideal from 200 random bootstrapping runs with a minimum correlation of R=0.7 (see PMF Diagnostics in Table S1) and an Fpeak = -0.1 was used to sharpen the differences of the factor contributions in the G-space. The regressions between the measured and modeled $TPAH_{22}$ and the average PAH concentrations for all the samples were R^2=0.97 and R^2=1.0 respectively (Figure S2). The multiple linear regression equation of the G matrix (factor contributions) against the $TPAH_{22}$ values (pg m^{-3}) was found to be:

$$TPAH_{22}=56.3G_1+21.9G_2+180G_3+134G_4 \quad (5)$$

Table 1: Method detection limits (MDL) and list of abbreviations of PAH quantified in this study.

PAH[a]	Abbreviation	MDL (pg m^{-3})
fluorene	FLUOR	5.8
phenanthrene	PHEN	43
anthracene	ANTH	3.6
3- methylphenanthrene	3MP	3.6 [b]
2- methylphenanthrene	2MP	3.6 [b]
9- methylphenanthrene	9MP	3.6 [b]
1-methylphenanthrene	1MP	3.6 [b]
4-methylphenanthrene	4MP	3.6 [b]
2,6-dimethylphenanthrene	2_6DMP	3.6 [b]
1,7(+2,5)-dimethylphenanthrene	1_7/2_5DMP	3.6 [b]
fluoranthene	FLUORT	12
pyrene	PYR	31
retene	RETENE	13
benz[*a*]anthracene	BENZOA	0.22
chrysene+triphenylene	CHRY	0.84
benzo[*b*]fluoranthene	B_B_F	0.24
benzo[*k*]fluoranthene	B_K_F	0.24
benzo[*e*]pyrene	B_E_P	0.21
benzo[*a*]pyrene	B_A_P	0.22
indeno[1,2,3-*cd*]pyrene	INDENO	0.31
dibenz[*a,h*]anthracene	DB_AH_A	0.28
benzo[*ghi*]perylene	B_GHI_P	0.71

[a] The order of PAH in the table is the order of elution on the DB-5ms column. [b] The MDL for anthracene (a low abundance 178 Da isomer) was used as a best match for the low abundance alkylated phenanthrenes since their concentrations were not determined initially with the other PAH.

From equation (5), the MLR-derived average contributions of these factors to the total PAH concentrations were 14% for Factor 1 (FAC_1), 6% for Factor 2 (FAC_2), 46% for Factor 3 (FAC_3) and 34% for Factor 4 (FAC_4) for the data set. The robust Q for the chosen random run was approximately 797. This value is in agreement with the approximate theoretical Q of 803, the expected value for a four-factor solution from a 95 x 13 matrix. Figure 1 shows the bootstrapped profiles of the four PMF factors (left) and their MLR-derived contributions to the $TPAH_{22}$ (right).

Factor 1 is dominated by particle-phase PAH, with higher mass PAH (i.e., ≥ 252 Da) accounting for 60-95% of the concentrations of these PAH in the entire data set. Bootstrapping the factor profile revealed it to be unique, as 100% of the bootstrapped factors of FAC_1 mapped to the base factor. The source contribution plot (Figure 1B) reveals a distinct seasonality in the profile, with highest levels occurring during winter and significantly lower levels during summer. The factor profile (Figure 1A) shows PAH with the highest levels are fluorene, phenanthrene, fluoranthene, chrysene+triphenylene and PAH with molecular masses ≥252 Da, a profile consistent with coal combustion. In a PMF-PSCF study on metals and PAH in NY, high levels of fluorene, phenanthrene, fluoranthene and pyrene were shown to be associated with emissions from coal-fired plants (Liu et al. 2003). We also compared the benzo[e]pyrene-normalized profile of this bootstrapped factor with that of NIST SRM 1597a (National Institute of Science and Technology Standard Reference Material PAH in Coal Tar; (Wise and Watters 2006). Within the uncertainty associated with FAC_1, the two profiles were similar (Figure S3) except for the absence of anthracene, benz[*a*]anthracene and benzo[*a*]pyrene in FAC_1; their absence are not surprising given that these PAH are known to be readily transformed in the atmosphere.

Factor 2 (Figure 1C) had ~91% of its bootstrapped factors mapped to its base factor with 6% being mapped to the third factor, indicating some similarity between these two factors. Factor 2 was also dominated by heavier PAH typically associated with the particle-phase, and had very low contributions from lighter PAH. Only Factors 1 and 2 had contributions from heavier PAH. Retene, a wood combustion marker, (Ramdahl 1983) and pyrene are the most abundant PAH in this factor profile. The source contribution plot (Figure 1D) shows a fairly flat baseline with occasional spikes during specific episodes. The highest concentration for Factor 2 was found as a one-week spike just prior to the coldest week recorded for 2008. This factor probably represented particle-phase wood combustion emissions. A comparison of the benzo[e]pyrene-normalized bootstrapped FAC_2 profile with that of particulate PAH from wood combustion emissions compiled by other researchers (Bzdusek et al. 2004) showed that within the uncertainty limits, they were essentially indistinguishable (see Figure S4).

The third factor (Figure 1E) was dominated by gas-phase PAH, with no PAH with molecular masses ≥ 228 Da. Factor 3 was unique, with ~100% of bootstrapped factors being mapped to the base factor. Retene was the most abundant PAH in FAC_3 and >70% of total retene in the data set is accounted for in this factor. The ratio of two dimethylphenanthrenes (1_7DMP and 2_6DMP) (Figure 1E) was consistent with reports that 1_7DMP was enriched in comparison with 2_6DMP in wood combustion emissions (Yunker et al. 2002; Benner et al. 1995). The source contribution plot (Figure 1F) shows a strong seasonal pattern, with episodic events in the 43rd and 49th weeks of 2008; the highest PAH levels for the entire data set were observed in the 43rd week of 2008 in both the gas-phase and particle-phase. This factor potentially reflected gas-phase wood combustion emissions. The seasonality of this profile was consistent with residential heating emissions.

The fourth factor (Figure 1G) was also unique with ~100% of its bootstrapped factors being mapped to its base factor. This factor was also dominated by gas-phase PAH; 76-98% of the levels of the five methylphenanthrenes and anthracene (a readily transformed PAH) were accounted for in this factor. Methylphenanthrenes were used to identify uncombusted petroleum in a factor analytic study (Kavouras et al. 2001) and to identify evaporative/uncombusted petroleum sources over the Hudson River airshed in a PMF study (Lee et al. 2004). The high level of anthracene points to a continuous, ubiquitous source. Figure 1H shows highest levels for this factor in the fall and winter months when atmospheric transformation is lowest. Any activity involving petroleum (e.g. incomplete combustion or volatilization) may give rise to this factor.

5.2. PSCF on the PMF Factors and Individual PAH: PSCF were computed for the factor contributions of the four PMF factors as well as individual PAH. In all cases, the threshold used for defining the m_{ij} was the arithmetic mean+1 standard deviation (S.D) of the species of interest, except for the fourth PMF factor (FAC_4), where the arithmetic mean was used as the threshold. This was done because only a few daily samples in 2007 had values above the more stringent arithmetic mean + 1S.D threshold for the fourth factor.

A stereographic projection was used to rectify the grid cells so that they covered equal areas as seen in Figures 2A – 2D. Due to the relatively large extent of 5° x 5° grid cells, the weighting factors applied to the raw PSCF (listed below) were not based directly on n_{ij} but on a parameter we defined to count the number of unique back trajectories (BT_{ij}; note that in this case one BT_{ij} can have endpoints ranging from 1 – 41 n_{ij}) in each grid cell.

$$W_{ij} = \begin{Bmatrix} 0.25, BT_{ij} = 1-4 \\ 0.50, BT_{ij} = 5-9 \\ 0.75, BT_{ij} = 10-14 \\ 1.0, BT_{ij} \geq 15 \end{Bmatrix} \qquad (6)$$

The PSCF maps for the first and second PMF factors (coal combustion emissions and particle-phase wood combustion emissions, Figures 2A & 2B) show that the strongest potential source regions of these particulate PAH-driven factors were primarily in Asia and northern Europe. It is important to note that the PSCF methodology identifies potential source regions and preferred pathways of emission contributions to the receptor site. Thus potential source regions of land-based anthropogenic emissions identified over oceans and uninhabited land do not indicate a source in these places but represent the preferred pathways of these emissions to the sampling site. The third factor (gas-phase wood combustion emissions) shows a major source region off the coast of southern California, USA (Figure 2C) with lesser contributions from sites in Asia. A sequence of large contributions to Factor 3 was observed in the fall of 2008 (Figure 1F). Some of the wildfires that occurred in the fall of 2008 were significant and well documented, e.g., the November 13–25, 2008 California fire period when ~170 km^2 burned (California Department of Forestry and Fire Protection 2009a; California Department of Forestry and Fire Protection 2009b). The overlay in Figure 3 for the PSCF map for FAC_3 (Figure 2C) with satellite images (MODIS; National Aeronautics and Space Administration 2010) of wildfires over British Columbia and western United States in the fall of 2008 reveals the potential source regions identified by the PSCF computations are realistic. Also, MODIS images of wildfire events in Asia in the fall of 2008 may help explain the source regions observed in north-eastern Asia in the 1° x 1° PSCF map (see example in Figure S5). Note that the source regions observed with the 1° x 1° grid cells are closer to the regions of the Asian wildfire event shown in Figure S5 but are unobservable in the 5° x 5° PSCF maps. This higher resolution observed with 1° x 1° grid cells is likely an effect of the smaller number of endpoints (n_{ij}) they contain compared to 5° x 5° grids. Transport of gas- and particle-phase PAH from Asia across the Pacific Ocean as forest fire emissions is consistent with transport observed to the north-western US (Genualdi et al. 2009). Trans-Pacific transport of pollutant species (including PAH) in episodic events has been reported to take 5-12 days (Bailey et al. 2000; Genualdi et al. 2009; Harner et al. 2005; Jaffe et al. 1999; Noël et al. 2009; Primbs et al. 2008; Wilkening, Barrie and Engle 2000; Holzer, McKendry and Jaffe 2003). Moreover, FAC_1 and FAC_3 (Figures 1B and 1F, respectively) have contributions that are very temperature dependent (see Figure S6), likely reflective of the seasonal usage of coal and wood as fuels for residential heating in Asia and the lesser degree of atmospheric transformation of PAH during winter months due to reduced sunlight in Polar regions.

The fourth factor, identified as evaporative/uncombusted fuels, is the most ubiquitous source type (Figure 2D). This factor is dominated by gas-phase PAH exclusively and shows prominent source regions in proximity to the sampling site, most likely reflective of petroleum use around the city of Whitehorse. An overlay of the PSCF plot for FAC_4 and a geo-referenced image of oil and gas production platforms in the Arctic (United Nations Environment Programme and GRID-Arendal 2009) is shown in Figure 4. This figure shows many oil and gas production regions overlap with the potential source regions identified by the PSCF computations for this factor.

PAH in the gas-phase can readily undergo atmospheric transformation reactions in the troposphere initiated primarily by hydroxyl radicals photolytically-generated from ozone (Brubaker and Hites 1998; Atkinson and Arey 1994). The rate of atmospheric transformation of gas-phase PAH is determined by the concentration of hydroxyl radicals, the intensity of sunlight and the nature of the PAH. The lifetimes of PAH in the atmosphere generally vary from a few hours to days depending on the reactivity of the PAH (Brubaker and Hites 1998; Atkinson and Arey 1994). The open oceans are the most important sinks for ozone (Bloss et al. 2005; Read et al. 2008) and thus the generation of the OH radical. The estimated diurnal levels of hydroxyl radicals over several sections of the Pacific Ocean have been reported (Cooper 1996) and are higher than recent averages reported for the entire northern hemisphere (Bloss et al. 2005). This suggests that atmospheric transformation of gas-phase PAH over the open waters of the Pacific Ocean may be important. Indeed, PAH are traditionally not considered as strong candidates for long-range transport. Killin et al. however argued and showed in their study that PAH can indeed undergo trans-Pacific transport (Killin et al. 2004). Trans-Pacific (and other types of long-range) PAH transport can occur during overcast periods, since atmospheric transformations of gas-phase PAH require sunlight.

Also, during transport, PAH may partition more into the particle-phase with a decrease in temperature. The reactivity of a PAH associated with particulate material is much less than same PAH in the gas-phase (see sorbed PAH lifetimes in (Behymer and Hites 1985); the darker the particulate and the higher the carbon content, the slower the PAH degradation (Behymer and Hites 1988).

PSCF maps for some individual gas- and particle- phase PAH selected for their source apportionment importance are shown in Figure S7. It appears that all PAH whether in the gas- or particle-phase have common source regions notably in Asia and northern Europe. Source regions for fluorene are strongest in Asia, (Figure S5A) particularly over China. The intensity of gas-phase PSCF over this region decreased from fluorene to pyrene.

In an un-weighted PSCF study of gas-phase pollutants over the Great Lakes region, the atmospheric lifetimes of PAH from fluorene to pyrene were found to decrease resulting in spatial inconsistencies in source regions even though all PAH were believed to originate from the same source (Hafner and Hites 2003). In other words, differences in atmospheric transformations of individual PAH may interfere with the proper identification of their potential source regions. Thus, it may be more meaningful to study entire source profiles rather than individual PAH. In this study, the use of factors from PMF analysis was found to give much better results than using individual PAH prior to PSCF computations.

The PSCF map for gas-phase retene (Figure S7B) shows its strongest potential source region off the coast of the southern part of California, USA presumably due to prevailing winds during the wildfire events which transported this PAH to the LFL sampling site.

The PSCF maps of two particle-associated PAH, chrysene+triphenylene and benzo[*ghi*]perylene (Figures S7C and S7D) are almost identical indicating that even after a significant period of long-range transport, the source regions of particle-phase compounds are more consistent due to their relative stabilization in particulate matter.

6. Summary and Conclusions

We have successfully applied the PMF-PSCF methodology to PAH data set collected at a sparsely populated, sub-Arctic site in Yukon Canada. The PMF analyses afforded four PAH source factors (coal combustion emissions, particle-phase wood combustion emissions, gas-phase wood combustion emissions, and unburned petroleum/petrogenic emissions). Application of the PSCF method using these source factors allowed us to identify major PAH source regions in northern North America (Canada & USA) and Asia. Two PMF factors (coal combustion and particle-phase wood combustion) were identified as being more associated with long-range atmospheric transport, having weak or no North American sources. Gas-phase wood combustion emissions were however found to have their strongest sources in North America while the petroleum emissions factor was ubiquitous, having both local and hemispherical source regions. The coal combustion and gas-phase wood combustion emissions factors were mainly observed during winter months, consistent with the time of year when there is maximal transport of anthropogenically-derived aerosols to the Arctic. These source profiles may also reflect seasonal usage of coal and wood as fuels for residential heating in these regions.

In conclusion, PMF-PSCF methodology is a useful receptor modeling hybrid that has general applicability to PAH data sets from studies in which the identification of potential source regions of PAH factors extracted by factor analysis is critical.

Acknowledgement

Funding for this project from the Government of Canada Program for International Polar Year (IPY) is hereby acknowledged.

Appendix A. Supplementary Data

PMF diagnostics and more PSCF maps can be found in Supplementary Information_PMF_PSCF_AftRev.doc.

References

Ara Begum, B., Kim, E., Jeong, C., Lee, D., Hopke, P.K., 2005. Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. Atmospheric Environment 39, 3719-3724.

Ashbaugh, L.L., Malm, W.C., Sadeh, W.Z., 1985. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. Atmospheric Environment (1967) 19, 1263-1270.

Atkinson, R. and Arey, J., 1994. Atmospheric chemistry of gas-phase polycyclic aromatic hydrocarbons: Formation of atmospheric mutagens. Environmental health perspectives 102 Suppl 4, 117-126.

Bailey, R., Barrie, L.A., Halsall, C.J., Fellin, P., Muir, D.C.G., 2000. Atmospheric organochlorine pesticides in the western Canadian Arctic: Evidence of transpacific transport. J. Geophys. Res. 105, 11805-11811.

Behymer, T.D. and Hites, R.A., 1988. Photolysis of polycyclic aromatic hydrocarbons absorbed on fly ash. Environmental Science & Technology 22, 1311-1319.

Behymer, T.D. and Hites, R.A., 1985. Photolysis of polycyclic aromatic hydrocarbons adsorbed on simulated atmospheric particulates. Environmental Science & Technology 19, 1004-1006.

Benner, B.A., Wise, S.A., Currie, L.A., Klouda, G.A., Klinedinst, D.B., Zweidinger, R.B., Stevens, R.K., Lewis, C.W., 1995. Distinguishing the contributions of residential wood combustion and mobile source emissions using relative concentrations of dimethylphenanthrene isomers. Environmental Science & Technology 29, 2382-2389.

Bloss, W.J., Evans, M.J., Lee, J.D., Sommariva, R., Heard, D.E., Pilling, M.J., 2005. The oxidative capacity of the troposphere: Coupling of field measurements of OH and a global chemistry transport model. Faraday discussions 130, 425-436.

Brubaker, W.W. and Hites, R.A., 1998. OH reaction kinetics of polycyclic aromatic hydrocarbons and polychlorinated dibenzo-p-dioxins and dibenzofurans. The Journal of Physical Chemistry A 102, 915-921.

Bzdusek, P.A., Christensen, E.R., Li, A., Zou, Q., 2004. Source apportionment of sediment PAHs in Lake Calumet, Chicago: Application of factor analysis with nonnegative constraints. Environmental Science & Technology 38, 97-103.

California Department of Forestry and Fire Protection, 2009a. 2008 fire summary. Accessed 04/09 2010. http://www.fire.ca.gov/downloads/redbooks/2008/02-wildland-statistic-all-agencies/11-2008-FIRE-SUMMARY.pdf

California Department of Forestry and Fire Protection, 2009b. CAL FIRE 2008 wildland fire summary. Accessed 04/09 2010. http://www.fire.ca.gov/communications/downloads/fact_sheets/2008Summary.pdf

Chen, L.-.A., Watson, J.G., Chow, J.C., Magliano, K.L., 2007. Quantifying PM2.5 source contributions for the san joaquin valley with multivariate receptor models. Environmental Science & Technology 41, 2818-2826.

Cheng, M.D., Hopke, P.K., Barrie, L., Rippe, A., Olson, M., Landsberger, S., 1993. Qualitative determination of source regions of aerosol in Canadian High Arctic. Environmental Science & Technology 27, 2063-2071.

Cheng, M. and Lin, C., 2001. Receptor modeling for smoke of 1998 biomass burning in Central America. J. Geophys. Res. 106, 22871-22886.

Cooper, D.J., 1996. Estimation of hydroxyl radical concentrations in the marine atmospheric boundary layer using a reactive atmospheric tracer. Journal of Atmospheric Chemistry 25, 97-113.

D'Amours, R. and Page, P., 2001. Atmospheric Transport Models for Environmental Emergencies.

Du, S. and Rodenburg, L.A., 2007. Source identification of atmospheric PCBs in Philadelphia/Camden using positive matrix factorization followed by the potential source contribution function. Atmospheric Environment 41, 8596-8608.

Environment Canada, 2009. Intercontinental atmospheric transport of anthropogenic pollutants to the Arctic: Http://www.ec.gc.ca.libaccess.lib.mcmaster.ca/api-ipy/default.asp?lang=En&n=100FA705-1. 2010.

Gao, N., Gildemeister, A.E., Krumhansl, K., Lafferty, K., Hopke, P.K., Kim, E., Poirot, R.L., 2006. Sources of fine particulate species in ambient air over Lake Champlain Basin, VT. Journal of the Air & Waste Management Association (1995) 56, 1607-1620.

Genualdi, S.A., Killin, R.K., Woods, J., Wilson, G., Schmedding, D., Simonich, S.L.M., 2009. Trans-pacific and regional atmospheric transport of polycyclic aromatic hydrocarbons and pesticides in biomass burning emissions to western North America. Environmental Science & Technology 43, 1061-1066.

Hafner, W.D. and Hites, R.A., 2003. Potential sources of pesticides, PCBs, and PAHs to the atmosphere of the great Lakes. Environmental Science & Technology 37, 3764-3773.

Halsall, C.J., Sweetman, A.J., Barrie, L.A., Jones, K.C., 2001. Modelling the behaviour of PAHs during atmospheric transport from the UK to the Arctic. Atmospheric Environment 35, 255-267.

Halsall, C.J., Barrie, L.A., Fellin, P., Muir, D.C.G., Billeck, B.N., Lockhart, L., Rovinsky, F.Y., Kononov, E.Y., Pastukhov, B., 1997. Spatial and temporal variation of polycyclic aromatic hydrocarbons in the Arctic atmosphere. Environmental Science & Technology 31, 3593-3599.

Harner, T., Shoeib, M., Kozma, M., Gobas, F.A.P.C., Li, S.M., 2005. Hexachlorocyclohexanes and endosulfans in urban, rural, and high altitude air samples in the Fraser Valley, British Columbia: Evidence for trans-pacific transport. Environmental Science & Technology 39, 724-731.

Heo, J.-., Hopke, P.K., Yi, S.-., 2009. Source apportionment of PM2.5 in Seoul, Korea. Atmospheric Chemistry and Physics 9, 4957-4971.

Holzer, M., McKendry, I.G., Jaffe, D.A., 2003. Springtime trans-pacific atmospheric transport from East Asia: A transit-time probability density function approach. J. Geophys. Res. 108, 4708.

Hopke, P., 2001. A Guide to Positive Matrix Factorization. http://www.epa.gov/ttn/amtic/files/ambient/pm25/workshop/laymen.pdf, USA, 16.

Hopke, P., 1985. Receptor Modeling in Environmental Chemistry. Wiley-Interscience, NY, Hopke, P.K., Zhou, L., Poirot, R.L., 2005. Reconciling trajectory ensemble receptor model results with emissions. Environmental Science & Technology 39, 7980-7983.

Hsu, Y., Holsen, T.M., Hopke, P.K., 2003. Locating and quantifying PCB sources in chicago: Receptor modeling and field sampling. Environmental Science & Technology 37, 681-690.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans and International Agency for Research on Cancer, 2010. Some Non-Heterocyclic Polycyclic Aromatic Hydrocarbons and some Related Exposures. World Health Organization, International Agency for Research on Cancer ; Geneva, Switzerland : Distributed by WHO Press, Lyon, France; Geneva, Switzerland, 853.

Jaffe, D., Anderson, T., Covert, D., Kotchenruther, R., Trost, B., Danielson, J., Simpson, W., Berntsen, T., Karlsdottir, S., BLake, D., Harris, J., Carmichael, G., Uno, I., 1999. Transport of Asian air pollution to North America. Geophysical Research Letters 26, 711-714.

Kavouras, I.G., Koutrakis, P., Tsapakis, M., Lagoudaki, E., Stephanou, E.G., Von Baer, D., Oyola, P., 2001. Source apportionment of urban particulate aliphatic and polynuclear aromatic hydrocarbons (PAHs) using multivariate methods. Environmental Science & Technology 35, 2288-94.

Killin, R.K., Simonich, S.L., Jaffe, D.A., DeForest, C.L., Wilson, G.R., 2004. Transpacific and regional atmospheric transport of anthropogenic semivolatile organic compounds to Cheeka Peak Observatory during the spring of 2002. J. Geophys. Res. 109, D23S15.

Kim, E., Hopke, P.K., Kenski, D.M., Koerber, M., 2005. Sources of fine particles in a rural midwestern U.S. area. Environmental Science & Technology 39, 4953-4960.

Kim, E., Hopke, P., Larson, T., Maykut, N., Lewtas, J., 2004. Factor analysis of Seattle fine particles. Aerosol Science & Technology 38, 724-738.

Larsen, R.K.,3rd and Baker, J.E., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. Environmental Science & Technology 37, 1873-81.

Lee, J.H., Gigliotti, C.L., Offenberg, J.H., Eisenreich, S.J., Turpin, B.J., 2004. Sources of polycyclic aromatic hydrocarbons to the Hudson River airshed. Atmospheric Environment 38, 5971-5981.

Legzdins, A.E., McCarry, B.E., Marvin, C.H., Bryant, D.W., 1995. Methodology for bioassay-directed fractionation studies of air particulate material and other complex environmental matrices. International Journal of Environmental Analytical Chemistry 60, 79.

Liu, W., Hopke, P.K., Han, Y., Yi, S., Holsen, T.M., Cybart, S., Kozlowski, K., Milligan, M., 2003. Application of receptor modeling to atmospheric constituents at Potsdam and Stockton, NY. Atmospheric Environment 37, 4997-5007.

Metcalfe, C.D., Balch, G.C., Cairns, V.W., Fitzsimons, J.D., Dunn, B.P., 1990. Carcinogenic and genotoxic activity of extracts from contaminated sediments in western Lake Ontario. Science of the total environment 94, 125-41.

National Aeronautics and Space Administration, 2010. Moderate resolution imaging spectroradiometer (MODIS) rapid response system. Accessed 04/09 2010. http://rapidfire.sci.gsfc.nasa.gov/subsets/

Noël, M., Dangerfield, N., Hourston, R.A.S., Belzer, W., Shaw, P., Yunker, M.B., Ross, P.S., 2009. Do trans-pacific air masses deliver PBDEs to coastal British Columbia, Canada? Environmental Pollution 157, 3404-3412.

Paatero, P. and Hopke, P.K., 2003. Discarding or downweighting high-noise variables in factor analytic models. Analytica Chimica Acta 490, 277-89.

Paatero, P., 1999. The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. Journal of Computational and Graphical Statistics 8, 854-88.

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemom Intell Lab Syst 18, 183-94; 183.

Paatero, P., Hopke, P.K., Begum, B.A., Biswas, S.K., 2005. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. Atmospheric Environment 39, 193-201.

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and Intelligent Laboratory Systems 18, 183-194.

Paterson, K.G., Sagady, J.L., Hooper, D.L., Bertman, S.B., Carroll, M.A., Shepson, P.B., 1999. Analysis of air quality data using positive matrix factorization. Environ Sci Technol 33, 635-41.

Pekney, N.J., Davidson, C.I., Zhou, L., Hopke, P.K., 2006. Application of PSCF and CPF to PMF-modeled sources of PM2.5 in Pittsburgh. Aerosol Science and Technology 40, 952.

Poirot, R.L., Wishinski, P.R., Hopke, P.K., Polissar, A.V., 2001. Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. Environmental Science & Technology 35, 4622-4636.

Polissar, A.V., Hopke, P.K., Paatero, P., Kaufmann, Y.J., Hall, D.K., Bodhaine, B.A., Dutton, E.G., Harris, J.M., 1999. The aerosol at Barrow, Alaska: Long-term trends and source locations. Atmospheric Environment 33, 2441-2458.

Polissar, A.V., Hopke, P.K., Paatero, P., Malm, W.C., Sisler, J.F., 1998. Atmospheric aerosol over Alaska 2. elemental composition and sources. Journal of Geophysical Research 103, 19,045-57.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001. Atmospheric aerosol over Vermont: chemical composition and sources. Environmental Science & Technology 35, 4604-4621.

Poster, D.L., Benner, B.A., Schantz, M.M., Sander, L.C., Wise, S.A., Vangel, M.G., 2003. Determination of methyl-substituted polycyclic aromatic hydrocarbons in diesel particulate-related standard reference materials. Polycyclic Aromatic Compounds 23, 113.

Primbs, T., Piekarz, A., Wilson, G., Schmedding, D., Higginbotham, C., Field, J., Simonich, S.M., 2008. Influence of Asian and western United States urban areas and fires on the atmospheric transport of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, and fluorotelomer alcohols in the western United States. Environmental Science & Technology 42, 6385-6391.

Ramdahl, T., 1983. Retene[mdash]a molecular marker of wood combustion in ambient air. Nature 306, 580-582.

Read, K.A., Mahajan, A.S., Carpenter, L.J., Evans, M.J., Faria, B.V.E., Heard, D.E., Hopkins, J.R., Lee, J.D., Moller, S.J., Lewis, A.C., Mendes, L., McQuaid, J.B., Oetjen, H., Saiz-Lopez, A., Pilling, M.J., Plane, J.M.C., 2008. Extensive halogen-mediated ozone destruction over the tropical Atlantic Ocean. Nature 453, 1232-1235.

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

United Nations Environment Programme and GRID-Arendal, 2009. Vital Arctic Graphics. Vital Graphics, 1-43.

United States Environmental Protection Agency, 2008. EPA Positive Matrix Factorization (PMF) 3.0: Fundamentals & User Guide. 3.0, 1-81.

Wang, R., Tao, S., Wang, B., Yang, Y., Lang, C., Zhang, Y., Hu, J., Ma, J., Hung, H., 2010. Sources and pathways of polycyclic aromatic hydrocarbons transported to Alert, the Canadian High Arctic. Environmental Science & Technology 44, 1017-1022.

Wilkening, K.E., Barrie, L.A., Engle, M., 2000. Trans-Pacific Air Pollution. Science 290, 65.

Wise, S.A. and Watters, R.L., 2006. Certificate of Analysis. Standard Reference Material 1597a. Complex Mixture of Polycyclic Aromatic Hydrocarbons from Coal Tar. National Institute of Standards & Technology, USA, 8.

Xie, Y., Hopke, P.K., Paatero, P., Barrie, L.A., Li, S., 1999a. Identification of source nature and seasonal variations of Arctic aerosol by positive matrix factorization. Journal of the Atmospheric Sciences 56, 249-60.

Xie, Y., Hopke, P.K., Paatero, P., Barrie, L.A., Li, S., 1999b. Identification of source nature and seasonal variations of Arctic aerosol by the multilinear engine. Atmospheric Environment 33, 2549-2562.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette, D., Sylvestre, S., 2002. PAHs in the Fraser River Basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry 33, 489-515.

Zeng, Y. and Hopke, P.K., 1989. A study of the sources of acid precipitation in Ontario, Canada. Atmospheric Environment (1967) 23, 1499-1509.

Zhou, L., Hopke, P.K., Zhao, W., 2009. Source apportionment of airborne particulate matter for the speciation trends network site in Cleveland, OH. Journal of the Air & Waste Management Association (1995) 59, 321-331.

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

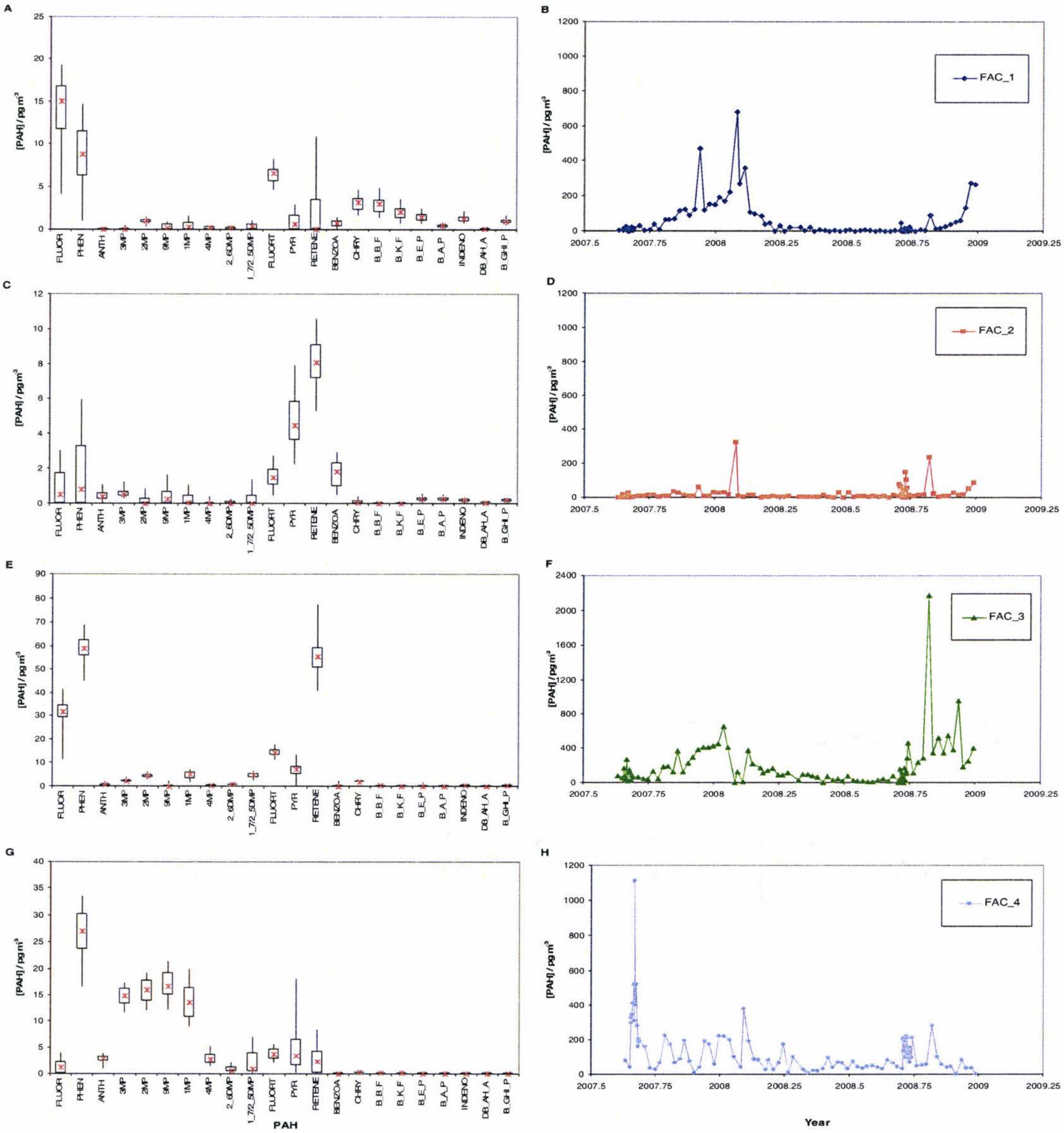


Figure 1: PMF generated factor profiles (left column) and their contributions to the TPAH$_{22}$ (right column). FAC_1: Coal combustion emissions factor profile (1A) and contributions (1B). FAC_2: Particle-phase wood combustion emissions factor profile (1C) and contributions (1D). FAC_3: Gas-phase wood combustion emissions factor profile (1E) and contributions (1F). FAC_4: Evaporative/uncombusted petroleum emissions factor profile (1G) and contributions (1H). Factor profiles have boxes and whiskers for the inter-quartile range and the 5th and 95th percentiles respectively.

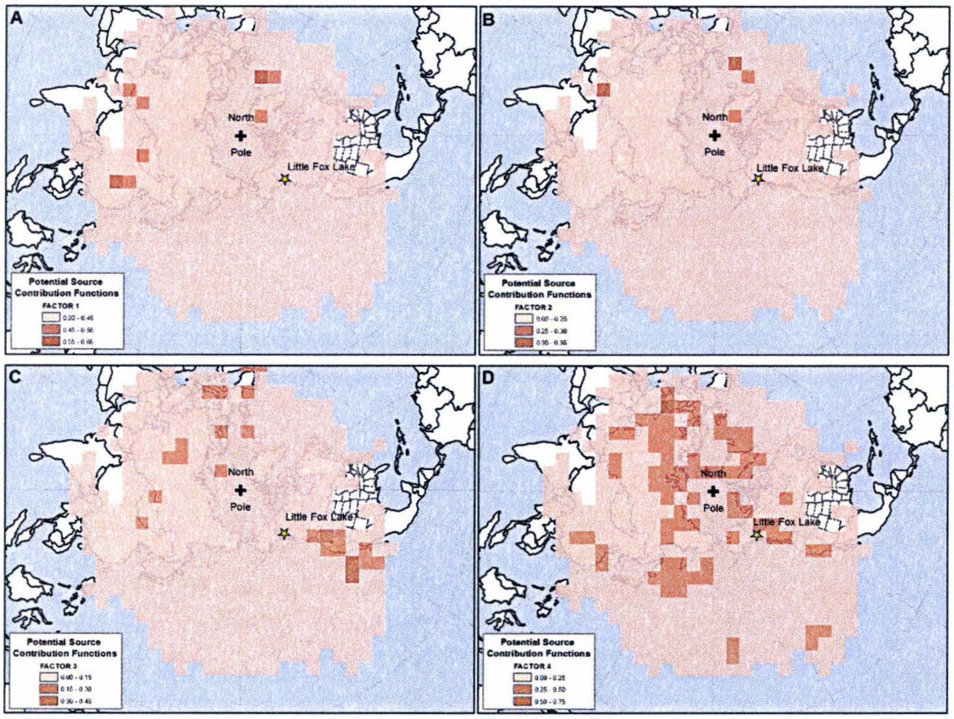


Figure 2: PSCF maps of PMF generated factors. A: FACTOR 1 (FAC_1; Coal combustion factor profile); B: FACTOR 2 (FAC_2; Particle-phase wood combustion emissions factor); C: FACTOR 3 (FAC_3; Gas-phase wood combustion emissions factor); D: FACTOR 4 (FAC_4; Evaporative/uncombusted petroleum emissions factor).

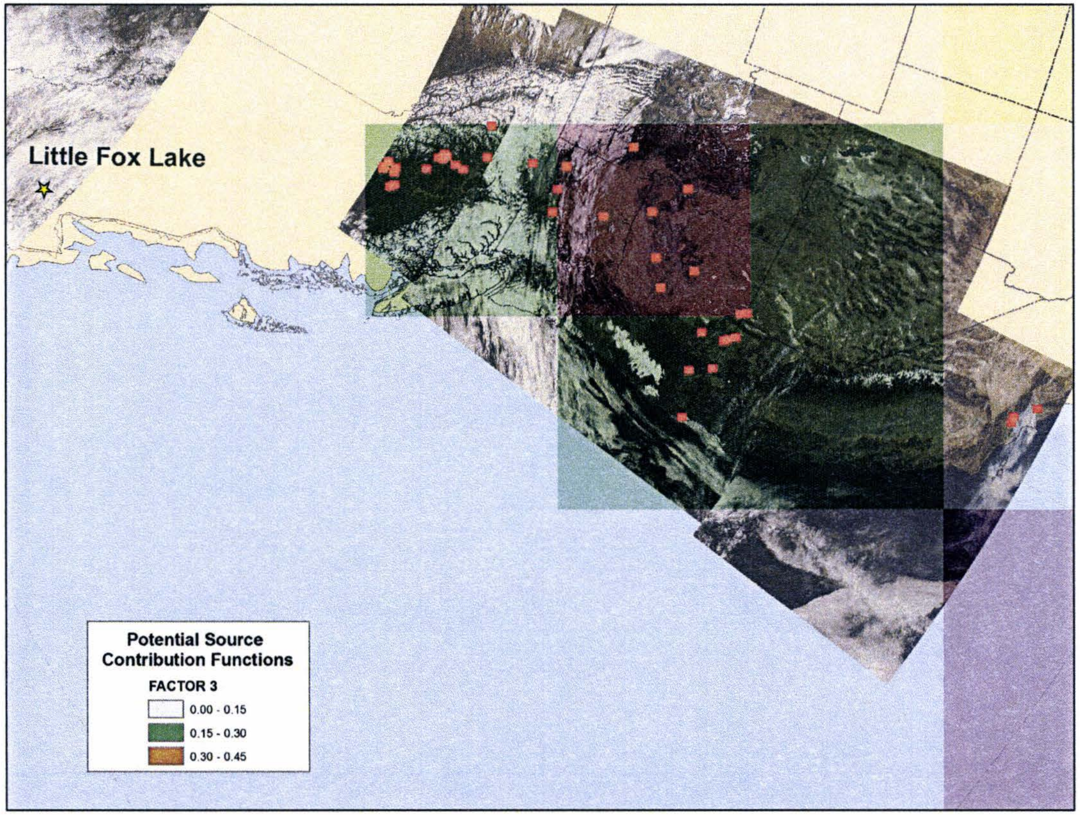


Figure 3: Overlay of 5° x 5° PSCF map for FAC_3 with satellite images (National Aeronautics and Space Administration 2010) of maximum wildfire detects over British Columbia, Canada and western United States in the fall of 2008.

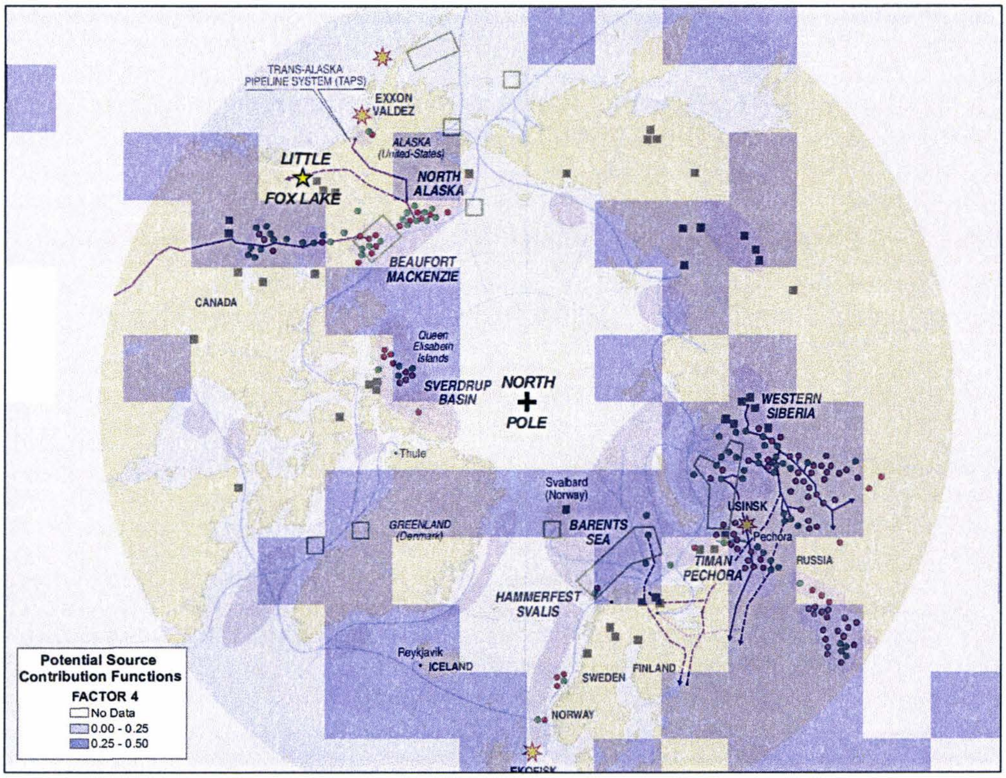


Figure 4: Overlay of 5° x 5° PSCF plot for FAC_4 and a geo-referenced image of oil and gas production platforms in the Arctic (United Nations Environment Programme and GRID-Arendal 2009).

Chapter Four

Evaluation of PAH Diagnostic Ratios as Source Apportionment Tools for Air Particulates Collected in an Urban-industrial Environment

The work in this chapter has been published in an article under the title: Uwayemi M. Sofowote[a], Laurie M. Allan[b], and Brian E. McCarry[a], 2010, Journal of Environmental Monitoring 12, 417-424 and is reproduced here with permission.

Author Contributions:

Uwayemi Sofowote: Performed data mining and interpretation, some integration of GC/MS chromatograms for alkylated phenanthrenes and wrote the first draft of the manuscript.

Dr. Laurie Allan: Performed original collection, extraction and cleanup of samples and quantification of most PAH.

Dr. Brian McCarry: Supervised key aspects of the data interpretation. Edited and wrote many portions of the final manuscript.

Supplementary Information:

Supplementary tables and figures for this manuscript can be found in the attached compact disc.

Author Affiliations:

[a]Department of Chemistry and Chemical Biology
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[b]Bruker Daltonics
555 Steeles Avenue, East Milton, Ontario, L9T 1Y6, Canada

Corresponding author email address: mccarry@mcmaster.ca; telephone: +1 (905) 525-9140

Evaluation of PAH Diagnostic Ratios as Source Apportionment Tools for Air Particulates Collected in an Urban-industrial Environment

Uwayemi M. Sofowote, Laurie M. Allan, and Brian E. McCarry

Abstract

A variety of polycyclic aromatic hydrocarbon (PAH) diagnostic ratios were examined as source apportionment tools in the analysis of a PAH data set associated with atmospheric particulate matter collected in an urban-industrial environment. Seventy-six PM_{10} samples were collected concurrently at 4 sampling sites over a one-month period in Hamilton, Ontario, Canada, a city of 500,000 people that is home to two integrated steel companies, associated industries and a network of roadways and major highways. Samples collected under well defined meteorological conditions were categorized as being 'upwind' or 'downwind' of the industrial sector. All sample extracts were analyzed for 48 parent PAH, methylphenanthrenes and sulphur-containing aromatics and showed a thousand-fold range of total PAH concentrations (0.23-172 ng m^{-3}). Of all PAH diagnostic ratios examined, the two most useful were the anthracene/(anthracene+phenanthrene) and benz[*a*]anthracene/(benz[*a*]anthracene+chrysene/triphenylene) ratios. These afforded the best discrimination of samples that had significant industrial impacts. This work is the first example of the use of a linear combination of PAH ratios, coupled with total PAH data and well defined local samples to determine the relative impacts of mobile and industrial emissions in an urban-industrial environment. Use of a linear combination of PAH ratios allowed us to categorize 95% of the data as 'upwind' or 'downwind' of the industrial sector. It is important to determine PAH ratio threshold values based on data from well defined local samples rather than relying on literature values alone.

Introduction

Many studies have shown significant correlations between industrial air pollution associated with particulate-bound pollutants, increased mortality rates and increased hospitalization rates for respiratory ailments and cardiovascular issues (Jerrett et al. 2005; Pope III et al. 2004; Pope III et al. 2002). There is still debate over whether these health impacts are caused by particulate matter alone, the chemical substances adsorbed to particulate matter or both (Harrison and Yin 2000). Polycyclic aromatic hydrocarbons (PAH), a class of planar fused-ring hydrocarbons that are products of combustion of a range of fuels inclusing gasoline, diesel fuel, coal, wood, etc. are released by industrial, residential, power generation, incineration, and mobile emissions sources (Grimmer, Naujack and Schneider 1982).

PAH exist in the environment as complex mixtures containing many structural isomers, alkylated derivatives and heterocyclic derivatives containing nitrogen, oxygen and sulfur. The latter are often referred to as polycyclic aromatic compounds (PAC). Higher mass PAH are found primarily adsorbed to particulate material (Serth and Hughes 1980). Several PAH, primarily with molecular masses of 228 Da and above, have been classified as carcinogens (Morris et al. 1995, Legzdins et al. 1995). Differentiation between PAH sources in complex environmental samples has been approached qualitatively using PAH ratios (Gogou et al. 1996) and quantitatively using a variety of multi-variate methods (Sofowote, McCarry and Marvin 2008).

PAH are not unique to a given pollution source, so apportionments based solely on 'marker' PAH are unreliable. Traditionally, PAH source apportionment studies have sought to distinguish between petrogenic sources and pyrogernic sources using molecular fingerprints of PAH isomers within the same molecular mass class, resulting in the development of several PAH ratios as source tracers (Yunker et al. 2002; Laflamme and Hites 1978; Budzinski et al. 1997). Commonly used PAH ratios include: a 178 mass PAH ratio, anthracene/(anthracene+phenanthrene) ratio (or An/Σ178); a methyl 178 mass PAH ratio, the phenanthrene/phenanthrene + sum of methylphenanthrenes ratio (or phen/phen+ΣMePhen); a 202 mass PAH ratio, fluoranthene/(fluoranthene+pyrene) or (Fl/Fl+Py); a 228 mass PAH ratio, benz[*a*]anthracene/(benz[*a*]anthracene+chrysene/triphenylene) or (BaA/Σ228); a 276 mass PAH ratio, indeno[1,2,3-*cd*]pyrene/(indeno[1,2,3-*cd*]pyrene+benzo[*ghi*]perylene) ratio or (IP/IP+B*ghi*P); in addition, a dimethylphenanthrene ratio, the 1,7/1,7+2,6-dimethylphenanthrene ratio or (1,7/1,7+2,6-DMP) has been used. Each PAH ratio has associated threshold values, proposed to aid in differentiating between source types. The PAH ratio thresholds proposed by Yunker et al. (Yunker et al. 2002) and Brandli et al. (Brandli et al. 2008) are widely used while the thresholds for the ΣMePhen/Phen ratio (reported herein as the phen/phen+ΣMePhen) were proposed by Gogou et al. (Gogou et al. 1996) and Poster et al. (Poster et al. 2003). Many studies have reported using diagnostic PAH ratios for source apportionment purposes, either alone (Gogou et al. 1996; Yunker et al. 2002; Laflamme and Hites 1978; Budzinski et al. 1997) or in conjunction with other source apportionment strategies (Brandli et al. 2008; Fang et al. 2006; Agarwal, Khillare and Shridhar 2006). For example, methylphenanthrene indices have been used to determine the maturity of crude oils (Radke, Welte and Willsch 1982; Stojanovic et al. 2001).

More recently, the use of a PAH ratio called CombPAH/TotalPAH (where CombPAH is the sum of fluoranthene, pyrene, benz[*a*]anthracene, chrysene, benzofluroanthenes, benzopyrenes, indeno[1,2,3-*cd*]pyrene and benzo[*ghi*]perylene) has been reported to be useful in differentiating between combustion sources and petrogenic sources (Gogou et al. 1996; Wang et al. 2007). PAH Ratios are not limited to traditional homocyclic PAH alone. Bence et al. reported the use of combination ratios of alkylated dibenzothiophenes and alkylated phenanthrenes to differentiate between Exxon Valdez crude oil and other crude oils (Bence, Kvenvolden and Kennicutt II 1996) while McCarry et al. used selected thia-arenes to differentiate between coal-derived emissions and diesel emissions (McCarry et al. 1996). Hegazi et al. have used ratios of dibenzothiophenes extensively in tracing sources of tar balls off the coast of Alexandria, Egypt (Hegazi et al. 2004). Recently, Bi et al. have reported hydroxylated PAH as markers for coal combustion in China (Bi et al. 2008; Simoneit et al. 2007). The number of PAH ratios used varies; Brandli et al. (Brandli et al. 2008) used a number of PAH ratios in their study of soils form Switzerland while Motelay-Massei et al. argued that the use of more than two ratios can lead to ambiguities in interpretation and a distortion of the source apportionment capabilities of diagnostic ratios (Motelay-Massei et al. 2007). PAH ratios have recently been combined successfully with carbon isotope measurements in a source apportionment study of harbor sediments (Yan et al. 2006).

A primary limitation of the PAH ratio approach is that it is qualitative, not quantitative. In the environment PAH may undergo degradations at widely varying rates (Galarneau 2008); the resultant 'weathering' of PAH may lead to alteration of PAH ratios, causing difficulties in source identitification and reduction in the reliability of the apportionment method. For example, anthracene, benz[*a*]anthracene and benzo[*a*]pyrene have significantly shorter lifetimes than their isomeric PAH when exposed to ozone Pitts Jr. et al. 1986). Weathered PAH source profiles have been reported by Stout et al. (Stout, Uhler and Boehm 2001). The factors that affect weathering of PAH bound to particulates were identified by Behymer and Hites (Behymer and Hites 1988) who showed that the darker particle colours and the higher carbon contents (traits typical of urban air particulate), the less the degradation of more reactive PAH; in these cases, PAH ratios were relatively unchanged. This viewpoint is consistent with the report by Gogou et al. who showed that PAH ratios were relatively stable in urban air particulate samples (Gogou et al. 1996). Thus, the source information preserved within PAH ratios need not be altered significantly when sampling air particulate in urban environments; however, PAH associated with particles arriving via long-range transport may have significantly altered PAH profiles due to atmospheric transformation reactions.

The City of Hamilton, with a population of over 500 000 (Statistics Canada 2006), is one of the most heavily industrialized cities in Canada. The city, which borders on Lake Ontario and sits on the south side of a large natural Harbour, is impacted by emissions from industrial sources amd mobile emissions from three major highways and a roadway network (Figure S1). As such Hamilton is a model urban-industrial centre, not unlike many cities in North America and Europe. This paper describes the application of PAH ratio methods for source discrimination while the accompanying paper (Sofowote et al. 2010) describes the application of two factor analytic (FA) methods to the same data set.

Experimental

Sample Collection and Sampling Sites

A map showing the locations of the air particulate sampling sites is provided (Figure S1, Supporting Information). Station 29000 is located just east of the downtown area of the city while Stations 29113, 29547 and 29531 are located to the south, east and west of the industrial area of the city. Respirable airborne particulate was collected using an Anderson PM_{10} modified high volume (Hi-Vol) air sampler (General Metal Works Ltd., Village of Cleves, OH) equipped with a flow controller operating at a flow rate of 1.13 m^3 min^{-1}. Flow controllers were calibrated before the sampling period and every two weeks during the sampling period. Air particulate samples were collected every day between July 20 and August 20, 1995 at four sites (Stations 29000 (Kelly), 29113 (Gertrude), 29547 (Pier 25), and 29531 (J.I. Case); Figure S1). Flow rates were calibrated using a rotometer once a week. Samples were collected from 8:00 a.m. to 6:00 a.m. the following morning on Teflon-coated glass fibre filters (Pallflex 20 cm x 25 cm, type TX40H120WW, Pall Corp., Putnam, Connecticut). Filters were pre-extracted in dichloromethane in a Soxhlet apparatus for 24 hours, dried and stored in a dessicator for five days prior to weighing on an analytical balance. After sample collection filters were stored in a desiccator for five days before weighing.

Meteorological Data

Wind direction and wind speed data was measured at the Hamilton Sewage Treatment Plant (Station 29026, Figure S1) and supplied by the Ontario Ministry of Environment (Allan and McCarry 1999). Twenty-two hour averages and standard deviations of the wind directions were calculated using Oriana software version 1.1 (Kovach Computing Services, Anglesey, Wales).

Table 1: TPAH and PAH diagnostic ratios (means ± standard deviation) in Hamilton air particulate samples collected at four sampling sites during study.

LOCATION	TPAH / ng m^{-3}	1,7/1,7+2,6-DMP[a]	Phen/ (Phen+ ΣMePhen)[b]	IP/ (IP+BghiP)[c]	BaA/∑228[d]	Fl/Fl+Py[e]	An/∑178[f]
GERTRUDE	10.8 ± 12.3	0.42 ± 0.19	0.59 ± 0.11	0.54 ± 0.09	0.37 ± 0.05	0.53 ± 0.02	0.16 ± 0.06
JICASE	37.1 ± 43.9	0.44 ± 0.15	0.57 ± 0.11	0.52 ± 0.13	0.39 ± 0.06	0.51 ± 0.03	0.18 ± 0.06
KELLY	8.7 ± 7.7	0.45 ± 0.16	0.61 ± 0.07	0.33 ± 0.06	0.39 ± 0.07	0.52 ± 0.02	0.15 ± 0.06
P25P10	14.5 ± 16.9	0.49 ± 0.07	0.53 ± 0.14	0.55 ± 0.08	0.35 ± 0.08	0.52 ± 0.02	0.17 ± 0.08

[a] 1,7/1,7+2,6-dimethylphenanthrene. [b] phenanthrene/phenanthrene+∑methylphenanthrenes. [c] indeno[1,2,3-*cd*]pyrene/(indeno[1,2,3-*cd*]pyrene + benzo[*ghi*]perylene). [d] benz[*a*]anthracene/(benz[*a*]anthracene + chrysene/triphenylene). [e] fluoranthene/(fluoranthene+pyrene). [f] anthracene/(anthracene + phenanthrene

Chemicals

PAH standards were purchased from Aldrich Chemical Company Inc. (Milwaukee, WI). Deuterated PAH standards were purchased from Cambridge Isotope Labs Ltd. (Woburn, MA). High purity helium carrier gas (>99.999%) was purchased from VitalAire (Hamilton, Ontario). HPLC grade solvents were purchased from Caledon Laboratories (Georgetown, Ontario). A Milli-Q purification system (Waters Associates, Milford, Massachusetts) was used to further purify distilled water.

Sample Extraction and Clean-up

A total of 76 respirable air particulate filters were selected, extracted and analyzed for this work. A Soxhlet extraction procedure adapted from Schuetzle et al. (Schuetzle, Jensen and Ball 1985) and McCalla et al. (McCalla et al. 1988) was used to extract filters individually using dichloromethane as the extracting solvent. Each extraction apparatus was spiked with 500 μL of a solution containing phenanthrene-d_{10} (732 ng mL^{-1}), chrysene-d_{12} (852 ng mL^{-1}) and dibenz[*a,h*]anthracene-d_{14} (820 ng mL^{-1}) in dichloromethane as surrogate standards. The extraction vessels were wrapped with aluminum foil and heating adjusted to obtain a rate of six cycles per hour with a total extraction time of 18-24 hours.

The extract was cleaned up using an open-column alumina method adapted from Later et al. (Later et al. 1981) Briefly, neutral alumina (2 g, Fisher, Brockman activity I, 80-200 mesh, activated for at least 48 hours at 170°C) was packed into a column (1 cm diameter) in hexane. The dichloromethane extract from one air particulate filter was added to 1 g of alumina, the solvent evaporated and the sample-adsorbed alumina placed atop the 2 g of alumina. Elution with a sequence of solvents afforded three fractions: hexane (20 mL), dichloromethane (25 mL) then methanol (20 mL). The dichloromethane eluent which contained the PAC fraction was further purified using Sephadex LH20 chromatography described previously (Allan and McCarry 1999; Marvin et al. 1993) to afford an aliphtic-free, 'aromatic' fraction. The eluate was reduced in volume, transferred to a GC vial with an insert, diluted with internal standards (70 μL of a toluene solution containing pyrene-d_{10} (5.00 μg mL^{-1}) and perylene-d_{12} (4.01 μg mL^{-1}) and the volume adjusted to approximately 70 μL using a stream of dry nitrogen gas.

GC/MS Analyses

GC/MS analyses were performed on a Hewlett-Packard Model 5890 Series II gas chromatograph equipped with a DB-17ht capillary column (50% phenyl methylpolysiloxane, 30 m x 0.25 mm i.d. x 0.15 μm film, J&W Scientific, Folson, CA) and an on-column injector coupled to a Hewlett-Packard Model 5971A mass selective detector operated in the selected ion monitoring mode (SIM; for operationing details, see Table S1). A 1 μL injection corresponded to about 20 m^3 of air. Samples with high PAH levels were diluted 10-fold prior to analysis. A calibration standard containing 45 polycyclic aromatic compounds and 5 deuterated PAH was used at six concentrations; R^2 values exceeded 0.998. Recoveries and standard deviations for deuterated PAH were: phenanthrene-d_{10}, 76±15%; chrysene-d_{12}, 89±2.9%; dibenz[*a,h*]anthracene-d_{14}, 90±4.4%. Detection limits (in pg m^{-3} of air sampled) ranged from 0.1 pg m^{-3} for phenanthrene to 0.5 pg m^{-3} for benz[*a*]anthracene-7,12-dione.

Results and Discussion

Total PAH values, defined as the sum of 48 PAH peaks, displayed a 1000-fold range for this 76-sample data set (0.23-172 ng m^{-3}, Table S2); this supplementary data table also lists all PAH ratio data. Table 1 shows the means and standard deviations of the total PAH data and six diagnostic PAH ratios at the four sampling sites; these PAH ratios have been used to differentiate between combustion sources, including petroleum, grass, wood and coal combustion (Yunker et al. 2002). The data in Table 1 indicates that there is a wide range of total PAH values at each site with a 4-fold range of mean values. The arithmetic means of the PAH ratios showed little variability between the sites with few exceptions. On an aggregate level, these data provided little insight into source apportionment contributions. We examined wind direction and wind speed data during the sampling period and found that winds blew primarily from two opposite directions; thus, a given sampling site would be 'upwind' or 'downwind' of the industrial area depending on the wind directions that day. The differentiation between 'upwind' and 'downwind' samples is critical to source apportionment because 'upwind' samples should have inputs from mobile sources while 'downwind' samples should have inputs from mobile and industrial sources. To obtain the best examples of the 'upwind' and 'downwind' samples, we identified five of the 22 days during which winds blew consistently from single directions with narrow standard deviations and wind speeds exceeding 0.8 m s^{-1} throughout the sampling period (Table 2). The sampling sites on these days were classified as being either upwind (UW) or downwind (DW) of the industrial area (Table 2, Figure S1). This categorization is consistent with the industrial sector being a major source of PAH; the total PAH values for the 'upwind' samples ranged from 0.73-2.9 ng m^{-3} (mean, 1.5; standard deviation, 0.7 ng m^{-3}) while the 'downwind' samples ranged from 6.6-172 ng m^{-3} (mean, 47; standard deviation, 55 ng m^{-3}, Table 2). These 18 meteorologically well-characterized 'upwind' and 'downwind' samples were regarded as 'boundary conditions' for the entire data set. The remaining samples should lie between these boundaries and be classified as having predominantly 'upwind' (mobile source), or 'downwind' (industrial source) characteristics.

Four bi-variate plots (Figures 1A-1D) show the data for the samples and 3 NIST Standard Reference Materials (SRM). The data for three SRMs were added to provide useful markers in all plots because they represent signatures for three different source types (diesel exhaust particulate, SRM 1650; urban dust, SRM 1649; coal tar/combustion emissions, SRM 1597). The reference material SRM 1649 is an integrated sample of urban dust collected in a city impacted primarily by vehicular emissions with little to no industrial emissions. The PAH ratio ranges that have been reported to correspond to different source types are included along the axes of each figure. Of the many bivariate plots examined, the four in Figure 1 provided the best disrimations of the sample and SRM data; additional plots can be found in the supplementary material (Figure S2). Plots in Figure 1 were created using four diagnostic PAH ratios discussed below. The first two ratios discussed were found to be the most useful.

Table 2: Data from air particulate samples collected on days when wind directions had narrow distributions

Location-Date	Mean Wind Direction (Std. Dev., degrees)	Mean Wind Speed (m s^{-1})	Upwind/ Downwind Designation	PAH Ratio Value BaA/∑228	PAH Ratio Value An/∑178	178-228 Mean Ratio Value	Total PAH (ng m^{-3})
GERT-Jul 20	238 (7)	2.9	UW	0.30	0.17	0.24	1.16
JICASE-Jul 20	238 (7)	2.9	UW	0.38	0.15	0.26	2.94
KELLY-Jul 20	238 (7)	2.9	UW	0.35	0.12	0.24	1.26
P25P10-Jul 20	238 (7)	2.9	DW	0.45	0.29	0.37	37.7
JICASE-Aug 02	61(12)	2.3	DW	0.46	0.28	0.37	101
KELLY-Aug 02	61(12)	2.3	DW	0.46	0.22	0.34	13.5
P25P10-Aug 02	61(12)	2.3	UW	0.31	0.17	0.24	2.16
GERT-Aug 07	96 (29)	3.4	DW	0.42	0.15	0.29	6.55
JICASE-Aug 07	96 (29)	3.4	DW	0.45	0.24	0.35	172
KELLY-Aug 07	96 (29)	3.4	DW	0.45	0.21	0.33	22.0
P25P10-Aug 07	96 (29)	3.4	UW	0.28	0.00	0.14	0.73
GERT-Aug 11	226 (26)	1.8	UW	0.34	0.11	0.23	1.16
JICASE-Aug 11	226 (26)	1.8	UW	0.27	0.10	0.19	1.67
KELLY-Aug 11	226 (26)	1.8	UW	0.29	0.08	0.19	0.98
P25P10-Aug 11	226 (26)	1.8	DW	0.44	0.25	0.34	40.7
GERT-Aug 18	76 (40)	1.6	DW	0.40	0.26	0.33	16.6
JICASE-Aug 18	76 (40)	1.6	DW	0.42	0.22	0.32	64.4
P25P10-Aug 18	76 (40)	1.6	UW	0.31	0.19	0.25	1.97

[a] UW = Upwind; DW = Downwind. None of the above samples were collected at wind speed values <0.8 m s^{-1}

The Anthracene/∑178 PAH (An/∑178) Ratio

The anthracene/∑178 PAH ratio is reported to provide discrimination between combustion and petroleum sources (Yunker et al. 2002). In Figures 1A and 1B, most samples lie in the 'combustion' region, indicating that all four sampling sites are impacted primarily by combustion sources. The NIST combustion reference standards (SRMs 1650 and 1649) have near identical values and lie on the border between 'petroleum' and 'combustion' sources. Samples identified as having been collected on days when the site was directly downwind of the industrial area (DW in Table 2) were found clustered near the coal tar reference material (SRM 1597, Figure 1A) while upwind samples (UW in Table 2) were found clustered near the urban dust/mobile emissions SRM (SRM 1649). The strong correlations between these SRMs and the 'upwind/downwind' samples indicated that this ratio functions as a useful source discriminator.

The Benz[*a*]anthracene/∑228 PAH (BaA/∑228) Ratio

The benz[*a*]anthracene/∑228 PAH (BaA/∑228) ratio is reported to provide discrimination between petroleum sources, combustion sources and mixed sources. Applying reported thresholds (Yunker et al. 2002), this data set had a 1:2 split between mixed sources and combustion sources (Figures 1B-1D) with no samples in the petroleum source range. The three SRMs were spread across the range of values in the data set, making this ratio potentially useful for source apportionment. While the plots against the 1,7/1,7+2,6-DMP and the IP/IP+BghiP ratios (Figures 1C and 1D) were useful, the plot of BaA/∑228 against the An/∑178 ratio plot (Figure 1B) provided the highest statistical correlation of all bi-variate plots ($R^2 = 0.52$). Samples classified as uniquely 'downwind' of industry sources were clustered in the upper right region of this figure while the unique 'upwind' samples were found in the lower left region. The remaining samples were spread between these extremes.

The IP/IP+BghiP Ratio

The IP/IP+BghiP ratio, reported to assist in the differentiation between petroleum, petroleum combustion and grass-wood-coal combustion sources, showed about two-thirds of the data set classified as grass-wood-coal combustion sources (Figure 1D). Oddly, the IP/IP+BghiP ratio showed no differentiation between the three SRMs while the BaA/∑228 ratio (x-axis) afforded a much greater degree of discrimination. The IP/IP+BghiP ratio was not considered to be useful in terms of source discrimination in this data set.

The 1,7/1,7+2,6-Dimethylphenanthrene (DMP) Ratio

The 1,7/1,7+2,6-dimethylphenanthrene (1,7/1,7+2,6-DMP) ratio has been reported to be useful in differentiating between vehicular emissions and wood combustion sources (Yunker et al. 2002). This ratio was plotted on the x-axis against the anthracene/∑178 PAH (An/∑178) ratio (Figure 1A) and the benz[*a*]anthracene/∑228 PAH (BaA/∑228) ratio (Figure 1C). Over 80% of the data had 1,7/1,7+2,6-DMP ratio values in the 'mixed sources' range (0.45-0.7). Table 3 shows the comparisons of 1,7/1,7+2,6-DMP ratio values for the three SRMs computed from the work of Benner et al. (Yunker et al. 2002; Benner, Gordon and Wise 1989), from the NIST certificates of analysis (Wise and Watters 2006a; Wise and Watters 2006b) and from values determined in this work. This ratio provided relatively little source discrimination.

Table 3: Comparisons of 1,7/1,7+2,6-Dimethylphenathrene ratios obtained for three NIST Standard Reference Materials

	Benner et al.[a]	NIST[b]	Current Work
SRM 1649	0.45	-	0.52
SRM 1650	0.37	0.38	0.40
SRM 1597	-	0.57	0.58

[a, b] The 1,7/1,7+2,6-DMP ratios were calculated from data in Benner et al. (Yunker et al. 2002; Benner, Gordon and Wise 1989) or NIST SRM certificates (Wise and Watters 2006a; Wise and Watters 2006b).

Overall, the BaA/∑228 and An/∑178 ratios provided the greatest source discrimination of all ratios examined. Additional bi-variate plots can be found in the supplementary material.

Diagnostic Ratios and Wind Directions

With a view to developing generic 'upwind' and 'downwind' PAH ratio criteria for the entire data set, we examined the BaA/∑228 and the An/∑178 ratio values for the 18 meteorologically classified 'upwind' and 'downwind' samples and compared these values to threshold values proposed by Yunker et al. to distinguish between petroleum and combustion sources (see stacked sidebars in Figures 2A and 2B). The proposed threshold values compiled by Yunker et al. did not align with the ratios of the 'upwind' and 'downwind' samples, likely due to the variable degrees of degradation of An and BaA in the samples examined by Yunker and others. Based on our analysis of these 18 samples, we proposed new threshold values for the two PAH ratios to discrimate between 'upwind' and 'downwind' samples. These criteria are listed in Table 4.

A clear differentiation between the 'upwind' and 'downwind' samples is apparent; 'upwind' samples had much lower An/∑178 and BaA/∑228 ratio values than 'downwind' samples. Next, the entire data set was classified into the four categories in Table 4 with the following results: (1) 22 samples met both 'downwind' criteria with An/∑178 ratios >0.21 and BaA/∑228 ratios >0.40 and were designated 'DW2'or downwind by 2 criteria; (2) 13 samples met the BaA/∑228 'downwind' criterion but not the An/∑178 criterion and were designated 'DW1;' (3) 32 samples met both 'upwind' criteria, i.e., BaA/∑228 ratio <0.38 and An/∑178 ratio ≤0.19, and were designated 'UW2;' (4) the 9 remaining samples were designated non-classified (NC). In total, 54 of the 76 samples (70%) satisfied either the upwind or the downwind criteria of both PAH ratios and were designated either DW2 or UW2. These threshold ratio values are new observations that are based on the approach taken by Yunker. The ratios used for the downwind samples have values that reflect a 'fresh' industrial source signature with little atmospheric transformation; if this approach is applied to other data sets, particularly where samples are collected at greater distances from sources, the thresholds values used here may not be applicable owing to losses of An and BaA due varying degrees of atmospheric degradation.

Since the BaA/∑228 and An/∑178 ratios afforded similar interpretations and their bi-variate plot had a reasonably good correlation ($R^2 = 0.52$, Figure 1B), a linear combination of these ratios called the 178-228 MeanRatio (defined as 0.5 BaA/∑228 + 0.5An/∑178) was calculated for use as a single parameter embodying the information content within the BaA/∑228 vs. An/∑178 bi-variate plot. The 178-228 MeanRatio values were calculated for the 18 meteoroloically defined samples and afforded more robust 'upwind' and 'downwind' assignments than either of the individual PAH ratios (see Figure 2C compared to Figures 2A and 2B). Samples with 178-228 MeanRatio values ≤0.265 were designated as 'upwind' of industry while those with 178-228 MeanRatio values ≥0.28, as 'downwind' of industry. The 178-228 MeanRatio values for the entire data set were plotted against total PAH (log scale, Figure 3). Data points were organized in the four categories described in Table 4. This plot showed an increase in 178-228 MeanRatio value with increasing total PAH and a log-linear correlation with an $R^2 = 0.59$. This correlation was better than the corresponding values of either PAH ratio ($R^2 = 0.49$ and 0.51, Figure S5).

The rectangles in Figure 3 designated as 'upwind of the industrial area' and 'downwind of the industrial area' were drawn based on the 18 meteorologically defined samples as discussed above. Only 4 of the 76 samples (5%) had 178-228 MeanRatio values between 0.265 and 0.28 and thus were not categorized by this metric. The rectangles were drawn deliberately with no overlap along the total PAH axis; the cross-over point was about 7 ng m^{-3}. Surprisingly, 64 of the 76 data points (84%) were contained within the two rectangles and identified using the DW2, DW1, UW2 or NC designations above. Of the remaining 12 data points the 'upwind' rectangle failed to capture 4 samples with total PAH values less than 7 ng m^{-3} while the 'downwind' zone failed to capture 8 samples with total PAH values greater than 7 ng m^{-3}. By extending the 'downwind' rectangle from 7 down to 3 ng m^{-3} or the 'upwind' rectangle from 7 up to 20 ng m^{-3}, all data points except those 4 data points which do not meet the 178-228 MeanRatio criteria would be included within the rectangles.

The robust performance of PAH ratios containing labile PAH (e.g., anthracene, benz[*a*]anthracene) is noteworthy. These particulate samples are black in colour and rich in carbon; adsorption to these dark particles reduces the rate of chemical transformation (Behymer and Hites 1988). Secondly, the industrial emissions were collected very soon after their release due to the proximity of the sampling sites, leaving little time for any significant transformation of labile PAH. For this reason, application of this approach to samples collected a some distance from industrial emissions may not see the same degree of correlation. On the other hand, for air monitoring applications proximate to industrial sources such as abatement studies, this approach should prove valuable.

Overall, the 'upwind' and 'downwind' designations as defined by the 178-228 MeanRatio criteria resulted in the classification of 95% of the ambient air samples. The accompanying paper (Sofowote et al. 2010) examines the use of factor analysis methods to provide detailed source apportionment assignments for all samples discussed in this paper; the source apportionment outcomes provide new insights into the reasons why some data points were outliers in the current analysis. For example, all five 'UW2' data points in Figure 3 with total PAH values $\geq$10 ng m^{-3} were shown to have significant contributions from diesel exhaust sources; hence, while these samples met the criteria of having an 'upwind' profile they were collected at sites where it was possible that diesel engine exhaust (e.g., from idling trucks) may have impacted the samplers.

Table 4: PAH ratio criteria for classification of samples as being 'upwind' or 'downwind' of industrial sources or non-classified (NC).

BaA/$\sum$228	An/$\sum$178	Classification	Number of Samples	TPAH / ng m^{-3} Mean±Std.Dev.
>0.40	>0.21	DW2	22	44±35
>0.40	$\leq$0.20	DW1	13	25±33
<0.38	$\leq$0.19	UW2	32	3.8±5.2
0.38 – 0.39	0.19 – 0.21	NC	9	9.4±5.3

Conclusions

A suite of diagnostic PAH ratios have been used to examine a data set of 76 samples collected in an urban-industrial environment with the goal of evaluating their source apportionment performance. PAH ratios are attractive tools because they can be calculated readily from available PAH data. This work showed that the threshold values compiled by Yunker et al. based on literature values reported for PAH sources and applied to sediments collected in the Fraser River delta in western Canada may not be applicable to all sources; thus, the application of reported threshold values should be done with great caution since, as noted previously by various authors (Yunker et al. 2002; Galarneau 2008), some PAH are subject to weathering in atmospheric, terrestrial and aquatic environments. The use of PAH ratios, particularly ratios involving readily transformed PAH, should be approached with care.

This work is the first example of use of a linear combination of PAH ratios, coupled with total PAH data and well defined local samples to determine the relative impacts of mobile and industrial emissions in an urban-industrial setting. Of all PAH ratios examined, the BaA/∑228 and An/∑178 ratios were found to be most useful. These two ratios when combined into a single ratio value (termed 178-228 MeanRatio) gave more robust discrimination of sources than either of the individual ratios. The well defined local samples were PAH data from 18 samples collected during nearly unidirectional wind conditions. The data defined PAH profiles that were typical of samples collected 'upwind' or 'downwind' of the industrial area. Application of the 178-228 MeanRatio criterion allowed classification of 95% of the samples as being collected at sites that were 'upwind' or 'downwind' of the industrial sector. When the 178-228 MeanRatio values were plotted against total PAH data, the resulting two-dimensional plot allowed the categorization of 85% of the data set as 'upwind' or 'downwind' using a single total PAH value as a cutoff (Figure 3). The approach outlined in this paper provides a new way of evaluating how to use diagnostic PAH ratio data for source apportionment in complex environmental settings.

Overall, PAH ratio approaches provide only crude and qualitative source apportionment assessments of environmental samples. More quantitative methods such as factor analysis approaches are needed to provide quantitative source apportionments (see accompanying paper (Sofowote et al. 2010).

Acknowledgements

We are grateful to the Ontario Ministry of the Environment for funding and sampling sites used in this study. We would also like to thank Ankit K. Rastogi for data mining and collation. Funding support from an NSERC Discovery Grant and the Great Lakes Sustainability Fund is acknowledged.

References

Statistics Canada, http://www12.Statcan.ca/english/census06/data/profiles/community/Details/Page.Cfm?Lang=E&Geo1=CSD&Code1=3525005&Geo2=PR&Code2=35&Data=Count&SearchText=Hamilton&SearchType=Begins&SearchPR=01&B1=All&Custom=. Statistics Canada, 2006, Canada,

Agarwal, T., Khillare, P.S., Shridhar, V., 2006. PAHs contamination in bank sediment of the Yamuna River, Delhi, India. Environmental Monitoring and Assessment 123, 151-166.

Allan, L.M. and McCarry, B.E., 1999. Thia-arenes as pollution source tracers in urban air particulate. Ph.D., McMaster University,

Behymer, T.D. and Hites, R.A., 1988. Photolysis of polycyclic aromatic hydrocarbons absorbed on fly ash. Environmental Science & Technology 22, 1311-1319.

Bence, A.E., Kvenvolden, K.A., Kennicutt II, M.C., 1996. Organic geochemistry applied to environmental assessments of Prince William Sound, Alaska, after the Exxon Valdez oil spill--a review. Organic Geochemistry 24, 7-42.

Benner, B.A., Gordon, G.E., Wise, S.A., 1989. Mobile sources of atmospheric polycyclic aromatic hydrocarbons: A roadway tunnel study. Environmental Science & Technology 23, 1269-1278.

Bi, X., Simoneit, B.R.T., Sheng, G., Fu, J., 2008. Characterization of molecular markers in smoke from residential coal combustion in China. Fuel 87, 112-119.

Brandli, R.C., Bucheli, T.D., Ammann, S., Desaules, A., Keller, A., Blum, F., Stahel, W.A., 2008. Critical evaluation of PAH source apportionment tools using data from the Swiss soil monitoring network. Journal of Environmental Monitoring : JEM 10, 1278-1286.

Budzinski, H., Jones, I., Bellocq, J., Piérard, C., Garrigues, P., 1997. Evaluation of sediment contamination by polycyclic aromatic hydrocarbons in the Tironde Estuary. Marine Chemistry 58, 85-97.

Fang, G.C., Wu, Y.S., Chang, C.N., Ho, T.T., 2006. A study of polycyclic aromatic hydrocarbons concentrations and source identifications by methods of diagnostic ratio and principal component analysis at Taichung Chemical Harbor near Taiwan Strait. Chemosphere 64, 1233-42.

Galarneau, E., 2008. Source specificity and atmospheric processing of airborne PAHs: Implications for source apportionment. Atmospheric Environment 42, 8139-8149.

Gogou, A.I., Stratigakis, N., Kanakidou, M., Stephanou, E.G., 1996. Organic aerosols in eastern Mediterranean: Components source reconciliation by using molecular markers and atmospheric back trajectories. Organic Geochemistry 25, 79-96.

Grimmer, G., Naujack, K., Schneider, D., 1982. Profile analysis of polycyclic aromatic hydrocarbons by glass capillary gas chromatography in atmospheric suspended particulate matter in the nanogram range collecting 10 m^3 of air. Fresenius' Journal of Analytical Chemistry 311, 475-484.

Harrison, R.M. and Yin, J., 2000. Particulate matter in the atmosphere: Which particle properties are important for its effects on health? The Science of The Total Environment 249, 85-101.

Hegazi, A.H., Andersson, J.T., Abu-Elgheit, M.A., El-Gayar, M., 2004. Source diagnostic and weathering indicators of tar balls utilizing acyclic, polycyclic and S-heterocyclic components. Chemosphere 55, 1053-65.

Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A.I.I.I., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J., 2005. Spatial analysis of air pollution and mortality in los angeles. Epidemiology 16, 727-736.

Laflamme, R.E. and Hites, R.A., 1978. The global distribution of polycyclic aromatic hydrocarbons in recent sediments. Geochimica et Cosmochimica Acta 42, 289-303.

Later, D.W., Lee, M.L., Bartle, K.D., Kong, R.C., Vassilaros, D.L., 1981. Chemical class separation and characterization of organic compounds in synthetic fuels. Analytical Chemistry 53, 1612-1620.

Legzdins, A.E., McCarry, B.E., Marvin, C.H., Bryant, D.W., 1995. Methodology for bioassay-directed fractionation studies of air particulate material and other complex environmental matrices. International Journal of Environmental Analytical Chemistry 60, 79.

Marvin, C.H., Allan, L., McCarry, B.E., Bryant, D.W., 1993. Chemico/biological investigation of contaminated sediment from the Hamilton Harbour area of western Lake Ontario. Environmental and molecular mutagenesis 22, 61-70.

McCalla, D.R., Quilliam, M.A., Kaiser-Farrell, C., Tashireo, C., Hoo, K., Gibson, E.S., Lockington, N.J., Kerr, A.A., Sheldrake, C. 1988.King, C.M., Romano, L.J., Schuetzle, D. (Eds.), Carcinogenic and Mutagenic Responses to Aromatic Amines and Nitroarenes, Elsevier Science Publishing Co., Inc., New York, 47-63.

McCarry, B.E., Allan, L.M., Legzdins, A.E., Lundrigan, J.A., Marvin, C.H., Bryant, D.W., 1996. Thia-arenes as pollution source tracers in urban air particulate. Polycyclic Aromatic Compounds 11, 75-82; 75.

Morris, W.A., Versteeg, J.K., Bryant, D.W., Legzdins, A.E., McCarry, B.E., Marvin, C.H., 1995. Preliminary comparisons between mutagenicity and magnetic susceptibility of respirable airborne particulate. Atmospheric Environment 29, 3441-3450.

Motelay-Massei, A., Ollivon, D., Garban, B., Tiphagne-Larcher, K., Zimmerlin, I., Chevreuil, M., 2007. PAHs in the bulk atmospheric deposition of the seine river basin: Source identification and apportionment by ratios, multivariate statistical techniques and scanning electron microscopy. Chemosphere 67, 312-321.

Pitts Jr., J.N., Paur, H., Zielinska, B., Arey, J., Winer, A.M., Ramdahl, T., Mejia, V., 1986. Factors influencing the reactivity of polycyclic aromatic hydrocarbons adsorbed on filters and ambient POM with ozone. Chemosphere 15, 675-685.

Pope III, C.A., Burnett, R.T., Thurston, G.D., Thun, M.J., Calle, E.E., Krewski, D., Godleski, J.J., 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. Circulation 109, 71-77.

Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132-1141.

Poster, D.L., Benner, B.A., Schantz, M.M., Sander, L.C., Wise, S.A., Vangel, M.G., 2003. Determination of methyl-substituted polycyclic aromatic hydrocarbons in diesel particulate-related standard reference materials. Polycyclic Aromatic Compounds 23, 113.

Radke, M., Welte, D.H., Willsch, H., 1982. Geochemical study on a well in the western Canada Basin: Relation of the aromatic distribution pattern to maturity of organic matter. Geochimica et Cosmochimica Acta 46, 1-10.

Schuetzle, D., Jensen, T.E., Ball, J.C., 1985. Polar polynuclear aromatic hydrocarbon derivatives in extracts of particulates: Biological characterization and techniques for chemical analysis. Environment International 11, 169-181.

Serth, R.W. and Hughes, T.W., 1980. Polycyclic organic matter (POM) and trace element contents of carbon black vent gas. Environmental Science & Technology 14, 298-301.

Simoneit, B.R., Bi, X., Oros, D.R., Medeiros, P.M., Sheng, G., Fu, J., 2007. Phenols and hydroxy-PAHs (arylphenols) as tracers for coal smoke particulate matter: Source tests and ambient aerosol assessments. Environmental Science & Technology 41, 7294-7302.

Sofowote, U.M., Allan, L.M., McCarry, B.E., 2010. A comparative study of two factor analytic models applied to PAH data from inhalable air particulate collected in an urban-industrial environment. Journal of Environmental Monitoring 12, 425-433

Sofowote, U.M., McCarry, B.E., Marvin, C.H., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. Environmental Science & Technology 42, 6007-6014.

Stojanovic, K., Jovancicevic, B., Pevneva, G.S., Golovko, J.A., Golovko, A.K., Pfendt, P., 2001. Maturity assessment of oils from the Sakhalin oil fields in Russia: Phenanthrene content as a tool. Organic Geochemistry 32, 721-731.

Stout, S.A., Uhler, A.D., Boehm, P.D., 2001. Recognition of and allocation among multiple sources of PAH in urban sediments. Environmental Claims Journal 13, 141.

Wang, Z., Chen, J., Qiao, X., Yang, P., Tian, F., Huang, L., 2007. Distribution and sources of polycyclic aromatic hydrocarbons from urban to rural soils: A case study in Dalian, China. Chemosphere 68, 965-71.

Wise, S.A. and Watters, R.L., 2006a. Certificate of Analysis. Standard Reference Material 1650b. Diesel Particulate Matter. National Institute of Standards & Technology, USA, 12.

Wise, S.A. and Watters, R.L., 2006b. Certificate of Analysis. Standard Reference Material 1597a. Complex Mixture of Polycyclic Aromatic Hydrocarbons from Coal Tar. National Institute of Standards & Technology, USA, 8.

Yan, B., Abrajano, T.A., Bopp, R.F., Benedict, L.A., Chaky, D.A., Perry, E., Song, J., Keane, D.P., 2006. Combined application of δ13C and molecular ratios in sediment cores for PAH source apportionment in the New York/New Jersey Harbor Complex. Organic Geochemistry 37, 674-687.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette, D., Sylvestre, S., 2002. PAHs in the Fraser River basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry 33, 489-515.

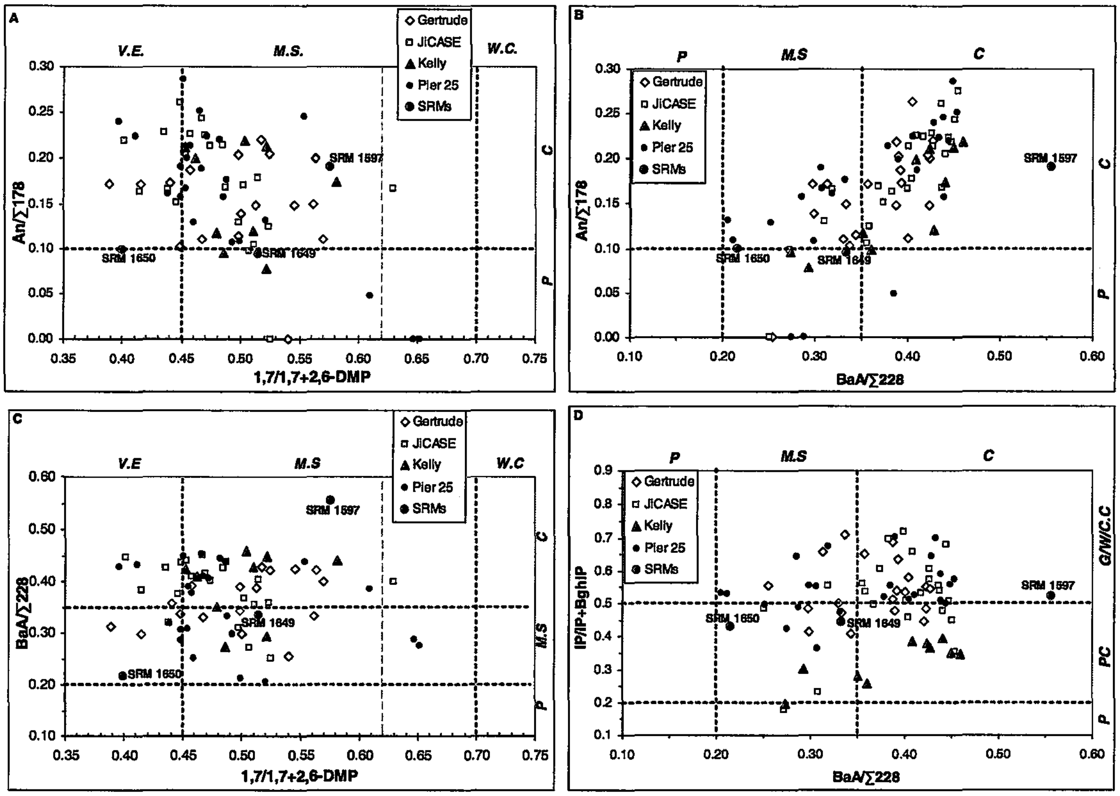


Figure 1: Bi-variate plots of PAH diagnostic ratios for Hamilton air samples; A: An/∑178 vs. 1,7/1,7+2,6-DMP; B: BaA/∑228 vs. 1,7/1,7+2, 6-DMP; C: An/∑178 vs. BaA/∑228; D: IP/IP+BghiP vs. BaA/∑228. Threshold values and source designations are those of Yunker et al. (Yunker et al. 2002): P = Petroleum; PC = petroleum combustion; C = combustion; WC = wood combustion; G/W/CC = grass; wood & coal combustion; MS = mixed sources; VE = vehicular emissions. Faint broken lines on 1,7/1,7+2,6-DMP axes of Figures 1A and 1C represent brown coal combustion.

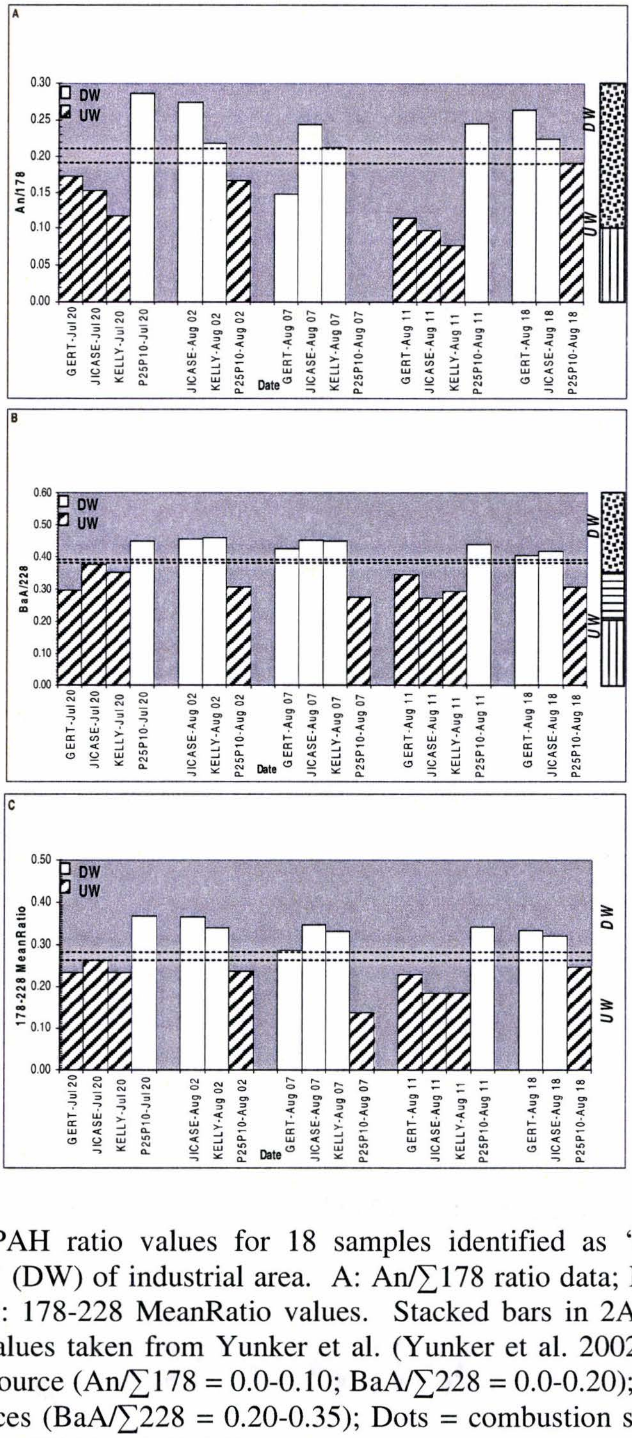


Figure 2: PAH ratio values for 18 samples identified as 'upwind' (UW) or 'downwind' (DW) of industrial area. A: An/∑178 ratio data; B: BaA/∑178 ratio data; and C: 178-228 MeanRatio values. Stacked bars in 2A and 2B represent threshold values taken from Yunker et al. (Yunker et al. 2002). Vertical lines = petroleum source (An/∑178 = 0.0-0.10; BaA/∑228 = 0.0-0.20); Horizontal lines = mixed sources (BaA/∑228 = 0.20-0.35); Dots = combustion source (An/∑178 = 0.10-1.0; BaA/∑228 = 0.35-1.0).

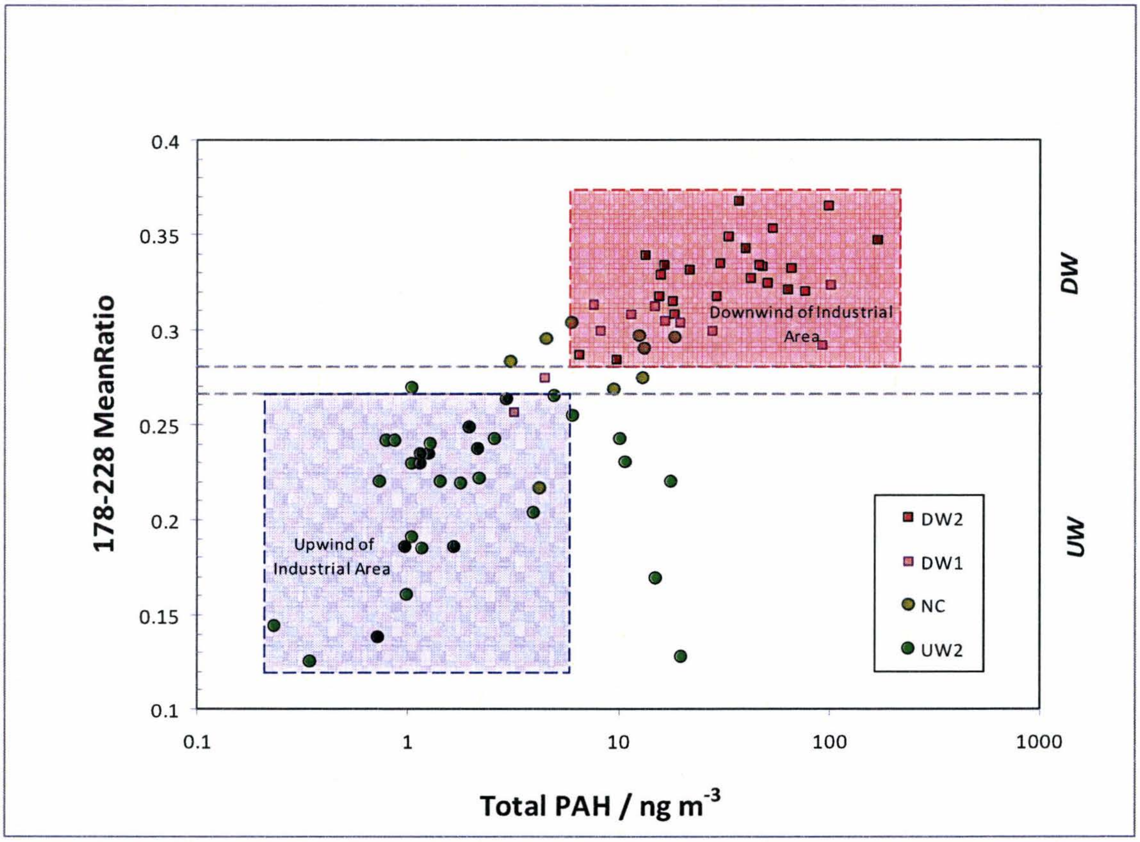


Figure 3: Plot of 178-228 MeanRatio values defined as 0.5(BaA/∑228 + An/∑178) as a function of total PAH (TPAH). Boxes encompass most of the samples classified 'upwind' and 'downwind' of the industrial area. DW2 = Downwind based on both PAH ratios; UW2 = Upwind based on both PAH ratios. DW1 = Downwind based only on BaA/∑228 ratio data; NC = non-classified; See text for ratio thresholds. Darkened points represent the 18 upwind and downwind samples as defined in Table 2.

Chapter Five

A Comparative Study of Two Factor Analytic Models Applied to PAH Data from Inhalable Air Particulate Collected in an Urban-industrial Environment

The work in this chapter has been published in an article under the title: Uwayemi M. Sofowote[a], Laurie M. Allan[b], and Brian E. McCarry[a], 2010, Journal of Environmental Monitoring 12, 425-433 and is reproduced here with permission.

Author Contributions:

Uwayemi Sofowote: Performed data mining, interpretation, statistical analyses, and wrote the first draft of the manuscript.

Dr. Laurie Allan: Performed original collection, extraction and cleanup of samples and quantification of most PAH.

Dr. Brian McCarry: Supervised many aspects of the data interpretation. Edited and wrote many portions of the final manuscript.

Supplementary Information:

Supplementary tables and figures for this manuscript can be found in the attached compact disc

Author Affiliations:

[a]Department of Chemistry and Chemical Biology
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[b]Bruker Daltonics
555 Steeles Avenue, East Milton, Ontario, L9T 1Y6, Canada

Corresponding author email address: mccarry@mcmaster.ca; telephone: +1 (905) 525-9140

A Comparative Study of Two Factor Analytic Models Applied to PAH Data from Inhalable Air Particulate Collected in an Urban-industrial Environment

Uwayemi M. Sofowote, Laurie M. Allan and Brian E. McCarry

Abstract

Two factor analysis-based receptor modeling methods (FA) were applied to a polycyclic aromatic hydrocarbon (PAH) data set from extracts of 75 PM_{10} air particulate samples collected concurrently at 4 sampling sites proximate to the urban-industrial area in Hamilton, Ontario, Canada. The total PAH concentrations of 48 target compounds ranged from 0.23 to 172 ng m^{-3}. Principal component analysis (PCA) and positive matrix factorization (PMF) analysis were followed by multilinear regression analyses to identify and quantify PAH source contributions, together with spatial and temporal trends. The correlations between predicted and observed total PAH levels were excellent in both models (R^2 >0.98). The PCA afforded large negative contributions in a number of samples, so further analysis was abandoned. The PMF analysis showed 3 factors which were identified as gasoline emissions, diesel emissions and coke oven emissions. Contributions of gasoline emissions and diesel emissions factors were surprisingly similar at all 4 sites indicative of a background of vehicle emissions across the city. The PMF coke oven emissions factor showed the greatest variability in total loadings, consistent with the large PAH emissions from the steel industries and the large influence of wind direction on PAH concentrations. The highest coke oven contributions were observed at sites closest to the industrial area on days when these sites were downwind of the industries. The PMF coke oven impacts factor showed good correlations with two commonly used PAH diagnostic ratios when the ratios were combined into a single ratio. This integrated approach allowed us to categorize >90% of the samples based on the wind direction of the impacting source.

Introduction

Source apportionment tools allow environmental scientists to determine the impacts of individual contaminant source types at sampling locations. Factor analyses (FAs) are a set of common receptor modeling techniques used for the apportionment of contaminants species and particulate loadings in sediments and aerosols. The value of FA models derives from their ability to delineate impacting sources of contaminants and their relative contributions at a receptor site without prior knowledge of the sources before analyses (Vega et al. 2000); the challenge is to identify the derived source profile with a source profile. Some receptor models including chemical mass balance approaches (Vega et al. 2000;Varadarajan and Kumar 2006) require prior knowledge of source contaminant profiles in order to undertake the apportionment exercise.

FA models typically employ eigen- or non-eigen value-based multivariate analytical methods to deduce the underlying factors responsible for variations observed in a data set. In this work, we compare two FA models as to their effectiveness in dealing with the source apportionment of a PAH data set from air particulate samples collected over 30 consecutive days at four sampling sites in Hamilton, Ontario, Canada. In an accompanying paper we examined this same data set using various diagnostic PAH ratios for source apportionment (Sofowote, Allan and McCarry 2010). The City of Hamilton, located at the western end of Lake Ontario with a population of over 500 000 (Statistics Canada 2006), is one of the most heavily industrialized cities in Canada. Two major integrated steel manufacturing companies with coking operations are situated in the northeast end of the city on the southern shore of a large natural Harbour. The industrial area is home to many smaller industries with the associated vehicular traffic and the city is bounded by major highways; the city is also impacted by the emissions from cars and trucks that are common to all major urban areas. Thus, Hamilton is a challenging, urban-industrial environment in which to compare these FA methods and the results of this work would be applicable to urban centres arounds the world.

Polycyclic aromatic hydrocarbons (PAH) exist in the environment as complex mixtures with a range of molecular mass classes, including many structural isomers, and alkylated derivatives; in addition, these mixtures contain structurally related species with oxygen, nitrogen and sulphur heteroatoms. PAH are widespread environmental contaminants produced by the combustion of fossil fuels, industrial sources, residential heating, power generation, incineration, and mobile and stationary engines.

In this paper we report the comparison of two factor analytical models, namely principal component analysis (PCA) and positive matrix factorization (PMF) used for the apportionment of PAH in PM_{10} air particulate material. The PMF results were compared to source predictions made using PAH diagnostic ratios (see accompanying article, Sofowote, Allan and McCarry 2010) and were used to explore trends observed in the data set.

Factor analytical models as source apportionment tools

Principal Component Analysis (PCA)

Factor analysis methods (FA) have been shown to be powerful tools in environmental assessment studies (Vega et al. 2000; Almeida et al. 2006). While several factor analytical models exist, FA has become synonymous with principal component analysis (PCA). PCA is perhaps the most common factor analysis model used for source apportionment (Almeida et al. 2006; Fang et al. 2006; Yunker et al. 1995). Concentrations of variables in individual samples can be regarded as linear combinations of a number of underlying factors (or sources). The task in PCA is to reproduce the correlation matrix as the sum of common factor coefficients, i.e., values in the loadings matrix (Hopke 1985).

For environmental applications, Varimax rotation is the preferred transformation so that loadings values could be clustered around 1 and 0, thus making them more physically interpretable (Hopke 1985; Abdi 2003). Kavouras et al. summarize the mathematical principles of PCA in a functional flow scheme (Kavouras et al. 2001). The commonest variation of the PCA method used in air particulate studies is the absolute principal component scores (APCS) approach developed by Thurston and Spengler (Thurston and Spengler 1985) in which a 'null' sample is introduced into the data set and then its scores are then subtracted out of the real samples. The APCS method facilitates the comparison of 'absolute' factor scores with measured concentrations of contaminants in a multilinear regression (MLR) analysis for mass apportionment. An example of a study employing this method is contained in the work of Kavouras et al. who used APCS for apportioning sources of aliphatic hydrocarbons and PAH in particulates in Chile (Kavouras et al. 2001). Other researchers (Almeida et al. 2006; Bruno et al. 2001; Guo, Wang and Louie 2004) have also employed this method for mass apportionment of particulate-bound contaminants.

It must be noted that factor analysis is not restricted to atmospheric monitoring. A common PCA variant which employs nonnegative constraints has been described in the literature (Rachdawong and Christensen 1997) and has been used for PAH apportionment in river sediments (Bzdusek et al. 2004) and soils (Wang et al. 2007). Christensen et al. have also applied this variant to sediment cores from the Ashtabula River in Ohio (Christensen and Bzdusek 2005). This variant was designed to eliminate negative and thus unrealistic PCA outputs in the loadings and/or scores (Rachdawong and Christensen 1997) matrices even after commonly applied transformations (e.g. Varimax rotations).

Positive Matrix Factorization

PMF is a receptor modeling tool developed in the early 1990s by Paatero and Tapper (Paatero and Tapper 1993) which has found use in many air quality/source apportionment studies (Anttila et al. 1995; Lee, Chan and Paatero 1999; Xie et al. 1999a; Kim, Hopke and Edgerton 2003; Lee et al. 2003; Larsen and Baker 2003), in water samples (Reinikainen et al. 2001; Du, Belton and Rodenburg 2008), more recently in soils (Vaccaro et al. 2007), and sediments (Bzdusek and Christensen 2006; Bzdusek, Lu and Christensen 2006; Sofowote, McCarry and Marvin 2008).

PMF was designed to overcome the shortcomings of PCA by treating factor analysis as a true least squares problem (Hopke 2001) i.e., for a data set denoted by X(n x m), X can be solved for as:

$$X = GF + E \quad (1)$$

then Q(E), an object function of the residual matrix E(n x m) which is to be minimized is defined as a constrained and weighted least squares function as follows:

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{e_{ij}}{s_{ij}} \right]^2 \quad (2)$$

where e_{ij} is the residual value defined as:

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \quad (3)$$

$\hat{x}_{ij}$ is the factor analytic modeled value of the measured x_{ij} value and s_{ij} is the uncertainty associated with the determination of x_{ij}. The values f_{kj} and g_{ik} are elements of the F and G matrices; the factor profile and contribution matrices respectively. The constraints arise because $F \geq 0$, $G \geq 0$, i.e., are not allowed to assume negative values. In a properly weighted analysis, the Q value should be approximately equal to the optimum Q defined as the dimensional product (n x m) of the data set (Eberly 2005).

Specifying the appropriate uncertainty estimates s_{ij} for the minimization of Q is an important step in PMF analysis. Guidelines exist and vary for the pre-assessment of data and their uncertainties before the PMF analysis depending on the PMF program used (Hopke 2001; Eberly 2005; Polissar et al. 1998). Xie et al. have treated missing data values differently from values below the detection limit by replacing missing data values with the geometric mean in the PMF analysis on Arctic aerosols using the multilinear engine (ME) program (Xie et al. 1999b). Polissar et al. have specified equations for dealing with these groups of data points when using the PMF2 program (Polissar et al. 1998). The ability to use built-in equations for computing the S matrix also is provided for common PMF software applications (Hopke 2001; Eberly 2005; Ke et al. 2008). The equations usually incorporate in some form uncertainties for concentration values below their method detection limits, thus reducing their weight in the factor analysis. The robust mode may be used in running the PMF analysis. The robust mode reduces the effect of very large variables by treating them as outliers so that they do not distort the model's fitting process (Hopke 2001). More details on the technical usage of PMF can be found elsewhere (Paatero and Tapper 1993). PMF offers the potential of more realistic source profiling and apportionment with better resolution. Comparative studies between PMF and chemical mass balance approaches for air particulates can be found in recent literature (Reinikainen et al. 2001; Ke et al. 2008; Pandolfi et al. 2008; Shrivastava et al. 2007; Miller et al. 2002). Fewer studies, notably by Huang et al. (Huang, Rahn and Arimoto 1999) and Qin et al. (Qin, Oduyemi and Chan 2002) have focused on the comparison of PMF and PCA for the apportionment of atmospheric particulates. Larsen et al. applied and compared three methods (PCA, UNMIX, and PMF) for the apportionment of PAH in an urban particulate sample set from Maryland, USA (Larsen and Baker 2003). Miller et al. have tested these three models and chemical mass balance on volatile organic contaminants and have indicated that PMF produced the most ideal apportionment (Miller et al. 2002; Anderson et al. 2002). We recently compared PCA and PMF in Hamilton Harbour sediments (Sofowote, McCarry and Marvin 2008) and found that PMF profiles were better resolved and more realistically interpretable (i.e., no negative values in factor profiles and very few in the factor contributions matrices). In this study, we focus on the use of these two models to delineate spatial and temporal trends of PAH sources in PM_{10} particulates affecting our receptor (sampling) sites. We compare summary statistics and use the PMF results to develop an approach on how PAH diagnostic ratios can be semi-quantitatively used in conjunction with factor analyses.

Table 1: Total PAH (TPAH) for particulate samples collected at four sampling sites during the period of study

TPAH^a (ng m^{-3})

Station Number (MOE)	Location Name	Arithmetic Mean	Standard Deviation	Maximum	Minimum
29113	GERTRUDE	10.8	12.3	47.2	0.73
29531	JICASE	37.1	43.9	172	0.34
29000	KELLY	8.7	7.7	20.0	0.88
29547	PIER25	14.5	16.9	54.9	0.23

[a] TPAH is defined as the sum of 51 polycyclic aromatic compounds contained in 48 chromatographic peaks

Experimental

Sample Collection and Sites

A map showing the locations of the air particulate sampling sites is provided (Figure S1, Supporting Information). Station 29000 (operated by the Ontario Ministry of the Environment) is located just east of the downtown area of the city while Stations 29547, 29531 and 29113 are located to the west, east and south of the city's industrial area, respectively. Respirable air particulate (PM_{10}) samples were collected for 22 hours every day (8 a.m. to 6 a.m.) between July 20th and August 20th, 1995. A total of 76 filters were selected, extracted and analyzed for this work. Details of the samplers and the sampling campaign can be found in the accompanying article (Sofowote, Allan and McCarry 2010).

Meteorological Data

Wind direction and wind speed data were obtained from a meteorological station located near the sampling area (at the Hamilton Sewage Treatment Plant, Station 29026, Figure S1) and supplied by of the Ontario Ministry of Environment. Standard deviations of the wind directions were calculated using Oriana software version 1.1 (Kovach Computing Services, Anglesey, Wales).
Chemicals

PAH standards were purchased from Aldrich Chemical Company Inc. (Milwaukee, WI) and Chiron (Trondheim, Norway). Thia-arene standards were a generous gift from Dr. M. L. Lee (Brigham Young University, Provo, UT). Deuterated PAH standards were purchased from Cambridge Isotope Labs Ltd. (Woburn, MA). High purity helium carrier gas (>99.999%) was purchased from Vitalaire (Hamilton, Ontario). HPLC grade solvents were purchased from Caledon Laboratories (Georgetown, Ontario). A Milli-Q purification system (Waters Associates, Milford, Massachusetts) was used to further purify distilled water.

Sample Extraction, Clean-up and GC/MS Analyses

These protocols have been described in greater detail in the accompanying work (Sofowote, Allan and McCarry 2010). A total of 48 parent and alkylated PAH (Table 2) were quantified in all samples by the use of internal standards. Recoveries and standard deviations for deuterated PAH were: phenanthrene-d_{10}, 75.7±15%; chrysene-d_{12}, 89.0±2.9%; dibenz[*a,h*]anthracene-d_{14}, 90.4±4.4%.

Method detection limits (in pg m^{-3} of air sampled) ranged from 0.1 pg m^{-3} for phenanthrene to 0.5 pg m^{-3} for benz(a)anthracene-7,12-dione.

Multivariate Analysis Software Programs

All PCA were carried out using SPSS version 15.0 for Windows (SPSS Inc., Chicago, IL). PMF analyses were carried out using the EPA PMF version 1.1 program http://www.epa.gov/heasd/products/pmf/pmf.htm. Multilinear regression (MLR) analyses were carried out on PC scores and the PMF G-matrix using the Microsoft Excel regression function (Microsoft Office Professional 2003).

Results and Discussion

Seventy-five of the 76 samples were selected for factor analyses. Forty-nine PAH peaks were routinely quantified. One analyte (4-methylphenanthrene) was removed from the data set because it was undetected in >80% of the samples. The sum of the 48 remaining peaks (all 45 peaks in Table 2 plus anthraquinone, benzanthrone and picene) was defined as total PAH (TPAH). Table 1 shows the measured TPAH values at the four sampling sites over the period of study. These concentrations were not blank corrected.

PCA of the Hamilton PAH Data Set

Principal component analysis was carried out on a data set containing 75 samples and 45 variables from all four sampling sites. The PAH concentrations were Varimax rotated with Kaiser-normalization. Three components (eigenvalues $\geq$ 1) were extracted and rotated, and accounted for >94% of the variance in the data set. The factor loadings matrix (Table 2) shows the loadings of variables in the 3 significant factors. Loading values $\geq$ 0.7 are highlighted in bold face. It should be noted that component loadings are not directly comparable to any known source profiles since they are correlation values generated from the eigenvalue analyses. The first rotated component is enriched in most PAH but not the dibenzothiophenes. The absence of alkylated dibenzothiophenes suggests that this component is not diesel-emissions related. The relatively high loadings of higher molecular mass PAH in rotated component 1 are suggestive of a pyrogenic source ((Stout, Uhler and Boehm 2001). The modest loadings of some alkylated phenanthrenes relative to the parent PAH in this factor is consistent with petrogenic sources even though these alkyl phenanthrenes may occur in coal and wood combustion sources (Yunker et al. 2002). In summary,the loading profile of PAH in the first rotated PCA component is consistent with a generic combustion source. The second rotated PCA component is exclusively loaded in dibenzothiophenes, phenanthrene and anthracene. This profile is suggestive of a petrogenic or a diesel-related source. Dibenzothiophenes have been used to trace crude oils (Bence, Kvenvolden and Kennicutt II 1996) and tar balls (Hegazi et al. 2004) and are important components of diesel emissions (McCarry et al. 1996). Thus, this PCA profile was assigned as a petrogenic source.

The third rotated PCA component is difficult to classify since it is solely enriched in 3-ethylphenanthrene and 3,9-dimethylphenanthrene. Indeed, the existence of a third profile may be in doubt since the loadings of these two alkylated phenanthrenes may be an artifact of the improper scaling that is inherent in PCA (Hopke 2001; Paatero and Tapper 1993).

Multiple linear regression analysis was performed on the PCA data for mass apportionment of the three factors to the total PAH burden in each particulate sample. A multiple linear regression (MLR) of elements in the factor scores matrix (t_k) against the standardized values of the TPAH was done using a spreadsheet to obtain the best correlation between observed and predicted TPAH for each sample. The three factor scores had coefficients (B_k) which were determined by the MLR analysis with a stipulated minimum 95% confidence level. The multiple linear regression equation for the standard deviate of the TPAH values was found to be:

$$\hat{z}_{TPAH} = 0.88t_1 + 0.42t_2 + 0.23t_3 \qquad (4)$$

Table 2: Rotated component matrix for the 3 significant factors extracted from PCA of Hamilton air particulates from four sampling sites. Values ≥ 0.7 are in boldface.

		Rotated Component		
Variable	**Variable Abbreviation**	**1**	**2**	**3**
phenanthrene	PHEN	0.613	**0.716**	0.224
anthracene	ANTH	0.618	**0.726**	0.150
1-methylphenanthrene	Me1PHEN	**0.713**	0.620	0.271
2- methylphenanthrene	Me2PHEN	0.673	**0.706**	0.194
3- methylphenanthrene	Me3PHEN	**0.688**	0.665	0.262
9- methylphenanthrene	Me9PHEN	**0.724**	0.622	0.261
3- ethylphenanthrene	Et3P	0.189	0.418	**0.861**
1,2-dimethylphenanthrene	DM12P	0.609	0.526	0.024
1,3-/2,10-/3,10-dimethylphenanthrene	DM13_210_310P	**0.743**	0.650	0.094
1,6-dimethylphenanthrene	DM16P	0.674	**0.691**	0.167
1,8-dimethylphenanthrene	DM18P	**0.779**	0.579	-0.011
1,9-dimethylphenanthrene	DM19P	**0.787**	0.585	0.044
2,5-/1,7-dimethylphenanthrene	DM25_17P	0.660	0.590	0.312
2,6-dimethylphenanthrene	DM26P	0.679	0.663	0.286
2,7-dimethylphenanthrene	DM27P	**0.692**	**0.698**	0.129
2,9-dimethylphenanthrene	DM29P	**0.701**	0.553	0.410
3,6-dimethylphenanthrene	DM36P	**0.716**	0.558	0.383
3,9-dimethylphenanthrene	DM39P	0.303	0.274	**0.863**
dibenzothiophene	DBT	0.352	**0.814**	0.391
2- methyldibenzothiophene	Me2DBT	0.330	**0.825**	0.435
3- methyldibenzothiophene	Me3DBT	0.256	**0.844**	0.390
4-methyldibenzothiophene	Me4DBT	0.320	**0.816**	0.449
fluoranthene	FLUORAN	**0.826**	0.460	0.277
pyrene	PYR	**0.834**	0.448	0.276
benzo[*a*]fluorene	BAFLUOREN	**0.866**	0.434	0.178
benzo[*b*]fluorene	BBFLUOREN	**0.872**	0.419	0.176
benzo[*ghi*]fluoranthene	BGHIF	**0.825**	0.413	0.360
benzo[*c*]phenanthrene	BCPHEN	**0.841**	0.447	0.269
benz[*a*]anthracene	BAA	**0.815**	0.487	0.264
cyclopenta[*cd*]pyrene	CPCDPYR	**0.818**	0.376	0.305
chrysene/triphenylene	CHRYS/TRIPH	**0.817**	0.474	0.307
benzo[*b*]naphtho[2,1-*d*]thiophene	BN21DT	**0.799**	0.471	0.352
benzo[*b*]naphtho[2,3-*d*]thiophene	BN23DT	**0.841**	0.398	0.322
benzo[*b*]fluoranthene	BBFLUORAN	**0.879**	0.392	0.241
benzo[*k*]fluoranthene	BKFLUORAN	**0.880**	0.399	0.211
benzo[*j*]fluoranthene	BJFLUORAN	**0.875**	0.414	0.223
benzo[*e*]pyrene	BEP	**0.871**	0.391	0.259
benzo[*a*]pyrene	BAP	**0.850**	0.405	0.290
perylene	PERY	**0.829**	0.456	0.242
benz[*a*]anthracene-7,12-dione	BAA712DIO	**0.716**	-0.015	0.291
indeno[1,2,3-*cd*]pyrene	ICDPYR	**0.785**	0.387	0.210
benzo[*ghi*]perylene	BGHIP	**0.890**	0.409	0.099
dibenz[*a,c*]anthracene	DBACANTH	**0.866**	0.406	0.203
coronene	COR	**0.882**	0.367	0.122
dibenzo[*a,e*]pyrene	DBAEPYR	**0.854**	0.436	0.205
% Explained Variance		***54.5***	***29.8***	***10.1***

Extraction Method: Principal Component Analysis. *Rotation Method:* Varimax with Kaiser Normalization. *Rotation converged in 9 iterations.*

The plot of measured and calculated TPAH values (see Figure S2a) showed a very high correlation ($R^2 = 0.988$). The percentage contributions to the mean for the three rotated factors (defined as $(B_k/\sum B_k)$ x 100% of each source) were 57% for t_1 (attributed to generic combustion), 28% for t_2 (identified as a petrogenic source), and 15% for t_3 (unknown source). In all, 85% of the PAH burden was ascribed to combustion or petrogenic sources. The contribution of each source *k* to the TPAH value in each particulate sample was calculated and a percentage normalized plot for each site over the period of sampling is shown subsequently. The sum of the three contributions was 100% in all cases. However, a number of samples had individual source contributions which exceeded or even greatly exceeded 100%; these large 'positive' source contributions were offset by large 'negative' source contributions in order to give net values of 100%. The large positive and negative contributions have no meaningful interpretation in a source apportionment context.

PMF Analysis of the Hamilton PAH data set

The PMF analysis was run on the same data used for the PCA using the EPA PMF 1.1 program which is based on the multilinear engine (ME) (Xie et al. 1999b; Paatero 1999). As discussed above, the specification of the uncertainty matrix *S* is essential to the PMF analysis. The EPA PMF 1.1 program allows users to specify uncertainties in a variety of ways: (i) observation-based uncertainties in which error estimates are given for each data point in the X matrix; (ii) equation-based uncertainties if the error associated with the X matrix is believed to derive from measurement uncertainties such as errors arising from instrumental fluctuations. The use of measurement uncertainties may not represent the total uncertainties in a given data set. Chen et al. argue that the use of measurement uncertainties may be an underestimation of the true s_{ij} (Chen et al. 2007). In this work, we applied the equation-based uncertainties, used the C3 function to add (20%) extra modeling uncertainty to the PMF runs all in the robust mode (Hopke 2001; Eberly 2005) and were able to get the robust Q (3938.91) in one of the random runs close to the optimum Q (3600.00).

Estimation of factors in PMF is subjective and there is no hierarchy of factors as with eigenvalue analyses. Three-factor to six-factor solutions were explored for this data set. The PMF analysis was re-run by reducing the number of factors from six to three in a step-wise manner with the criteria that the regression co-efficient between any two factors should not be better than 0.6, i.e., $R^2 < 0.6$.

Each PMF run was made with 30 initial random starting points. Nine variables were downweighted based on the model diagnostics. To provide uncertainities for source profiles, a bootstrapping technique (Eberly 2005) was applied in the EPA PMF 1.1 program with the selection of 100 random iterations and a minimum R=0.6 on the three base factors in the initial solution. The regression between the measured and modeled average variable concentrations for all the samples was $R^2 = 0.991$ (Figure S3).

The factor profiles (F matrix) for the 3 extracted factors and their comparisons with known PAH sources are displayed graphically in Figure 1. PAH:BeP normalizations were used where possible for the comparisons as a means of accessing the weathering of individual PAH compared to the relatively stable benzo[*e*]pyrene. In the case of the second factor however, the fractional composition of each PAH was used. Bootstrapped factor profiles are provided in the supplementary material (Table S2 and Figure S4). The first PMF factor, PMF Factor 1 (F_1, Figure 1A) was identified as gasoline emissions due to its high correlation to the PAH profile in the National Institute for Standards and Technology (NIST) Urban Dust Standard Reference Material (SRM 1649a, $R^2 = 0.77$). The relatively high levels of phenanthrene, fluoranthene and pyrene are consistent with a vehicular emissions/combustion source profile. This reference sample is an integrated source of mobile PAH emissions as it was collected in an area impacted primarily by mobile/vehicular emissions with little to no industrial impacts. PMF Factor 2 (F_2, Figure 1B) is relatively enriched in benzo[*b*]fluoranthene, benzo[*ghi*]perylene, and chrysene/triphenylene. A similar pattern ($R^2 = 0.61$) was reported previously and identified as a profile of diesel emissions (Sofowote, McCarry and Marvin 2008). Comparison of the PAH profile of NIST SRM 1650b (diesel exhaust reference material) and PMF Factor 2 showed a good agreement with the particulate-bound, higher molecular mass PAH (Figure S5).

Figure 1C shows comparisons of the PAH profile of PMF Factor 3 (F_3) with a Hamilton 'coal breeze' profile (unpublished data) ($R^2 = 0.75$) and the 'weathered PAH profile' (Sofowote, McCarry and Marvin 2008) identified in suspended sediment samples collected in Hamilton Harbour ($R^2 = 0.65$). Similar PAH profiles ($R^2 = 0.57$ & 0.75, Figures 2 & S9) were reported by Bzdusek et al. (Bzdusek et al. 2004) in sediment samples from Lake Calumet in Chicago. Their profiles showed a general depletion in low molecular mass PAH and relatively large enrichments in benzo[*ghi*]perylene, indeno[1,2,3-*cd*]pyrene, benzo[*a*]pyrene and benzo[*b*+*k*]fluoranthene. Thus, PMF Factor 3 was identified as coke oven impacts. These data, together with our previous findings in Hamilton Harbour suspended sediments (Sofowote, McCarry and Marvin 2008) strongly support direct atmospheric deposition of coke oven emissions into Hamilton Harbour in areas near steel-making operations. Mass apportionment was done by multilinear regression (MLR) of the elements in the G (source contibution) matrix against the corresponding TPAH for each sample. A minimum 95% confidence level was stipulated for the MLR. The correlation between the modeled and measured TPAH for the data set was $R^2 = 0.982$ (Figure S2b) and the multiple linear regression equation for the TPAH values (ng m^{-3}) was found to be:

$$y_{TPAH} = 1.97G_1 + 3.52G_2 + 13.2G_3 \quad (5)$$

Comparisons of PMF and PCA Models

Figure 3 shows the comparison of the percentage source contributions of the three factors identified in the PCA and PMF analyses in the 75 air particulate samples in this study. The data set was divided into approximate quartiles corresponding to the following total PAH concentration ranges: 0.23-2 ng m^{-3}, 20 samples; 2-10 ng m^{-3}, 18 samples; 10-20 ng m^{-3}, 19 samples; 20-172 ng m^{-3}, 18 samples. The percentage composition data for the PCA and PMF analyses are shown in Figures 3A and 3B, respectively, while the factor contributions to total PAH are shown in Figures 3C and 3D, respectively. The PCA percentage composition data (Figure 3A) show a number of samples with large positive and large negative percentage composition values with negative source contributions across the whole data set (Figure 3C). Negative source contributions are problematic since source apportionment interpretations are precluded.

On the other hand, the PMF percentage composition data always added to 100% (Figure 3B) with no negative contributions (Figure 3D). At low total PAH, mobile emissions sources (G_1 and G_2) were predominant while at high total PAH G_3 (coke oven emissions contributions) was predominant. Some samples with high total PAH showed significant contributions from diesel sources. Coke oven emissions impacts were discernible in almost all samples with the percent contribution increasing with increasing total PAH: first total PAH quartile (0.23-2.0 ng m^{-3} TPAH), mean = 21% coke oven contribution; second quartile (2-10 ng m^{-3}), 42%; third quartile (10-20 ng m^{-3}), 60%; fourth quartile (20–172 ng m^{-3}), 73%. Contributions due to the two mobile emissions sources were observed in almost all samples, consistent with a background of mobile emissions across the city, including the industrial area.

The PMF method is superior for source apportionment of this data set (see Table S1) yielding better resolved factors with realistic source contributions (Figure S8). The PCA failed to provide meaningful source contributions information even though the regression between the modeled and observed TPAH values was very high ($R^2 = 0.988$). A centered log ratio transformation of this data set before PCA neither produced more interpretable components nor yielded good correlations with MLR. The huge negative contributions in the PCA model are due to the model's inability to properly error-scale each datum before eigenvalue analysis. This problem has been noted previously by other researchers (Kiers 1997; Jansen et al. 2004). Weighted PCA (Kiers 1997) and PCA with nonnegativity constraints (Rachdawong and Christensen 1997) are PCA variants which have been devevloped to address these problems. We chose PMF because it is an FA method designed specifically to minimize negativity by properly error-scaling the data.

Comparison of PAH Diagnostic Ratios and PMF Factors as Source Apportionment Approaches

There is good agreement between the PMF source contribution plot and the classification of samples as 'upwind' or 'downwind' as discussed in the accompanying article (Sofowote, Allan and McCarry 2010). In that article, two PAH diagnostic ratios (An/∑178 and BaA/∑228 ratios) were shown to provide useful discrimination between samples collected upwind or downwind of industrial operations. In this work, each PMF factor was plotted against five PAH diagnostic ratios to afford 15 plots. Only two plots showed meaningful correlations: the BaA/∑228 and An/∑178 ratios plotted against PMF Factor 3 contributions (%G_3, Figure 4A). These plots had nearly identical slopes and correlation coefficients (R^2= 0.58 and 0.54). The other thirteen plots showed no meaningful correlations. In the accompanying paper (Sofowote, Allan and McCarry 2010), samples were classified as 'upwind' or 'downwind' based on their An/∑178 and BaA/∑228 ratios; 53 samples met both PAH ratio criteria and were classified as UW2 and DW2 respectively. Thirteen samples met the BaA/∑228 downwind criterion and were called DW1 while the remainder (9) were non-classified (NC); these classifications are used in Figure 4B. This figure shows a plot of a linear combination of the two PAH diagnostic ratios (termed 178-228 MeanRatio and defined as 0.5BaA/∑228 + 0.5An/∑178) against the percentage of PMF contributions (%G_3, Figure 4B) with an R^2 value of 0.65.

Samples with 178-228 MeanRatio values ≥0.28 were classified as downwind of industry while samples with 178-228 MeanRatio values ≤ 0.265, as upwind of industry (see horizontal dashed lines in Figure 4B). The 178-228 MeanRatio approach was surprisingly robust in that only 4 of 75 samples were unclassified (values between 0.265 and 0.28) and only one DW1 sample was misclassified.

The data set was divided into quartiles based on %G_3, values with break points 26% and 54% and 69% (as shown in Figure 4B). There is a strong correlation between samples designated as UW2 and DW2 and these quartiles. All UW2 samples save one are found in the bottom two quartiles while all DW2 samples save one, are in the upper two quartiles. In other words, samples with significant industrial impacts had > 50% G_3 contributions. The plots of 178-228 MeanRatio value against total PAH (log scale) proved very valuable for source interpretation (Figure 4C and 4D). In Figure 4C, the data points are categorized into four G_3 quartiles (Q1- Q4). Samples categorized as 'downwind' of the industrial area are found in the rectangle in the upper right portion of the figure while 'upwind' samples are in the area in the bottom left of the figure. A dashed diagonal line indicative of the gap between data points in the second and third %G_3 quartiles (see Figure 4B) divides the 'upwind' and 'downwind' bounded areas; a dashed arrow drawn perpendicular to this line goes diagonally from the lower left to the upper right of Figure 4C in the direction of increasing %G_3. Overall, the %G_3 contribution quartiles were excellent predictors of 'upwind' or 'downwind' source categories; 68 of the 75 samples (91%) showed the expected alignment. That is, within the 'downwind' rectangle are contained 35 of the 38 data points in the upper %G_3 quartiles; within the 'upwind' area are contained 33 of the 37 data points in the two lower %G_3 quartiles.

The data in Figure 4D is plotted identically to Figure 4C except that data points are categorized depending on the percent contribution of each one of the three PMF factors. The diagonal dashed line in Figure 4C that divided the 'upwind' and 'downwind' areas has been included in Figure 4D. Of the 39 samples with >50% contributions (shown as diamonds) all but 3 lie above the dashed line in the area identified as 'downwind' in Figure 4C. Of the data points below the line, 19 have %G_1 contributions >50% (circles), while 8 have %G_2 contributions >50% (triangles). Six of the 7 data points which did not fit these 3 categories (square boxes) also lie below the line. Thus, the sources of the 'upwind' samples were primarily mobile emissions while the 'downwind' samples were primarily coke oven emissions.

Conclusions

We have assessed the performance of two factor analysis methods, principal component analysis (PCA) and positive matrix factorization (PMF), on a 75 sample PAH data set from respirable air particulate material collected in an urban-industrial area. Both methods showed very high correlations between measured and predicted total PAH values. The PMF analysis performed significantly better than the PCA in identifying sources and providing meaningful source contributions to the PAH burden in each sample. The PCA was abandoned because it gave large negative contributions which offered no meaningful interpretations. The PMF analysis identified three factors which were linked to three sources: gasoline emissions, diesel emissions and coke oven emissions. The linear regression of the PMF data (equation 5) showed mean contributions of the three factors to be 1.97, 3.52 and 13.2 ng m^{-3}, respectively. Samples collected downwind of the industrial area showed high to very high total PAH levels, dominated by PMF Factor 3 (coke oven emissions source). Samples collected upwind of the industrial area had low total PAH levels and had significant contributions due to mobile emissions (PMF Factors 1 and 2).

We explored relationship between the PAH diagnostic ratio approach and PMF analysis. The only meaningful correlations found were between one PMF factors (coke oven impact factor) and two PAH ratios, the BaA/∑228 ratio and the An/∑178 ratio. These two ratios were linearly combined into a single value (defined as 178-228 MeanRatio) which showed a higher correlation than either individual ratio. Plots of 178-228 MeanRatio versus % coke oven emissions and versus total PAH allowed us to classify samples according to sources, wind directions and impacts at sampling sites. This approach is of general applicability to samples collected in complex urban-industrial environments.

Acknowledgements

We are grateful to the Ontario Ministry of the Environment for funding and sampling sites used in this study. We would also like to thank Ankit K. Rastogi for data mining and collation. Funding support from an NSERC Discovery Grant and the Great Lakes Sustainability Fund is acknowledged.

References

Abdi, H. 2003. Factor Rotations in Factor Analyses. In: Lewis-Beck, M., Bryman, A., Futing, T. (Eds.), Encyclopedia of Socical Sciences Research Methods, Sage, Thousand Oaks (CA),

Almeida, S.M., Pio, C.A., Freitas, M.C., Reis, M.A., Trancoso, M.A., 2006. Approaching PM(2.5) and PM(2.5-10) source apportionment by mass balance analysis, principal component analysis and particle size distribution. Science of the Total Environment 368, 663-74.

Anderson, M.J., Daly, E.P., Miller, S.L., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds: II. application of receptor models to TEAM study data. Atmospheric Environment 36, 3643-3658.

Anttila, P., Paatero, P., Tapper, U., Jarvinen, O., 1995. Source identification of bulk wet depostion in Finland by positive matrix factorization. Atmospheric Environment 29, 1705-18.

Bence, A.E., Kvenvolden, K.A., Kennicutt II, M.C., 1996. Organic geochemistry applied to environmental assessments of Prince William Sound, Alaska, after the Exxon Valdez oil spill--a review. Organic Geochemistry 24, 7-42.

Bruno, P., Caselli, M., de Gennaro, G., Traini, A., 2001. Source apportionment of gaseous atmospheric pollutants by means of an absolute principal component scores (APCS) receptor model. Fresenius' Journal of Analytical Chemistry 371, 1119-23.

Bzdusek, P.A. and Christensen, E.R., 2006. Comparison of a new variant of PMF with other receptor modeling methods using artificial and real sediment PCB data sets. Environmetrics 17, 387-403.

Bzdusek, P.A., Lu, J., Christensen, E.R., 2006. PCB congeners and dechlorination in sediments of sheboygan river, wisconsin, determined by matrix factorization. Environmental Science & Technology 40, 120-9.

Bzdusek, P.A., Christensen, E.R., Li, A., Zou, Q., 2004. Source apportionment of sediment PAHs in Lake Calumet, Chicago: Application of factor analysis with nonnegative constraints. Environmental Science & Technology 38, 97-103.

Chen, L.-.A., Watson, J.G., Chow, J.C., Magliano, K.L., 2007. Quantifying PM2.5 source contributions for the San Joaquin Valley with multivariate receptor models. Environmental Science & Technology 41, 2818-2826.

Christensen, E.R. and Bzdusek, P.A., 2005. PAHs in sediments of the Black River and the Ashtabula River, Ohio: Source apportionment by factor analysis. Water research 39, 511-24.

Du, S., Belton, T.J., Rodenburg, L.A., 2008. Source apportionment of polychlorinated biphenyls in the tidal Delaware River. Environmental Science & Technology 42, 4044-4051.

Eberly, S., 2005. EPA PMF 1.1 User's Guide, http://www.aqnis.pcd.go.th/project/TApce1/Tapce51/CMB%20WORKSHOP%20Prsentations/March6_2008PRESENTATIONS/EPA%20PMF%201_1%20Users%20Guide%20June%202005.pdf, 2005.

Fang, G.C., Wu, Y.S., Chang, C.N., Ho, T.T., 2006. A study of polycyclic aromatic hydrocarbons concentrations and source identifications by methods of diagnostic ratio and principal component analysis at Taichung Chemical Harbor near Taiwan Strait. Chemosphere 64, 1233-42.

Guo, H., Wang, T., Louie, P.K., 2004. Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model. Environmental pollution 129, 489-98.

Hegazi, A.H., Andersson, J.T., Abu-Elgheit, M.A., El-Gayar, M., 2004. Source diagnostic and weathering indicators of tar balls utilizing acyclic, polycyclic and S-heterocyclic components. Chemosphere 55, 1053-65.

Hopke, P., 2001. A Guide to Positive Matrix Factorization. http://www.epa.gov/ttn/amtic/files/ambient/pm25/workshop/laymen.pdf, USA, 16.

Hopke, P., 1985. Receptor Modeling in Environmental Chemistry. Wiley-Interscience, NY, Huang, S., Rahn, K.A., Arimoto, R., 1999. Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island. Atmospheric Environment 33, 2169-2185.

Jansen, J.J., Hoefsloot, H.C.J., Boelens, H.F.M., van der Greef, J., Smilde, A.K., 2004. Analysis of longitudinal metabolomics data. Bioinformatics 20, 2438-2446.

Kavouras, I.G., Koutrakis, P., Tsapakis, M., Lagoudaki, E., Stephanou, E.G., Von Baer, D., Oyola, P., 2001. Source apportionment of urban particulate aliphatic and polynuclear aromatic hydrocarbons (PAHs) using multivariate methods. Environmental Science & Technology 35, 2288-94.

Ke, L., Liu, W., Wang, Y., Russell, A.G., Edgerton, E.S., Zheng, M., 2008. Comparison of PM2.5 source apportionment using positive matrix factorization and molecular marker-based chemical mass balance. Science of the Total Environment 394, 290-302.

Kiers, H.A.L., 1997. Weighted least squares fitting using ordinary least squares algorithms. Pyschometrika 62, 251-266.

Kim, E., Hopke, P.K., Edgerton, E.S., 2003. Source identification of Atlanta aerosol by positive matrix factorization. Journal of the Air & Waste Management Association (1995) 53, 731-9.

Larsen, R.K.,3rd and Baker, J.E., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. Environmental Science & Technology 37, 1873-81.

Lee, E., Chan, C.K., Paatero, P., 1999. Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong. Atmospheric Environment 33, 3201-12.

Lee, P.K., Brook, J.R., Dabek-Zlotorzynska, E., Mabury, S.A., 2003. Identification of the major sources contributing to PM2.5 observed in Toronto. Environmental Science & Technology 37, 4831-40.

McCarry, B.E., Allan, L.M., Legzdins, A.E., Lundrigan, J.A., Marvin, C.H., Bryant, D.W., 1996. Thia-arenes as pollution source tracers in urban air particulate. Polycyclic Aromatic Compounds 11, 75-82; 75.

Miller, S.L., Anderson, M.J., Daly, E.P., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds. I. evaluation of receptor models using simulated exposure data. Atmospheric Environment 36, 3629-3641.

Paatero, P., 1999. The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. Journal of Computational and Graphical Statistics 8, 854-88.

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and Intelligent Laboratory Systems 18, 183-194.

Pandolfi, M., Viana, M., Minguillón, M.C., Querol, X., Alastuey, A., Amato, F., Celades, I., Escrig, A., Monfort, E., 2008. Receptor models application to multi-year ambient PM10 measurements in an industrialized ceramic area: Comparison of source apportionment results. Atmospheric Environment 42, 9007-9017.

Polissar, A.V., Hopke, P.K., Paatero, P., Malm, W.C., Sisler, J.F., 1998. Atmospheric aerosol over Alaska 2. elemental composition and sources. Journal of Geophysical Research 103, 19,045-57.

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

Rachdawong, P. and Christensen, E.R., 1997. Determination of PCB sources by a principal component method with nonnegative constraints. Environmental Science & Technology 31, 2686-2691.

Reinikainen, S.P., Laine, P., Minkkinen, P., Paatero, P., 2001. Factor analytical study on water quality in Lake Saimaa, Finland. Fresenius' journal of analytical chemistry 369, 727-32.

Shrivastava, M.K., Subramanian, R., Rogge, W.F., Robinson, A.L., 2007. Sources of organic aerosol: Positive matrix factorization of molecular marker data and comparison of results from different source apportionment models. Atmospheric Environment 41, 9353-9369.

Sofowote, U.M., Allan, L.M., McCarry, B.E., 2010. Evaluation of PAH diagnostic ratios as source apportionment tools for air particulates collected in an urban-industrial environment. Journal of Environmental Monitoring 12, 417-424.

Sofowote, U.M., McCarry, B.E., Marvin, C.H., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. Environmental Science & Technology 42, 6007-6014.

Statistics Canada, 2006. Http://www12.Statcan.ca/english/census06/data/profiles/community/Details/Page.Cfm?Lang=E&Geo1=CSD&Code1=3525005&Geo2=PR&Code2=35&Data=Count&SearchText=Hamilton&SearchType=Begins&SearchPR=01&B1=All&Custom=. Statistics Canada, Canada,

Stout, S.A., Uhler, A.D., Boehm, P.D., 2001. Recognition of and allocation among multiple sources of PAH in urban sediments. Environmental Claims Journal 13, 141.

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

Vaccaro, S., Sobiecka, E., Contini, S., Locoro, G., Free, G., Gawlik, B.M., 2007. The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. Chemosphere 69, 1055-63.

Varadarajan, C. and Kumar, A., 2006. Application of source-receptor models for air quality management. Environmental Progress 25, 180-5.

Vega, E., Mugica, V., Carmona, R., Valencia, E., 2000. Hydrocarbon source apportionment in Mexico City using the chemical mass balance receptor model. Atmospheric Environment 34, 4121-4129.

Wang, Z., Chen, J., Qiao, X., Yang, P., Tian, F., Huang, L., 2007. Distribution and sources of polycyclic aromatic hydrocarbons from urban to rural soils: A case study in Dalian, China. Chemosphere 68, 965-71.

Xie, Y., Hopke, P.K., Paatero, P., Barrie, L.A., Li, S., 1999a. Identification of source nature and seasonal variations of Arctic aerosol by positive matrix factorization. Journal of the Atmospheric Sciences 56, 249-60.

Xie, Y., Hopke, P.K., Paatero, P., Barrie, L.A., Li, S., 1999b. Identification of source nature and seasonal variations of Arctic aerosol by the multilinear engine. Atmospheric Environment 33, 2549-2562.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette, D., Sylvestre, S., 2002. PAHs in the fraser river basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry 33, 489-515.

Yunker, M.B., Macdonald, R.W., Veltkamp, D.J., Cretney, W.J., 1995. Terrestial and marine biomarkers in a seasonally ice-covered Arctic estuary- integration of multivariate and biomarker approaches. Marine Chemistry 49, 1-50.

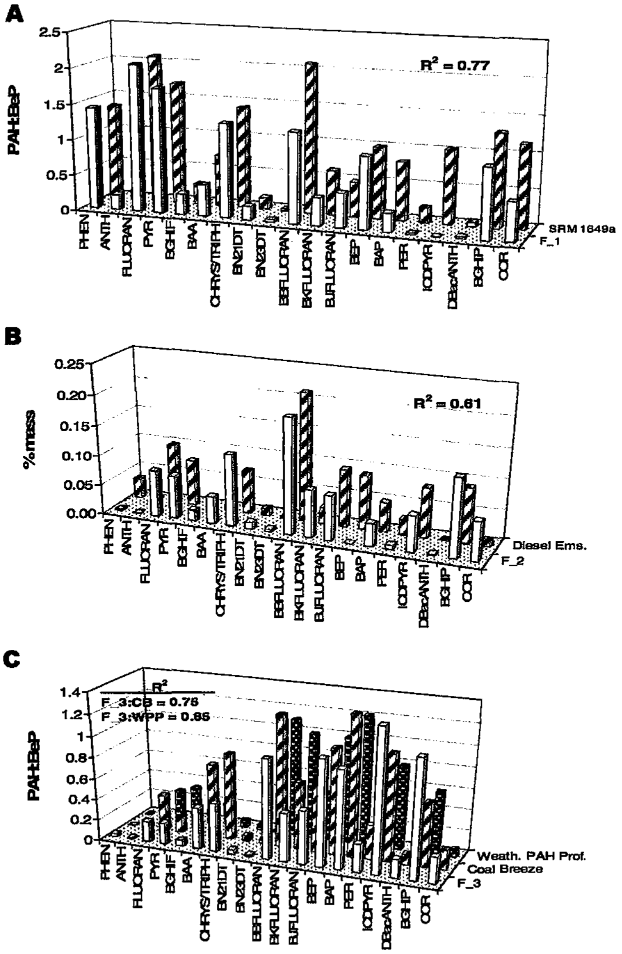


Figure 1: Factor profile comparisons of the three extracted PMF factors (F_1: gasoline emissions, F_2: diesel emissions, F_3: coke oven impacts) with known or previously reported factor profiles: A: Factor F_1 and SRM 1649a = NIST urban dust standard reference material; B: Factor F_2 and 'Diesel Ems' = Diesel emissions profile in Hamilton Harbour sediments (Sofowote, McCarry and Marvin 2008); C: Factor F_3, 'Coal Breeze' = PAH profile of a sample of coal dust obtained from a local company (unpublished data) and 'Weathered PAH Prof.' = PAH profile identified previously in suspended Hamilton Harbour sediments (Sofowote, McCarry and Marvin 2008).

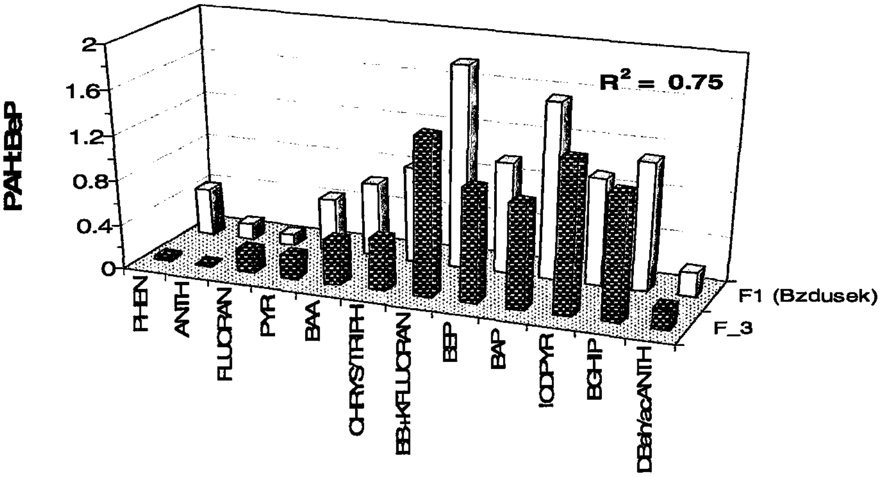


Figure 2: Comparison of PAH profile of PMF Factor 3 (F_3) with one of the 'Coke Oven Impacts' factors reconstructed from Bzdusek et al. (Bzdusek et al. 2004)

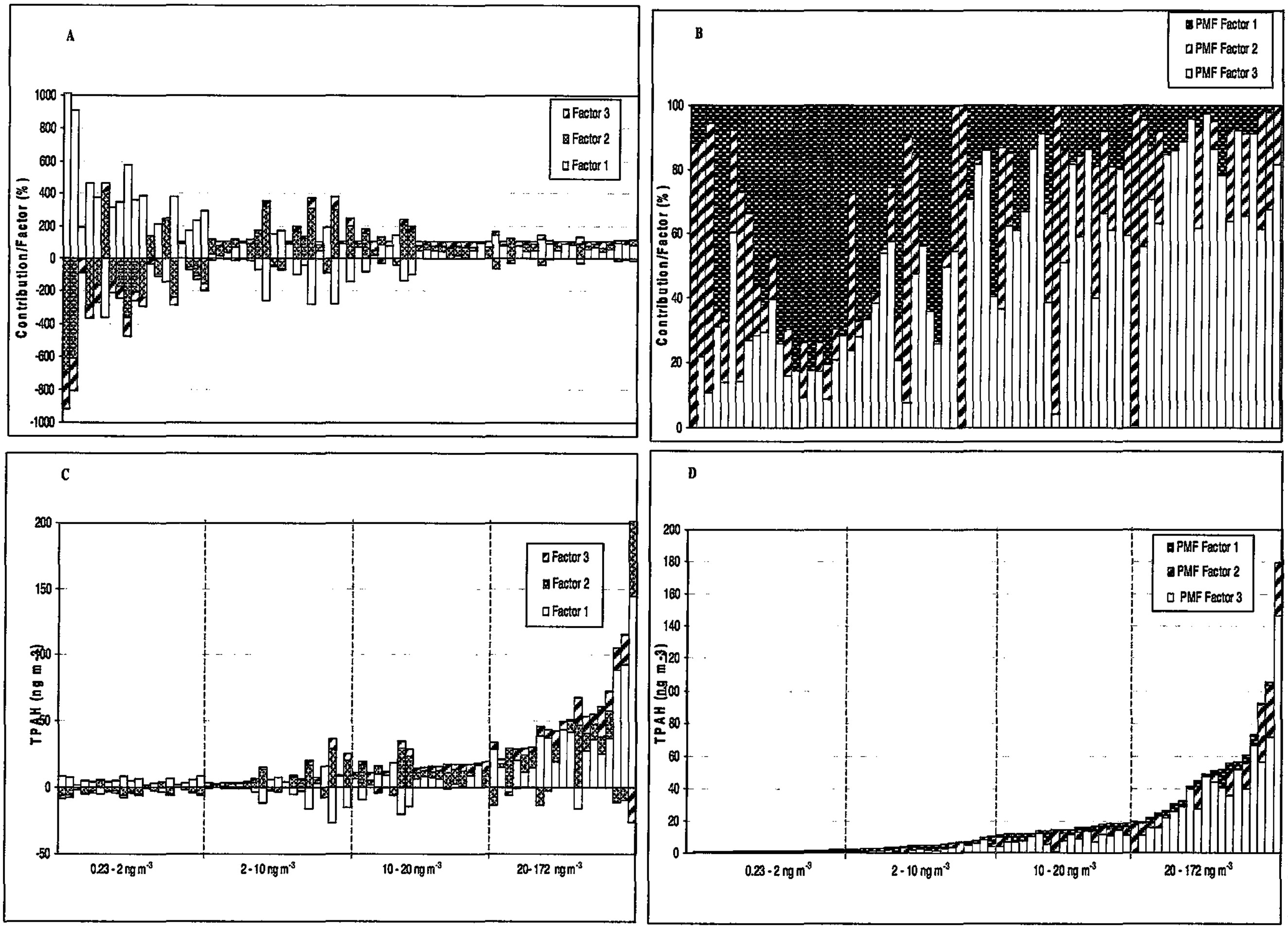


Figure 3: A: Percentage PCA factor contribution plot; B: Percentage PMF factor contribution plot; C: Absolute PCA factor contribution to the TPAH/ng m^{-3}; D: Absolute PMF factor contribution to the TPAH/ng m^{-3}. Left column: PCA; Factor 1 = 'Combustion'; Factor 2 = 'petrogenic'; Factor 3= unknown. Right column: PMF; PMF Factor 1 = Gasoline emissions; PMF Factor 2 = Diesel emissions; PMF Factor 3 = Coke oven impacts.

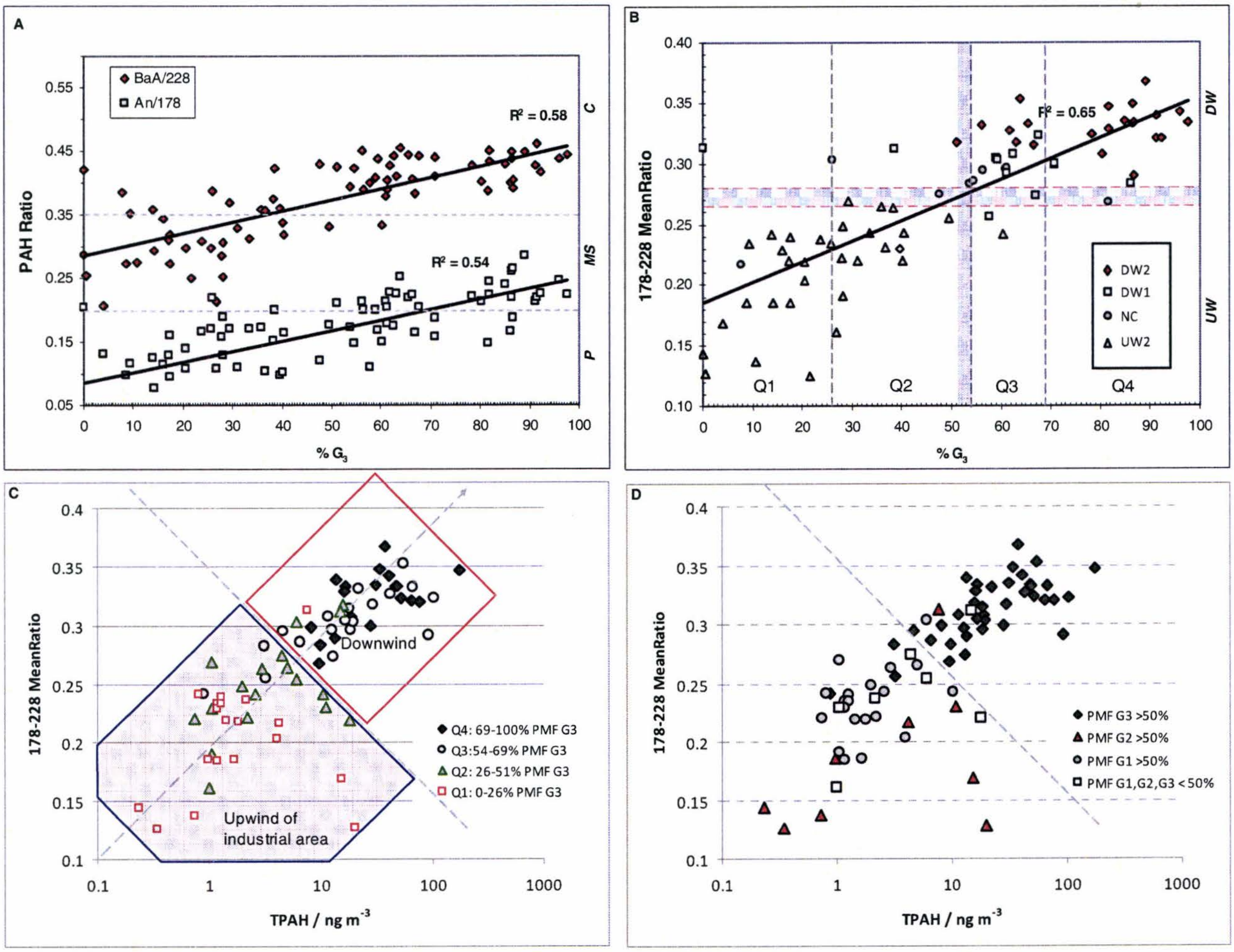


Figure 4: A: Correlation of two PAH diagnostic ratios ($BaA/\sum 228$ and $An/\sum 178$) with percentage of PMF Factor 3 ($\%G_3$) in 75 air particulate samples. *P* (petroleum), *MS* (mixed sources) and *C* (combustion) apportionments applicable to $BaA/\sum 228$ only; B: Correlation of the combined PAH diagnostic ratios (178-228 MeanRatio) with percentage of PMF Factor 3 ($\%G_3$; sectioned into quartiles). Shaded vertical region indicates gap between samples in second and third quartiles. Shaded horizontal region, 'upwind' (UW), 'downwind' (DW), 'non-classified' (NC) classifications are discussed in the main text; C: 178-228 MeanRatio variation with TPAH / ng m^{-3} sectioned into percentage PMF factor contribution $\%G_3$ quartiles ($\%G_3$ increases generally in the direction of dashed grey arrow); D: Summary of variations of all major percentage PMF factor contributions ($\%G_1$ -$\%G_3$) with the 178-228 MeanRatio and TPAH/ ng m^{-3} for the entire sample dataset.

Chapter Six

Source Apportionment of PAH in Hamilton Harbour Suspended Sediments: Comparison of Two Factor Analysis Methods

The work in this chapter has been published in an article under the title: Uwayemi M. Sofowote[1], Brian E. McCarry[1] and Chris H. Marvin[2], 2008, Environmental Science & Technology 42, 6007 – 6014 and is reproduced here with permission.

Author Contributions:

Uwayemi Sofowote: Performed some of the sample collection, all the sample extractions, cleanups and PAH quantification, data mining, interpretation, statistical analyses, and wrote the first draft of the manuscript.

Dr. Brian McCarry: Supervised many aspects of the data interpretation. Edited and wrote many portions of the final manuscript.

Dr. Chris Marvin: Supplied and supervised the collection of all the samples used. Co-authored and edited the manuscript.

Supplementary Information:

Supplementary tables and figures for this manuscript can be found in the attached compact disc

Author Affiliations:

[1]Department of Chemistry & Chemical Biology
McMaster University
1280 Main Street West, Hamilton, Ontario, L8S 4M1, Canada

[2]Water Science and Technology Directorate
Environment Canada
867 Lakeshore Road, Burlington, Ontario, L7R 4A6, Canada

Corresponding author email address: mccarry@mcmaster.ca; telephone: +1 (905) 525-9140

Source Apportionment of PAH in Hamilton Harbour Suspended Sediments: Comparison of Two Factor Analysis Methods

Uwayemi M. Sofowote, Brian E. McCarry, and Christopher H. Marvin

Abstract

A total of 26 suspended sediment samples collected over a 5-year period in Hamilton Harbour, Ontario, Canada and surrounding creeks were analyzed for a suite of polycyclic aromatic hydrocarbons and sulfur heterocycles. Hamilton Harbour sediments contain relatively high levels of polycyclic aromatic compounds and heavy metals due to emissions from industrial and mobile sources. Two receptor modeling methods using factor analyses were compared to determine the profiles and relative contributions of pollution sources to the Harbour; these methods are principal component analyses (PCA) with multiple linear regression analysis (MLR) and positive matrix factorization (PMF). Both methods identified four factors and gave excellent correlation coefficients between predicted and measured levels of 25 aromatic compounds; both methods predicted similar contributions from coal tar/coal combustion sources to the Harbour (19% and 26%, respectively). One PCA factor was identified as contributions from vehicular emissions (61%); PMF was able to differentiate vehicular emissions into two factors, one attributed to gasoline emissions sources (28%) and the other to diesel emissions sources (24%). Overall, PMF afforded better source identification than PCA with MLR. This work constitutes one of the few examples of the application of PMF to the source apportionment of sediments; the addition of sulfur heterocycles to the analyte list greatly aided in the source identification process.

Introduction

Hamilton Harbour is a 22 km^2 embayment on the western end of Lake Ontario in Canada (Hall, O'Connor and Ranieri 2006), bordered by the Niagara Escarpment, and the cities of Hamilton and Burlington; the harbour is connected to Lake Ontario via a shipping canal (Figure 1). The Hamilton-Burlington area is highly urbanized with a combined population of about 750,000. Roadway run-off and sewage treatment plant effluents are primary dischargers to the harbour. The southern shore is intensively industrialized, including the presence of two large steel mills.

In 1985, Hamilton Harbour was designated as an Area of Concern (AOC) by the International Joint Commission (Fox, Khan and Thiessen 1996) due, in part, to high levels of contaminants in harbour sediments, including polycyclic aromatic hydrocarbons (PAH) and heavy metals. An area within the harbour containing grossly contaminated bottom sediments is known as Randle Reef. The sum of EPA priority PAH concentrations (minus naphthalene) in bottom sediments in this region of the harbour can exceed 1400 μg/g (Environment Canada). These contaminated sediments act as an intra-harbour source of PAHs due to re-suspension into the water column. A containment facility based on *in situ* sequestration of these sediments is currently scheduled for construction. (Environment Canada). Suspended sediments within the harbour will be monitored before, during, and after implementation to obtain pre- and post-containment sediment profiles; we propose the use of source apportionment tools to assess contributions of coal tar-derived PAH.

Methods for source apportionment of PAH include diagnostic PAH ratio approaches (Yunker et al. 2002), multi-variate analysis methods (Larsen and Baker 2003) and more recently, isotope ratio mass spectrometry methods (Philip 2007). Diagnostic ratios of specific PAH give useful results primarily when there are only one or two primary sources. The diagnostic ratio approach is not limited to PAH; ratios of 3- and 4-ringed polyaromatic sulfur heterocycles (PASH or thia-arenes) have been used to distinguish between coal tar pollution and diesel exhaust emissions (McCarry et al. 1996) in air particulate and sediments (Marvin et al. 2000). Receptor modeling has been used primarily in air quality monitoring studies to describe a set of mass apportionment models to account for sources of pollution and their relative contributions at a receptor site (Hopke 1985). The two main approaches to receptor modeling are factor analyses and chemical mass balance approaches the latter being the method of choice when the impacting sources and their profiles have been pre-determined (Bi et al. 2007). Factor analyses, on the other hand, identify and quantify factors contributing to pollution data sets without prior knowledge of source profiles or source characteristics

(Guo, Wang and Louie 2004). This approach has been applied in studies involving atmospheric pollutants, both organic (Larsen and Baker 2003) and inorganic (Almeida et al. 2006), to sediments (Christensen and Bzdusek 2005) and soils (Wang et al. 2007).

The two factor analysis methods explored herein are principal component analysis (PCA) and positive matrix factorization (PMF). PCA is a multivariate analytical tool used widely for receptor modeling in environmental source apportionment studies (Almeida et al. 2006); for mass apportionment applications, multiple linear regression (MLR) of the standard deviate of the measured mass or concentration is applied to principal component scores (Larsen and Baker 2003). Positive matrix factorization (PMF), a receptor modeling tool developed in the early 1990s by Paatero and Tapper (Paatero and Tapper 1993), has been applied to source apportionment of aerosols in cities, (Lee et al. 2003), in bulk wet deposition (Anttila et al. 1995) and in characterization of contaminated soils (Vaccaro et al. 2007) and sediments (Bzdusek, Lu and Christensen 2006). In this paper we report the levels of 25 selected PAH and thia-arenes in 26 suspended sediment samples collected over a 5-year period from Hamilton Harbour and surrounding creeks. These data were analyzed using three source apportionment methods: diagnostic PAH ratios, principal component analysis with multiple linear regression and positive matrix factorization.

Experimental Section

Chemicals

Some deuterated standards (dibenzothiophene-d_8, phenanthrene-d_{10}, pyrene-d_{10}, chrysene-d_{12}, perylene-d_{12}, benzo[ghi]perylene-d_{12}, and dibenz[a,h]anthracene-d_{14}) were purchased from Cambridge Isotope Laboratories Inc, Massachusetts, pristane-d_{40} and 29$\alpha\alpha\alpha$(20R)ethylcholestane-d_2 from Chiron AS, Norway and hexacosane-d_{54} from CDN Isotopes, Quebec, Canada. A solution containing 48 PAH standards ranging from naphthalene to coronene (Sigma-Aldrich, MO), and 6 thia-arenes (Chiron AS, Norway) was prepared in toluene from authentic standards to give final concentrations of approximately 200 pg/μL for each compound.

Sample Collection

Suspended sediments from Hamilton Harbour (Figure 1, Table S1) were collected using sediment traps. Traps were open-bore polyacrylate columns fitted with removable base cups deployed for one-month periods at variable depths. Monthly suspended sediment samples were combined to afford composite seasonal samples representing 3 to 4 months of collection. A total of 19 within-harbour composite samples spanning 5 years (2002-2006) were included in this study. In addition, suspended sediment samples were collected in November 2006 using a Westfalia flow-through centrifuge from seven creeks discharging to Hamilton Harbour; Indian Creek (#2), Grindstone Creek (#5), and Red Hill Creek (#7) and the Desjardin Canal (#4) have direct inputs to the harbour (Figure 1, Table S1). Chedoke Creek (#6), Spencer Creek (#1) and Borer's Creek (#3) flow into a marsh which subsequnely flows into the harbour via the Dejardins Canal (#4). About 1000 L of water were centrifuged from each creek to afford 2-3 g of suspended sediment

Sample Extraction and Clean-up

The clean-up scheme followed a standard protocol from the National Laboratory for Environmental Testing Method 03-3751 (NLET 1997). Air-dried sediment samples (1.0-1.5 g dry weight) were loaded into extraction cells, spiked with recovery standards (pristane-d_{40}, hexacosane-d_{54}, ethylcholestane-d_2, phenanthrene-d_{10}, chrysene-d_{12}, dibenz[a,h]anthracene-d_{14} in amounts ranging from 10-40 ng each). Solvent extractions were carried out using dichloromethane in a Dionex ASE 200 instrument. Raw extracts were evaporated to 1 mL under nitrogen and transferred unto a silica gel column (4.5g, activated overnight at 170°C). Elution with pentane (25 mL) afforded Fraction A while elution with pentane/dichloromethane (1:1, 60 mL) afforded Fraction B. Fractions A and B were evaporated, reduced under nitrogen, and solvent-exchanged into iso-octane (1 mL).

GC/MS Analysis, Quality Control and Multivariate Analysis

A 100 µL aliquot of Fraction B was reduced under nitrogen, dissolved in a 100 µL iso-octane containing dibenzothiophene-d_8, pyrene-d_{10}, perylene-d_{12}, benzo[ghi]perylene-d_{12}, and analyzed by GC/MS (1 µL injection). Analyses were performed on a Varian CP3800/1200L system using electron impact ionization (EI^+) in selected ion monitoring (SIM) mode on a J&W DB-17ht column with He as the carrier gas. Method blanks were run with each sample batch. The PAH recoveries and standard deviations were: phenanthrene-d_{10}, 63.6±21.5%; chrysene-d_{12}, 97.7±2.2%; dibenz[a,h]anthracene-d_{14}, 101±1%. Method detection limits (in ng/g of dry sediment extracted) were 19.6 ng/g for dibenzothiophene-d_8, 24.6 ng/g pyrene-d_{10}, 147 ng/g for perylene-d_{12} and 96.6 ng/g benzo[ghi]perylene-d_{12}. SPSS 15.0 for Windows (SPSS Inc., Chicago) was used for PCA; PMF analyses were carried out using EPA PMF 1.1 program (http://www.epa.gov/heasd/products/pmf/pmf.htm).

Results and Discussion

Target PAH and thia-arenes used in statistical analyses along with their abbreviations are listed in Table 1 in order of elution from the GC column. The six thia-arenes included dibenzothiophene (DBT), 4-methyldibenzothiophene (4MeDBT), 2-methyldibenzothiophene (2MeDBT), 3-methyldibenzothiophene (3MeDBT), benzonaphtho[2,1-d]thiophene (BN21T), and benzonaphtho[2,3-d]thiophene (BN23T). The sum of the concentrations of the above target analytes was defined as SumPAH. The diagnostic PAH ratios which showed the greatest differentiation were the benz[a]anthracene/Σ(benz[a]anthracene + chrysene/triphenylene) ratio and the indeno[cd]pyrene/Σ(indeno[cd]pyrene + benzo[ghi]perylene) ratio. The plot of these two ratios (Supplementary Material, Figure S7) showed that there were primary contributions from coal combustion and vehicular/urban sources but it was not possible to draw any quantitative conclusions as to their relative contributions.

Principal Component Analysis

Principal component analysis (PCA) is a multivariate analytical tool used for receptor modeling in environmental source apportionment studies (Almeida et al. 2006; Bruno et al. 2001). The major advantage of this form of factor analysis is that all sources and source profiles need not be known or pre-determined. Concentrations of variables can be regarded as linear combinations of a number of underlying factors (or sources). The task in PCA analysis is to reproduce the correlation matrix as the sum of common factor coefficients, i.e., values in the loadings matrix (Hopke 1985). Sediment concentrations of 25 PAH and thia-arenes (Table 1) were used as variables for the PCA. The concentrations were Kaiser-normalized and Varimax rotation was used as the preferred transformation so that loadings values could be clustered around 1 and 0, making them more physically interpretable (Hopke 1985; Abdi 2003). More information on PCA analysis in environmental studies can be found in the Hopke (Hopke 1985) and in the supplementary data section. Four components were extracted and rotated, accounting for >97% of the variance in the data set. The factor loadings matrix is shown in Table 1; loading values exceeding 0.5 are highlighted in bold face.

Table 1: Varimax-rotated component matrix following principal component analysis of Hamilton Harbour suspended sediments. Values >0.5 are highlighted.

Variable	Variable Abbreviation	Rotated Component Number			
		1	2	3	4
phenanthrene	Phen	.196	**.557**	**.778**	.098
anthracene	Anth	.136	.367	**.884**	.191
fluoranthene	Fluoranth	**.829**	.139	**.529**	-.051
pyrene	Pyrene	**.722**	.099	**.638**	.013
benzo[ghi]fluoranthene	BghiFlu	**.951**	.123	.146	.056
benz[a]anthracene	BaA	**.798**	.161	**.541**	.161
cyclopenta[cd]pyrene	CcdPyr	.472	.031	.264	**.837**
chrysene/triphenylene	Chrys/triph	**.932**	.124	.258	.115
benzo[b]fluoranthene	BbF	**.962**	.213	.040	.127
benzo[k]fluoranthene	BkF	**.969**	.145	.079	.100
benzo[j]fluoranthene	BjF	**.968**	.164	.075	.131
benzo[e]pyrene	BeP	**.966**	.203	.089	.119
benzo[a]pyrene	BaP	**.944**	.135	.180	.208
perylene	Per	**.973**	.109	.125	.125
indeno[1,2,3-cd]pyrene	IcdPyr	**.932**	.232	.181	.123
dibenz[a,h]anthracene	DBahA	**.917**	.225	.184	.036
picene	Pic	**.873**	.242	.322	.244
benzo[ghi]perylene	BghiPer	**.946**	.234	.192	.093
coronene	Cor	**.894**	.293	.283	.118
dibenzothiophene	DBT	.191	**.766**	**.595**	.093
4-methyldibenzothiophene	4MeDBT	.251	**.950**	.143	-.024
2-methyldibenzothiophene	2MeDBT	.212	**.949**	.198	.035
3-methyldibenzothiophene	3MeDBT	.218	**.958**	.151	.022
benzonaphtho[2,1-d]thiophene	BN21T	**.934**	.262	.179	.070
benzonaphtho[2,3-d]thiophene	BN23T	**.954**	.182	.161	.107
Percentage Variance Explained		62.24	17.64	13.49	4.13

The first rotated component showed very high correlations for 16 PAH and 2 thia-arenes. Perylene, chrysene, the benzofluoranthenes, the benzopyrenes, DBahA, IcdPyr, BghiPer, BN21T, BN23T and Cor showed higher loadings than fluoranth, pyrene and BaA. This profile is highly indicative of vehicular sources (i.e., a combination of gasoline and diesel emissions sources). Coronene has been suggested as a gasoline mobile emissions tracer (Larsen and Baker 2003; Nielsen 1996) while perylene is strongly associated with gasoline derived from diagenetic petroleum sources (Kavouras et al. 2001). High loadings of benzo[k]fluoranthene

(Larsen and Baker 2003; Venkataraman, Lyons and Friedlander 1994) and benzonaphtho[2,1-d]thiophene (McCarry et al. 1996; Marvin et al. 2000) are consistent with diesel emissions.

The second rotated component is highly loaded in the three methyl dibenzothiophenes, and to a lesser extent in dibenzothiophene and phenanthrene, suggesting a generic petroleum (i.e., unburned fossil fuel) source. Crude oils and tar balls have been traced by patterns of dibenzothiophene and its alkylated homologues (Hegazi et al. 2004; Andersson, Hegazi and Roberz 2006). Phenanthrene and its methylated derivatives have also been used to identify unburned fossil fuels (Kavouras et al. 2001; Gogou et al. 1996); the relatively low loadings of the benzonaphthothiophenes made it clear that this source was not a diesel-like source. This rotated component was thus attributed tentatively to unburned petroleum.

The third rotated component showed good loadings for phenanthrene, anthracene, fluoranthene, pyrene, benz[a]anthracene, and dibenzothiophene, consistent with sources related to coal combustion (Larsen and Baker 2003; Kulkarni and Venkataraman 2000; Kulkarni and Venkataraman 2000; Fang et al. 2006). Relative loadings of dibenzothiophene to methylated dibenzothiophenes agree with data reported for coke oven condensate (McCarry et al. 1996; Marvin et al. 2000). The fourth rotated component showed a good correlation with only one PAH (cyclopenta[cd]pyrene) but remains unassigned due to insufficient chemical information to make a source interpretation possible. The existence of a fourth source is in doubt, even though it had an eigenvalue of >1 after varimax rotation; this source may be due to the inability of PCA to correctly scale each data point before analysis.

Component Scores Regression (PCA-MLR): Multiple linear regression analysis was performed on the PCA scores in order to obtain mass apportionment of the four sources to the total PAH burden in each sediment sample. The PCA-MLR methodology builds upon the PCA analysis and follows the model described elsewhere (Larsen and Baker 2003). A multiple linear regression (MLR) of elements in the factor scores matrix (t_k) against the normal standard deviate of the SumPAH values was carried out using a spreadsheet (Microsoft Excel) to obtain the best correlation between observed and predicted SumPAH values for each sample. The four factor scores had coefficients (B_k) which were determined by the MLR analysis with a stipulated minimum 95% confidence limit. The multiple linear regression equation for the standard deviate of the SumPAH values was determined to be:

$$\hat{z}_{\mathrm{SumPAH}} = 0.94t_1 + 0.21t_2 + 0.30t_3 + 0.11t_4 \qquad (1)$$

The plot of measured and calculated SumPAH values (see Figure S1) had an excellent correlation ($R^2 = 0.998$). The PCA-MLR model is an excellent predictor of source contributions in these samples over a wide range of PAH values. The percentage contributions to the mean for the four factors (defined as $(B_k/\Sigma B_k) \times 100\%$ of each source) were 61% for t_1 (attributed to vehicular emissions), 13% for t_2 (tentatively identified as unburned fossil fuels), 19% for t_3 (coal tar/coal combustion), and 7% for t_4 (unknown source). Overall, 80% of the PAH burden was ascribed to either mobile emissions (t_1) or coal-related sources (t_3). The contribution of each source k to the SumPAH value in each suspended sediment sample was calculated by re-defining the individual contribution I (in μg/g) of source k (Larsen and Baker 2003) as:

$$I_k = \text{mean[SumPAH]} \times (B_k/\Sigma B_k) + B_k \sigma_{\text{SumPAH}} t_k \qquad (2)$$

where mean[SumPAH] is 24.3μg/g and σ_{SumPAH} is 18.9μg/g for the Hamilton Harbour suspended sediment samples. The distributions of the four sources are shown in Figure 2. Contributions due to coal sources were relatively low except for samples collected near Randle Reef (Stations 9081 and 9083) and the centre of the harbour (Station 9032); these data point to Randle Reef as a significant source of coal tar contamination. The PAH levels in the Red Hill Creek and Windermere samples were high and showed substantial impacts due to mobile sources. The source tentatively attributed to unburned fossil fuels was a contributor to most harbour samples, particularly those sites in the east (e.g., Station 9033); this may be consistent with spills of fuels.

While the PCA-MLR analysis showed very high correlations, the source contribution plots (Figure 2) showed negative source contributions in some samples, particularly in the creek samples. These negative contributions cannot be explained rationally and are the outcomes of improper variable scaling inherent in eigenvalue-based (e.g., PCA) methods. We sought a factor analysis method that would allow for proper variable scaling and impose non-negativity constraints on both the source profiles and the resulting contributions to individual samples; positive matrix factorization fits these criteria.

Positive Matrix Factorization

The principal advantage of positive matrix factorization (PMF) over PCA methods stems inherently from the model employed in factor analysis. PMF treats the fundamental receptor modeling equation as a least squares problem (Paatero and Tapper 1993) and does not employ eigen-based analyses. In factor analysis, a data set X (n x m) consisting data points x_{ij} can be denoted as:

$$X = GF + E \qquad (3)$$

where G is the factor contribution matrix, F is the factor profile matrix and E (n x m) is the residual error matrix of elements e_{ij} indicating the residual value not accounted for by the modeled data value $\hat{x}_{ij}$, i.e.,

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \qquad (4)$$

where g_{ik} is the kth source's contribution to sample i and f_{kj} is the jth element's concentration in source k. By specifying an object function Q of the residual matrix which is to be minimized, PMF is better able to weight individual points in a data set based on uncertainties while simultaneously imposing non-negativity constraints on the contribution and profile matrices. We allowed the analysis to be run with small negative contributions in the G matrix, i.e., alowlim = -0.1 by default in the PMF 1.1 program, where Q is defined as:

$$Q(E) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{e_{ij}}{s_{ij}}\right]^2 \qquad (5)$$

where s_{ij} is the uncertainty in the x_{ij} measurement and is related to the method detection limit e_j for the variable in question. The EPA PMF program also allows an equation-based mode for specifying the uncertainties in x_{ij}. The uncertainty equations used by the EPA PMF program are shown below (Anttila et al. 1995; Eberly 2005).

$$S_{ij} = \sqrt{e_j^{\,2} + (d_j x_{ij})^2} \quad \text{if } x_{ij} > e_j$$

or, (6)

$$S_{ij} = 2(e_j) \qquad \text{if } x_{ij} \leq e_j$$

where d_j is the percentage uncertainty in determination of the variable which we estimated as standard deviations of deuterated surrogates recoveries (see Tables S2 and S4). Estimation of uncertainty from the equations above can be problematic for measured values clustering tightly around e_j A huge decrease in uncertainty is observed in going from $\leq e_j$ to even $1.1e_j$. However, we used the equations without modifications since our goal was to effectively down-weight noisier variables and force the data to be more strongly fitted by more reliable variables. Moreover, our concentration matrix did not show a narrow range of values around the method detection limit for most variables (>80%) in the concentration matrix thus the equations were deemed appropriate for this application.

This approach of weighting individual data points in PMF has been argued to be superior to eigen-based analysis (e.g., PCA) as it better deals with values below the method detection limit (MDL) and with missing data values (Anttila et al. 1995). Indeed, PCA has been shown to be based on a crude estimation of standard deviations of data points that invariably leads to scaling of data by column or row and leading to distortions in the analyses' (Hopke 2001). A number of algorithms have been written to solve the Q function depending on dimensions of linearity. The bilinear PMF2 (Ramadan et al. 2003) and trilinear PARAFAC (Xie et al. 1999) algorithms are well known. The multilinear engine (ME) was developed to expand applicability of PMF to more complex multi-linear data sets without the restriction to fixed linear models (Paatero 1999). The ME-2 engine allows the input of n-dimensions of linearity, which can be specifically tailored by the user to a data set, and used to carry out modeling (Anttila et al. 1995; Eberly 2005).

The EPA program PMF1.1 is based on the ME-2 engine (Eberly 2005) and was used in a robust, bilinear manner for analysis of Hamilton Harbour suspended sediment data. The multi-linear engine pre-screens data inputs and provides diagnostics for optimization before the PMF analysis is run. One such diagnostic tool is the signal-to-noise ratio (S/N (Eberly 2005)) which allows individual variables or points to be weighted as good, bad or weak. S/N for species j is estimated by the EPA PMF 1.1 program as;

$$\mathrm{S/N} = 0.5\left(\sum_{i=1}^{n} x^{2}_{ij} \Big/ \sum_{i=1}^{n} s^{2}_{ij} \right)^{0.5} \qquad (7)$$

This represents half the value extrapolated from the definition of S/N described by Paatero and Hopke for uncensored data (Paatero and Hopke 2003). From this, it is understood that a variable is weak if its S/N lies between 0.1 and 1. Nine variables were downweighted based on their S/N ratios among other things. More details on the data handling before PMF analysis can be found in the Supplementary Information section. The robust Q for 7 of 15 runs (all convergent) was 215.62. Random run #3 (robust Q = 215.62) was selected for further processing because it had the lowest true Q (216.74) and was converged in the fewest number of steps.

Estimation of PMF Factors and Mass Apportionment: Because PMF is not an eigen-based analysis, there is no apparent order in the arrangement of factors. There are also no strict rules for determining the number of factors to retain; however, Hopke has suggested that summing up multiple linear regression coefficient-scaled source profiles (F matrix elements) can be helpful in determining if too few factors have been extracted (Hopke 2001). In our analysis, a trial run with 5 random runs and five factors was conducted. Two of the five factors were found to have identical source profiles; a more detailed analysis was then performed in robust mode with four retained factors using 15 convergent random runs to afford equation 8 below. The correlation between the measured and calculated SumPAH values obtained from the MLR analysis on the source contribution matrix (G matrix) was very high ($R^2 = 0.995$). The sum of the MLR coefficient-scaled profiles for the four factors ranged from 92-101%.

$$y_{SumPAH} = 5.78G_1 + 6.85G_2 + 6.31G_3 + 5.49G_4 \qquad (8)$$

The profiles (in µg/g) of the 25 compounds (F matrix) in each of the four PMF factors are listed in Table 2. The source contribution matrix (G matrix) output is dimensionless since its elements g_{ik} represent the mass contribution (grams) of source k to the total mass (grams) of sample i. The G matrix can be further processed to reflect the fraction of each source k in each sample i by scaling elements in each row so their combined sum equals 1. Alternatively, the actual concentration contribution in µg/g of each source k to the SumPAH value of each sample i can be obtained by multiplying the MLR coefficient (µg/g) of each source k by the corresponding G matrix element (g/g). The sum of products of these MLR coefficients and corresponding G matrix elements in each row gives the modeled SumPAH value for each sample (Figure 3).

Table 2: Factor profiles (F matrix) of individual PAH and thia-arenes (in µg/g) in each of the four PMF factors identified in the analysis of Hamilton Harbour suspended sediment data set.

Compound	Diesel Emissions Factor f_1	Weathered PAH Profile Factor f_2	Gasoline Emissions Factor f_3	Coal Tar/ Combustion Factor f_4
phenanthrene	0.19	0.0020	0.17	**0.84**
anthracene	0.017	0.000	0.028	**0.15**
fluoranthene	0.58	0.24	**0.90**	**0.81**
pyrene	0.44	0.29	**0.62**	**0.55**
benzo[ghi]fluoranthene	0.013	**0.12**	0.029	**0.056**
benz[a]anthracene	0.054	0.19	**0.31**	**0.41**
cyclopenta[cd]pyrene	**0.011**	0.0070	**0.014**	0.0090
chrysene/triphenylene	**0.42**	0.058	**0.52**	0.084
benzo[b]fluoranthene	**1.2**	**0.95**	0.78	0.033
benzo[k]fluoranthene	0.15	**0.86**	0.16	0.30
benzo[j]fluoranthene	**0.55**	**0.38**	**0.48**	0.016
benzo[e]pyrene	**0.52**	**0.86**	0.33	0.26
benzo[a]pyrene	0.28	**1.05**	0.42	0.54
perylene	**0.15**	**0.35**	0.17	0.11
indeno[1,2,3-cd]pyrene	0.48	**0.68**	0.34	0.31
dibenz[a,h]anthracene	0.0030	**0.11**	0.000	**0.059**
picene	0.027	**0.13**	0.041	**0.11**
benzo[ghi]perylene	**0.53**	**0.51**	0.40	0.26
coronene	**0.055**	**0.056**	0.050	0.051
dibenzothiophene	**0.019**	0.0010	0.0040	**0.032**
4-methyledibenzothiophene	**0.027**	0.0010	0.000	0.0070
2-methyledibenzothiophene	**0.011**	0.000	0.000	0.0030

3-methyledibenzothiophene	**0.015**	0.000	0.000	0.0050
benzonaphtho[2,1-d]thiophene	**0.066**	0.034	**0.047**	0.015
benzonaphtho[2,3-d]thiophene	**0.011**	**0.010**	**0.012**	0.0030

The identification of sources using PMF was based on source profiles (Table 2). However unlike PCA, high values of a variable in a source factor do not mean the variable is necessarily highly correlated with the source. To decipher which variables are highly loaded, explained variation plots (Anttila et al. 1995) in the form of percentage mass of variables obtained from the PMF model output are helpful (Figure S4). It is important to note that explained variation plots are not source profiles and must never be interpreted as such. High loadings of phenanthrene, anthracene, fluoranthene and pyrene identified factor 4 (f_4) as a coal-related source; the loading of DBT relative to the methyl DBTs in factor 4 is consistent with data reported for coal-related sources (McCarry et al. 1996; Marvin 2007). Moreover, a comparison of the PAH profile of this factor (obtained by scaling F matrix elements in the column corresponding to factor 4) with a coal tar reference material NIST SRM 1597a (Figure S5b) gave a high correlation (R=0.89).

Factor 3 (f_3) was identified as gasoline emissions, based on loadings of perylene, chrysene and cyclopenta[cd]pyrene; the latter has been reported as a unique tracer for gasoline emissions (Larsen and Baker 2003; Daisey, Cheney and Lioy 1986). A comparison of the PAH loading profile of this factor with NIST SRM 1649a (Urban Dust) gave a good correlation (Figure S5a, R=0.89).

Factor 1 (f_1) was identified as related to diesel emissions based on loadings of thia-arenes. DBT and its methylated homologues are known to be highly loaded in diesel emissions sources (McCarry et al. 1996). The DBT:$\sum$DBT ratio of 0.36 for this factor is consistent with the value of 0.37 for the NIST diesel exhaust reference material SRM 1650 (McCarry et al. 1996). Indeno[1,2,3-cd]pyrene shows some loading in this factor and it has been reported as a tracer for diesel emissions (Kavouras et al. 2001).

The profile in factor 2 is dominated by 252 Da PAH (less perylene) and 276 Da PAH and represent >70% SumPAH) in this factor (Figure S6). The predominance of PAH with more than 4 rings has been attributed to a weathered PAH profile (Stout, Uhler and Boehm 2001). A PAH profile similar to that in factor 3 was reported in pre-1940 sections of sediment cores from Kamloops Lake; the authors argued that PAH in oils or on particle surfaces are not protected and will be subjected to weathering associated with solubilization and biodegradation (Yunker et al. 2002). Finally, the maximal contributions of this factor were in samples collected in the south-east part of the harbour (e.g., in Windermere Arm) proximate to a 50 ha confined disposal facility (CDF, Figure 1) which has been used to house PAH-containing sediments dredged from the harbour over many years. While this CDF was designed to prevent re-entry of pollutants back to the harbour (Rodgers et al. 1992a), the CDF is uncapped (Rodgers et al. 1992b) and when water levels are low, sediments are exposed to the air, allowing weathering of PAH in the CDF. This site may be a source of PAH with a weathered profile.

Comparisons of Mean Contributions of Factors by the Two Factor Models

Both factor analysis techniques were found to be very useful in quantifying source contributions and identifying the chemical profiles of sources without any prior knowledge of the number or types of sources impacting the harbour. Both PCA with MLR and PMF gave comparable mean contributions from a coal tar/coal combustion source to Hamilton Harbour suspended sediments (19% vs. 26%, respectively). The PAH contributions attributed to vehicular emissions sources (combined gasoline and diesel sources) identified by PCA-MLR (61%) is comparable to the sum (52%) of gasoline and diesel emissions identified as separate factors by PMF (28% and 24%, respectively, Table S4). Thus, coal-derived and vehicular sources account for 80% and 78% of the PAH burden in harbour sediments. The remainder of the PCA-MLR contributions (20%) was associated with two factors which could not be attributed to PAH sources with any degree of certainty. The fourth PMF factor (22%) was attributed to a weathered PAH profile, presumably from legacy sources. The addition of a suite of thia-arenes to the PAH target list greatly assisted both factor analysis methods in source identification.

PMF analysis not only differentiated between gasoline and diesel emissions, but overcame the problems of negative contributions to profiles that are inherent to the PCA approach to a large extent. The identification of sources with PCA is suspect as there is no scaling of data based on uncertainty estimates before analysis (as seen with rotated component 4 which was solely loaded in CcdPyr). The other sources identified with PCA may not be entirely accurate since interpretation of PAH profiles was undertaken with noisy variables included at equal strength with more reliable ones. Overall, the PMF analysis gave more complete and clearer source apportionment results on this data set than did the PCA-MLR approach and is the preferred method for this data set.

Relevance of Current Factor Analytic Work to Hamilton Harbour

Both the PCA-MLR and PMF analyses confirmed that coal-derived sources of PAH are significant contributors to suspended sediment samples collected near Randle Reef while coal-derived sources have lesser impacts elsewhere; on the other hand, mobile sources are important contributors to PAH contamination throughout the harbour and in the creeks. Earlier research we have undertaken on sediments from the harbour employing thia-arene ratios has shown that coal-tar inputs generally decrease farther away from Randle Reef (Marvin et al. 2000). The inclusion of the creek samples confirmed that the coal-derived source(s) of PAH contamination were within-harbour source(s). The project to remediate the Randle Reef area of the harbour should reduce the contribution of coal-derived PAH contamination to harbour sediments dramatically, leaving mobile sources as the primary sources of PAH contamination. The challenge for Hamilton Harbour and many PAH-impacted, urbanized areas throughout the world is the development of strategies to reduce the PAH inputs from mobile sources in general, and particularly once industrial-derived sources of PAH contamination have been contained or eliminated.

The PMF model indicates the influence of a diesel source at Station 9033. We speculate this observation may be the result of increased large-ship traffic which uses heavy diesel fuels. Gasoline emissions are ubiquitous within the harbour and are strongest in the Windermere Bridge samples. This is consistent with the highly urbanized nature of this area with its substantial road networks and vehicular traffic. The presence of this source at Randle Reef is also consistent with many combined sewer overflows (CSO) that discharge into the harbour at the reef (Pozza, Boyce and Morris 2004).

Acknowledgments

Funding for this project was provided through Environment Canada's Great Lakes Sustainability Fund. Samples were collected under the auspices of the National Water Research Institute's Technical Operations department. Hamilton Harbour map inset courtesy of Great Lakes Website http://www.epa.gov/glnpo/lakeont/2002highlight/index.html.

Supporting Information

Supplementary figures and tables are freely available at http://pubs.acs.org.

References

Abdi, H. 2003. Factor Rotations in Factor Analyses. In: Lewis-Beck, M., Bryman, A., Futing, T. (Eds.), Encyclopedia of Socical Sciences Research Methods, Sage, Thousand Oaks (CA),

Almeida, S.M., Pio, C.A., Freitas, M.C., Reis, M.A., Trancoso, M.A., 2006. Approaching PM(2.5) and PM(2.5-10) source apportionment by mass balance analysis, principal component analysis and particle size distribution. Science of the Total Environment 368, 663-74.

Anderson, M.J., Daly, E.P., Miller, S.L., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds: II. application of receptor models to TEAM study data. Atmospheric Environment 36, 3643-3658.

Andersson, J.T., Hegazi, A.H., Roberz, B., 2006. Polycyclic aromatic sulfur heterocycles as information carriers in environmental studies. Analytical and Bioanalytical Chemistry 386, 891-905.

Anttila, P., Paatero, P., Tapper, U., Jarvinen, O., 1995. Source identification of bulk wet depostion in finland by positive matrix factorization. Atmospheric Environment 29, 1705-18.

Bi, X., Feng, Y., Wu, J., Wang, Y., Zhu, T., 2007. Source apportionment of PM10 in six cities of northern China. Atmospheric Environment 21, 903-12.

Bruno, P., Caselli, M., de Gennaro, G., Traini, A., 2001. Source apportionment of gaseous atmospheric pollutants by means of an absolute principal component scores (APCS) receptor model. Fresenius' Journal of Analytical Chemistry 371, 1119-23.

Bzdusek, P.A., Lu, J., Christensen, E.R., 2006. PCB congeners and dechlorination in sediments of sheboygan river, wisconsin, determined by matrix factorization. Environmental Science & Technology 40, 120-9.

Christensen, E.R. and Bzdusek, P.A., 2005. PAHs in sediments of the black river and the ashtabula river, ohio: Source apportionment by factor analysis. Water research 39, 511-24.

Daisey, J.M., Cheney, J.L., Lioy, P.J., 1986. Profiles of organic particulate emissions from air pollution sources: Status and needs for receptor source apportionment modeling. Journal of the Air Pollution Control Association 36, 17-33.

Eberly, S., 2005. EPA PMF 1.1 User's Guide.

Environment Canada, http://sustainabilityfund.gc.ca/Randle_Reef_Sediment_Remediation_Project-WS4F110F61-1_En.htm.(Accessed August 2007)

Fang, G.C., Wu, Y.S., Chang, C.N., Ho, T.T., 2006. A study of polycyclic aromatic hydrocarbons concentrations and source identifications by methods of diagnostic ratio and principal component analysis at Taichung Chemical Harbor near Taiwan Strait. Chemosphere 64, 1233-42.

Fox, M.E., Khan, R.M., Thiessen, P.A., 1996. Loadings of PCBs and PAHs from Hamilton Harbour to Lake Ontario. Water Quality Research Journal of Canada 31, 593-608; 593.

Gogou, A.I., Stratigakis, N., Kanakidou, M., Stephanou, E.G., 1996. Organic aerosols in eastern mediterranean: Components source reconciliation by using molecular markers and atmospheric back trajectories. Organic Geochemistry 25, 79-96.

Guo, H., Wang, T., Louie, P.K., 2004. Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model. Environmental pollution 129, 489-98.

Hall, J.D., O'Connor, K., Ranieri, J., 2006. Progress toward delisting a great lakes area of concern: The role of integrated research and monitoring in the Hamilton Harbour remedial action plan. Environmental monitoring and assessment 113, 227-43.

Hegazi, A.H., Andersson, J.T., Abu-Elgheit, M.A., El-Gayar, M., 2004. Source diagnostic and weathering indicators of tar balls utilizing acyclic, polycyclic and S-heterocyclic components. Chemosphere 55, 1053-65.

Hopke, P., 2001. A Guide to Positive Matrix Factorization. http://www.epa.gov/ttn/amtic/files/ambient/pm25/workshop/laymen.pdf, USA, 16.

Hopke, P., 1985. Receptor Modeling in Environmental Chemistry. Wiley-Interscience, NY,

Kavouras, I.G., Koutrakis, P., Tsapakis, M., Lagoudaki, E., Stephanou, E.G., Von Baer, D., Oyola, P., 2001. Source apportionment of urban particulate aliphatic and polynuclear aromatic hydrocarbons (PAHs) using multivariate methods. Environmental Science & Technology 35, 2288-94.

Kulkarni, P. and Venkataraman, C., 2000. Atmospheric polycyclic aromatic hydrocarbons in Mumbai, India. Atmospheric Environment 34, 2785-90.

Larsen, R.K.,3rd and Baker, J.E., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. Environmental Science & Technology 37, 1873-81.

Lee, P.K., Brook, J.R., Dabek-Zlotorzynska, E., Mabury, S.A., 2003. Identification of the major sources contributing to PM2.5 observed in Toronto. Environmental Science & Technology 37, 4831-40.

Marvin, C.H., McCarry, B.E., Villella, J., Allan, L.M., Bryant, D.W., 2000. Chemical and biological profiles of sediments as indicators of sources of genotoxic contamination in Hamilton Harbour. part I: Analysis of polycyclic aromatic hydrocarbons and thia-arene compounds. Chemosphere 41, 979-988.

McCarry, B.E., Allan, L.M., Legzdins, A.E., Lundrigan, J.A., Marvin, C.H., Bryant, D.W., 1996. Thia-arenes as pollution source tracers in urban air particulate. Polycyclic Aromatic Compounds 11, 75-82; 75.

Nielsen, T., 1996. Traffic contribution of polycyclic aromatic hydrocarbons in the center of a large city. Atmospheric Environment 30, 3481-90.

NLET, 1997. National Laboratory for Environmental Testing NLET Method 03-3751: Method for the Analysis of 9 Chlorobenzenes, 17 Organocholrine Pesticides, Total Polychlorinated Biphenyls and 16 Polynuclear Aromatic Hydrocarbons in Sediment using Ultrasonic Extraction and Gas Chromatrography with Electron Capture and Mass Selective Detectors. 3, Organics, Organics in Sediments.

Paatero, P. and Hopke, P.K., 2003. Discarding or downweighting high-noise variables in factor analytic models. Analytica Chimica Acta 490, 277-89.

Paatero, P., 1999. The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. Journal of Computational and Graphical Statistics 8, 854-88.

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and Intelligent Laboratory Systems 18, 183-194.

Philip, R.P., 2007. The emergence of stable isotopes in environmental and forensic geochemistry studies: A review. Environmental Chemistry Letters 5, 57-66.

Pozza, M.R., Boyce, J.I., Morris, W.A., 2004. Lake-based mapping of contaminant sediment distribution, Hamilton Harbour, Lake Ontario, Canada. Journal of Applied Geophysics 57, 23-41.

Ramadan, Z., Eickhout, B., Song, X., Buydens, L.M.C., Hopke, P.K., 2003. Comparison of postive matrix factorization and multilinear engine for the source apportionment of particulate pollutants. Chemometrics and Intelligent Laboratory Systems 66, 15-28.

Rodgers, K., Vogt, J., V, C., Boyd, D., Simser, L., Lang, H., Murphy, T., Painter, S., 1992a. Environmental Conditions and Problem Definition. , 73-9.

Rodgers, K., Vogt, J., V, C., Painter, S., Boyde, D., Lang, H., Simser, L., Murphy, T., 1992b. Goals, Options and Recommendations. 2- Main Report, 114.

Stout, S.A., Uhler, A.D., Boehm, P.D., 2001. Recognition of and allocation among multiple sources of PAH in urban sediments. Environmental Claims Journal 13, 141.

Vaccaro, S., Sobiecka, E., Contini, S., Locoro, G., Free, G., Gawlik, B.M., 2007. The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. Chemosphere 69, 1055-63.

Venkataraman, C., Lyons, J.M., Friedlander, S.K., 1994. Size distributions of polycyclic aromatic haydrocarbons and elemental carbon. 1. sampling, measurement methods and source characterization. Environmental Science & Technology 28, 555-62.

Wang, Z., Chen, J., Qiao, X., Yang, P., Tian, F., Huang, L., 2007. Distribution and sources of polycyclic aromatic hydrocarbons from urban to rural soils: A case study in Dalian, China. Chemosphere 68, 965-71.

Xie, Y., Hopke, P.K., Paatero, P., Barrie, L.A., Li, S., 1999. Identification of source nature and seasonal variations of arctic aerosol by positive matrix factorization. Journal of the Atmospheric Sciences 56, 249-60.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette, D., Sylvestre, S., 2002. PAHs in the Fraser River Basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. Organic Geochemistry 33, 489-515.

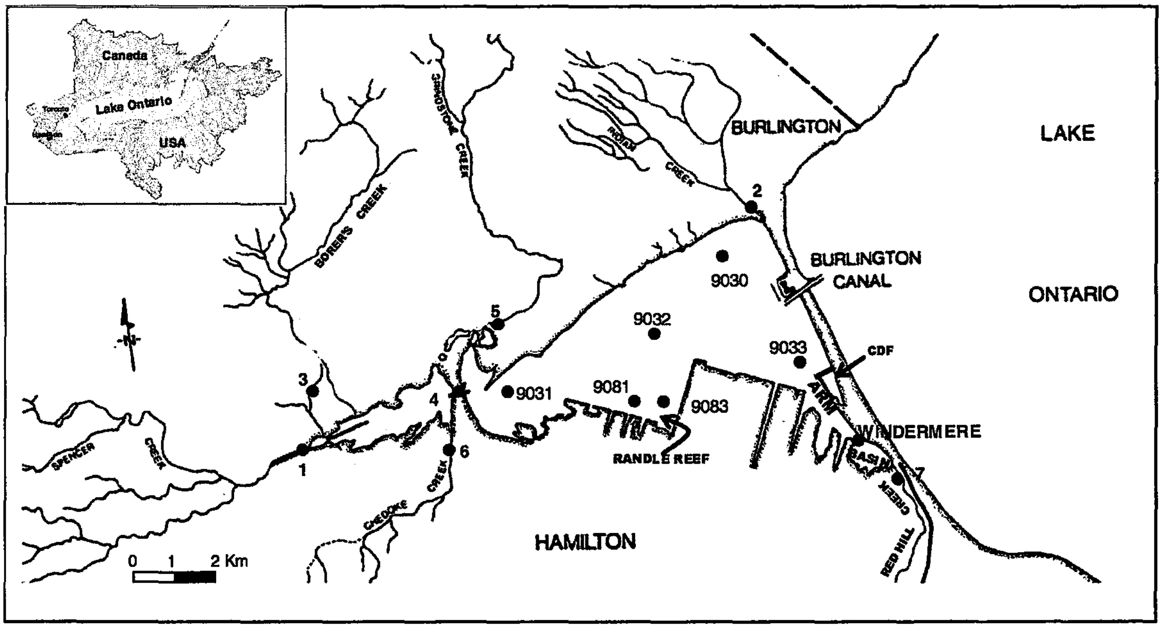


Figure 1: Map of Hamilton Harbour showing harbour sampling stations and sampling sites in surrounding creeks. Creek sampling sites are numbered 1-7 (Table S1).

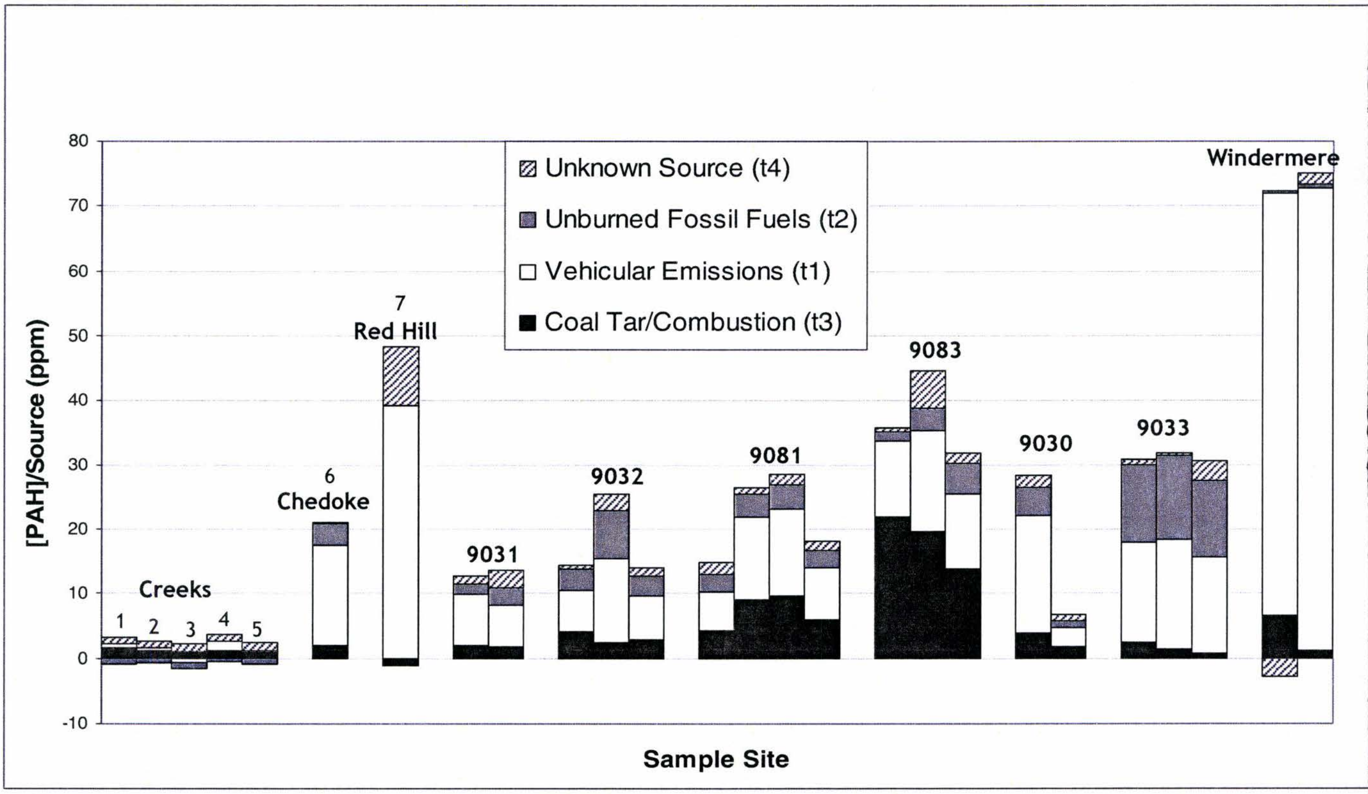


Figure 2: PCA-MLR source contributions plot for Hamilton Harbour suspended sediments and creek sediments (µg/g). Creeks are numbered 1-7 as in Figure 1.

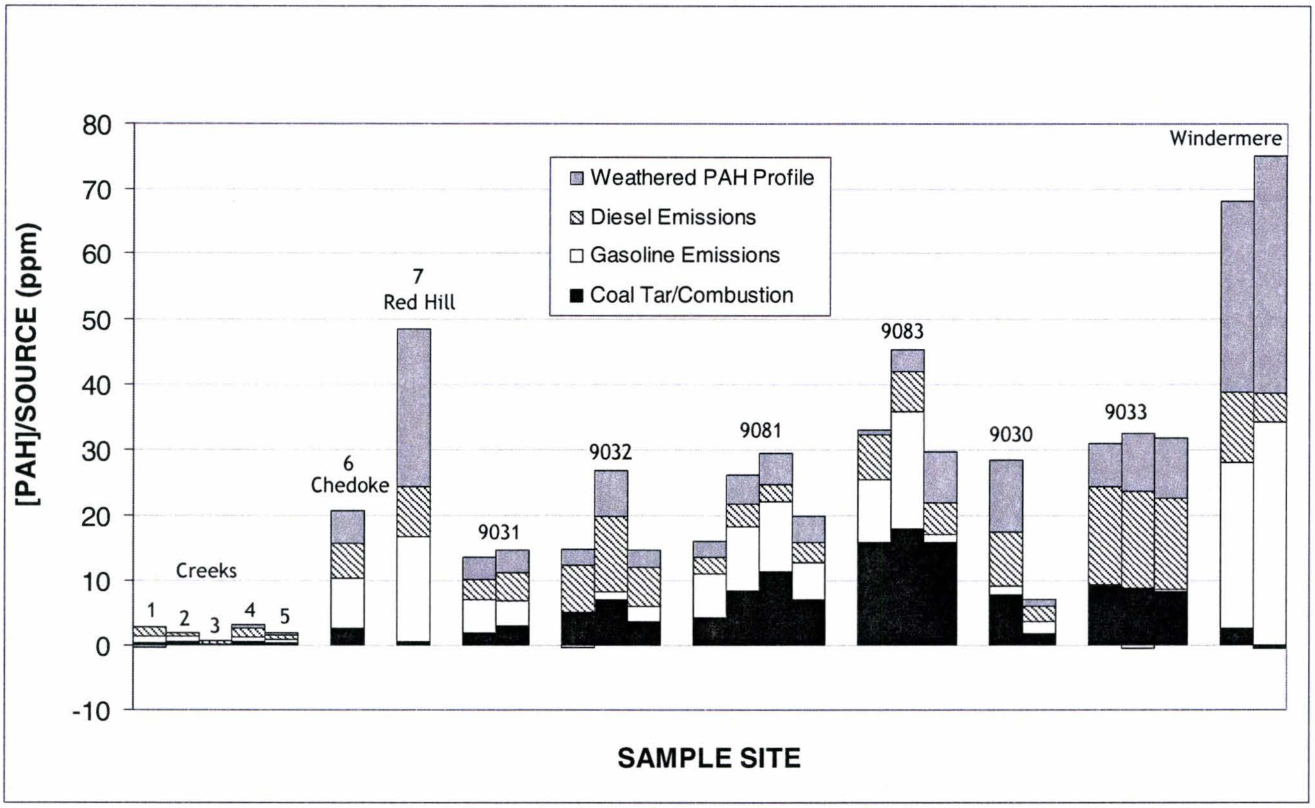


Figure 3: Contributions of four PMF factors (in μg/g) to the SumPAH (μg/g) of each suspended sediment sample collected in Hamilton Harbour and surrounding creeks (numbered 1-7 as in Figure 1).

Chapter Seven

Conclusions

7.1. Overview

Analytical chemistry is the branch of chemistry that is concerned with the identification and quantification of species of interest (analyte) in a well-defined matrix. The fundamental questions that must be answered by any analysis can be summed up as a. What is it; b. How much of it is in the matrix; and c. What does the result mean? Analytes of interest may vary from simple elements to complex molecules. In environmental studies as with other analysis-oriented specializations, very sensitive detection methods are now readily available so that the first two questions are now routinely answered with good accuracy and precision to generate reliable concentration data. The third question deals with the bigger picture which is usually extremely challenging to grasp by the human mind. The solution to this question may require source apportionment techniques for delineating the impacting sources of pollution and determining their relative contributions. Thus, source apportionment goes beyond contaminant monitoring in the environment.

Uni-variate analyses of analyte/matrix relationships have been used as representative estimators of underlying causal factors but this is an overly simplistic interpretation of complex data. The multivariate nature of environmental pollution data requires the use of models that reveal multi-dimensional relationships for thorough source apportionment. The simultaneous contributions of causal factors are best revealed by the application of multivariate statistics. Factorization models coupled with regression techniques offer the best approach to tackling quantitative source apportionment without a prior knowledge of causal agents and have been applied extensively to data from air and sediments from polluted waters in this thesis. Factorization models are usually successful because no matter how complex a pollution data set appears to the human mind, only a relatively few factors in their varying forms and contributions may be responsible for the distribution of contaminant species observed in a data set.

Factor analytical results are not incontrovertible. Firstly, models vary in their approach to factorization of complex data sets and model results vary in their ability to fully represent the variability inherent in real world data. A recurring finding in this work was that even after suitable rotations, PCA yielded unrealistic (negative) source profiles and contributions whenever it was used, thus reducing the accuracy of the factor analysis (see chapters 5 and 6). Positive matrix factorization of data overcomes the negativity shortcomings of PCA and was the factor analytic model of choice for the more recent work presented in Chapter 3. PMF has been shown to approach chemical mass balance results in many studies but is substantially more complicated to understand and use compared to common factor analysis such as PCA. The subjective nature of interpreting factors often requires a keen knowledge of the underlying chemical principles responsible for the formation and modification of factor profiles. PMF has the added complexity of not being an eigenvalue analysis, thus, the enumeration and order of causal factors are arbitrary. In short, the outcome of the interpretation is dependent on the experience and knowledge of the environmental researcher.

An additional limitation encountered with the use of factor analytic receptor modeling is the inability of these models to correct for fate processes during long-range transport when dealing with atmospheric organic contaminants with relatively short lifetimes such as PAH. It must however be noted that no known atmospheric transport model has yet effectively accounted for the complicated degradation and deposition processes PAH undergo during long-range transport even when these parameters have been factored in the model assumptions. Thus, this is a generic problem in long-range transport modeling with no simplified solution.

This thesis in all its chapters shows that source apportionment of PAH is possible in diverse environmental media (in this case, environmental samples ranged from suspended sediments to atmospheric (gaseous and particulate) samples). The use of factor analysis as a stand-alone research tool or in conjunction with other non-multivariate apportionment methods is a strong theme that permeates this body of work (with the exception of Chapters 2 and 4). A few chapters dealt with comparing PMF and PCA factor analytic models for their source apportionment outcomes (Chapters 5 and 6) or showed the difference in the quality of outcomes obtained with different types of source apportionment techniques notably diagnostic ratios and PMF (Chapters 4 and 5). After rigorous testing of the two factor analytic models used in this work, PMF was identified as the more realistic model to apply to the sub-Arctic data set shown in Chapter 3. This finding was consistent with what was observed in Chapters 5 and 6. Chapter 3 focused on the application of PMF in conjunction with PSCF on a hemispherical scale. This chapter detailed the first instance of the use of the PMF-PSCF hybridized methodology for organic pollutants on a hemispherical scale. Chapter 2 showed how temporal variations (seasonality) and meteorological changes affect the sources of PAH (local versus long-range sources) which in turn can affect the gas-/particle-partitioning of PAH in sub-Arctic samples collected to investigate long-range transport of pollutants to Arctic regions.

The discussion below summarizes the contributions and major findings of each chapter of my thesis.

7.2. Summary

Chapter 2

Gas- and particle-phase PAH were collected in polyurethane foam plugs and on glass fibre filters respectively at a fairly remote air monitoring site in the Yukon Territory, Canada from August 2007 to October 2009. PAH concentrations were found to be highest in the winter months and lowest in summer, also, an enrichment of particle-phase PAH in samples collected in winter was observed. $PM_{2.5}$ were also found to be highest in winter, though many incidences of relatively high levels were observd in spring and summer. Summertime partioning of lower molecular mass PAH were found to have a preference for the particle-phase, thus resulting in shallower slopes in the log-log correlation plots between the partition coefficient and the sub-cooled vapour pressures. This irregular partitioning behaviour suggests that lower molecular mass PAH in summer samples are affected by non-exchangeability, a property of their impacting source. Air mass back trajectories for an intensive fourteen-day summer sampling period at the receptor site show the densest aggregation of endpoints in relative proximity of the site suggesting that local impacting sources may be more important during this period. Local summer sources of PAH may arise from camping activities involving increased vehicular transport, wood and fuel combustion. The wintertime increase in gas- and particle-phase PAH concentrations observed at this sparsely populated sub-Arctic site may be generally reflective of the increased residential heating requirements in the northern hemisphere occurring during this time. In summary, both summer and wintertime variations in PAH concentrations and gas/particle partitioning are believed to be source-dependent. The source characteristics encountered in this sample set were explored further in Chapter 3 using receptor models.

Chapter 3

Receptor modeling using positive matrix factorization (PMF) was applied to a PAH data set from the same sparsely populated sub-Arctic environment in the Yukon Territory, Canada between August 2007 and December 2008 discussed above to afford four factors. These factors were identified as coal combustion emissions, particle-phase wood combustion emissions, gas-phase wood combustion emissions, and unburned petroleum/petrogenic emissions. Potential source contribution functions (PSCF) for 1° x 1° grid cells (latitude-longitude coordinates) for the entire northern hemisphere were computed using the source contributions of individual PAH and the four PMF factors. Multi-height, ten-day back trajectories were used to assess the potential for long-range transport and sources of these PAH compounds. Mapping the computed PSCF values for the four PMF factors revealed different source regions in the northern hemisphere for each PMF factor. Atmospheric transport of PAH occurred from both relatively short and long distances with both continental (North American) and trans-oceanic (Asian) sources contributing significantly to the total PAH. This study provided evidence of the transport of coal and wood combustion emissions from eastern Asia to northern Canada (and by extension to the Canadian Arctic) primarily during cooler months. This study demonstrated for the first time that the combined PMF-PSCF methodology can be used to identify geographically-disperse PAH source contributors on a hemisphericalal scale.

Chapter 4

A variety of polycyclic aromatic hydrocarbon (PAH) diagnostic ratios were examined as source apportionment tools in the analysis of a PAH data set associated with atmospheric particulate matter collected in an urban-industrial environment. Seventy-six PM_{10} samples were collected concurrently at 4 sampling sites over a one-month period in Hamilton, Ontario, Canada, a city of 500,000 people that is home to two integrated steel companies, associated industries and a network of roadways and major highways by Dr. L. M. Allan in 1995 for her PhD project . She extracted all samples and reported concentration data for 32 polycyclic aromatic compounds including sulphur-containing aromatics. For this work and the one following it, the samples were re-processed for 16 alkylated phenanthrenes and total concentration levels showed a thousand-fold range (0.23-172 ng m^{-3}). Of all PAH diagnostic ratios examined, the two most useful were the anthracene/(anthracene+phenanthrene) and benz[*a*]anthracene/(benz[*a*]anthracene+chrysene/triphenylene). These afforded the best discrimination of samples that had significant industrial impacts. This work is the first example of the use of a linear combination of PAH ratios, coupled with total PAH data and well defined local samples to qualitatively determine the relative impacts of mobile and industrial emissions in an urban-industrial environment. Use of a linear combination of PAH ratios made it possible to categorize 95% of the data as 'upwind' or 'downwind' of the industrial sector. This work showed how diagnostic PAH ratios can be safely used for source apportionment. It is important to determine PAH ratio threshold values based on data from well defined local samples rather than relying on literature values alone.

Chapter 5

Principal component analysis (PCA) and positive matrix factorization (PMF) analysis were applied to a polycyclic aromatic hydrocarbon (PAH) data set from extracts of 75 PM_{10} air particulate samples collected concurrently at 4 sampling sites proximate to the urban-industrial area in Hamilton, Ontario, Canada in 1995 by Dr. L. M. Allan (see more details above). These FA methods were followed by multilinear regression to identify and quantify PAH source contributions, and ultimately to reveal spatial and temporal trends that could not be observed by the use of diagnostic PAH ratios. The correlations between predicted and observed total PAH levels were excellent in both models (R^2 >0.98). The PCA afforded large negative contributions in a number of samples, so further analysis was abandoned. The PMF analysis showed 3 factors which were identified as gasoline emissions, diesel emissions and coke oven emissions. Contributions of gasoline emissions and diesel emissions factors were surprisingly similar at all 4 sites indicative of a background of vehicle emissions across the city. The PMF coke oven emissions factor showed the greatest variability in total loadings, consistent with the large PAH emissions from the steel industries and the large influence of wind direction on PAH concentrations. The highest coke oven contributions were observed at sites closest to the industrial area on days when these sites were downwind of the industries. The PMF coke oven impacts factor improved the apportionment value of two commonly used PAH diagnostic ratios when the ratios were combined into a single ratio as discussed in Chapter 5. This integrated approach made it possible to categorize >90% of the samples based on the wind direction of the impacting source.

Chapter 6

Hamilton Harbour sediments are known to contain relatively high levels of polycyclic aromatic compounds and heavy metals presumably due to emissions from industrial and mobile sources. Principal component analysis (PCA) with multiple linear regression analysis (MLR) and positive matrix factorization (PMF) were applied to PAH and sulfur heterocycles pollution data from a total of 26 suspended sediment samples collected over a 5-year period in Hamilton Harbour, Ontario, Canada and its surrounding creeks to determine the profiles and relative contributions of pollution sources to the harbour. Both methods identified four factors and gave excellent correlation coefficients between predicted and measured levels of 25 aromatic compounds; both methods predicted similar contributions from coal tar/coal combustion sources to the harbour (19% and 26%, respectively). One PCA factor was identified as contributions from vehicular emissions (61%); PMF was able to differentiate vehicular emissions into two factors, one attributed to gasoline emissions sources (28%) and the other to diesel emissions sources (24%). Overall, PMF afforded better source identification than PCA with MLR. This work constitutes one of the few examples of the application of PMF to the source apportionment of sediments; the addition of sulfur heterocycles to the analyte list greatly aided in the source identification process.

7.3. Outcomes and Suggested Future Directions

The work in this thesis has focused exclusively on polycyclic aromatic hydrocarbons and related compounds. Many other classes of contaminants of organic and inorganic compositions are known. In principle, the model integrations discussed in this work should be applicable to those other contaminant classes. Indeed, the application and integration of receptor modeling techniques to compound classes which are less susceptible to fateful processes such as atmospheric transformation and degradation reactions should yield assessments of high quality. As an example, data sets of halogenated contaminants which are known to be capable of undergoing long-range transport (over hemispherical and global scales) to pristine regions can be factorized to determine their underlying sources and source contributions can be spatially represented using a suitable mapping model.

The PMF-PSCF methodology is a useful integration of two receptor models which was first deployed on a hemispherical scale for organic pollutants in Chapter 3 and should be applicable globally for source apportionment and spatial mapping of pollutant source regions. It must however be noted that in studies concerned with the assessment of long-range transport of contaminants, the PMF-PSCF methodology is critically limited by the accuracy of the chemical and meteorological inputs. Chemical inaccuracies affect the PMF portion of the methodology; it is often very difficult and sometimes impossible to estimate the true uncertainty associated with each concentration data point. This critical step which is one of the most important factors that differentiate PMF from common factor analysis can gravely affect the outcomes of the factorization. Inconsistencies in the meteorological data which are in the form of air mass back trajectories will affect the PSCF computations. After a certain length of time, back trajectories are rarely useful due to grave limitations in positional accuracy typically after three days. Indeed, simple trajectory models can yield errors in the range of thousands of kilometers if they regress in time for five days or more. While more complex (and presumably more reliable) trajectory models have been designed and are now deployed (e.g., the Hybrid Single-Particle Lagrangian Integrated Trajectory HYSPLIT model) regularly, there is still no clear consensus on their absolute uncertainties. Thus, identified sources and potential regions may need to be independently verified if possible. Satellite images and other forms of anecdotal evidence are extremely useful in corroborating results from the receptor modeling integration discussed herein and are strongly encouraged. Indeed, for combustion factors with large spatial ranges such as wildfires, episodic emissions may be detected by receptor modeling and verified by satellite images as discussed in Chapter 3. This is a clear advantage over the use and absolute reliance on dispersion models which use global emission inventories since these inventories rarely account for instantaneous/episodic events.

In conclusion, the use of receptor modeling for source apportionment of contaminants from multivariate datasets aids the environmental analyst to grasp the true complexity of source of pollutions and their contributions in the studied environmental compartment.

Appendix I.

Methods

Method Used for the Cleanup of Little Fox Lake (IPY) Sample Extracts for Organic Contaminants including OC, cPCB, PBDE and PAH at NLET

1. Reagents

2,2,4- Trimethylpentane (Iso-octane)
Dichloromethane (DCM)
Hexane

2. Glassware

10 or 15 mL disposable glass centrifuge tubes (calibrated to 1 mL)
Pasteur pipettes
Glass columns (320 mm x 1.1 mm) with overhead reservoirs
150 mL, 250 mL flat bottom flasks
Hamilton Syringes – 1 mL, 100 µL, 250 µL

3: Preparation of the Raw Sample Extracts

Raw extracts were received in brown vials.

3.1. The samples were transferred into a previously calibrated centrifuge tube (calibrated to 1 mL mark) suitably labeled with 2 x 2 mL rinses of hexane

3.2. The transferred samples were capped and shaken vigorously on a vortex mixer

3.3. The samples were then blown down to 1 mL using the Turbovap (refer to Protocol OI-02). Just prior to the silica gel clean-up, the following NLET surrogate standards were added to the blown down sample extracts to monitor clean-up recovery:
- 100 µL of CB-OC Lab Surrogate
- 100 µL of PCB 30+204 Lab Surrogate
- 250 µL of PBDE Lab surrogate
- 100 µL of PAH Lab Surrogate

3.4. The spiked extracts were then capped and mixed on a vortex mixer

4: Preparation of Quality Control (QC) samples

QC samples include the following:

4.1. Hexane method blank (called OC BLK or MB): 1 mL Hexane was dispensed into a clean disposable calibrated centrifuge tube. The hexane was then spiked with the same laboratory surrogates used for the raw sample extracts (see Section 3)

4.2. Reference sample and duplicate (called OC1SPK or RF1 and RD1) were made by adding the following into a disposable calibrated centrifuge tube:
- 100μL of OC CAL 4
- 100 μl of CPCB CAL STD 4
- 250μL of PBDE CAL 4
- 100μL of PAH CAL STD 5

The contents are then made up to the 1 mL mark with iso-octane.

5: Preparation of Adsorbents

5.1. Silica Gel: For each new batch of silica gel, preparation included adding a large quantity of silica gel to a ceramic dish which was then placed in a muffle furnace at 160°C overnight. The silica gel was then transferred to amber glass jars and stored in a dessicator the next day. (For each new lot number of the silica gel used, a verification check was performed as described in NLET Protocol OC-14).

5.2. Sodium Sulphate: added to a ceramic dish and this was placed in a muffle furnace at 600°C overnight (Refer to NLET Protocol OC-07 for more details on preparing the sodium sulphate).

6: Clean-up and Vialling for GC/MS:

6.1. For OC, PCB, PAH and PBDE in Raw Sample Extracts: When the 100% activated silica gel was needed, an appropriate quantity of the cooled silica gel (Section 5) was placed in a suitable beaker and heated in an oven to 130°C overnight. The next day, the beaker was retrieved and covered and left to cool in the desiccator before use.

A 50 mL aliquot of hexane was added to each fritted silica gel column with the stopcock shut. The columns were wet-packed with sodium sulphate to a height of 1 cm, followed by 8 grams of the 100% activated silica gel. The columns were gently tapped and hexane rinses were used when necessary during the wet-packing process. Finally add some 1 cm of sodium sulphate was added above the settled silica gel. A waste beaker was then placed under the column and the hexane was drained till the level was just above the silica gel. This waste hexane was then discarded into the appropriate waste container.

A clean 150 mL (refer to NLET Protocol 01-01 for cleaning procedure) and suitably labeled flat bottom flask was then placed below the column to collect the subsequent eluate.

The spiked sample extracts were then added to the columns which were drained to the top of the silica gel. Two 2 mL hexane rinses of the sample tube followed sequentially and these rinses were transferred to the column. The column was then drained to the top of the silica gel.

Elution of the column with 70-90 mL hexane (the silica gel verification check; NLET Protocol OC-14 determines the exact amount required). This first eluate is Fraction A.

The 150 mL flask containing Fraction A at the bottom of the column was then replaced by another clean 250mL labeled flat bottom flask. The column was then eluted with 130 mL 1:1 Hexane/DCM (v/v). This is Fraction B.

A 10 mL iso-octane aliquot was then added to each of the fractions in the two separate flasks. Solvent volume reduction on each faction was then carried out on a Rotovap till the final volume was approximately 2-3 mL. Each reduced fraction was then transferred to a separate disposable centrifuge tube (pre-calibrated to 1 mL). Each flask was rinsed with 2 x 2 mL iso-octane and the rinses were transferred to the appropriate centrifuge tubes. The tubes were then capped and the contents mixed on a vortex mixer. Further solvent volume reduction was then initiated on each fraction with a N_2-Turbo-vap till a 1mL final volume was obtained.

For PAH analyses by GC/MS at NLET a 45 μL aliquot of Fraction A and a 45 μL of Fraction B were combined and spiked with a 10 μL aliquot of PAH Internal Standard (prepared by NLET) directly in an insert in a GC vial which was then capped and shaken vigorously on a vortex mixer before submission to the GC technician. GC/MS analyses were performed on 5% phenylmethylsilicone columns (30m x 0.25mm x

0.25μm film; J&W DB-5ms) using a mass selective quadrupole detector (Agilent 6890 GC/ 5973 MSD).

6.2. For OC, PCB, PAH and PBDE in QC Samples: The procedure discussed above was followed for each of the QC samples. A final check was always done to ensure that each typical batch run had vialed samples corresponding to 12 raw sample extracts, 1 method blank sample (MB) and 2 surrogate samples (RFl, and RDl) along with six PAH calibration standard (PAH CAL; see table below) solutions with concentration ranging from ~10 – 2000 pg/μL, three continuing calibration standard (PAH CC; vialed from CAL Level 3) solutions and a control standard solution (PAH CS; vialed from CAL Level 3).

PAH CalStd Calibration Standards							**PAH Lab Surrogate**
	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	
	final conc	final conc	final conc	final conc	final conc	final conc	
	ng/mL	ng/mL	ng/mL	ng/mL	ng/mL	ng/mL	
Indene	10.1	50.3	101	503	1005	2010	
Tetrahydronaphthalene	10.1	50.3	101	503	1005	2010	
Naphthalene-D8	21.4	107	214	1070	2140	4281	1070
Naphthalene	10.0	50.0	100	500	1000	2000	
2-methylnaphthalene	10.1	50.3	101	503	1005	2010	
1-methylnaphthalene	10.1	50.3	101	503	1005	2010	
2-chloronaphthalene	10.0	50.0	100	500	1000	2000	
Acenaphthylene	10.0	50.0	100	500	1000	2000	
Acenaphthene	10.0	50.0	100	500	1000	2000	
Fluorene-D10	19.0	95.1	190	951	1901	3802	951
Fluorene	10.0	50.0	100	500	1000	2000	
Dibenzothiophene	15.0	75.0	150	750	1500	3000	
Phenatherene	10.0	50.0	100	500	1000	2000	
Anthracene	10.1	50.3	101	503	1005	2010	
Fluoranthene	5.00	25.0	50.0	250	500	1000	
Pyrene-D10	10.4	52.1	104	521	1043	2085	521
Pyrene	5.00	25.0	50.0	250	500	1000	
Retene	20.0	100	200	1000	2000	4000	
Benzo(a)anthracene	10.0	50.0	100	500	1000	2000	
Chrysene	5.00	25.0	50.0	250	500	1000	
Benzo(b)fluoranthene	10.1	50.3	101	503	1005	2010	
Benzo(k)fluoranthene	10.0	50.0	100	500	1000	2000	
Benzo(a)pyrene	10.0	50.0	100	500	1000	2000	
Benzo(e)pyrene	10.0	50.0	100	500	1000	2000	
Perylene-D12	29.4	147	294	1468	2936	5872	1468
Perylene	15.0	75.0	150	750	1500	3000	
Indeno(123-c,d)pyrene	15.0	75.0	150	750	1500	3000	
Dibenzo(ah)anthracene	15.0	75.0	150	750	1500	3000	
Benzo(g,h,i)perylene	15.1	75.3	151	753	1505	3010	
Solvent	Iso-octane	Iso-octane	Iso-octane	Iso-octane	Iso-octane	Iso-octane	Iso-octane
Expiry Date	30-Apr-11	30-Apr-11	30-Apr-11	30-Apr-11	30-Apr-11	30-Apr-11	21-Apr-11
Code	09std008	09std008	09std008	09std008	09std008	09std008	09std001
spiking amount 100 µL/FV 1 mL					RF & RD		MB, Samples

All other NLET internal protocols used for this project are given with their full numbers and headings below.
The National Laboratory for Environmental Testing

NUMBER: PROTOCOL OI-01

Title: PROTOCOL FOR CLEANING LABORATORY GLASSWARE USED BY THE ORGANIC ANALYSIS LABORATORY (OAL)

Version Number: **1.2** **Date Revised:** **May 2006**

Background Information:

The National Laboratory for Environmental Testing (NLET) insists that good laboratory practices are followed by all analysts. Contamination of glassware can result in undesirable analytical results. In analysis dealing with residues in the parts per billion range, the preparation of scrupulously clean glassware is mandatory. Failure to do so can lead to many problems in the interpretation of the final measurements. These problems are due to the presence of extraneous peaks resulting from contamination.

Glassware cleanliness is monitored by method blank analysis.

Safety Note:
Safety glasses and rubber gloves must always be worn. Solvents should be handled in a fume hood or vented canopy.

Scope: The protocol is valid for most routinely used laboratory glassware. This includes labware used trace organic analyses in water, sediment and biota matrices. The protocol is not valid for trace analysis work, for labware used with highly contaminated samples.

Purpose:
The protocol describes the procedure for cleaning glassware used by the OAL in the trace analysis of organic contaminants.

Procedure:

Trace Analytical Glassware

1. Soak glassware in hot water containing laboratory grade detergent

(Biodegradable Sparkleen or Micro Liquid Laboratory cleaner). For separatory funnels or glassware that has a stopcock, remove the stopcock before soaking.

2. Depending on the glassware, choose an appropriate scrub brush. Gently scrub the glassware, BUT do not scratch any surfaces.

3. Rinse five times under hot running water to remove soap, drain and empty between rinses. Place the washed glassware on a clean surface until you are ready to solvent rinse. Note: Teflon separatory funnels should be solvent rinsed immediately to prevent water spot formation.
4. The first solvent rinse is done with acetone, reagent grade. Agitate and circulate the acetone within the glassware to disperse any water and remove possible contaminating film. Empty the rinse acetone into a waste beaker for later disposal.

5. Repeat step 4.

6. Follow the same procedure but now rinse with petroleum ether so that it contacts the entire inside surface of the container. Dump waste into the same beaker for proper disposal.

7. Repeat step 6.

8. Place glassware in an oven set to 200°C, and heat overnight. Do not heat Teflon labware.

9. Store cooled glassware in an appropriate dust free area.

10. Before using this glassware for analysis, give it a preliminary rinse with the first solvent that will be used in the first step of the procedure.

Non-Trace Glassware

Glassware that is discoloured or has knowingly been used to process heavily contaminated samples are cleaned as follows. ***Do not mix trace and non-trace glassware.***

1. Rinse the contaminated glassware first with acetone.

2. Soak overnight in concentrated Extran solution.
3. Rinse glassware with tap water then acetone, reagent grade.
4. Follow all of the steps listed above for lightly contaminated glassware from 1-9.

Graduated Glassware

Note: **DO NOT HEAT** graduated glassware. Graduated tubes, cylinders and volumetric flasks will no longer be accurate if heated. Allow these to dry at room temperature in a fume hood or canopy. Graduated centrifuge tubes may be heated but will require calibration prior to use.

SIGNATURE AND DATE AUTHORISED:

Appendix II.

Mathematical Principles of the Receptor Models Used

1. Principal Component Analysis – Multilinear Regression (PCA-MLR)

The PCA model in this context was run thus:

For a data matrix X_{nxm}

where n=number of rows indicating observations/samples
m = number of columns indicating species/elements

First, the x_{ij} matrix elements are standardized such that a new z matrix is produced in which

$$z_{ij} = (x_{ij} - \bar{x}_j)/\sigma_j \qquad (1)$$

where z_{ij} is the standard deviate of element x_{ij}

x_{ij} is the concentration value of the jth element in the ith sample

$\bar{x}_j$ is the mean concentration of the jth element over all observations

σ_j is the standard deviation of the concentrations of the jth element

The PCA is run such that

$$Z_{ij} = \sum_{k=1}^{p} P_{ik} W_{kj} \qquad (2)$$

where; k = 1…., p the pollutional source components
i = 1…., n the number of observations/samples
j = 1…., m the number of species/elements
W_{kj} is the value of the component in the coefficient matrix (elements in transpose of loadings matrix)
P_{ik} is the kth component's value for observation/sample i (scores)

Thurston and Spengler (1985) showed that inversion of equation 2 yields

$$P_{ik} = \sum_{j=1}^{m} Z_{ij} \frac{W_{jk}}{\lambda_k} \qquad (3)$$

λ_k is an eigenvalue associated with the principal component P_k and it is derived from eigenvalue analysis of R_{jxj} the correlation matrix, i.e., find an eigenvector matrix U_{jxj} such that R_{jxj} is diagonalized

$$U_{jxj}^{-1} R_{jxj} U_{jxj} = \Lambda_{jxj} \qquad (4)$$

since PCA assumes that components are not correlated, the eigenvector matrix must satisfy the square orthonormal rule i.e.,

$$U_{jxj}^{-1} U_{jxj} = U_{jxj} U_{jxj}^{-1} = I \qquad (5)$$

for orthogonality of the components

The diagonal elements of Λ_{jxj} are the eigenvalues which are ordered in descending magnitude since each preceding eigenvector is chosen such that its eigenvalue from the diagonalization of the correlation matrix explains more variance in the data set than the succeeding one.

The elements in each column of the U_{jxj} matrix (i.e., eigenvectors) by definition are equal to $\dfrac{W_{jk}}{\lambda_k}$

Substituting the eigenvectors into equation 3 gives P_{ik} (scores)

A key feature of PCA is data dimension reduction. This is possible because the eigenvalues are ordered in descending magnitude, thus the number of pollutional components p is usually less than the number of species m, i.e., $p<m$

Multilinear regression MLR was done such that the scores P_{ik} were regressed against the standard deviate elements of the y column vector i.e.,

$$Z_{i_y} = \sum_{k=1}^{p} Z_{ik} = \sum_{k=1}^{p} P_{ik} B_k \qquad (6)$$

where B_k are the MLR coefficients for each column k of P_{ik} over all i.

To determine the individual contribution (I) of k to y, we re-arrange equation 1 for y

Thus, from equation 1; $y = \bar{y} + Z_{i_y}\,\sigma_y$ (7)

The fraction k contributes to $\bar{y} = B_k/\Sigma B_k$ (8)

The individual contribution (I) of k to y is defined as follows;

$I_k = (B_k/\Sigma B_k)\, y$

multiplying both sides of equation 7 by $B_k/\Sigma B_k$ gives

$I_k = (B_k/\Sigma B_k)\, y = (B_k/\Sigma B_k)\, \bar{y} + Z_{i_y}\, \sigma_y\, B_k/\Sigma B_k$ (9)

But $(B_k/\Sigma B_k)\; Z_{i_y} = Z_{ik}$ (10)

Equation (6) shows that $Z_{ik} = P_{ik}\, B_k$, therefore,

$Z_{i_y}\, \sigma_y\, B_k/\Sigma B_k = B_k\, P_{ik}\, \sigma_y$. Thus equation (9) reduces to;

$I_k = (B_k/\Sigma B_k)\, \bar{y} + B_k\, \sigma_y\, P_{ik}$, (11)
the individual contribution of source to sample is useful in generating temporal or spatial variations of extracted factors (Larsen and Baker 2003; Sofowote, McCarry and Marvin 2008)

Rotations are done such that a square orthonormal matrix $T_{(j\mathrm{x}j)}$ can transform the P and W matrix as follows -
$P'_{ik} = P_{ik} T_{jj}$ and $W'_{kj} = T_{jj}^{-1} W_{kj}$ (12)
Therefore,
$P'_{ik} W'_{kj} = P_{ik} T_{jj} \times T_{jj}^{-1} W_{kj} = P_{ik} \times 1 \times W_{kj} = P_{ik} W_{kj}$ (13)
Thus, equation (2) can be re-written as;

$$Z_{ij} = \sum_{k=1}^{p} P_{ik} W_{kj} = \sum_{k=1}^{p} P'_{ik}\, W'_{kj} \qquad (14)$$

2. Positive Matrix Factorization (PMF)

In factor analysis, a data set X (n x m) consisting data points x_{ij} can be denoted as:
X = GF + E (15)
where G is the factor contribution matrix, F is the factor profile matrix and E (n x m) is the residual error matrix of elements e_{ij} indicating the residual value not accounted for by the modeled data value $\hat{x}_{ij}$, i.e.,

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} \qquad (16)$$

where g_{ik} is the kth source's contribution to sample i and f_{kj} is the jth element's concentration in source k. By specifying an object function Q of the residual matrix which is to be minimized, PMF is better able to weight individual points in a data set based on uncertainties while simultaneously imposing non-negativity constraints on the contribution and profile matrices, i.e., $F \geq 0$, $G \geq 0$, where Q is defined as:

$$Q(E) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{e_{ij}}{s_{ij}} \right]^2 \quad (17)$$

where s_{ij} is the uncertainty in the x_{ij} measurement and is related to the method detection limit e_j for the variable in question. The uncertainties in x_{ij} are preferably observation-based, otherwise, an algorithm can be applied in most PMF software packages for specifying the uncertainties based on the method detection limit and the uncertainty in determination of variables. For most variables, the uncertainty in their determination can be estimated as the standard deviation in the analytical determination of deuterated surrogates' recoveries.

A multilinear regression analysis typically follows such that

$$y_i = \sum_{k=1}^{p} B_k g_{ik} \quad (18)$$

Temporal or spatial variations of extracted factors are directly observable by graphing the observation/sample against the elements of the weighted contributions of k i.e., products of $B_k g_{ik}$.

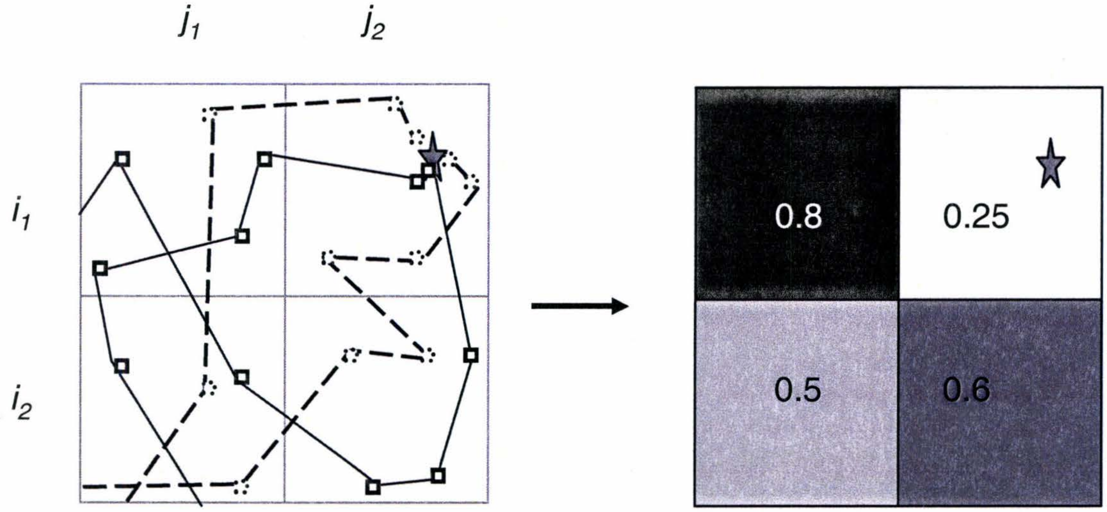
Note: 'Factor' and 'source' in many literature texts including this one are used interchangeably.

3. Potential Source Contribution Function (PSCF)

PSCF is the conditional probability that an air mass trajectory passing over a fixed grid will reach the receptor site and be laden with pollutant levels above a pre-defined threshold. For a given *ij*th cell;

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \quad (19)$$

where n_{ij} is the total number of endpoints of back trajectories that reach the receptor site and m_{ij} for a given pollutant or contaminant species is a subset of n_{ij} whose back trajectory endpoints are associated with concentration values higher than the pre-defined threshold for the pollutant. Detailed discussions on the formulation of PSCF can be found elsewhere (Cheng et al. 1993). The picture in Figure 1 shows how a simple PSCF calculation is done. The use of back trajectories starting at different heights has been encouraged since they not only better approximate the vertical motion of air parcels (Cheng et al. 1993) but they also increase the number of back trajectory endpoints used in the analysis compared to single layer trajectories. The use of back trajectories in ensembles has been noted to give a better estimation of the position of air parcels since random errors will generally be averaged out (Hopke, Zhou and Poirot 2005).



Colour Codes

0.75 -1.0 = Black

0.50- 0.75 = Dark Grey

0.25 - 0.50 = Grey

0.00 - 0.25 = White

Figure AII.1: Simple PSCF computations using hypothetical back trajectories and grid cells. The grey star represents the receptor site, the solid lines are for back trajectories associated with levels of a given contaminant that are higher than a pre-defined threshold and the broken lines are for back trajectories associated with levels of the pollutant below the threshold. Back trajectory segments or endpoints are boxed in squares with solid (higher than threshold) or broken (lower than threshold) lines. Darker colours and higher PSCF values in the image on the right indicate a greater probability of a potential source region of the contaminant to the receptor site.

Weighting factors are usually applied to raw PSCF values to reduce the potential for spurious identification of source regions and 'tailing' effects (Cheng et al. 1993; Cheng and Lin 2001; Zhou, Hopke and Liu 2004). These weighting factors are a function of n_{ij} and are subjective but the effect is to eliminate the possibility that grid cells with relatively small n_{ij} values have a PSCF value approaching 1.

While PSCF computations are helpful in the spatial study of source regions of contaminant species, they cannot replace emissions inventories; as noted by other researchers, they should be viewed as providing qualitative information only (Cheng et al. 1993). Also, PSCF will reveal the locations of potential source regions of contaminant species of interest only if these species are borne by air masses that impact the sampling site after passage through the source region, thus, low PSCF values do not necessarily mean an absence of source regions (Polissar, Hopke and Poirot 2001).

References

Cheng, M.D., Hopke, P.K., Barrie, L., Rippe, A., Olson, M., Landsberger, S., 1993. Qualitative determination of source regions of aerosol in Canadian high Arctic. Environmental Science & Technology 27, 2063-2071.

Cheng, M. and Lin, C., 2001. Receptor modeling for smoke of 1998 biomass burning in Central America. J. Geophys. Res. 106, 22871-22886.

Hopke, P.K., Zhou, L., Poirot, R.L., 2005. Reconciling trajectory ensemble receptor model results with emissions. Environmental Science & Technology 39, 7980-7983.

Larsen, R.K.,3rd and Baker, J.E., 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. Environmental Science & Technology 37, 1873-81.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001. Atmospheric aerosol over vermont: chemical composition and sources. Environmental Science & Technology 35, 4604-4621.

Sofowote, U.M., McCarry, B.E., Marvin, C.H., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. Environmental Science & Technology 42, 6007-6014.

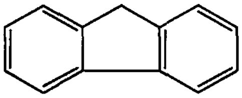
Thurston, G. and Spengler, J., 1985. A quantitiative assessment of source contributions to inhalable particulate matter pollution in Metropolitan Boston. Atmospheric Environment 19, 9-25.

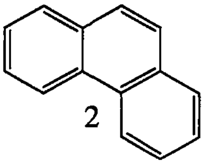
Zhou, L., Hopke, P.K., Liu, W., 2004. Comparison of two trajectory based models for locating particle sources for two rural New York sites. Atmospheric Environment 38, 1955-1963.

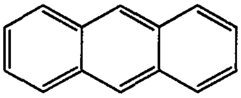
Appendix III.

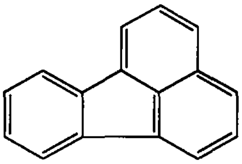
Identities and structures of some PAH and related compounds

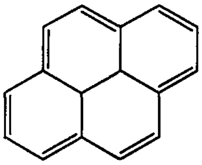
No.	Compound	Molecular mass
1	Fluorene	166
2	Phenanthrene	178
3	Anthracene	178
4	Fluoranthene	202
5	Pyrene	202
6	Cyclopenta[*cd*]pyrene	226
7	Benzo[*c*]phenanthrene	228
8	Benzo[*a*]anthracene	228
9	Chrysene	228
10	Benzo[*b*]fluoranthene	252
11	Benzo[*j*]fluoranthene	252
12	Benzo[*k*]fluoranthene	252
13	Benzo[*e*]pyrene	252
14	Benzo[*a*]pyrene	252
15	Perylene	252
16	Dibenz[*a,h*]anthracene	278
17	Dibenz[*a,c*]anthracene	278
18	Indeno[1,2,3-*cd*]pyrene	276
19	Benzo[*ghi*]perylene	276
20	Coronene	300
21	Benzanthrone	230
22	Benz[*a*]anthracene-7,12-dione	258
23	Dibenzothiophene	184
24	Benzo[*b*]naphtha[2,1-*d*]thiophene	234
25	Benzo[*b*]naphtha[2,3-*d*]thiophene	234

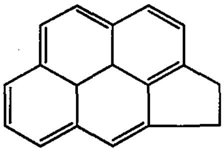

1

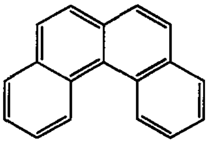

2

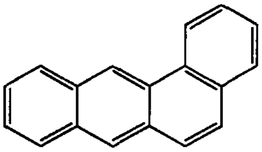

3

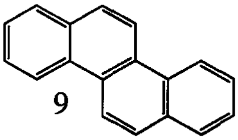

4

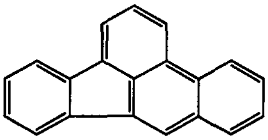

5

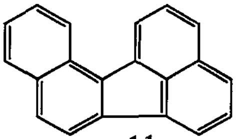

6

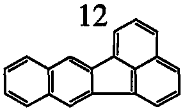

7

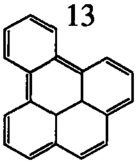

8

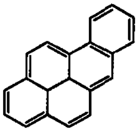

9

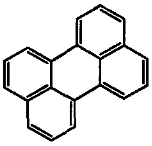

10

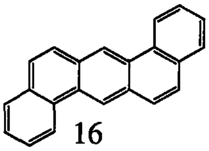

11

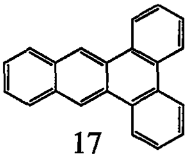

12

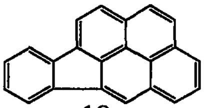

13

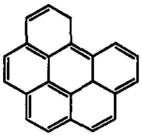

14

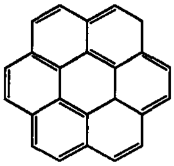

15

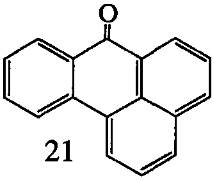

16


17

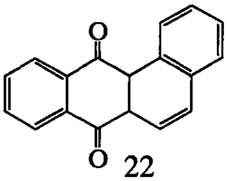

18


19

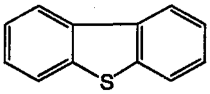

20



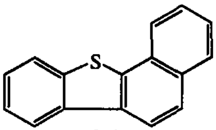
21



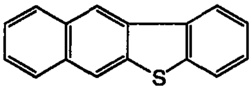
22



23



24



25

12637