DATA HANDLING IN SCIENCE AND TECHNOLOGY

Advisory Editors: B.G.M. Vandeginste and O.M. Kvalheim

Other volumes in this series:

Volume 1 Microprocessor Programming and Applications for Scientists and Engineers by R.R. Smardzewski
Volume 2 Chemometrics: A textbook by D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman
Volume 3 Experimental Design: A Chemometric Approach by S.N. Deming and S.L. Morgan
Volume 4 Advanced Scientific Computing in BASIC with Applications in Chemistry, Biology and Pharmacology by P. Valkó and S. Vajda
Volume 5 PCs for Chemists, edited by J. Zupan
Volume 6 Scientific Computing and Automation (Europe) 1990, *Proceedings of the Scientific Computing and Automation (Europe) Conference, 12–15 June, 1990, Maastricht, The Netherlands.* Edited by E.J. Karjalainen
Volume 7 Receptor modeling for air quality management. Edited by P.K. Hopke

DATA HANDLING IN SCIENCE AND TECHNOLOGY — VOLUME 7

Advisory Editors: B.G.M. Vandeginste and O.M. Kvalheim

Receptor modeling for air quality management

edited by

PHILIP K. HOPKE
Department of Chemistry, Clarkson University, Potsdam, NY 13699-5810,USA

ELSEVIER

Amsterdam — Oxford — New York — Tokyo 1991

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
655, Avenue of the Americas
New York, NY 10010, U.S.A.

ISBN 0-444-88218-9

© Elsevier Science Publishers B.V., 1991

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, Elsevier Science Publishers B.V./ Physical Sciences & Engineering Division, P.O. Box 330, 1000 AH Amsterdam, The Netherlands.

Special regulations for readers in the USA – This publication has been registered with the Copyright Clearance Center Inc. (CCC), Salem, Massachusetts. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the USA. All other copyright questions, including photocopying outside of the USA, should be referred to the publisher.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

This book is printed on acid-free paper.

Printed in The Netherlands

Preface

This book has been written because of the continuing rapid growth and development of the field of receptor modeling. When I finished writing the first integrated book on this field in 1984, I knew it would require revision and updating, but as the time came to do that, it was also clear that I did not want to try to cover the whole subject alone. Thus, a new book using contributions from a number of people who have been active in the development of the field seemed to be a more appropriate approach. Particularly, because of the explicit incorporation of receptor modeling into the guidance documents that have been developed to accompany the new airborne particulate matter air quality standard, I wanted to incorporate the viewpoints of both state/local air quality agencies that have to actually develop the air quality management plans as well as the federal viewpoint where these management strategies are reviewed. Thus, I believe that together we have been able to bring together a useful compendium of material including source and ambient sampling and analysis methods, statistical analysis of data, receptor model study planning, and several expanding areas of application including airborne organic compounds and automated electron microscopy. We hope that such a book will be both a useful introduction to the field and will contain enough depth of information that it will allow those who wish to pursue the subject beyond straightforward applications to do so. It also serves an another benchmark on where receptor modeling is in its development and maturation as a part of the field of air resources management.

I wish to express my deep thanks to the chapter authors for participating in this venture and I particularly wish to thank Dr. Houck for joining the project in its late stages and writing an excellent chapter in a very short time. I also wish to thank Annette Green for all her help in assembling this volume into a coherent whole.

Philip K. Hopke

Potsdam, New York

DEDICATION

During the final stages in the preparation of this volume, Dr. Thomas G. Dzubay succumbed to the cancer he had been fighting for some time. Tom was a fine scientist who had been an integral part of the development of receptor modeling during the past 15 years as well as a good friend and he will be missed. Thus, the rest of us who participated in preparing this volume would like to dedicate it to Dr. Thomas G. Dzubay.

Chapter 1

AN INTRODUCTION TO RECEPTOR MODELING

Philip K. Hopke
Department of Chemistry
Clarkson University
Potsdam, N.Y. 13699-5810

1. INTRODUCTION

The management of ambient air quality is a difficult but important problem. In general, it involves the identification of the sources of materials emitted into the air, the quantitative estimation of the emission rates of the pollutants, the understanding of the transport of the substances from the sources to downwind locations, and the knowledge of the physical and chemical transformation processes that can occur during that transport. All of those elements can then be put together into a mathematical model that can be used to estimate the changes in observable airborne concentrations that might be expected to occur if various actions are taken. Such actions could include the initiation of new sources as new industries are built and begin to function and the imposition of emission controls of existing facilities in order to reduce the pollutant concentrations.

However, the atmosphere is a very complex system, and it is necessary to greatly simplify the descriptions of reality in order to produce a mathematical model capable of being calculated on even the largest and fastest computers. Thus, although significant improvements have been made over the past twenty years in the mathematical modeling of dispersion of pollutants in the atmosphere, there are still many instances when the models are insufficient to permit the full development of effective and efficient air quality management strategies. Thus, it is necessary to have alternative methods available to assist in the identification of sources and the apportionment of the observed pollutant concentrations to those sources. Such methods are called receptor-oriented or receptor models since they are focussed on the behavior of the ambient environment at the point of impact as opposed to the source-oriented dispersion models that focus on the transport, dilution, and transformations that begin at the source and then follow the pollutants to the sampling or receptor site.

2. PRINCIPLE OF MASS CONSERVATION

All of the currently used receptor models are based on the assumption of mass conservation and the use of a mass balance analysis. For example, let us assume that the total airborne particulate lead concentration (ng/m^3) measured at a site can be considered to be the sum of contributions from independent source types such as motor vehicles, incinerators, smelters, etc.

$$Pb_T = Pb_{Auto} + Pb_{Smelter} + Pb_{Refuse} + Pb_{Soil} + \dots \tag{1}$$

However, a motor vehicle burning leaded gasoline emits particles containing materials other than lead. Therefore, the atmospheric concentration of lead from automobiles in ng/m^3, Pb_{Auto}, can be considered to be the product of two cofactors; the gravimetric concentration (ng/mg) of lead in automotive particulate emissions, $a_{Pb,Auto}$, and the mass concentration (mg/m^3) of automotive particles in the atmosphere, f_{Auto}.

$$Pb_{Auto} = a_{Pb,Auto} f_{Auto} \tag{2}$$

The normal approach to obtaining a data set for receptor modeling is to determine a large number of chemical constituents such as elemental concentrations in a number of samples. The mass balance equation can thus be extended to account for all *m* elements in the *n* samples as contributions from the *p* independent sources

$$x_{ij} = \sum_{k=1}^{p} a_{ik} f_{kj} \tag{3}$$

where x_{ij} is the *i*th elemental concentration measured in the *j*th sample (nm/m^3), a_{ik} is the gravimetric concentration of the ith element in material from the *k*th source (ng/mg), and f_{kj} is the airborne mass concentration of material from the kth source contributing to the jth sample (mg/m^3). There are several different approaches to receptor model analysis that have been successfully applied including chemical mass balance (CMB) and multivariate receptor models. Recently, the use of single particle data obtained using computer-controlled scanning electron microscopy has proven to be a valuable tool in providing more information about the airborne particle samples and thereby improve the resolution and precision of the source apportionment. This book covers the air sampling, chemical analysis, and mathematical procedures necessary to perform a receptor modeling study. The following section will provide an introduction to these various receptor models and an introduction to the organization of this volume.

3. OVERVIEW

3.1 Chemical Mass Balance

Receptor models date back to the late 1960's when both mass balance analysis and multivariate statistical methods were first applied to air quality data. The concept of a atmospheric mass balance model was suggested independently by Miller *et al.* (1972) and by Winchester and Nifong (1971). In these initial models, specific elements were associated with particular source types to develop a mass balance for airborne particles. Subsequently, more chemical species than sources were used in a least-squares fit to provide estimates of the mass contributions of the sources (Friedlander, 1973).

There were a number of these early application of the mass balance analysis including Ghent, Belgium (Heindryckx and Dams, 1974), Heidelberg, Germany (Bogen, 1973), and Chicago, Illinois (Gatz, 1975). Several major research efforts have subsequently resulted in substantially better source data. The source emission studies led to much improved resolution of the particle sources in Washington, D.C. (1978; 1982). In the first of these studies, Kowalczyk *et al.* (1978) used a weighted least-squares regression analysis to fit six sources with eight elements for ten ambient samples.

Subsequently, Kowalczyk *et al.* (1982) examined 130 samples using 7 sources with 28 elements included in the fit. They obtained an excellent fit of the ambient concentration data and a quite good understanding of the major sources of airborne particles in the Washington, D.C. area.

Mayrsohn and Crabtree (1976) presented the use of an iterative least-squares approach to apportion six sources of airborne hydrocarbon compounds in the Los Angeles basin. The sources were automotive exhaust, volatilization of gasoline and release of gasoline vapor, commercial natural gas, geological natural gas, and liquified petroleum gas. They performed the least-squares fit to the hydrocarbon compound concentrations using gas chromatography to determine the concentrations of eight compounds. Their ordinary least-squares source reconciliation algorithm recognized that not all sources may contribute to every sample, and, if negative contributions were obtained, a different configuration of sources was employed with certain qualifying assumptions. Each possible configuration with positive coefficients was considered and the one with the lowest standard error was chosen as the optimum solution. On the average, automotive exhaust was the source of almost 50% of observed hydrocarbons. Gasoline and its vapor contributed 30 to 30% by weight and the balance resulted from commercial and geological natural gas. Thus, automobiles and other highway related sources were responsible for the majority of these hydrocarbons.

A similar study utilizing this mass balance approach for resolving hydrocarbon sources has been made by Nelson *et al.* (1983) in Sydney, Australia. Thus, it is possible to identify the impact of emission sources on gaseous as well as particulate pollutants.

In 1979, Watson (1979) and Dunker (1979) independently suggested a mathematical formalism called effective variance weighting that included the uncertainty in the measurement of the source composition profiles as well as the uncertainties in the ambient concentrations. As part of this analysis, a method was also developed to permit the calculation of the uncertainties in the mass contributions. Effective-variance least squares has been incorporated into the standard personal computer software developed by the U.S. EPA for receptor modeling. The most extensive use of effective-variance fitting has been made by Watson and colleagues in their work on data from Portland, OR (Watson, 1979;

Cooper *et al.*, 1984). Since that study, a number of other applications of this approach have been made in a wide variety of locations and extensive libraries of compositional profiles of emission sources have been developed to be used in the mass balance models. These models are described in detail in the chapter by Watson *et al.*

3.2 Multivariate Methods

Actually, the first type of receptor modeling analysis reported in the literature was factor analysis using eigenvector methods that had been developed in the social sciences for interpreting large data sets. Blifford and Meeker (1967) used a principal component analysis with several types of axis rotations to examine particle composition data collected by the National Air Sampling Network (NASN) during 1957-61 in 30 U.S. cities. They were generally not able to extract much interpretable information from their data. Since there are a very wide variety of particle sources among these 30 cities and only 13 elements were measured, it is not surprising that they were not able to provide much specificity to their factors. Prinz and Stratmann (1967) examined both the aromatic hydrocarbon content of the air in 12 West German cities and data from Colucci and Begeman (1965) on the air quality of Detroit using factor analysis methods. In both cases, they found solutions that yielded readily interpretable results.

However, there was not further use of factor analysis until it was reintroduced in the mid-1970's by Hopke *et al.* (1976) and Gaarenstroom *et al.* (1977) in their analyses of particle composition data from Boston, MA and Tucson, AZ, respectively. In the Boston data for 90 samples at a variety of sites, six common factors were identified that were interpreted as soil, sea salt, oil-fired power plants, motor vehicles, refuse incineration and an unknown manganese-selenium source. In the study of Tucson (Gaarenstroom *et al.* 1977), whole filter data were analyzed separately at each site. They find factors that are identified as soil, automotive, several secondary aerosol materials such as $(NH_4)_2SO_4$ and several unknown factors. They also discovered a factor that represented the variation of elemental composition in their aliquots of their neutron activation standard containing Na, Ca, K, Fe, Zn, And Mg. This finding illustrates one of the important uses of factor analysis; screening the data for noisy variables or analytical artifacts. The value of this type of analysis in screening large data sets to identify errors was demonstrated by Roscoe *et al.* (1982). With the use of atomic and nuclear methods to analyze environmental samples for a multitude of elements, very large data sets have been generated. Because of the ease in obtaining these results with computerized systems, the elemental data acquired are not always as thoroughly checked as they should be, leading to some, if not many, bad data points. It is advantageous to have an efficient and effective method to identify problems with a data set before it is used for further studies. Eigenvector analysis described later in this book in the chapter by Henry can provide useful insight into several possible problems that may exist in a data set including incorrect single values and some types of systematic errors.

Gatz (1977) used a principal components analysis of aerosol composition and meteorological data for St. Louis, MO. In addition to the 20 elements analyzed in the filter samples by Particle Induced X-ray Excitation (PIXE), Gatz introduced additional parameters in his analysis including day of the week, mean wind speed, percent of time with the wind from NE, SE, SW, or NW quadrants or variable, ventilation rate, rain amount and duration. At several sites the inclusion of wind data permitted the extraction of additional factors that allowed identification of motor vehicle emissions in the presence of specific point sources of lead such as a secondary copper smelter.

A problem that exists with these forms of factor analysis is that they do not permit quantitative source apportionment of particle mass or of specific elemental concentrations. In an effort to find alternative methods that would provide information on source contributions when only the ambient particulate analytical results are available, various investigators (Thurson and Spengler, 1985; Hopke, 1988; Henry and Kim, 1989) have developed alternative methods based on eigenvector methods. In these analyses, resolution similar to that obtained from a CMB analysis can be obtained. However, a CMB analysis can be made on a single sample if the source data is known while multivariate methods requires a series of samples with varying impacts by the same sources, but does not require *a priori* knowledge of the source characteristics. These methods are also described in the chapter on multivariate methods. These methods provide a useful parallel analysis with CMB to help insure that the profiles used are reasonable representations of the sources contributing to a given set of samples.

3.3 Sampling and Analysis Methods

For both mass balance models and multivariate methods, the analysis is based on the chemical species that are determined in samples of particulate matter. Elemental analysis methods are described in the chapter by Dzubay and Stevens along with a review of ambient sampling methods. Source sampling is review by Houck. However, even with the best of sampling and analysis methods, only a limited number of elements are generally available for the receptor modeling analysis. Thus, in a complex airshed, the resolution and precision of the analysis may be limited by the number of chemical species on which the mathematical analysis is based.

To provide additional information on particle samples in an effort to improve the identification of specific emission sources, automated single particle analysis methods have been developed (Casuccio *et al.*, 1983). These methods provide a large base of data that needs to be analyzed to produce quantitative mass apportionment results. Both the analytical and mathematical methods that have been developed to make automated electron microscopy a practical receptor modeling technique are presented in the chapter by Hopke and Casuccio.

Although a mass balance approach has been previously applied to gaseous hydrocarbons as discussed previously, there has been considerable recent work to extend the sampling, analysis, and mathematical methods to examine the problem of source apportionment of volatile organic compounds (VOCs). With the increased interest in airborne toxic compounds from industry and waste sites, these methods described in the chapter by Scheff and Wadden are likely to become much more widely used in the near term future.

3.4 Air Quality Management Using Receptor Modeling

On the basis of the results of these studies showing the value of receptor modeling to air quality planning, receptor modeling was explicitly approved as an acceptable part of the development of State Implementation Plans for Particulate Matter 10 μm (PM10) (EPA, 1986). Thus, receptor modeling can be an effective part of the development of air quality management strategies. To illustrate how this is done, this volume includes the view of a state air quality manager who has extensively used receptor models in developing the management plans for the state of Oregon. The chapter by Core provides his perspective on the development and use of these tools from the viewpoint of the state that has the most extensive experience in employing them. In the chapter by Pace, the perspective of an air quality manager in the U.S. Environmental Protection Agency toward the use of these methods is presented. Thus, we have the ideas of individuals who are involved in both the preparation and the review of air quality management plans.

4. SUMMARY

Thus, a variety of topics related to the management of air quality through the use of receptor models are presented in this volume. It was planned as a practical guide to persons who may be given the task of implementing receptor modeling as a part of some air quality management problem and it is the intention of all of the chapter authors that this volume provide both the basic information needed to begin doing receptor modeling as well as some insight into some of the problems related to the use of these models. These tools like any others used in solving complex technological problems are not a panacea, but do represent powerful aides in data analysis that can lead to insights as to how an airshed functions and thus, to effective and efficient air quality management strategies.

5. REFERENCES

Blifford, I.H. and G.O. Meeker, A Factor Analysis Model of Large Scale Pollution, *Atmospheric Environ.*, 1:147-157 (1967).

Bogen, J. Trace Elements in Atmospheric Aerosol in the Heidelberg Area Measured by Instrumental Neutron Activation Analysis, *Atmospheric Environ.* 7:1117-1125 (1973).

Casuccio, G.S., P.B. Janocko, R.J. Lee, J.F. Kelly, S. L. Dattner and J. S. Mgebroff, The use of computer controlled scanning electron microscopy in environmental studies, *JAPCA*, 33:937-943 (1983).

Colucci, J.M. and C.R. Begeman, The Automotive Contribution to Air-Borne Polynuclear Aromatic Hydrocarbons in Detroit, *J. Air Pollut. Control Assoc.*, 15:113-122 (1965).

Cooper, J.A., J.G. Watson, and J.J. Huntzicker, The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model, *Atmospheric Environ.*, 18:1347-1355 (1984).

Dunker, A.M., *A Method for Analyzing Data on the Elemental Composition of Aerosols*, General Motors Research Laboratories Report MR-3074 ENV-67, Warren, MI, 1979.

EPA, *PM10 SIP Development Guideline*, Report No. EPA450/2-86-001, U.S. Environmental Protection Agency, Research Triangle Park, NC, 1986.

Friedlander, S.K., Chemical Element Balances and Identification of Air Pollution Sources, *Environ. Sci Technol.* 7:235-240 (1973).

Gaarenstroom, P.D., S.P. Perone, and J.P. Moyers, Application of Pattern Recognition and Factor Analysis for Characterization of Atmospheric Particulate Composition in Southwest Desert Atmosphere, *Environ. Sci. Technol.*, 11:795-800 (1977).

Gatz, D.F., Relative Contributions of Different Sources of Urban Aerosols: Application of a New Estimation Method to Multiple Sites in Chicago, *Atmospheric Environ.*, 9:1-18 (1975).

Gatz, D.F., Identification of Aerosol Sources in the St. Louis Area Using Factor Analysis, *J. Appl. Met.*, 17:600-608 (1978).

Heindryckx, R. and R. Dams, Continental, Marine, and Anthropogenic Contributions to the Inorganic Composition of the Aerosol of an Industrial Zone, *J. Radioanal. Chem.* 19:339-349 (1974).

Henry, R.C. and B.M. Kim, Extension of Self-Modeling Curve Resolution to Mixtures of More Than Three Components. Part 1. Finding the Basic Feasible Region, *Chemometrics and Intelligent Laboratory Systems* 8:205-216 (1989).

Hopke, P.K., E.S. Gladney, G.E. Gordon, W.H. Zoller, and A.G. Jones, The Use of Multivariate Analysis to Identify Sources of Selected Elements in the Boston Urban Aerosol, *Atmospheric Environ.*, 10:1015-1025 (1976).

Hopke, P.K., Target Transformation Factor Analysis as an Aerosol Mass Apportionment Method: A Review and Sensitivity Analysis, *Atmospheric Environ.* 22:1777-1792 (1988).

Kowalczyk, G.S., C.E. Choquette, and G.E. Gordon, Chemical Element Balances and Identification of Air Pollution Sources in Washington, D.C., *Atmospheric Environ.*, 12:1143-1153 (1978).

Kowalczyk, G.S., G.E. Gordon, and S.W. Rheingrover, Identification of Atmospheric Particulate Sources in Washington, D.C. Using Chemical Element Balances, *Environ. Sci. Technol.*, 16:79-90 (1982).

Mayrsohn, H. and J.H. Crabtree, Source Reconciliation of Atmospheric Hydrocarbons, *Atmospheric Environ.*, 10:137-143 (1976).

Miller, M.S., S.K. Friedlander, and G.M. Hidy, A Chemical Element Balance for the Pasadena Aerosol, *J. Colloid Interface Sci.* 39:65-176 (1972).

Nelson, P.F., S.M. Quigley, and M.Y. Smith, Sources of Atmospheric Hydrocarbons in Sydney: A Quantitative Determination Using a Source Reconciliation Technique, *Atmospheric Environ.*, 17:439-449 (1983).

Prinz B. and H. Stratmann, The possible Use of Factor Analysis in Investigating Air Quality, *Staub-Reinhalt Luft*, 28:33-39 (1968).

Roscoe, B.A., P.K. Hopke, S.L. Dattner, and J.M. Jenks, The Use of Principal Components Factor Analysis to Interpret Particulate Compositional Data Sets, *J. Air Pollut. Control Assoc.*, 32:637-642 (1982).

Thurston, G.D. and J.D. Spengler, A Quantitative Assessment of Source Contributions to Inhalable Particulate Matter Pollution in Metropolitan Boston, *Atmospheric Environ.*, 19:9-26 (1985).

Watson, J.G., *Chemical Element Balance Receptor Model Methodology for Assessing the Source of Fine and Total Suspended Particulate Matter in Portland, Oregon*, Ph.D. Thesis, Oregon Graduate Center, Beaverton, OR, 1979.

Winchester J.W. and G.D. Nifong, Water Pollution in Lake Michigan by Trace Elements from Pollution Aerosol Fallout, *Water, Air, Soil Pollut.* 1:50-64 (1971).

Chapter 2

SAMPLING AND ANALYSIS METHODS FOR AMBIENT PM-10 AEROSOL

Thomas G. Dzubay and Robert K. Stevens
Atmospheric Research and Exposure Assessment Laboratory
U. S. Environmental Protection Agency
Research Triangle Park, NC 27711

1. INTRODUCTION

In an early application of receptor models, Friedlander (1973) used data on atmospheric concentrations of Na, Mg, Al, K, Ca, V and Pb to apportion mass concentrations into components related to five different types of emission sources. Improvements in measurement technology enable many more species to be measured, and this can improve the accuracy and resolution of receptor model results. This chapter is intended to provide information for designing a measurement program to obtain ambient aerosol data for use in receptor models. Other chapters in this book show how receptor models are used in air quality management and how species concentrations are used in receptor models.

Primary emphasis in this chapter is given to particles in the PM-10 size range (aerodynamic equivalent diameter less than 10 μm). Characteristics of PM-10 sampling devices and filter media are presented, and the laboratory methods listed in Table 1 are described.

It is shown that some combinations of filter medium, air sampler and analysis method are incompatible. Information presented here will enable one to avoid incompatibility and to devise optimum measurement strategies. Measurement artifacts that can alter ionic species concentrations in particles are described, and an annular denuder system for avoiding such artifacts and for measuring gaseous species is described.

Table 1. Laboratory Methods to Measure Aerosol Composition.

Method	Abbreviation	Species Measured[a]
X-Ray Fluorescence	XRF	Elements with $Z > 12$
XRF with Energy-Dispersive Detector	EDXRF	Elements with $Z > 12$
XRF with Wavelength Dispersive Detector	WDXRF	Elements with $Z > 12$
Particle-Induced X-ray Emission	PIXE	Elements with $Z > 12$
Instrumental Neutron Activation Analysis	INAA	More than 25 Elements
Atomic Absorption	AA	One Element per Lamp
Inductively Coupled Plasma	ICP	More than 12 Elements
Pyrolysis in Helium		Volatilizable Carbon
Pyrolysis in Helium with Oxygen		Total Carbon
Accelerator Mass Spectrometry	AMS	^{13}C and ^{14}C
Ion Chromatography		Ions in Solution
Colorimetry		Ions in Solution

[a] The symbol Z represents atomic number.

2. AIR SAMPLING

The monitoring requirements for determining attainment of the U. S. Environmental Protection Agency (EPA) National Ambient Air Quality Standard for PM-10 are defined so that sampling devices that meet certain performance specifications can be designated as reference methods (Purdue, 1988). By 1987, PM-10 high volume (hi vol) sampling systems made by Wedding and Associates, Inc. (Fort Collins, CO) and Andersen Samplers, Inc. (Atlanta, GA) had been designated as reference methods (Purdue, 1988). In 1989, the dichotomous sampler made by Andersen Samplers, Inc., was designated as a PM-10 reference method (EPA, 1989). Table 2 shows some of the characteristics of these reference methods and indicates that the devices differ in inlet type and flow rate. A feature unique to the dichotomous sampler is the use of a virtual impactor to collect particles in fine (0 to 2.5 μm) and coarse (2.5 to 10 μm aerodynamic diameter) fractions.

Several field studies have been conducted to compare performances of the samplers shown in Table 2 (Rodes *et al.*, 1985; Purdue *et al.*, 1986; Hoffman *et al.*, 1988). John (1988) evaluated data from Phoenix, AZ (Purdue *et al.*, 1986) and found that measured PM-10 mass concentrations can be ranked by sampler as follows:

Wedding Hi Vol < Dichotomous Sampler < Andersen Hi Vol (1)

with deviations from dichotomous sampler results being as large as 30%. John (1988) attributed the higher Andersen values to bounce and reentrainment of particles in the impactor inlet — an effect that diminished when the impactor plate was coated with oil. He attributed the lower Wedding values to blocking of the airflow by buildup of collected particles in the cyclone — an effect that diminished when cyclone surfaces were cleaned daily. Two models of dichotomous samplers gave equivalent PM-10 values (fine and coarse fraction data were summed).

Table 2. Characteristics of PM-10 reference method samplers.

Sampler	Size Ranges (μm)	Inlet Type	Flow Rate (Lm^{-1})	Filter Size
Wedding PM-10	< 10	cyclone	1133	203 x 254
Andersen PM-10	< 10	impactor	1133	203 x 254
Andersen Dichotomous	<2.5 and 2.5-10	impactor	16.7	37 mm diameter

When an oiled Andersen PM-10 hi vol and a daily-cleaned Wedding PM-10 hi vol were compared with a dichotomous sampler, the ranking in Equation (1) still applied, but average deviations were less than 13% (John, 1988; Hoffman *et al.*, 1988). Consequently, oiling and cleaning of hi vol inlets have become a necessary part of the PM-10 measurement process (Purdue, 1988).

The dichotomous sampler's ability to collect particles in the fine and coarse fractions is useful in receptor modeling because such size separation discriminates between particles generated by combustion and mechanical processes. Although it may seem that the flow rate in a dichotomous sampler is too low to collect sufficient sample for analysis, this is generally not the case because many analytical methods do not require large samples, and high-purity Teflon filters work best in a low-flow-rate sampler.

An automated device named "SCISAS" has been designed to collect up to six samples of particles in two size ranges for applications of receptor models to visual air quality in western United States (Rogers *et al.*, 1988). Another sampler is the "Streaker", which collects a time-sequence of 1-mm by 8-mm samples on a strip of Nuclepore filter for element analysis by PIXE (Nelson, 1977); this sampler can be equipped with an impactor to remove particles larger than 10 μm.

3. FILTER MEDIA

Filter media must be compatible with requirements of both the air sampler and analysis method. The characteristics of filter media commonly used in air sampling are shown in Table 3, and their compatibility with various air samplers and analysis methods is described in Table 4.

Because thin Teflon membrane is difficult to handle, it is either mounted on annular polymethylpentene rings and sold as Teflo, or it is mounted on porous Teflon mat and sold as Zefluor (Gelman Sciences, Ann Arbor, MI). Despite the 2-μm pore size, the efficiency is very high for particles as small as 0.03 μm (Liu *et al.*, 1983). Ring-mounted Teflon filters are excellent for XRF because their low mass per unit area and high purity enable good detection limits to be obtained. An exception is Lot 225, which is contaminated with Pb (Cooper, 1988). Teflon

Table 3. Characteristics of several filter media used in aerosol sampling.

	0.3-μm Nuclepore	2 μm Ringed Teflon	2 μm Teflon Zefluor	Quartz Type QMA (Whatman)	Quartz Microquartz (Gelman)
Face Velocity^a	10.6	64.1	86.7		41.6
Mass/Area^b, mg/cm^2	1.0	0.6	10.8	8.4	8.7
E(0.035 μm)^c, %	99.3	99.9	94.6		99.9
E(0.1 μm)^c, %	98.6	>99.9	99.0		99.7
E(0.3 μm)^c, %	>99.9	>99.9	99.9		>99.9

[a] Nuclepore sometimes contains troublesome amounts of Si and Br.

[b] Zefluor is fair for XRF only if sample is collected on the membrane side.

[c] Quartz is good for gravimetric analysis when RH is properly controlled.

filters are not very good for PIXE because of a γ-ray background due to interactions between the proton beam and the fluorine in the Teflon.

The Teflon mat in Zefluor can cause problems for XRF and wet chemical analysis. The mat adds to the total mass per unit area, which increases X-ray blank levels and raises the minimum detection limits by about a factor of 3. Also, if particles are sampled on the mat side, they are collected deep within its structure, which attenuates soft X rays from light elements such as Al, Si, S and K and causes a large, difficult to determine bias. The membrane side of Zefluor is not easily discerned at first, but personnel can locate it with training and a magnifying glass. Postsampling validation by examination with a light microscope (100 to 400X) is needed to verify that the deposits are on the proper side. Another reason to require that particles be collected on the membrane side is that some wet extraction procedures do not efficiently extract particles imbedded in the Teflon mat.

Zefluor filters are available that will fit in hi vol samplers, and Table 3 indicates that they have low flow resistance when clean. However, the Teflon

Table 4. Compatibility of Filter Media with Sampling and Analysis Methods.

	0.3 μm Nuclepore	Ringed Teflon	Zefluor	Quartz
Sampling				
PM-10 Hi Vol	poor	none	fair	excellent
Dichotomous	excellent	excellent	excellent	excellent
Resistance to Clogging	poor	fair	fair	excellent
Analysis				
Gravimetric	good	good	good	fair-good[c]
XRF	good[a]	excellent	poor-fair[b]	poor
PIXE	excellent	fair	poor	poor
INAA	good	excellent	excellent	difficult
ICP	good	excellent	excellent	good
Pyrolysis	none	none	none	fair-good[d]

[a] Nuclepore sometimes contains troublesome amounts of Si and Br.

[b] Zefluor is fair for XRF only if sample is collected on the membrane side.

[c] Quartz is good for gravimetric analysis when RH is properly controlled.

[d] Because quartz filters adsorb carbon vapor, quartz filters are good for particulate carbon only if an appropriate correction is made for adsorbed carbon.

membrane in Zefluor clogs so quickly that after a few hours of sampling, the pump in a typical hi vol cannot maintain a steady flow rate. Clogging is less of a problem with Teflon membrane in a dichotomous sampler, because its pump can deliver a low but constant flow rate for fine particle loadings up to 150 $\mu g/cm^2$ on a dry weight basis. However, higher fine particle mass concentrations in humid air can cause the flow-rate controller in a dichotomous sampler to fail to maintain constant flow. Quartz filters are much less prone to clogging than Teflon and are therefore best to use in hi vol samplers. However, quartz filters have high concentrations of Si, Ca, Fe and other elements and are not good for XRF and INAA. A further complication for XRF is that a quartz filter collects particles within its volume, which causes soft X rays from light elements to be attenuated by amounts that are difficult to determine. Quartz filters are best analyzed for trace elements by extracting the sample in strong acids. However, an extraction procedure designed not to dissolve the quartz filter may not be efficient for elements in silicate minerals (Al, Si, Ti and Fe).

Whatman 41 is a type of cellulose filter that has high loading capacity, works well in a hi vol sampler, and has very low blank values (Dams *et al.*, 1972). Laboratory tests show that these filters have efficiencies of 43, 52 and 92% for 0.03-, 0.1- and 0.3-µm-diameter particles, respectively (Liu *et al.*, 1983). Although such low efficiencies exclude this filter for monitoring to meet the U. S. PM-10 standard, Lowenthal and Rahn (1987) demonstrate that the efficiency of Whatman 41 filters exceeds 90% for 24-hour atmospheric samples.

4. MASS AND ELEMENTAL ANALYSIS METHODS

4.1 Sample Handling and Transport

When quartz filters are folded, they splinter, and their tare masses change. To maintain stable tare masses and avoid particle loss, quartz filters should be transported from sampling site to laboratory in cassettes oriented deposit-side-up. After mass measurements are completed, quartz filters may be folded to avoid particle loss during further shipment or storage.

Ring-mounted Teflon filters are held in circular cassettes in a dichotomous sampler and may be stored in those cassettes in petri dishes before and after sampling. To avoid particle loss, samples should be deposit-side-up during transport from field to laboratory. When filters are shipped by mail, one has little control over orientation, but losses can be greatly reduced by applying a thin layer of mineral oil (25 $\mu g/cm^2$) to coarse fraction filters in the laboratory before sampling (Dzubay and Barber, 1983). No oil is needed for fine particles. Accurate gravimetric analysis is possible if the mineral oil is applied before tare mass is determined. Small amounts of mineral oil do not interfere with XRF analyses.

4.2 Gravimetric Analysis

Mass is determined by weighing each filter before and after sampling. Filters are stored in an atmosphere of 20 to 45% relative humidity (RH) for 24 hours before each mass measurement. Because the mass of a quartz filter is sensitive to relative humidity, RH must be constant to ±5% between tare and final mass determinations. Ring-mounted Teflon filters require a microbalance with 1-μg precision. Electrostatic charge, which accumulates on Teflon filters, must be removed by exposing each side of a filter for a few seconds to alpha particle radiation from a ^{210}Po source (Model 2U500 Staticmaster, Nuclear Products Co., El Monte, CA). If a wire balance pan is used a ^{210}Po source can be mounted below the pan so that charge neutralization occurs while each filter is weighed. Because the half-life of ^{210}Po is 138 days, the source should be replaced at least every 12 months.

4.3 X-Ray Fluorescence Spectrometers and Sample Requirements

In XRF analysis a beam of X rays irradiates the sample and causes elements to emit characteristic X rays, which are detected by a solid state detector or a crystal spectrometer. The X-ray source may be direct emissions from an X-ray tube, or it may be the nearly monochromatic X rays from a secondary target irradiated by an X-ray tube.

Figure 1 shows an X-ray spectrum for fine particles on a Teflon filter irradiated by a Mo secondary target. The detector is a silicon diode (1-cm diam, 0.3-cm thick) containing small amounts of lithium. The detector and related electronic components constitute an EDXRF spectrometer, which enables several elements to be determined simultaneously (Jaklevic *et al.*, 1981). Optimum detection limits for a wide range of elements are obtained when several different secondary targets and X-ray tube voltages are used to analyze each sample. In WDXRF analysis, a crystal spectrometer is used to measure elements sequentially (Wagman *et al.*, 1977). In PIXE analysis, a beam of protons or alpha particles

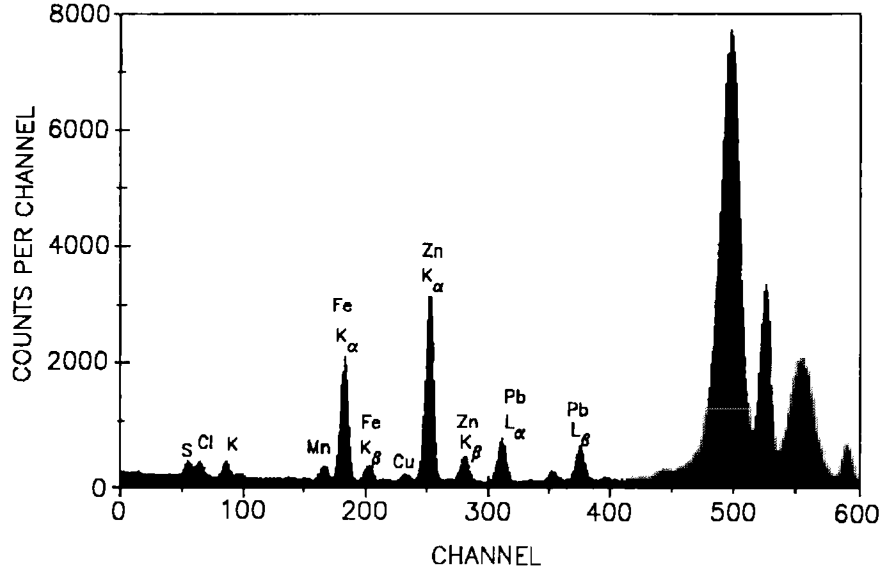

Figure 1. EDXRF spectrum of fine particle sample using excitation by a Mo secondary target. X rays above channel 450 are due to the Mo target.

excites the sample (Nelson, 1977).

Table 5 compares element concentrations measured in Camden, NJ, and Great Smoky Mountains with 2σ measurement precision for EDXRF, WDXRF, and PIXE. The units for 2σ in Table 5 are ng/cm^2 and are equivalent to ng\m^3 for a 7-hour sample collected at 16 L/min on a 6.5-cm^2 area in a dichotomous sampler. The 2σ data pertain to analysis of low element concentrations on a blank filter and include propagation of error due to statistical fluctuations in the X-ray background of element measured and interfering elements. Much of the background is due to X-ray scattering by the filter.

In EDXRF analysis Cd, Sn, Sb and Ba are best determined by their K_α X rays, which are free of interferences, but the lower energy L_α and L_β X rays from those elements interfere with the K_α and K_β X rays measured for K, Ca, Ti and V.

Table 5. Typical atmospheric concentrations and 2σ measurement precision.

	Concentrations, ng/m^3				2σ measurement precision[a] in ng/cm^2			
	Camden, NJ[b]		Elkmont, TN[c]		EDXRF[d]	EDXRF[e]	WDXRF[f]	PIXE[g]
	Fine	Coarse	Fine	Coarse	DRI	EPA	EPA	EAC
Al	53	550	20	195	13	101	4	24
Si	103	1610	38	580	6	36	7	20
S	4200	230	3700	200	3	13	5	15
Cl	3	69	<10	7	7	10	26	14
K	101	151	40	110	4 (3)	8 (5)	13	(10)
Ca	40	360	16	190	4 (3)	10 (5)	2	(9)
Ti	15	65	<2	<2	18 (2)	14 (3)	2	(7)
V	13	7	<2	<2	7 (1)	5 (3)	8	(6)
Cr	2	3			2	2	26	2
Mn	6	11			1	2	23	2
Fe	91	490	28	120	1	3	9	1
Ni	11	4	1	2	1	2	9	1
Zn	82	30	2	<1	1	1	11	1
As	1	0	2	<1	2	2	13	1
Se	2	0	2	<1	1	1	16	1
Br	29	15	18	5	1	1	38	3
Cd	2	0			9	9	2	46
Sn	5	1			14	16	11	75
Sb	79	181			16	14	5	99
Ba	<30	<30			59	33	7	(24)
Pb	249	54	97	14	3	3	94	4

[a] Filter medium is 2-μm pore size ringed Teflon for XRF and Nuclepore for PIXE. Data in parentheses represent assumption of no interferences between K_α X rays from K, Ca, Ti and V and L_α and L_β X rays from Cd, Sn, Sb and Ba.

[b] One-month average for summer 1982 (Dzubay *et al.*, 1988).

[c] Data for September 1978 in Great Smoky Mountains (Stevens *et al.*, 1980).

[d] Desert Research Institute (DRI) excitation conditions (and X-ray tube voltage) are cellulose-filtered direct excitation (8 kV) for Al - S; Ti secondary target (30 kV) for Cl - Ca; Ge secondary target (30 kV) for Ti - Zn; Rh-filtered direct excitation (35 kV) for As - Mo and Pb; Mo-filtered direct excitation (60 kV) for Pd - La. Analysis time is 18 min. (Frazier, 1989).

[e] Secondary targets (and X-ray tube voltages) are Ti (41 kV) for Al - Ca; Co (41 kV) for S - Mn; Mo (50 kV) for Fe - Sr and Pb; Sm (66 kV) for Sr - Ba. Analysis time is 14 min. (Kellogg, 1989).

[f] Excitation by Cr X-ray tube operated at 54 kV and 44 mA; Al, Si, P, S, Cl, K, Br, Sr, Mo, Cd and Pb are determined simultaneously in 100 seconds using fixed channels. All other listed elements are determined sequentially by scanning the X-ray spectrum for 20 sec per element. (Kellogg, 1989).

[g] Proton beam energy is 3.2 MeV; analysis time is 6 min. Data are from brochure by Element Analysis Corporation (EAC), Tallahassee, FL.

In most EDXRF analyses the needed interference corrections are based on Cd, Sn, Sb and Ba concentrations, which are measured poorly by XRF, and this causes the interference corrections to have poor precision. Table 5 shows two sets of precision data for K, Ca, Ti and V. The larger values are more realistic, for they include the precision with which the interference corrections can be made. The smaller values (in parentheses) are based on the assumption that interfering elements are absent.

Although XRF is considered to be nondestructive, this is not strictly true. During XRF analysis, the sample must be in vacuum or in helium to eliminate air and argon, which attenuate and interfere with X rays from light elements. If a sulfuric acid sample is analyzed in vacuum, loss by volatilization may occur. Also, during WDXRF analysis a sample receives a large radiation dose, which makes the filter brittle and may volatilize Br and Cl. Radiation damage during EDXRF analysis is negligible. Data in Table 5 suggest that detection limits can be improved if a sample is first analyzed by EDXRF for all elements and then by WDXRF for Al, Ti, Cd, Sn, Sb and Ba. By limiting WDXRF to those few elements, precision can be improved by using longer counting time per element while total radiation dose is reduced. The precision of XRF varies inversely as the square root of counting time, so precision improves by a factor of 2 for every factor of 4 increase in counting time.

The sensitivity profile for a typical XRF analyzer is nonuniform across a sample's diameter. For such an instrument, quantitative analysis is possible only if the standards and samples have identical deposit profiles. This requirement is satisfied most easily if both sample and standards have uniform deposit profiles. Thus, most cascade impactor samples, which are nonuniform, are unsuitable for XRF. An exception is the micro orifice impactor, which deposits size-fractionated particles uniformly on surfaces (Hasan and Dzubay, 1987). Uniformity of deposit profile is less a concern for PIXE because the proton beam used by that method can be made small and uniform.

4.4 Quantitative XRF analysis

In an X-ray spectrum such as the one shown in Figure 1, peak areas are proportional to the concentrations of the elements they represent. One method to determine peak areas in a multielement spectrum is to use a least squares fitting method to represent the spectrum as a superposition of spectra for pure element standards and a blank filter (Arinc *et al.*, 1977). This method automatically determines the peak area for each element in the presence of background and interferences, and it provides a good way to estimate measurement precision. Another method is to sum the spectrum over specified intervals to determine peak areas and background; arithmetic corrections are made to account for interferences.

A practical way to relate spectral peak areas to element concentrations is to use thin, single element standards and make mathematical correction for attenuation of X rays within unknown samples (Dzubay *et al.*, 1987). A single concentration is needed for each element if it is in the instrument's linear range and if background is subtracted appropriately. A seemingly good alternative is to use multielement standards that exactly duplicate the attenuation effects of all unknown samples, but such standards do not exist. Reliable single element standards are made by Micromatter Co. (Deer Harbor, WA) by vacuum deposition of pure metals or compounds onto Nuclepore filters. Deposit masses are determined gravimetrically and range typically from 20 to 80 $\mu g/cm^2$. Another type of standard that is rugged and reliable is polymer film with known amounts of organometallic compounds; element concentrations range from 2 to 12 $\mu g/cm^2$ in the polymer (Dzubay *et al.*, 1987). The substrates for standards and samples need not be identical if background is properly subtracted from the spectra for standards and unknowns.

Because no standard has perfect accuracy, one should not rely on any single standard. We use a semiempirical model to derive a calibration constant for each element based on data for all standards and all elements. This approach not only improves the calibration for elements that have accurate standards, but it provides accurate calibration by interpolation for P, Ga, As, Se and Br, which have poor standards because of volatility or uncertain stoichiometry. Results are plotted in Figure 2.

The National Institute of Standards (NIST, formerly NBS) has developed standard reference materials (SRMs) for XRF which are thin, uniform glass deposits on Nuclepore filters. Certified elements in SRM 1832 are Al, Si, Ca,V, Mn, Co and Cu; certified elements in SRM 1833 are Si, K, Ti, Fe, Zn and Pb. We use these standards to validate the calibration constants derived from thin film standards and to provide daily quality control checks on accuracy of EDXRF data. Fine particles are typically collected as layers of particles on the surface of a membrane filter (Loo *et al.*, 1977), and reliable corrections for attenuation can be made assuming a uniform deposit. The correction for fine particles is usually small since the membrane filter will clog before enough mass is collected to cause large attenuation.

In a dichotomous sampler, coarse particles are collected mainly as a monolayer, and the correction for attenuation is made by assuming that each particle is a sphere (Dzubay and Nelson, 1975). Such corrections are uncertain when size distributions and compositions of individual coarse particles are unknown. Thus, we base the attenuation correction on the range of results calculated for a wide variety of compositions and size distributions. Table 6 shows attenuation correction factors calculated for excitation by Ti, Co and Mo secondary targets. Uncertainty is largest for Al, Si and P, which emit soft X rays that are easily absorbed within each particle. The attenuation correction factor for Al excited by Ti X rays is 0.56 ± 0.15 and represents the diverse values 0.41, 0.58 and 0.70 for almandine, muscovite and kaolinite, respectively. For samples collected by using a PM-10 inlet, the dependence on size distribution is not large. For example, the attenuation factors for muscovite are 0.58 and 0.54 for a size distribution with geometric standard deviation of 2 and mass median aerodynamic diameters of 10 and 15 μm, respectively.

When XRF data are reported, both concentrations and uncertainties should be reported. Measurement uncertainty due to the random nature of counting X rays can be derived from Poisson statistics and should include count rates due to

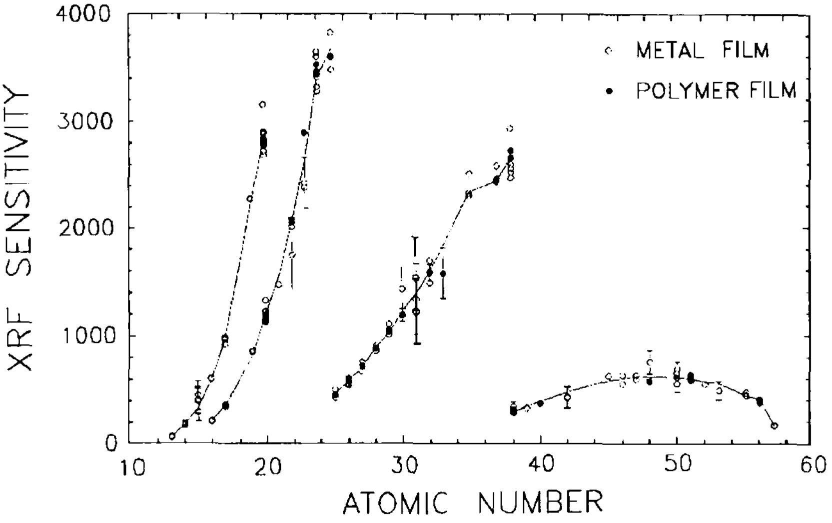

Figure 2. XRF calibration constants based on semiempirical model applied to 203 different polymer film and evaporated film standards. (Many data points close to the fitted line overlap each other and are therefore not seen.)

Table 6. Attenuation correction factors for XRF analysis of coarse particles[a].

Element Analyzed	Ti target 4.5-keV excitation	Co target 7-keV excitation	Mo target 18-keV excitation
Al	0.56 ± 0.15	0.59 ± 0.16	
Si	0.57 ± 0.13	0.61 ± 0.12	
P	0.65 ± 0.20	0.68 ± 0.19	
S	0.81 ± 0.13	0.85 ± 0.10	
Cl	0.81 ± 0.10	0.84 ± 0.09	
K	0.84 ± 0.10	0.88 ± 0.08	0.90 ± 0.08
Ca	0.81 ± 0.05	0.88 ± 0.04	0.91 ± 0.04
Ti		0.81 ± 0.08	0.90 ± 0.06
V		0.87 ± 0.08	0.92 ± 0.05
Cr		0.87 ± 0.08	0.93 ± 0.05
Mn		0.87 ± 0.08	0.94 ± 0.05
Fe			0.95 ± 0.04
Ni			0.93 ± 0.05
Zn			0.93 ± 0.05

[a] Size range is 2.5- to 10-μm aerodynamic equivalent diameter.

background and interfering elements as well as the element itself. Some investigators also include systematic error. If the true concentration is zero, negative and positive values are equally probable because background, which has random error, is subtracted to obtain a final value. Thus, -1.1 ± 0.8 ng/m^3 is a valid result that is not significantly different from zero and does not imply a meaningless negative concentration. It is common to replace negative values with zero or a positive number, but that introduces bias that complicates use of the data to compute statistical parameters.

4.5 Instrumental Neutron Activation Analysis

In INAA a sample absorbs thermal neutrons in a nuclear reactor and undergoes nuclear transitions that result in γ-ray emissions characteristic of nuclei in the sample (Zoller and Gordon, 1970; Dams *et al.*, 1972; Ondov *et al.*, 1975). After samples are withdrawn from the reactor, they are allowed to "cool" to diminish activities of short-lived interfering isotopes, and γ-ray spectra are obtained with lithium-drifted or intrinsic germanium detectors. Near optimal results are obtained when irradiation, cooling and counting periods are about equal to the half life of the isotope measured. Because half-lives of isotopes of interest range from 2 min to more than 12 years, it is not practical to satisfy such criteria for every element, but very good results can be obtained for more than 25 elements if two irradiation periods and four counting periods are used as shown in Table 7.

Gordon *et al.* (1984) compare detection limits for INAA with typical ranges of element concentrations in the atmosphere in order to predict which elements can be measured. Gordon *et al.* also characterize INAA by noting whether an element was measured strongly or moderately in aerosol samples collected in St. Louis, and their results are summarized in Table 7. The results are based on the NBS heavy water moderated reactor at Gaithersburg, MD. The neutron flux was 2 to 5 x 10^{13} cm^2/sec. These results illustrate what can be measured by INAA, but they may not generally apply to other laboratories, because each laboratory optimizes its measurement system in a different way. For example, there are fewer interferences among elements if a heavy water reactor is used, but Ni can be

Table 7. Typical INAA Analysis Conditions and Performance[a].

	Group 1	Group 2	Group 3	Group 4
Product half-life	<12 min	12-1440 min	1-3 d	>3 d
Irradiation time	5 min	5 min	4 h	4 h
Cooling time	3-6 min	15 min	2-5 d	20-30 d
Counting time	6 min	30 min	1-4 h	2-24 h
	+ Mg (^{27}Mg)	+ + +Na (^{24}Na)	+ + +As (^{76}As)	+ + +Sc (^{46}Sc)
	+ + Al (^{28}Al)	+ + +Cl (^{38}Cl)	+ +Br (^{82}Br)	+ + +Cr (^{51}Cr)
	+ S (^{37}S)	+ + +K (^{42}K)	+Mo (^{99}Mo)	+ + +Fe (^{59}Fe)
	+ + Ca (^{49}Ca)	+ + +Mn (^{56}Mn)	+ +Cd (^{115}In)	+ + +Co (^{60}Co)
	+ + Ti (^{51}Ti)	+Ga (^{72}Ga)	+Sb (^{122}Sb)	Ni (^{58}Co)
	+ + + V (^{52}V)	+ +Br (^{80}Br)	+ + +La (^{140}La)	+ + +Zn (^{65}Zn)
	+ Cu (^{66}Cu)	+Br (^{80m}Br)	+ + +Sm (^{153}Sm)	+ + +Se (^{75}Se)
		+Sr (^{87m}Sr)	+ +W (^{187}W)	+ +Rb (^{86}Rb)
		+In (^{116m}In)	+Au (^{198}Au)	+Zr (^{95}Zr)
		+ +I (^{128}I)		+Ag (^{110m}Ag)
		+Ba (^{139}Ba)		+ + +Sb (^{124}Sb)
		+Dy (^{165}Dy)		+ +Cs (^{134}Cs)
		+ +W (^{187}W)		+Ba (^{131}Ba)
				+ + +Ce (^{141}Ce)
				+ +Eu (^{152}Eu)
				+Tb (^{160}Tb)
				+Yb (^{169}Yb)
				+ +Lu (^{177}Lu)
				+ +Hf (^{181}Hf)
				+ +Ta (^{182}Ta)
				+ +Th (^{233}Pa)

[a] The isotope measured for each element is shown in parentheses. Performance, based on results of Gordon *et al.* (1984) for analysis of St. Louis samples, is represented as follows:

+ + + elements strongly measured in fine and coarse fractions.
+ + elements strongly measured in fine or coarse fraction.
+ elements moderately measured in fine or coarse fraction.

detected only if a light water reactor is used (Gordon *et al.*, 1984).

4.6 Element Analysis by Optical Spectroscopy

Elements extracted from filters can be measured by AA or ICP. A typical AA instrument can analyze only one element at a time, although newer instruments can analyze two or even four elements simultaneously. The ICP system used by EPA to determine trace elements in hi vol samples analyzes 48 elements per sample simultaneously (Cummings *et al.*, 1984). Harper *et al.* (1983) have developed an optimized method that extracts more of the particulate matter and less of the contaminants in a glass fiber filter than was possible with an earlier EPA method. In the optimized method, a 2.5- x 20-cm portion of filter is extracted in a heated ultrasonic bath in 2.2 M HCl and 1 M HNO_3 (Harper *et al.*, 1983).

Table 8 compares ICP results for 195 hi vol samples extracted by the optimized and earlier extraction methods. In every case the optimized method has equal or improved extraction efficiency. Several elements (Ti, V, Mn, Fe, Cu, Zn and Pb) were detected in almost all of the samples, but concentrations of other elements (Cr, Co, Ni, As, Mo, Cd and Ba) were significantly greater than the blank

Table 8. Performance of ICP for analysis of ambient samples and standard reference material[a].

	Optimized vs earlier extraction method for 195 hi vol samples			Extraction efficiency for NBS SRM 1648 (Urban particulate matter standard)	
	Number Detected	Slope[b]	Correlation Coefficient	Mean Recovery	Relative Standard Deviation (%)
Ti	192	0.33	0.87	12	5.0
V	138	0.91	1.00	79	1.9
Cr	27	0.95	0.93	23	1.4
Mn	195	0.91	0.98	88	1.6
Fe	175	0.34	0.87	68	1.4
Co	0			96	5.4
Ni	33	0.88	0.98	90	9.0
Cu	194	0.92	0.99	100	1.4
Zn	170	1.13	0.73	97	3.8
As	15	1.00	0.99	130	2.2
Mo	14	0.88	0.95		
Cd	19	0.95	1.00	114	8.5
Ba	51	0.17	0.49	80	0.8
Pb	192	1.00	1.00	95	1.1

[a] Results reported by Harper *et al.* (1983).

[b] Regression results for: earlier method = intercept + slope x optimized method.

in less than 30% of the samples. Newer, more sensitive ICP instruments applied to purer quartz filters may yield better detection for more elements.

Table 8 also shows extraction efficiencies for elements in a standard reference material for urban particulate matter. Extraction is quantitative for 10 elements, but Ti and Cr are not extracted efficiently. The acids used in the extraction are not expected to provide quantitative recovery of elements in silicate minerals. Thus, if ICP is used to provide data for receptor models, it may be necessary also to use INAA to determine elements that are difficult to extract. Although the Na in glass fiber filters causes high background that impairs INAA performance, Lambert and Wilshire (1979) were able to obtain reliable INAA results for several elements.

Table 9 shows typical 2σ detection limits for a modern ICP instrument. Data in Tables 5 and 9 indicate that ICP is superior to XRF for Li, Be, B, Na, Mg, Ca, Ti, V, Mn, Sr, Y, Zr, Ag, Cd and Ba collected in a dichotomous sampler. However, to take full advantage of the capabilities of ICP for analyzing Teflon filters, it will be necessary to develop an extraction procedure that does not introduce contamination and is quantitative for all elements.

Table 9. Limits of Detection for ICP Instrument in ng/mL of Sample Extract[a,b].

Li	0.8	Ca	0.07	Se	14.
Be	0.13	Ti	0.4	Sr	0.07
B	1.1	V	0.8	Y	0.5
Na	5.0	Cr	1.0	Zr	2.5
Mg	0.07	Mn	0.2	Ag	1.2
Al	4.7	Fe	1.0	Cd	0.6
Si	3.3	Cu	0.9	Sn	6.7
P	13.	Ni	2.7	Sb	7.4
S	20.	Zn	0.5	Ba	0.7
K	20.	As	7.4	Pb	10.

[a] Model JY 70 Plus by Instruments S. A. Inc., Optical Systems Div., Edison, NJ.

[b] The tabulated data are equivalent to ng/m^3 for 10-hour air samples collected in a dichotomous sampler at 16.7 L/min and extracted in 10 mL of solution.

4.7 Carbon Analysis

Volatilizable carbon (C_v) and nonvolatilizable carbon (C_{nv}) on quartz filters can be determined by pyrolysis. To remove carbon contamination before sampling, quartz filters are heated in air at 700°C for several hours and stored in aluminum-foil-lined petri dishes. During analysis, C_v is determined first by heating the sample to temperature T_1 in pure helium and measuring volatilized carbon as CO_2 by infrared spectroscopy after oxidation or as methane by flame ionization after reduction. Then, C_{nv} is determined by heating the sample to temperature T_2 in a helium-oxygen mixture and measuring oxidized carbon. Pyrolysis temperatures used by various laboratories are in the ranges $350 < T_1 < 950°C$ and $600 < T_2 < 950°C$ (Hering *et al.*, 1990). To account for charring so that C_v and C_{nv} can be related to organic and elemental carbon, some investigators measure light absorption as samples are heated during pyrolysis (Johnson *et al.* 1981).

Several laboratories participated in a field study to compare methods for measuring carbon (Hering *et al.*, 1990). When participants analyzed identical samples, excellent agreement was obtained for total carbon ($C_v + C_{nv}$), but results were variable for C_{nv}, which comprised about 25% of total carbon (Hering *et al.*, 1990). Larger variability for both C_v and C_{nv} was obtained when both sampling and analysis methods were compared. Some of the variability was attributed to adsorption of organic vapors by quartz filters.

A correction for adsorption of organic vapors can be made if carbon is analyzed on two quartz filters that have been sampled in series. One approach is to use two quartz filters in series in the same airstream. Another approach is to collect carbon on quartz filters in two parallel sampling streams where quartz is preceded by a Teflon filter in one of the streams. With either approach, particulate carbon is calculated as the difference between amounts of carbon on the two quartz filters. Field study results showed that quartz behind quartz collected less organic carbon vapor than quartz behind Teflon (Hering *et al.*, 1990), which indicates that quartz collects more organic carbon vapors than Teflon. Until it becomes known which approach is best, data for C_v and C_{nv} should be used with caution. If one must use such data in a chemical mass balance, some of the ambiguity can be avoided if the same sampling protocol and laboratory are used to analyze all source signature and ambient aerosol samples. That may preclude the use of carbon data in source signature libraries.

4.8 ^{14}C Analysis

Cosmic rays continuously bombard atoms in the upper atmosphere and produce ^{14}C, a radioactive isotope with 5730-year half-life. Carbon atoms containing this isotope are oxidized in the atmosphere to form CO_2 that becomes incorporated into living matter. Although fossil fuel originated as living matter, its ^{14}C has completely decayed away, and it contains only stable isotopes ^{12}C and ^{13}C. Thus, through measurement of the $^{14}C/^{13}C$ ratio, fossil fuel combustion products can be distinguished from modern carbon-containing particles such as pollen and combustion products of wood and paper. In past years, ^{14}C in aerosol samples was measured by oxidizing the sample to CO_2 and incorporating it into a gas-filled radiation counting tube (Currie, 1979). About 5 mg of carbon was needed for analysis, an amount easily collected in a hi vol sampler. Cooper *et al.* (1981) determined ^{14}C in <2.5-μm particles collected in a hi vol equipped with an impactor inlet and were able to estimate contributions due to slash burning and field burning. Lewis *et al.* (1988) collected <2.5-μm particles during winter in Albuquerque and demonstrated that ^{14}C in extractable organic matter was proportional to the amount of wood smoke.

Accelerator mass spectrometry enables ^{14}C atoms to be measured directly. An AMS instrument consists of a sample ionizer, a tandem Van de Graaff that accelerates carbon ions to an energy of 12 MeV, a magnetic separator that selects mass 13 and 14 ions, and a particle detector that identifies and counts ^{13}C and ^{14}C ions. Verkouteren *et al.* (1987) have developed a sample preparation technique that enables ^{14}C to be determined by AMS in samples that have 10 to 800 μg of carbon. This technique is sensitive enough to analyze 24-hour samples collected in a dichotomous sampler. Because there is a limited number of laboratories capable of AMS analysis, there are large backlogs, and it is difficult to obtain timely analysis.

5. IONIC ANALYSIS OF PARTICLES COLLECTED ON FILTERS

5.1 Extraction

Analysis of ionic species collected on a filter requires extraction of the ions into an aqueous solution. Quantitative extraction of soluble ions on Teflon filters requires the use of ethanol to overcome the hydrophobic nature of Teflon. The following procedure has been developed for 37- and 47-mm Teflon filters and is similar to one used by Koutrakis *et al.* (1988a): Immediately after sampling, a filter is placed in an extraction vial (vial diameter must exceed filter diameter) containing 10 mL of extraction solution and stored in the dark at 5°C to retard bacteria growth until time of analysis. When ready for analysis, the filter is removed from its vial, and 100 μL of ethanol is pipetted onto the filter to wet it uniformly. The filter is immersed again in the vial and agitated in an ultrasonic bath for 15 min. If H^+ is to be measured, the extraction solution should be 10^{-4} M $HClO_4$, which keeps CO_2 and weak acids from disassociating. Otherwise, the extraction solution may be either deionized water or the eluent solution used in ion chromatography.

5.2 Ion Chromatography

Ion chromatography is a technique for separation and automatic measurement of ions in solution (Mulik and Sawicki, 1979; Gjerde and Fritz, 1987). Measurement consists of injecting a sample into an eluent that passes through an ion exchange column followed by a suppressor unit and a conductivity detector. For anion analysis, the eluent typically consists of a mixture of 2.8 mM $NaHCO_3$ and 2.2 mM Na_2CO_3 (Gjerde and Fritz, 1987). The ion exchange column retards the flow of sample by a different amount of time for each ion, and the conductivity detector responds linearly to each anion. The suppressor unit eliminates the high conductivity of the eluent by converting it to H_2CO_3. The separation characteristics of the ion exchange column can be optimized for certain ions by adjusting the ratio of $NaHCO_3$ to Na_2CO_3 in the eluent (Speitel *et al.*, 1979). Ion chromatography can also be used to measure cations (Gjerde and Fritz, 1987).

Analytical detection limits below 1 µg/mL are obtained routinely by using an instrument made by Dionex Corp. (Sunnyvale, CA). This corresponds to detection limits below 0.5 $\mu g/m^3$ for a 24-hour fine particle sample collected at 15 L/min and extracted in 10 mL. Anions routinely measured in aerosol samples include sulfate and nitrate. Additional anions seen in some samples include bromide, chloride, nitrite, sulfite, formate and acetate. An automated ion chromatograph can analyze three to five samples per hour. Ethanol, which serves as a wetting agent in the extraction, is unlikely to damage the IC column. Reasonably long life for the ion exchange column can be obtained if an upstream filter is used to remove bacteria in the eluent and if a relatively fresh "guard column" is used upstream to remove polyvalent metals that might contaminate the main column (Gjerde and Fritz, 1987).

5.3 Ion Analysis by Colorimetry

Sulfate, nitrite, nitrate and ammonium in sample extracts can be analyzed rapidly by colorimetry. Sulfate is determined by measuring absorbance at 460 nm in a sample processed as follows: (1) passage through a cation exchange column to remove interferences, (2) reaction with barium chloride in acidic solution to form barium sulfate, (3) reaction of excess barium with methylthymol blue in alkaline solution to form a blue chelate. Nitrite is determined by adding an acid and sulfanilamide to the extraction solution to form a soluble dye that is measured by absorbance at 520 nm. Nitrate is determined similarly after it is reduced by a copper catalyst to nitrite in an alkaline solution of hydrazine sulfate. Ammonium is determined by measuring absorbance at 660 nm of a blue dye formed by reaction of ammonium with sodium phenoxide followed by addition of sodium hypochlorite. Interfering calcium and magnesium ions are removed by complexing with EDTA. An automated "TRAACS" colorimetric instrument by Technicon Industries can analyze up to 40 unknowns plus 30 standards per hour, which is considerably faster than the 3 to 5 per hour capability of an ion chromatograph.

5.4 Hydrogen Ion Measurement

Stevens *et al.* (1978) describe a Gran titration procedure by which H^+ from strong acid is determined from a plot of pH vs amount of NaOH titrant. As explained in Section 5.1, the extraction solution should contain $HClO_4$, but the portion analyzed for H^+ should also contain 0.04 M of KCl in order to increase the ionic strength and speed the response of the pH meter. Koutrakis *et al.* (1988a) deduce H^+ concentration from a single pH measurement without the need for a Gran titration. Such measurements determine strong acid because the $HClO_4$ prevents disassociation of weak acids.

5.5 Artifacts in Measurement of Ionic Species

Although techniques described above are useful for determining ions collected in a PM-10 sampler, three types of artifacts can cause measured ions not to represent atmospheric particles: (1) filter artifacts, (2) reactions between gases and particles, and (3) reactions among particles.

Alkalinity of glass fiber filters can cause gaseous sulfur and nitrogen compounds to react and produce sulfate and nitrate artifacts. Such filter artifacts are minimal for Teflon filters or the low-alkalinity quartz filters required for PM-10 sampling.

Two kinds of reactions between gases and particles can alter the aerosol composition. (a) Acidic gases can react with alkaline particles, and (b) ammonia gas can react with acidic particles. Such reactions can occur in the atmosphere or on the filter. Examples are

$$\mathbf{2HNO_3 + CaCO_3 \rightarrow Ca(NO_3)_2 + H_2O + CO_2\uparrow} \tag{2}$$

$$\mathbf{2HCl + MgCO_3 \rightarrow MgCl_2 + H_2O + CO_2\uparrow} \tag{3}$$

$$\mathbf{HNO_3 + NaCl \rightarrow NaNO_3 + HCl\uparrow} \tag{4}$$

$$\mathbf{NH_3 + NH_4HSO_4 \rightarrow (NH_4)_2SO_4} \tag{5}$$

Reaction (2) can explain the occurrence in Camden, NJ, of nitrate in coarse particles, which contain calcium carbonate (Dzubay *et al.*, 1988). Reaction (5) can cause hydrogen ion collected on a filter to be neutralized by ambient ammonia during sampling and analysis. Stevens *et al.* (1978) prevented neutralization by ammonia during sampling by using a diffusion denuder to remove ammonia upstream of the filter, and Koutrakis *et al.* (1988a) developed a procedure for preventing neutralization in the laboratory during storage and analysis.

Particles collected on a filter can also react with each other. This is especially true for fine particles, which can liquify and coalesce if they are hygroscopic and the relative humidity is high (Mamane and Dzubay, 1990). If both acid sulfate and anions of volatile acids are present, reactions such as the following can occur:

$$\mathbf{H_2SO_4 + 2NH_4NO_3 \rightarrow (NH_4)_2SO_4 + 2HNO_3 \uparrow} \tag{6}$$

$$\mathbf{NH_4HSO_4 + NH_4NO_3 \rightarrow (NH_4)_2SO_4 + HNO_3 \uparrow} \tag{7}$$

$$\mathbf{NH_4HSO_4 + NaNO_3 \rightarrow NH_4NaSO_4 + HNO_3 \uparrow} \tag{8}$$

$$\mathbf{NH_4HSO_4 + NaCl \rightarrow NH_4NaSO_4 + HCl \uparrow} \tag{9}$$

$$\mathbf{NH_4HSO_4 + CH_3COONa \rightarrow NH_4NaSO_4 + CH_3COOH \uparrow} \tag{10}$$

$$\mathbf{NH_4NO_3 \rightarrow NH_3 + HNO_3 \uparrow} \tag{11}$$

Because acidic sulfate is abundant in many parts of United States (Stevens *et al.*, 1978; Stevens *et al.*, 1980; Koutrakis *et al.*, 1988a), reactions (7-11) can cause loss of hydrogen ion, nitrate, halides and salts of organic acids. Reaction (6) was observed in laboratory studies (Harker *et al.*, 1977), and Slater *et al.* (1989) demonstrated significant loss of H^+ on ambient samples collected during winter in eastern United States. Reaction (11) expresses the fact that ammonium nitrate can decompose. All of these reactions indicate that, except for sulfate and metallic ions, ions measured on a filter may not represent the ionic composition of atmospheric particles.

5.6 Diffusion Denuder Systems for Measuring Gases and Particles

The artifacts described in Section 5.5 may be minimized through the use of diffusion denuders and a filter pack for collecting the various vapors released from a filter. A filter pack consists of two or more filters in series. A typical diffusion denuder is a chemically coated cylindrical tube that removes gases that diffuse to the walls and react with the coating. Particles are not removed because their diffusion coefficients are too low to cause significant diffusion to the tube walls. After sampling, the denuders can be extracted and analyzed for their ionic content. In past years, denuders were made with hollow tubes whose gas collection efficiencies are large only at very low flow rates. To obtain both high collection efficiency and high flow rate with hollow tube denuders, it is necessary to operate several denuder tubes in parallel (Stevens *et al.*, 1978).

Possanzini *et al.* (1983) made a substantial improvement in denuder performance by using an annular denuder which collects gaseous pollutants by moving air through an annular space between two concentric glass cylinders coated with an appropriate chemical substrate. The cylinder walls are etched by sand blasting to increase their capacity to support the denuding chemical substrate. Collection efficiency E_h for a hollow tube denuder of length L is

$$E_h = 0.82\exp[-3.66(\pi DL/F)] \tag{12}$$

and efficiency E_a for an annular denuder of length L is

$$E_h = 0.82\exp\left[-3.66(\pi DL/F)\frac{(d_2 + d_1)}{(d_2 - d_1)}\right] \tag{13}$$

where D = species diffusion coefficient; d_2 = internal diameter of tube; d_1 = diameter of internal cylinder; F = flow rate (Possanzini *et al.*, 1983). For a given tube length, an annular denuder can operate at 20 times the flow rate and yet have the same efficiency as a hollow tube. Moreover, laminar flow conditions are maintained up to fairly high flow rate in an annular denuder (Possanzini *et al.*, 1983).

Figure 3 and Table 10 represent an annular denuder system and filter pack that can measure a variety of gaseous and particulate species. The system has a cyclone on the inlet to remove coarse (>2.5 μm) particles. Three chemically treated denuders are used for selectively collecting the gases listed in Table 10. Denuders 1, 2 and 3 are quantitative for HNO_3, HNO_2 and NH_4, respectively, and the collection of SO_2 occurs on denuders 1 and 2. The filter pack consists of three filters in series. Vapors that volatilize from the first filter via reactions (6-11) are collected on filters 2 and 3 and enable the original ionic content of the sample

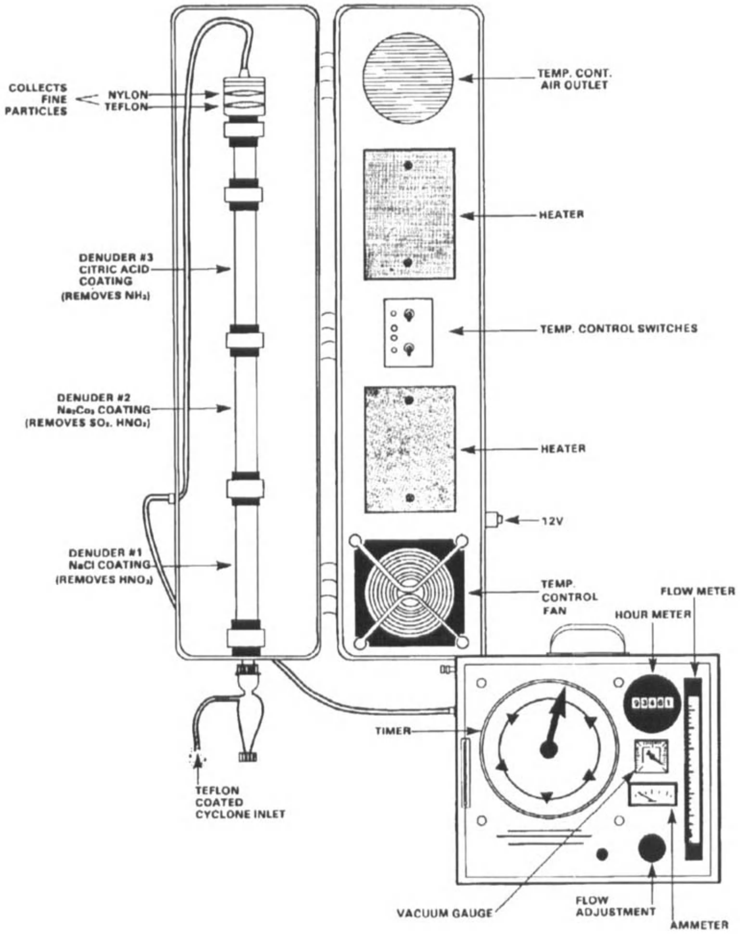

Figure 3. Schematic Description of Annular Denuder System with Filter Pack.

to be deduced (Slater *et al.*, 1989). Earlier systems did not use the first NaCl-coated denuder and attempted to distinguish HNO_2 from HNO_3 on the basis of the nitrite/nitrate ratio (Vossler *et al.*, 1988; Koutrakis *et al.*, 1988b), but this is unreliable because the collected nitrite can oxidize to nitrate during sampling.

Table 10. Annular Denuder and Filter Pack System for Collecting Particulate and Gaseous Ionic Species.

Stage	Substrate	Species Collected	Species Analyzed by Ion Chromatography
Denuder 1	NaCl + glycerine	HNO_3, SO_2	NO_3^-, $SO_4^=$
Denuder 2	$CaCO_3$ + glycerine	HNO_2, SO_2	NO_2^-, NO_3^-, $SO_4^=$
Denuder 3	Citric acid + glycerine	NH_3	NH_4^+
Filter 1	Ringed Teflon filter	particles	H^+, NH_4^+, NO_3^-, $SO_4^=$
Filter 2	Na_2CO_3 on glass filter	volatilized HNO_3	NO_3^-
Filter 3	Citric acid on glass	volatilized NH_3	NH_4^+

Ionic species in denuder and filter extracts are determined by techniques described in sections 5.2 - 5.4. Analysis of SO_2 is simplified if any collected sulfite is oxidized to sulfate by adding 50 μL of 0.3% fresh H_2O_2 to each 2.5 mL of denuder extract before it is analyzed by ion chromatography.

The system described in Table 10 provides a reliable means to obtain concentrations of both gaseous and particulate species for use in receptor models. For example, such a system is useful for obtaining simultaneous SO_2 and $SO_4^=$ data in studies to determine the origins and age of sulfate aerosol. Also, such a system enables particulate and gas phase nitrate to be measured without bias due to nitrate evaporation in studies to determine contributions of particulate nitrate to visibility degradation.

6. CONCLUSIONS

6.1 Analysis of Quartz Filters from PM-10 Hi Vol Samplers

If the only available samples were collected on quartz filters, optical spectroscopy is the preferred method for determining element concentrations. The extraction and ICP methods developed by Harper *et al.* (1983) provide quantitative data for V, Mn, Fe, Co, Ni, Cu, Zn, As, Cd, Ba and Pb. Such data can be used in receptor models to determine many sources, but additional data for Al and perhaps Ca and Sc are needed to determine soil. It is difficult to extract the Al from soil particles on a quartz filter, but Al can be measured on quartz by INAA with 5-min activation time. A portion of the quartz filter can be extracted in deionized water and analyzed by ion chromatography for sulfate and nitrate. Carbon can be determined by pyrolysis. Although sampling artifacts may cause the collected carbon and nitrate not to represent true atmospheric concentrations, the analyses will determine those components as they are represented in the measured PM-10 mass concentration.

6.2 Analysis of Filters from Dichotomous Samplers

The use of dichotomous samplers may be preferred for collecting particles on Teflon filters since they can be analyzed inexpensively by XRF. Usually Al, Si, S, K, Ca, Ti, Mn, Fe, Zn, Br and Pb can be determined by EDXRF. Depending on the airshed and XRF procedure, Cl, V, Cr, Ni, Cu, Se, Sr, Cd, Sb and Ba can often be determined. Excellent precision for V can be obtained by EDXRF if precise data for Ba are available so that the Ba - V interference correction can be made algebraically. Such Ba data can be obtained by WDXRF. If better precision and additional elements are needed, WDXRF can be used to determine Al, Ti, Cd, Sn and Sb, and INAA can be used to determine Na, Al, Sc, V, Cr, Mn, Co, As, Se, Mo, Sb, La, Sm and Ce. Gordon *et al.* (1984) suggest that the cost of INAA can be reduced by using only a single 5-min irradiation, which can determine most of the elements in Groups 1, 2 and 3 in Table 7. Because this strategy does not determine Ce, which is emitted by refineries and automobile catalysts, selected samples could also be analyzed for elements in Group 4 by the more expensive 4-hour INAA irradiation (see Table 7).

It is not always necessary to measure sulfate because XRF data for S have been shown to be equivalent to ion chromatography data for sulfate in many airsheds (Stevens *et al.*, 1978; Dzubay *et al.*, 1988). Ion analysis is needed to determine nitrate, ammonium and hydrogen ion, but sampling artifacts can cause filter samples not to represent atmospheric concentrations of those ions. Use of an annular denuder system with a filter pack will enable those ions to be measured quantitatively, along with SO_2, HNO_3, and HNO_2 gases. If carbon data are needed, additional samples need to be collected on quartz filters. However, measurement uncertainty due to adsorption of organic vapors on quartz filters and ambiguity in measuring the C_e/C_v ratio discourages the use of carbon data in the chemical mass balance receptor model.

6.3 Quality Assurance

To establish the precision and accuracy of the data, several of the following steps should be included in any measurement program: (1) Determine overall precision by analyzing samples collected in simultaneously operated duplicate samplers. (2) Determine XRF precision by reanalysis of 5 to 10% of the samples. Precision for ICP or INAA can be established by splitting samples and analyzing both portions. To avoid bias, the samples can be given new identification numbers known only to the one conducting the evaluation. (3) Determine the standard deviation for several blanks to provide an estimate of precision for elements with low concentration. Some of the blanks could be filters that were installed in a sampler that was not operated. (4) Analyze a subset of samples by two separate methods such as EDXRF and WDXRF or EDXRF and INAA to assess accuracy. (5) Determine accuracy by analyzing standard reference materials by a procedure identical to the one used for unknown samples.

6.4 Selection of Analytical Laboratories

When one lacks equipment or expertise to perform all needed analyses, assistance from another laboratory is required. Because receptor models require data with known precision and accuracy, the selected laboratory must be able to validate its results for air samples. Attention to the following criteria can be helpful in selecting a laboratory: (1) The laboratory has experience in successfully analyzing your kind of samples. (2) Measurement uncertainty is routinely reported with each reported datum. (3) Effects of interferences are included in all uncertainty estimates. (4) Measurement uncertainties have been previously validated by a quality assurance program like the one described in Section 6.3. (5) Adequate precision and accuracy have been demonstrated for your type of sample.

7. ACKNOWLEDGEMENTS

We thank the following individuals for valuable suggestions: Glen Gordon, Larry Purdue, Sharon Harper, Robert Kellogg, Clifton Frazier and Frank McElroy. The information in this chapter has been funded in part by the United States Environmental Protection Agency. It has been subjected to Agency review and approved for Publication. Mention of commercial products and company names does not constitute endorsement by the EPA.

8. REFERENCES

Arinc, F., L. Wielopolski, and R.P. Gardner, The Linear Least-Squares Analysis of X-Ray Fluorescence Spectra of Aerosol Samples Using Pure Element Library Standards and Photon Excitation, in *X-Ray Fluorescence Analysis of Environmental Samples*, T.G. Dzubay, ed, Ann Arbor Science Publishers, Inc., Ann Arbor, MI, 227-240 (1977).

Cooper, J.A., L.A. Currie, and G.A. Klouda, Assessment of Contemporary Carbon Combustion Source Contributions to Urban Air Particulate Levels Using Carbon-14 Measurements, *Environ. Sci. Technol.* 15:1045-1050 (1981).

Cooper, J.A., Personal Communication (1988).

Cummings, S.L., S.L. Harper, W.A. Loseke, and L.J. Pranger, *Standard Operating Procedure for the ICP-OES Determination of Trace Elements in Suspended Particulate Matter Collected on Glass-Fiber Filters*, Report No. EMSL/RTP-SOP-EMD-002, Office of Research and Development, U.S. EPA, Research Triangle Park, NC (1984).

Currie, L., Origin of Carbonaceous Species in the Atmosphere Deduced by Radiocarbon, in *Sampling and Analysis of Toxic Organics in the Atmosphere*, S.S. Verner, Symposium Chairman, American Society for Testing and Materials, Philadelphia, PA, 120-138 (1979).

Dams, R., K.A. Rahn, and J.W. Winchester, Evaluation of Filter Materials and Impaction Surfaces for Nondestructive Neutron Activation Analysis of Aerosols, *Environ. Sci. Technol.* 6:441-448 (1972).

Dzubay, T.G. and R.O. Nelson, Self Absorption Corrections for X-Ray Fluorescence Analysis of Aerosols, in *Advances in X-Ray Analysis*, W.L. Pickles, C.S. Barrett, J.B. Newkirk, and C.O. Ruud, eds, Plenum Publishing Corporation, New York, 619-631 (1975).

Dzubay, T.G. and R.K. Barbour, A Method to Improve the Adhesion of Aerosol Particles on Teflon Filters, *J. Air Pollut. Contr. Assoc.* 33:692-695 (1983).

Dzubay, T.G., N. Morosoff, G.L. Whitaker, H. Yasuda, F. Bazan, R.L. Bennett, J. Cooper, W.J. Courtney, C.A. Frazier, R.C. Gatti, M. Germani, G. Gordon, S. Hanamura, R.B. Kellogg, J.R. Rhodes, and J.S. Schindler, Polymer Film Standards for X-Ray Fluorescence Spectrometers, *J. Trace and Microprobe Techniques* 5:327-341 (1987).

Dzubay, T.G., R.K. Stevens, G.E. Gordon, I. Olmez, A.E. Sheffield, and W.J. Courtney, A Composite Receptor Method Applied to Philadelphia Aerosol, *Environ. Sci. Technol.* 22:46-52 (1988).

EPA, Ambient Air Monitoring Reference and Equivalent Methods; Reference Method Designation, *Federal Register* 54:31247 (1989).

Frazier, C.A., Personal Communication (1989).

Friedlander, S.K., Chemical Element Balances and Identification of Air Pollution Sources, *Environ. Sci. Technol.* 7:235-240 (1973).

Gjerde, D.T. and J.S. Fritz, *Ion Chromatography*, Verlag, Heidelberg (1987).

Gordon, G.E., W.R. Pierson, J.M. Daisey, P.J. Lioy, J.A. Cooper, J.G. Watson, Jr., and G.R. Cass, Considerations for Design of Source Apportionment Studies, *Atmospheric Environ.* 18:1567-1582 (1984).

Harker, A.B., L.W. Richards, and W.E. Clark, The Effect of Atmospheric SO_2 Photochemistry Upon Observed Nitrate Concentrations in Aerosols, *Atmospheric Environ.* 11:87-91 (1977).

Harper, S.L., J.F. Walling, D.M. Holland, and L.J. Pranger, Simplex Optimization of Multielement Ultrasonic Extraction of Atmospheric Particulates, *Anal. Chem.* 55:1553-1557 (1983).

Hasan, H. and T.G. Dzubay, Size Distributions of Species in Fine Particles in Denver Using a Microorifice Impactor, *Aerosol Sci. Technol.* 6:29-39 (1987).

Hering, S.V., B.R. Appel, W. Cheng, F. Salaymeh, S.H. Cadle, P.A. Mulawa, T.A. Cahill, R.A. Eldred, M. Surovik, D. Fitz, J.E. Howes, K.T. Knapp, L. Stockburger, B.J. Turpin, J.J. Huntzicker, X.Q. Zhang, and P.H. McMurry, Comparison of Sampling Methods for Carbonaceous Aerosols in Ambient Air, *Aerosol Sci. Technol.* 12:200-213 (1990).

Hoffman, A.J., L.J. Purdue, K.A. Rehme, and D.M. Holland, 1987 PM_{10} Sampler Intercomparison Study, in *PM-10: Implementation of Standards*, C.V. Mathai and D.H. Stonefield, eds, Transactions No. TR-13, Air & Waste Management Association, Pittsburgh, PA, 138-149 (1988).

Jaklevic, J.M., R.C. Gatti, F.S. Goulding, B.W. Loo, and A.C. Thompson, *Aerosol Analysis for the Regional Air Pollution Study*, Report No. EPA-600/4-81-006, Environmental Sciences Research Laboratory, U.S. EPA, Research Triangle Park, NC (1981).

John, W., Thoracic and Respirable Particulate Mass Samplers: Current Status and Future Needs, in *Advances in Air Sampling*, Lewis Publishers, Inc., Chelsea, MI, 25-38 (1988).

Johnson, R.L., J.J. Shah, R.A. Cary, and J.J. Huntzicker, An Automated Thermal-Optical Method for the Analysis of Carbonaceous Aerosol, in *Atmospheric Aerosol: Source/Air Quality Relationships*, E.S. Macias and P.K. Hopke, eds, ACS Symposium Series No. 167, American Chemical Society, Washington, DC (1981).

Kellogg, R.B., Personal Communication (1989).

Koutrakis, P., J.M. Wolfson, and J.D. Spengler, An Improved Method for Measuring Aerosol Strong Acidity: Results from a Nine-Month Study in St. Louis, Missouri and Kingston, Tennessee, *Atmospheric Environ.* 22:157-162 (1988a).

Koutrakis, P., J.M. Wolfson, J.L. Slater, M. Brauer, and J.D. Spengler, Evaluation of an Annular Denuder/Filter Pack System to Collect Acidic Aerosols and Gases, *Environ. Sci. Technol.* 22:1463-1488 (1988b).

Lambert, J.P.F. and F.W. Wilshire, Neutron Activation Analysis for Simultaneous Determination of Trace Elements in Ambient Air Collected on Glass-Fiber Filters, *Anal. Chem.* 51:1346-1350 (1979).

Lewis, C.W., R.E. Baumgardner, R.K. Stevens, L.D. Claxton, and J. Lewtas, Contribution of Woodsmoke and Motor Vehicle Emissions to Ambient Aerosol Mutagenicity, *Environ. Sci. Technol.* 22:968-971 (1988).

Liu, B.Y.H., D.Y.H. Pui, and K.L. Rubow, Characteristics of Air Sampling Filter Media, in *Aerosols in the Mining and Industrial Work Environments*, Vol. 3, V.A. Marple and B.Y.H. Liu, eds, Ann Arbor Science, Ann Arbor, MI, 989-1038 (1983).

Loo, B.W., R.C. Gatti, B.Y.H. Liu, C-S Kim, and T.G. Dzubay, Absorption Corrections for Submicron Sulfur Collected in Filters, in *X-Ray Fluorescence Analysis of Environmental Samples*, T.G. Dzubay, ed, Ann Arbor Science Publishers Inc., Ann Arbor, MI, 187-202 (1977).

Lowenthal, D.H. and K.A. Rahn, Comments on Use of Whatman 41 Filter Papers in Particle Sampling, *Atmospheric Environ.* 21:2732-2734 (1987).

Mamane, Y. and T.G. Dzubay, Dichotomous Samplers Modified for Use with Electron Microscopy, *Aerosol Sci. Technol*, 13:241-248 (1990).

Mulik, J.D. and E. Sawicki, *Ion Chromatographic Analysis of Environmental Pollutants*, Vol. 2, Ann Arbor Science Publishers, Inc., Ann Arbor, MI, 203-209 (1979).

Nelson, J.W., Proton-Induced Aerosol Analyses: Methods and Samplers, in *X-Ray Fluorescence Analysis of Environmental Samples*, T.G. Dzubay, ed, Ann Arbor Science Publishers Inc., Ann Arbor, MI, 19-34 (1977).

Ondov, J.M., W.H. Zoller, I. Olmez, N.K. Aras, G.E. Gordon, L.A. Rancitelli, K.H. Abel, R.H. Filby, K.R. Shah, and R.C. Ragaini, Elemental Concentrations in the National Bureau of Standards' Environmental Coal and Fly Ash Standard Reference Materials, *Anal. Chem.* 47:1102-1109 (1975).

Possanzini, M., A. Febo, and A. Liberte, New Design of a High-Performance Denuder for the Sampling of Atmospheric Pollutants, *Atmospheric Environ.* 17:2605-2610 (1983).

Purdue, L.J., C.E. Rodes, K.A. Rehme, D.M. Holland, and A.E. Bond, Intercomparison of High-Volume PM_{10} Samplers at a Site with High Particulate Concentrations, *J. Air Pollut. Control Assoc.* 36:917-920 (1986).

Purdue, L.J., EPA PM_{10} Methodology Requirements, in *PM-10: Implementation of Standards*, C.V. Mathai and D.H. Stonefield, eds, Transactions No. TR-13, Air & Waste Management Association, Pittsburgh, PA, 85-92 (1988).

Rodes, C.E., D.M. Holland, L.J. Purdue, and K.A. Rehme, A Field Comparison of PM_{10} Inlets at Four Locations, *J. Air Pollut. Control Assoc.* 35:345-354 (1985).

Rogers, C.F., J.G. Watson, and C.V. Mathai, Design and Testing of a New Size Classifying Isokinetic Sequential Aerosol Sampler (SCISAS), *J. Air Pollut. Control Assoc.* 39:1569-1576 (1989).

Slater, J.L., P. Koutrakis, G.J. Keeler, and J.M. Wolfson, Determination of Aerosol Strong Acidity Losses Due to Interactions of Collected Particles: Results from Laboratory and Field Studies. In *EPA/AWMA Symposium on Measurement of*

Toxic and Related Air Pollutants, Proceedings No. VIP-13, Air & Waste Management Association, Pittsburgh, PA, 248-255 (1989).

Speitel, L.C., J.C. Spurgeon, and R.A. Filipczak, Ion Chromatographic Analysis of Thermal Decomposition Products of Aircraft Interior Materials, in *Ion Chromatographic Analysis of Environmental Pollutants*, Vol. 2, J.D. Mulik and E. Sawicki, eds, Ann Arbor Science Publishers Inc., Ann Arbor, MI, 75-87 (1979).

Stevens, R.K., T.G. Dzubay, G. Russwurm, and D. Rickel, Sampling and Analysis of Atmospheric Sulfates and Related Species, *Atmospheric Environ.* 12:55-68 (1978).

Stevens, R.K., T.G. Dzubay, R.W. Shaw, W.A. McClenny, C.W. Lewis, and W.E. Wilson, Characterization of the Aerosol in the Great Smoky Mountains, *Environ. Sci. Technol.* 14:1491-1498 (1980).

Verkouteren, R.M., G.A. Klouda, L.A. Currie, D.J. Donahue, A.J.T. Jull, and T.W. Linick, Preparation of Microgram Samples on Iron Wool for Radiocarbon Analysis Via Accelerator Mass Spectrometry: A Closed-System Approach, *Nucl. Instrum. Methods* B29:41-44 (1987).

Vossler, T.L., R.K. Stevens, R.J. Paur, R.E. Baumgardner, and J.P. Bell, Evaluation of Improved Inlet and Annular Denuder Systems to Measure Inorganic Air Pollutants, *Atmospheric Environ.* 22:1729-1736 (1988).

Wagman, J., R.L. Bennett, and K.T. Knapp, Simultaneous Multiwavelength Spectrometer for Rapid Elemental Analysis of Particulate Pollutants, in *X-Ray Fluorescence Analysis of Environmental Samples*, T.G. Dzubay, ed, Ann Arbor Science Publishers Inc., Ann Arbor, MI, 35-55 (1977).

Zoller, W.H. and G.E. Gordon, Instrumental Neutron Activation Analysis of Atmospheric Pollutants Utilizing Ge(Li) Gamma-Ray Detectors, *Anal. Chem.* 42:257-265 (1970).

Chapter 3

SOURCE SAMPLING FOR RECEPTOR MODELING

James E. Houck
OMNI Environmental Services, Inc.
10950 S.W. Fifth Street, Suite 160
Beaverton, Oregon 97005

1. INTRODUCTION

The receptor modeling scientific community is in agreement that one of the largest impediments to receptor modeling today is the dearth of accurate, precise, and comparable chemical profiles ("fingerprints") for major particulate emitters (Javitz et al., 1988). These source profiles are needed quantitatively as input for the chemical mass balance (CMB) receptor model, and they are needed qualitatively by the principal components and multiple linear regression receptor models. Receptor models for particulate source apportionment require compatible ambient and source data sets. While techniques for ambient sampling and subsequent analyses have been well established and are nearly routine, sampling and analysis of particulate sources require customized procedures for nearly every source type and, in some cases, are virtually source-specific. In an effort to improve the quality of source data sets and hence the overall quality of receptor modeling results, the requirements of receptor modeling have been reviewed and, in response to these requirements, source sampling instruments, protocols, and analyses have been developed (Core and Houck, 1987; Houck *et al.*, 1989a; Watson *et al.*, 1988).

Particulate sources can be categorized into four groups for the purposes of sampling: (1) high-temperature point sources, (2) low-temperature point sources, (3) process fugitive sources, and (4) passive fugitive sources. High-temperature point sources include industrial-type "smoke-stacks" and can include such sources as vehicular exhaust and woodstoves which are often considered area sources in emission inventories. Low-temperature sources are, as the name implies, simply ducted sources which have temperatures within a few tens of

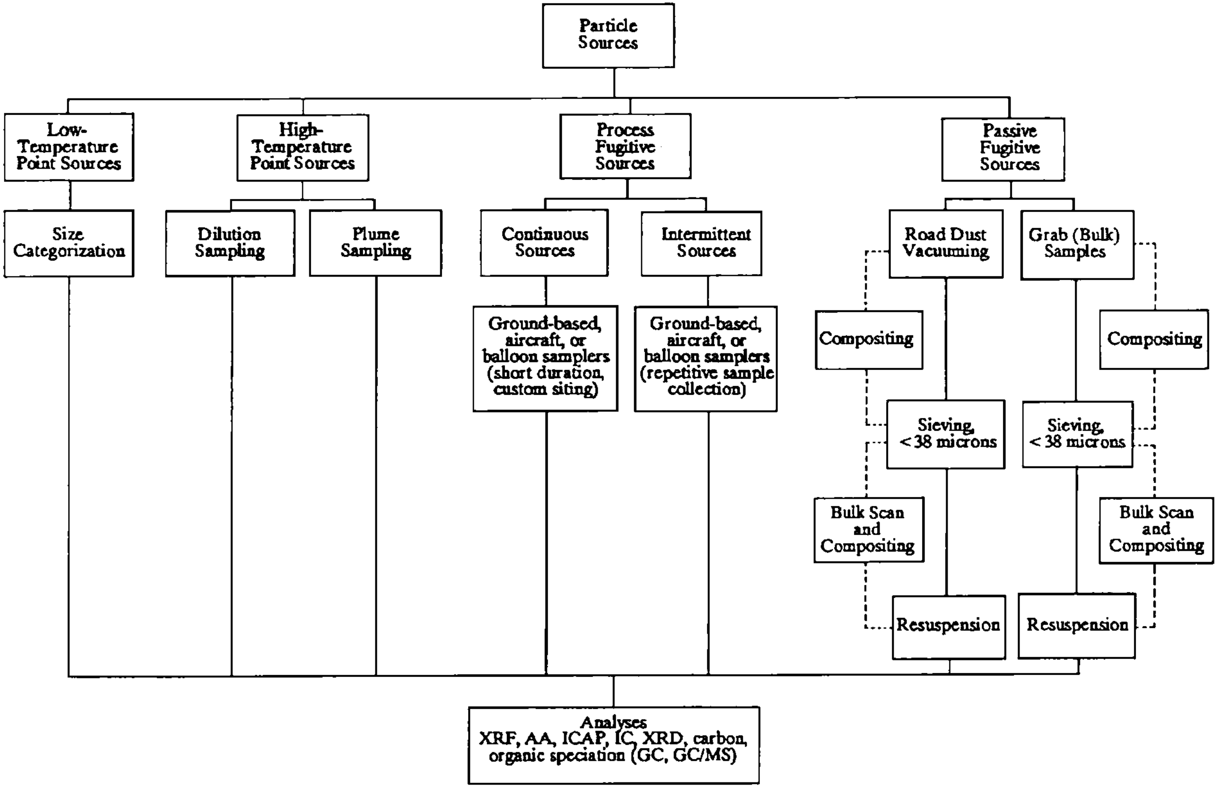

Figure 1. Flow Diagram of Particulate Source Sampling Techniques.

degrees of ambient temperature. The process fugitive category covers a wide range of industrial, agricultural, forestry, and area sources. Examples of process fugitive emission sources include field-burning, slag-pouring, unducted industrial roof-top emissions, construction and demolition activities, stockyards, ore crushing, and forest fires. The passive fugitive category includes wind-blown dust from roadways, parking lots, agricultural fields, dry lake beds, desert regions, and industrial storage piles. The passive fugitive category also includes dust suspended by vehicular traffic simply because sampling procedures used for its collection are similar to those used for wind-blown dust.

A number of methods have been used to obtain samples with chemical and physical properties similar to those found at a receptor (Core and Houck, 1987). Inherent in these methods is the requirement that the samples be collected on substrates that are compatible with ambient (albeit modified) methods of chemical analyses. Figure 1 is a flow diagram of the most appropriate source sampling techniques for the major particulate source categories. For high-temperature point sources, either sampling with a dilution/cooling system (a plume simulator) or plume sampling are the state-of-the-art techniques. For low-temperature point sources, source sampling is much simpler and size-categorization is generally the only principal concern. Both ground-based and airborne samplers are appropriate for process fugitive emissions. A major subdivision in the process fugitive category is between sources with continuous or near-continuous emissions and those which are intermittent. A variety of specially designed fugitive samplers as well as commercially available ambient samplers have been successfully used for collecting samples from process fugitive sources. The key considerations for sample collection are sampler siting, sampling duration, composting emissions variability, and sample contamination by other nearby sources. The final source category, passive fugitive emissions, does not require sophisticated instrumentation to obtain sample material. However, a key and often complex consideration in passive fugitive source sampling is the selection of appropriate samples and compositing schemes to produce chemical profiles representative of the material impacting the receptor sites.

Descriptions of source sampling strategies, instrumentation, protocols, and data management procedures are provided in this chapter. In addition, representative source profiles from selected sources are presented.

2. HIGH-TEMPERATURE POINT SOURCES - DILUTION SAMPLING

High-temperature point source emissions represent a special problem for chemical receptor model source sampling. Condensation, vaporization, agglomeration, and secondary chemical reactions can all alter particulate chemistry and size distribution when emissions cool and mix with ambient air. These alterations in chemistry and size reduce the validity of using source profiles obtained from in-stack samplers or from more traditional U.S. Environmental Protection Agency (EPA) compliance sampling equipment (e.g., U.S. EPA Method 5). The desire to obtain samples of particles in the form that they will have after they are emitted into the atmosphere has stimulated research into the development of dilution/cooling systems (Houck *et al.*, 1982a; Pan, 1986).

The principal objective of dilution sampling is to obtain chemical data representative of particulate emissions after mixing and cooling with the atmosphere has occurred. Unlike more traditional source sampling, in-stack gravimetric emission rates are only of secondary importance. This fundamental difference in objectives requires modification of criteria which are typically considered most important for obtaining "good" stack test data.

Dilution samplers draw hot exhaust gases into a chamber where they are mixed with filtered ambient air. After an aging period, the particles are drawn through a size-selective inlet and onto the filter substrates for subsequent analysis. Particulate material on multiple substrates for different chemical analyses can be obtained simultaneously or through sequential sampling of the same gas stream. Houck *et al.* (1982a) have developed a dilution system which draws the diluted sample through an impactor or virtual impactor to provide particle size fractionation. McCain and Williamson (1984) performed tests on this sampler which showed losses of large particles owing to inertial impaction and electrostatic charging. They recommended design changes to minimize these losses and these changes have been implemented in the more recent design presented here (Houck *et al.*, 1989a; Watson *et al.*, 1988). Harris (1986), Huynh *et al.* (1984), Heinsohn *et al.* (1980), Stiles (1983), and Cooke *et al.*, (1984) offer variations of the dilution sampler approach.

Figure 2 is a schematic of a dilution source sampling system that has been used extensively (Goulet *et al.*, 1989a, 1989b; Houck *et al.*, 1981, 1982b, 1989a, 1990; Olmez *et al.*, 1988; Watson *et al.*, 1988). Its actual physical appearance varies, as several different dilution chamber and inlet geometries are necessary to position the sampler adjacent to each specific source since it is desirable to minimize the inlet probe length. It has been found that the principal point of particle loss is within the sampling probe and inlet line (McCain and Williamson,

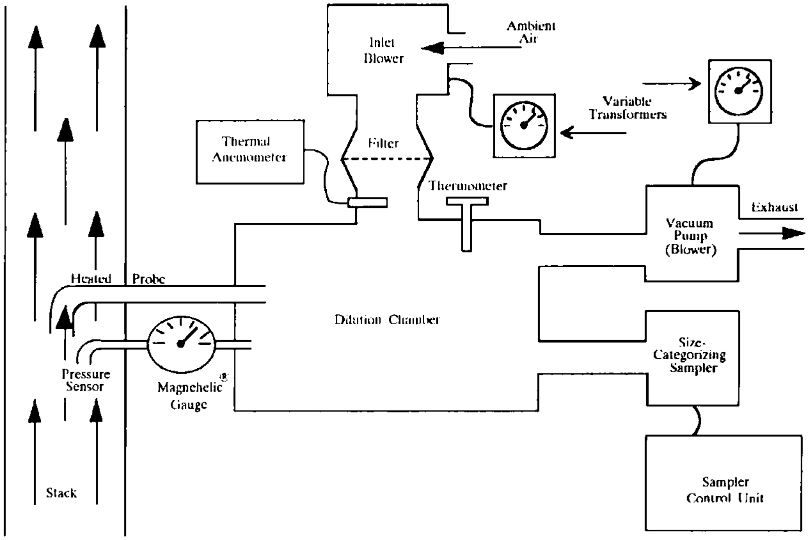

Figure 2. Schematic of a Dilution/Cooling System.

1984). The system was designed to be "broken down" to be easily transported and cleaned in the field. The system also has interchangeable dilution chamber lengths and "elbows", as well as various diameters and lengths of inlet probes, so that the system can be used under various spatial restrictions. The dilution chamber components are constructed of light gauge stainless steel to minimize sample contamination and weight.

Characteristic temperatures, flow rates, particulate concentration, and water vapor content (condensed water is deleterious to sample collection) vary dramatically with source type; consequently, the dilution ratio is adjustable (approximately 10:1 to 100:1) for general application. Additionally, because the sampler is often inherently in an area of high ambient particulate concentration, and because of the high ambient air/sample ratio, the dilution air needs to be well filtered to prevent sample contamination. Inlet air is filtered with a standard high-volume 8-by-10-inch filter. The dilution ratio is adjustable at any reasonable inlet flow by the combined control of an inlet blower and outlet vacuum pump. Both can be controlled by variable transformers. The dilution ratio can be set at any predetermined value, since the inlet airflow rate is monitored with a thermal anemometer and the pressure difference between the interior of the dilution chamber and the source is monitored with a pressure gauge or manometer. The flow-versus-pressure difference relationship is determined in the laboratory prior to field deployment. The dilution chamber temperature is monitored to ensure that the chamber temperature is a few degrees within ambient, and for documentation of the aerosol sampling environment. One or more size-categorization samplers withdraw samples from the dilution chamber; if two are used, generally one is for Teflon filters and one is for quartz filters. (Teflon membrane filters are well-suited for elemental analyses. Quartz filters are appropriate for organic compound, elemental carbon, and wet chemical analyses.) The diameter of the inlets of the collimating tubes can be restricted to achieve isokinetic sampling conditions. This is generally not essential, since the majority of the particles originating from most high-temperature sources are less than 5μ in aerodynamic diameter. Similarly, the flow in the inlet to the dilution chamber can be adjusted to remove the aerosol from the source isokinetically, although as mentioned, it is not critical for most high-temperature sources. Table 1 is a summary of dilution sampler design and operation parameters.

The transfer of particulate-bearing stack gases through the heated probe to the dilution chamber is accomplished by maintaining a pressure differential between the dilution chamber and the interior of the stack. From Bernoulli's equation of continuity, it can be shown that the linear velocity of gas entering the inlet is dependent only on the pressure drop (ΔP) and the density of the source gas (ρ), i.e.,

$$V = \sqrt{\Delta P/0.5\rho} \tag{1}$$

Bernoulli's equation is only strictly applicable to idealized fluids but is illustrative for design considerations. Since the inlet will collect gas parallel with the direction of flow, the pressure value used to calculate ΔP in Equation 1 must take into account the effect of velocity pressure, i.e.,

$$\Delta P = [P_{S,S} + 0.5\rho_S V_S^2] - P_{d,S} \tag{2}$$

where $P_{S,S}$ is the static pressure within the source;

ρ_S is the density of gas within the source;

V_S is the linear velocity of gas within the source; and

$P_{d,S}$ is the static pressure within the dilution chamber.

Measurement of ΔP can be accomplished with a manometer or Magnehelic® gauge.

Reduced pressure and flow within the dilution chamber is produced by an outlet vacuum pump (blower). If the inlet blower is removed, each flow rate across the 8-by-10-inch inlet filter has a corresponding pressure drop associated with it which is determined by the filter medium and the outlet vacuum pump. The addition of a variable inlet blower reduces the pressure drop and permits a wide range of combinations of dilution chamber pressure and flow rate. For example, if

Table 1. Dilution Sampler Design and Operation Parameters

Parameter	Explanation
Variable dilution ratio	Response to variable temperature, water content, and particulate concentration
Size-categorized samples	PM_{10}, $PM_{2.5}$, TSP
Filtered ambient air	High ambient air to stack gas ratio, ambient air, with high particulate concentration
Multiple simultaneous samples	Multiple size ranges, multiple filter substrates for different analytical techniques
Integration of emission variability	Batch process: ability to slow sampling rate or intermittently sample
Material compatibility	High-temperature corrosive gases, weight of sampler
Heated inlet	Prevent condensation of particulate material and water vapor before reaching dilution chamber
Isokinetic sampling	Accurate size representation
Stack gas physio-chemical measurements	Gas composition, water vapor content, gas velocity, particulate concentrations.
Variable chamber geometry and ease of disassembly/assembly	Spatial restrictions, field cleaning

a high dilution flow rate (i.e., high dilution ratio) and a low pressure drop (low linear velocity in the sampling inlet) are desired, the vacuum pump would be operated at near-maximum power and the inlet blower would be adjusted until the pressure drop across the high-volume filter was lowered to the point where low inlet velocities were obtained.

Some limited source data are generally collected prior to sample collection. Stack flow rate, temperature, water vapor content, and particulate concentration are helpful in estimating appropriate dilution ratios, for selection of inlets, and in estimating the duration of sample collection. Adequate data are frequently obtained from records of previous compliance tests or typical characteristics of the source type. Sampling periods are generally between 15 minutes and 2 hours. Sampling periods as short as five minutes have been encountered (coal-fired power plant operating without emission controls), and in some cases periods as long as 14 hours (efficient baghouses) have been necessary. In most cases the proper mass loading on the filters (approximately 0.4 to 2 mg for XRF, carbon, and IC analyses) dictates the length of the sampling period required. However, in some cases a long sampling duration has been required because a very high dilution ratio was needed to prevent water condensation from occurring.

Besides direct in-field sampling, dilution sampling lends itself to laboratory simulations of emissions from individual sources. Dynamometer simulations of motor vehicle driving with exhaust sampled from a dilution tunnel can provide examples of aggregate emissions from a large number of separate vehicles. Similarly, wood stoves and fireplaces can be operated under different conditions with emissions sampled from a dilution tunnel. Because of their ubiquity and the magnitude of particulate contribution to many airsheds, the chemical quantification of the particulate emissions from residential wood combustion (RWC) and vehicles has been and will continue to be necessary for receptor modeling.

3. RESIDENTIAL WOOD COMBUSTION

Residential wood combustion appliances present special problems for dilution sampling, and producing representative source profiles is a complex task due to the inherent number of variables associated with them. Notable among these are (1) appliance types and installation factors, (2) fuels, (3) fueling practices, and (4) burn conditions. Table 2 summarizes these and other variables.

To further complicate the development of representative source profiles, woodburning appliances are difficult to sample because (1) the emissions are tar-like, (2) the average stack gas velocity is low, (3) the average concentration of particulate material in the stack gas is very high, (4) there is a high water vapor content in the stack gas, and (5) emission rates and gas velocities are very variable. In addition, the difference between particulate (solid and liquid) emissions and gaseous emissions is a matter of definition, since many of the chemical compounds contained in wood smoke are semi-volatile. Woodburning appliances are also pragmatically difficult to sample as their stacks obviously have no sampling ports and in-field sampling requires the positioning of heavy, cumbersome, and noisy sampling equipment in private residences.

Table 2. Residential Wood Combustion Appliance Variables.

Source Type	Key Variables
Appliance types and installation factors	Woodstove, fireplaces, pellet stoves Woodstove firebox size Model (fundamental design) Woodstove technology type (catalyst versus noncatalyst) Damper (draft) control (excess air) Airtight versus non-airtight woodstove (excess air) Chimney system (draft)
Fuels	Species Moisture content Seasoned versus non-seasoned Size of fuel pieces Density Extent of decomposition
Fueling practice	Burn rate Burn duration (all day versus evenings only) Fuel load amount Frequency of fueling Kindling (start-up) procedure Household trash
Burn conditions	Kindling phase Main burn (dampered-down cool burn versus hot burn with excess air) Charcoal phase (end of burn) Damper (draft) settings

Due to the inherent variability among woodburning appliances and the difficulty of obtaining samples in the field from a meaningful number of appliances, a laboratory sampling methodology has been developed (Houck *et al.*, 1989a, 1989d; Watson *et al.*, 1988). The approach entails (1) determination of the most abundant cordwood type(s), burn rates, appliance types, and burn cycles for a given geographical area from existing literature, surveys, or from other studies; and (2) long-term, in-laboratory sampling of emissions from the most representative woodburning appliance types, operating under the mean burn rate and most common cycles, and using the most abundant wood type(s).

Since particulate emission rates, stack velocities, and apparently the chemical character of particles, vary dramatically over the course of a normal burn cycle (Burnet *et al.*, 1986; Shelton and Gay, 1986), the dilution sampler system used for industrial sampling is not appropriate for residential woodburning appliance application. Rapid changes in velocity and particulate loading levels over short time periods would be nearly impossible to follow in a proportionate manner with the industrial dilution sampler. In addition, the very high particulate concentration characteristic of residential wood combustion as compared to most controlled industrial sources makes the direct use of the industrial dilution sampler less than ideal.

To rectify these problems, a modified U.S. EPA reference Method 5G dilution tunnel system (U.S. Environmental Protection Agency, 1987) has been used to dilute the entire appliance emissions with ambient air. From the diluted stream a fixed-flow aliquot can be removed and passed through a size-categorizing sampler. Figure 3 is a schematic of the tunnel system. The most significant modification made in the Method 5G protocol is the increase in flow rates when fireplaces are tested to compensate for increased stack gas volumes

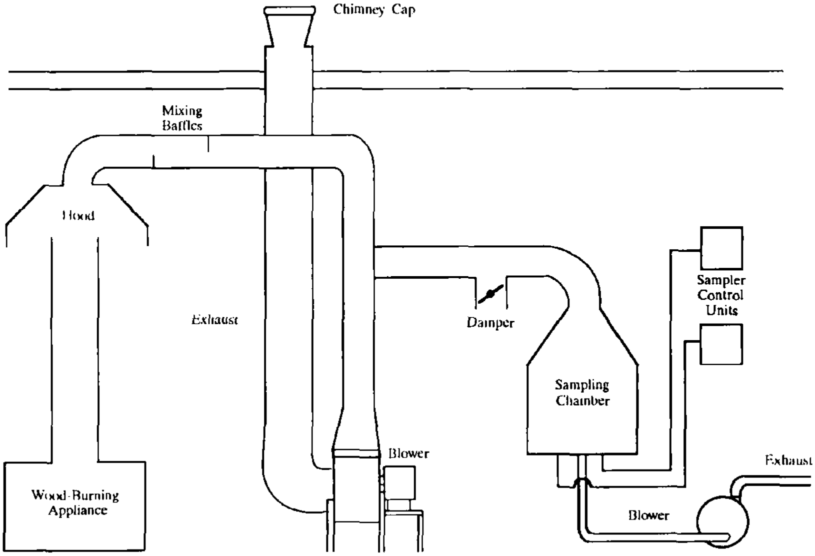

Figure 3. Schematic of Modified Method 5G Dilution tunnel Interfaced with a Size-Categorizing Sampler.

characteristic of residential fireplaces as compared to woodstoves (Shelton and Gay, 1987). Since unfiltered ambient air is used for dilution/cooling, particulate samples of the dilution air need to be taken simultaneously with the source tests so that the contribution of the dilution air can be accounted for in the chemical composition of the samples collected. Ground-based size-categorizing samplers can be used to sample the dilution air.

The moisture content of the cordwood should be measured with a commercial moisture meter. A spring scale can be used to pre-weigh appropriate amounts of each species of wood prior to the beginning of a test. Wood addition (target and actual), wood moisture content, wood species, and draft control/door positions should be taken into consideration in the test design.

4. HIGH-TEMPERATURE POINT SOURCES - PLUME SAMPLING

For some high-temperature point sources it is possible to collect plume samples which, in effect, permit dilution and cooling to occur naturally. An inherent problem associated with this approach is the presence of background pollutants in the ambient air. It must be confirmed that the impact of background air for all chemical species of interest is negligible compared to the concentrations originating from the point source. For large area sources and stacks, airborne sampling is required. Source sampling from airborne platforms to characterize the chemical and physical properties of emissions has been performed from airplanes (Small *et al.*, 1981; Richards *et al.*, 1981, 1985), tethered balloons (Armstrong *et al.*, 1981; Shah *et al.*, 1988) and helicopters. It has also been proposed that model airplanes be used to carry ultra-light sampling payloads. Sampling components of appropriate weight and packaging are elevated above the emissions, usually on the order of 100 to 500 meters, to draw samples of the effluent. Airborne sampling is not practical in most cases, due to its high cost and unproven techniques. Plume sampling is more appropriate for small industrial stacks and sources such as residential wood combustion and diesel train exhaust (Houck *et al.*, 1987).

5. LOW-TEMPERATURE POINT SOURCES SAMPLING REQUIREMENTS

Emissions from point sources which are near ambient temperature generally do not require dilution sampling, as condensation of particles is not an issue and low temperatures permit direct collection onto ambient filter media. Sampling of low-temperature point sources is relatively simple compared to high-temperature sources. In some cases, however, a dilution sampling system still needs to be used with low-temperature point sources due to a high condensed water droplet content which can be deleterious to samples collected on filter substrates. The ambient air introduced by the dilution system permits the water droplets to evaporate prior to sample collection.

For collection of particulate samples from low-temperature point sources, a size-categorizing device is positioned into the flow stream; isokinetic sampling is approximated (generally by inlet diameter, since size-categorizing devices require a fixed flow); and a flow collimating tube (30 cm or longer) is used with an inlet diameter that is the same size as the exposed face of standard filters (e.g., 37- or 47-mm diameter) which are being employed. This ensures even loading across the face of the filter, which is necessary for most analytical techniques in which a small portion of the filter is analyzed and assumed to be representative of the entire filter surface. Figure 4 is a schematic of an example low-temperature point source sampler. Integration of process variability can be accomplished with a cycle timer.

6. MOBILE SOURCES

Mobile source particulate emissions are difficult to measure with respect to emission rate and chemical composition. This difficulty arises from (1) the large number of mobile source types (passenger cars, light duty trucks, diesel trucks, diesel buses, trains, aircraft, motorcycles, etc.); (2) a large number of individual emitters within each types; (3) changing fuel characteristics; (4) a large variety of undefined (and probably undefinable) operating conditions; (4) a variety of emission points on each vehicle (e.g., tail pipe, resuspended dust, evaporation, tire wear, brake wear); and (6) a large fraction of emissions which are semi-volatile or are secondary particulate precursors. This discussion will focus primarily on primary particulate emissions from the engine exhausts. Emissions from overall vehicle activities can best be sampled in tunnel-like studies using techniques discussed for process fugitive emissions (section 7).

The most widely used source profile for apportioning ambient particulate matter concentrations to motor vehicles is an emissions inventory-weighted composite of chemical compositions in different vehicle categories (Core, 1986). This profile is becoming of limited utility owing to great reductions in the lead content of leaded fuels and to the preponderance of vehicles which operate on unleaded fuels. In additional, elevation and ambient temperatures have been shown to significantly impact emissions requiring area-specific source profiles to be developed.

Vehicles must be grouped into categories both for source characterization purposes and in order to composite source profiles which are too similar to allow their source compositions to be resolved. For example, in a recent study in the Denver, Colorado area (Watson *et al.*, 1988), the categories were divided into (1) leaded gasoline, (2) unleaded, non-oxygenated gasoline with oxidation and three-way catalysts, (3) unleaded oxygenated gasoline with oxidation and three-way

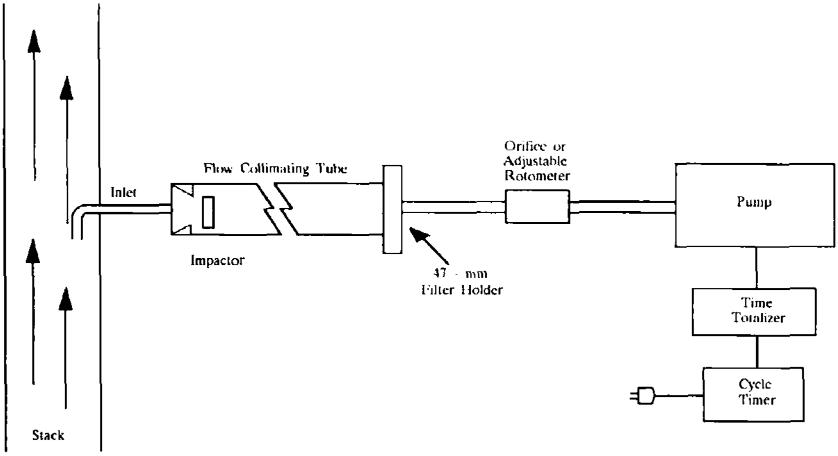

Figure 4. Schematic of Low-Temperature Point Source Sampler.

catalysts, (4) diesel passenger cars, (5) heavy duty diesel trucks, and (6) heavy duty diesel buses.

Even through the lead emitted by vehicles using leaded gasoline is much lower than it was previously, and even though the percentage of the entire fleet they occupy is much lower, they still provide the best indicator of overall motor vehicle contributions since they are the major source of lead in urban atmosphere. Though the number of leaded vehicles is lower than the number of unleaded vehicles, these older vehicles have fewer emission controls and higher particle emission rates.

Diesel trucks generally have four stroke engines, while diesel buses used for urban transportation have two stroke engines. The two stroke engines have higher mass and organic emissions and have ratios of other chemical species which differ from those of four stroke engines. Motorcycles have two stroke engines, but their population is generally low when compared to other vehicles.

Any tests of motor vehicle emissions should provide a sampling of vehicles from each of these categories. Naturally, large variations are expected based on vehicle age, maintenance, model, and many other differences between individual vehicles.

The most important fuel characteristic with respect to exhaust emissions is its chemical composition. Lubricating oils are included with diesel fuel and gasoline in this discussion. Fuels of the same type differ in composition because of (1) different refining techniques, (2) cross-contamination among fuels in storage and transport, and (3) different additives.

A National Petroleum products survey is published twice a year for gasoline and once a year for diesel fuel. This survey identifies the additives reported by different refiners by geographical area. While this is a good guide, it may not reflect the actual composition of the fuel. The best approach to fuel characterization is to use the survey as a guide to differences among fuels sold in an area for the period of time closest to ambient sampling for receptor modeling. Samples from different gas stations should be taken and analyzed for elemental and organic content to determine what species might be present.

Both the quantity and chemical character of motor vehicle exhaust emissions change with different operating conditions. This is further complicated by the fact that emissions at a given operating condition are often functions of pervious operating conditions.

On catalyst-equipped cars, emission rates are highest and have their largest organic content when the car is first started. In modern spark ignition passenger vehicles, hydrocarbons are not properly converted until the temperature of the catalyst reaches its specified operating condition. In some catalytic converters, the aluminum oxide absorbs sulfur (as SO_3) at low speeds, only to release it at higher speeds. Several thousand miles of emissions can be stored with low-speed driving, only to be emitted rapidly under fast highway driving conditions.

Lead emissions increase with vehicle speed, both because more lead is being burned and because the higher exhaust velocities sweep deposited lead out of the exhaust system. For diesel systems, particulate emission rates increase substantially when the vehicles accelerate and the proportion of organic carbon decreases.

Changing operating conditions are generally handled in source tests through a driving cycle of starting, stopping, accelerating, cruising at various velocities, and idling. A large number of driving cycles are simulated in dynamometer studies. For example, the LA cycle simulates congested freeway driving. The New York City cycle simulates east/west stop-and-go driving in Manhattan. The Federal Test Procedure (FET), Sulfate Emissions Test, Highway Fuel Economy Test (HWFET), Motor Vehicle Manufacturers Association (MVMA) Test, and various steady state driving cycles have also been devised. None of these cycles fully represents changes in emissions characteristics. They do, however, have the advantage of providing a reproducible testing method.

Dilution and ground-based sampling have been applied to motor vehicle source characterization. Each of these methods has distinct advantages and disadvantages, and they are more complementary than exclusive of each other.

Dynamometer tests are conducted in a controlled laboratory setting. Each vehicle tested is run through a series of different operating conditions which are intended to simulate different types of driving. The exhaust from these vehicles is generally diluted prior to sampling. Detailed procedures have been established for NO_x, CO, hydrocarbon, and particulate mass measurements on dynamometers. Detailed chemical and size speciation of particles and their precursors have been accomplished by a number of different procedures. The common elements of successful dynamometer chemical measurements are:

- All of the exhaust is diluted prior to sampling. The volumetric flow of exhaust varies throughout the cycle, but the flow of diluted exhaust (i.e., exhaust plus dilution air) is maintained constant.
- Dilution ratios are around ten to one. This generally cools the exhaust to approximately 37° C, which is still higher than normal ambient temperature.
- Two tests on each vehicle are run and compared for equivalency prior to validating the data.
- Dilution tubes are usually made of stainless steel. It is good practice to employ separate dilution tubes for diesel and spark-ignition vehicles to avoid cross-contamination (e.g., the elemental carbon from diesel exhaust may deposit in the dilution tube and adsorb other species).

In the ideal case, every motor vehicle would be tested under every possible driving cycle. Source profiles would then be constructed from weighted averages of these individual profiles. The weights would be derived from the emissions per mile times the miles travelled by each vehicle and the frequency distribution of the different operating cycles. This is not a practice approach. A smaller number of vehicles which is statistically representative of a larger population needs to be selected. The number of vehicle tests which are needed to represent an entire fleet is a question that is difficult to answer, and is generally controlled by budget rather than technical considerations.

The advantages of dynamometer testing are:

- A number of different vehicle types and defined operating conditions can be tested; and

- Motor vehicle exhaust can be separated from other motor vehicle emissions.

The disadvantages are:

- Dilution ratios are relatively small and ambient temperatures are not achieved.
- Residence times are on the order of only a few seconds and chemical reactions may not have taken place.
- Driving cycles are not necessarily representative of actual operating conditions.
- There is no opportunity for interaction among different exhausts or with the ambient aerosol.
- These tests are generally expensive.
- The worst-maintained vehicles (which are generally the highest pollutant emitters) are rarely available for such tests.
- A large number of test results are necessary to form a composite source profile.

Ground-based measurements of motor vehicle emissions are generally acquired from tunnels, major highways, and chase-car sampling. Other ground-based sampling settings have included truck weigh-stations, bus parking lots, underground bus transfer stations, and underground parking garages (Houck *et al.*, 1989a; Watson *et al.*, 1988). The most commonly used source profiles in CMB calculations are traceable to tunnel samples, especially those performed by Pierson and Brachaczek (1976, 1983). Pierson *et al.* (1978) devised a linear regression technique which separates emission rates and compositions into two categories, usually gasoline- and diesel-powered vehicles. A series of chemical concentration measurements and vehicle counts are required.

Each tunnel study must be designed specifically for the tunnel which is used. Desirable tunnel characteristics include (1) greater than 1000 m length, (2) two lanes or wider, (3) a place to mount sampling instruments, (4) ventilation and exhaust systems, and (5) minimal changes in driving patterns for periods which exceed minimum sample durations. Vehicle characterization using a tunnel should contain the following elements:

- Air flow rates into and out of the tunnel are measured at all points. Wind speed traverses can be measured just inside the entrance and exit portals to determine homogeneity. Intake and exhaust flow rates are generally well specified and monitored.
- Traffic volume is counted for each sampling period within the classes of diesel trucks, diesel buses, diesel passenger cars, gasoline trucks, and gasoline passenger cars. It is difficult to distinguish among leaded, unleaded, and different catalyst categories, through such a division would be desirable if it were practical.
- Approximate vehicle speed is estimated.
- Geological material is sampled from roads, shoulders, and soils at the tunnel entrance and exit, and inside the tunnel. These compositions are used to determine soil entrainment contributions from motor vehicle movements. Brake-shoe and tire residue are obtained from a selection of vehicles to estimate these compositions. Semi-metal brake linings are suspected of being a source of several metals found in the Sepulveda tunnel in the South Coast Air Basin (Cooper *et al.*, 1987).
- Air at the intake fans and/or entrance portals is sampled to obtain ambient concentrations for subtraction.
- Sampling periods are chosen to represent different proportions of vehicle categories (e.g., diesel trucks often comprise a larger fraction on weekdays than they do at night or on weekends).
- Species and size fractions are chosen to correspond to those of the ambient particulate data. This assures that the source profile is compatible with receptor measurements in CMB analysis.

The advantages of tunnel measurements are:

- A tunnel presents the closes approximation to the real world in a controlled environment.
- The dilution of exhaust and the temperatures obtained are realistic. Condensible species will be measured as particulate matter.
- Exhausts from different vehicles are mixed so that fast chemical reactions (several minutes) can take place.

- All motor vehicle emissions, not just motor vehicle exhaust, can be estimated.
- A mixture of vehicles which is somewhat similar to the entire fleet is obtained. Since a significant number of these vehicles is probably badly maintained, or may have had their emissions controls disengaged, the mix is more realistic than that obtained from "randomly" sampling cars in laboratory tests.
- Measurement programs are relatively inexpensive compared to other alternatives.

The disadvantages of tunnel sampling are:

- The driving conditions, when they can be meaningfully defined at all, are generally limited to a high speed, steady state condition. Many operating variables are not represented by this condition.
- The distribution of vehicles may not be the same as that which is contributing most heavily at the receptors.

Roadside sampling is nearly identical to tunnel sampling, but in a less controlled environment. Typically, ambient sampling equipment is placed in an area highly dominated by traffic and samples are taken during periods when contributions from other sources are presumed to be at a minimum. Local geological material is sampled and analyzed to account for resuspended dust. This method is probably less expensive than a tunnel study, but it is also much less accurate. This method has not been adequately developed or evaluated for the measurement of receptor model source profiles. Dzubay *et al.* (1979) provide an example of this method in Southern California.

In chase car sampling, a sampling device is located on a vehicle which follows the vehicle being tested. The inlet to the sampler and the distance between cars is adjusted to maximize the quantity of exhaust collected and to minimize contamination from other sources. A "self-chasing" sampler has been developed at the Warren Springs Laboratory which can be located in the truck on a car and takes samples from the exhaust of the same car. The chase car methods have not been adequately developed or evaluated for the measurement of receptor model source profiles.

While both ground-based sampling and dilution sampling using a dynamometer have their limitations, a carefully designed study can develop adequate mobile source profiles using either approach. Because they are complementary in many ways, state-of-the-art profiles have been developed using a combination of both techniques. In a recent study conducted in Denver, Colorado, results of source sampling with a dilution tunnel using a dynamometer were combined with ground-based sampling results from an Interstate Highway tunnel, three underground apartment garages (to obtain cold-start profiles), and from a municipal bus transfer station tunnel (Watson *et al.*, 1988, 1990).

7. PROCESS FUGITIVE EMISSIONS

Process fugitive emissions are defined for the purposes of sample collection techniques as those emissions which are produced by active processes, in contrast to passive fugitive emissions, which are produced by wind or vehicular traffic. There is generally enough localization to most process fugitive emissions sources so that their particles can be sampled near their point of entry into the atmosphere. Examples of constant process fugitive sources are industrial roof vents, material handling areas, milling and grinding operations, open agricultural or forestry burning, stockyard dust, slag pouring, and construction and demolition activities. Process fugitive emissions sources can be sub-categorized into those which have continuous or near-continuous emissions and those with intermittent emissions.

Typically, the most appropriate technique for the sampling of continuous process fugitive sources is to utilize a ground-based sampler. In some cases, ultra-lightweight samplers tethered to helium balloons have been used (Armstrong *et al.*, 1981; Shah *et al.*, 1988), as have aircraft. With several exceptions, ground-based fugitive source samples are collected in an analogous fashion to ambient samples and in general the strategy for ground-based sampling is relatively simple. The positioning of the ground-based sampler inlet in the plume or path of source particles is often accomplished with custom-made brackets, extension tubes, or platforms. In some cases, simply situating the samplers on a roof top or on a topographic high point is adequate. In each case, the samplers are situated in the aerosol plume, with care being taken to minimize contamination from other sources by carefully selecting the sampling location and time. If no power is available at the sampling site, as may be the case on a moving crane platform over an active work area or near a vegetative burning area, battery-powered stacked filter units (Cahill *et al.*, 1979; Houck and Pritchett, 1985) or battery-powered impactor systems (Shah *et al.*, 1988) can be used to collected size-categorized samples, or a portable generator with heavy-gauge extension cords can be used to operate standard AC-powered samplers. Conversion of generators to use propane fuel reduces their emissions, and hence reduces the possibility of exhaust contaminating the aerosol samples. A long run of heavy gauge power cord can also be used to position the generator at a distance downwind of the sampling area.

A major difference which is encountered between sampling ambient air and fugitive emissions is the heavier particulate concentrations near the emission sources. To achieve optimal filter loading, a shorter sampling duration is required. Consequently, to obtain the integrated character of an emission source, a number of short-duration samples can be taken and analyzed, followed by the mathematical averaging of analytical data. Cycle timers can also be used to integrate process variability. In practice, fugitive sampling times have generally varied from 15 minutes to four hours.

Ground-based sampling has been successfully used for mobile sources (Section 6), a variety of industrial sources (Houck *et al.*, 1982b), and agricultural and forestry burning sources (Houck *et al.*, 1989a, 1989c; Ward, 1983; Ward and Core, 1984; Ward and Hardy, 1988; Ward *et al.*, 1979, 1982). Agricultural burning of crop residues has been successfully sampled with ground-based instruments with particulate inlets located 2 to 5 meters above ground level. For forestry burning, samplers have been located 10 to 15 meters above the fuel. Measurements for TSP and PM_{10} have typically shown great variability, even when fuels are relatively evenly distributed such as in fields of cereal stubble. Raising the sampling inlet allows dilution and mixing of emissions from various locations in the fuel bed, thus providing a cooler and more representative sample. The variability in emissions has also resulted in the use of multiple samplers at various sampling points and other integration techniques in attempts to ensure that representative samples are collected from these large, non-homogeneous sources. Ward and Hardy (1988) developed regional forestry burning profiles by combining profiles for individual forest species, burn types, and seasons.

A review of agricultural crop census data, and forest and rangeland wildfire and prescribed burn records is needed to determine the largest open-burning sources affecting critical receptor areas. Freeburn (1985) found that much of the emissions from agricultural sources was from the burning of non-crop wastes and land-clearing debris which were not routinely monitored. It is suggested that these sources be investigated through interviews with fire officials and extension agents in the affected areas prior to developing a sampling strategy.

Some important process fugitive emissions are episodic and of short duration, e.g., slag pouring, dynamite explosions, furnace upsets, etc. To collect an adequate sample for analyses, usually a number of emission episodes are collected onto a single filter set by turning the sampler on at the beginning of an emission period and turning if off again immediately thereafter. Slag pouring and blast furnace upset profiles developed by Houck *et al.* (1982b) are examples of profiles produced by the analyses of samples comprised of repetitive sampling of a series of short-term episodes.

Dust from agricultural tillage and wind-blown urban unpaved areas can either be sampled by ground-based sampling or by grab bulk sampling followed by laboratory resuspension, and the decision on which procedure to follow should be made on a case-by-case basis. The advantages of ground-based sampling are that a more integrated fingerprint is likely to be obtained, particularly for sources like dust from wind-blown urban unpaved areas where many small chemically non-homogenous areas may contribute to the source profile. Ground-based sampling may also provide slightly more representative size distribution than would grab sampling followed by resuspension. A major disadvantage of ground-based sampling is that fingerprint contamination may occur from other sources in the airshed. A discussion on grab-sampling of these sources is presented in the following section.

8. PASSIVE FUGITIVE EMISSIONS

Passive fugitive sources are defined, for the purposes of sampling strategy, as those sources where wind or vehicular activities produce dust from exposed surfaces. Road dust, wind-blown raw materials, agricultural soil dust from tilling, and wind-blown desert soil fit into this category. In addition to ground-based sampling as described in the preceding section, two types of sampling techniques have been used to sample passive fugitive sources. These are grab sampling and vacuum sampling. Once a sample is collected by either technique, laboratory drying, sieving, and resuspension are conducted.

Grab sampling, followed by laboratory sieving and resuspension, has been used to characterize ambient particles originating from a variety of passive fugitive sources (e.g., Houck and Pritchett, 1985, 1988; Houck *et al.*, 1981, 1982b, 1989a, 1989b, 1989c, 1990; Watson *et al.*, 1988). The collection of representative grab samples from friable industrial storage piles, from unpaved roads, and from exposed soil surfaces such as agricultural fields generally produces many more samples than can analyzed within the scope of a given project. To reduce the total number of individual samples, bulk samples are often composited. Two approaches to compositing have been taken. For very large storage or waste piles, a transect sampling plan is developed and a large number of samples are collected, composited to form an integrated sample representative of the material as a whole, then sieved and resuspended. If, on the other hand, a number of samples are collected from various locations which cannot be assumed *a priori* to be similar in composition (such as soil samples collected from different fields), the samples are first sieved to less than 38μ and a semi-quantitative XRF scan is then conducted on the bulk samples. Based on the similarities or dissimilarities detected by the XRF scan, the similar samples are composited before resuspension and before detailed analyses are conducted.

While the field and laboratory procedures for collection and resuspension of the grab sample material are relatively simple, less simple is the selection of the location and number of samples needed to adequately represent material, e.g., soils, from a large area. Selection of locations within a several-kilometer radius of the receptor site is one consideration, since dust transport drops off rapidly with distance. Another consideration is the soil type and usage. Soils vary chemically due to their geological origin and addition of agricultural amendments. Local U.S. Department of Agriculture Soil Conservation Service and Extension Service officials and maps need to be consulted to ensure that major soil types are identified. Once they are identified, a transect-like grid similar to what might be used on a smaller scale with industrial storage areas needs to be established and bulk surface samples should be collected from selected grids. Only the top 1 or 2 cm should be collected, since this constitutes the wind-blown material and may be chemically different from deeper material. In some cases, multiple samples should be resuspended and analyzed to assess the chemical variability of the source. In most cases, the bulk samples will be composited prior to laboratory preparation, resuspension, and analysis. When sampling unpaved roads, multiple samples should be collected. Each grab sample from unpaved roads should consist of portions removed across the road from shoulder to shoulder at predetermined intervals to obtain the multiple samples that, as with the soil samples, can be analyzed individually to assess source variability, or can be composited to produce an average source fingerprint.

The suspension of road dust by vehicular traffic has been shown to be a major source of coarse particles (>2.5μ) and a minor source of respirable particles (<2.5μ) in numerous studies. Road dust is particularly important in and around industrial complexes and urban areas due to the increased vehicular traffic and track-out of commercial materials. The characteristic chemistry and morphology of resuspended road dust originating from commercial materials and from areas heavily impacted by traffic is typically quite distinct as compared to surrounding native soils. The sampling strategy for road dust collection usually entails collection of samples from each road type (e.g., industrial haul roads, employee parking lots, urban streets, highways, and rural roads) near the receptor.

A high-volume "vacuum" sampler has been used to collect road dust samples (Houck *et al.*, 1989a; Watson *et al.*, 1988). Using the high-volume sampler, samples have been collected from parking lots, paved roads, unpaved and oiled roads which have well-packed substrates, coarse gravel roads, and railroads. As with bulk grab samples, once road dust samples have been collected, the samples are sieved and resuspended in the laboratory onto size-categorizing samplers. Commercial high-volume 8×10-inch filters make excellent particulate traps and their particle penetration efficiencies are well defined. Typically about one centimeter of material is collected onto the filter. This material is easily shaken off the filters for the laboratory work-up.

The sampling of dust from paved surfaces is relatively simple. The vacuum sampling device, rather than simple grab sampling, is frequently required due to the fact that a relatively small amount of dust often remains on the pavement surface, which makes it physically impossible to collect a grab sample. Each road dust sample is comprised of material captured by numerous traverses across the pavement until an adequate sample is obtained. The criteria for road selection for sampling includes proximity to receptor sites, traffic patterns, industrial activity, wintertime sanding or salting, and the chemical composition of the surrounding soils, road-fills and cuts, and shoulder gravel characteristics. Figure 5 is a schematic of the road dust sampling system.

9. PROFILE DEVELOPMENT

Source profiles, in the simplest of terms, consist of (1) mass fractions of selected chemical species and (2) estimates of the variabilities of these fractions. Compilations of such source profiles have been prepared for use with receptor models (U.S. Environmental Protection Agency, 1984, 1988). The chemical species which have comprised source profiles are those which are routinely measured by standard analytical methods used with ambient samples, as there is little value in terms of receptor modeling in quantifying chemical species not in the

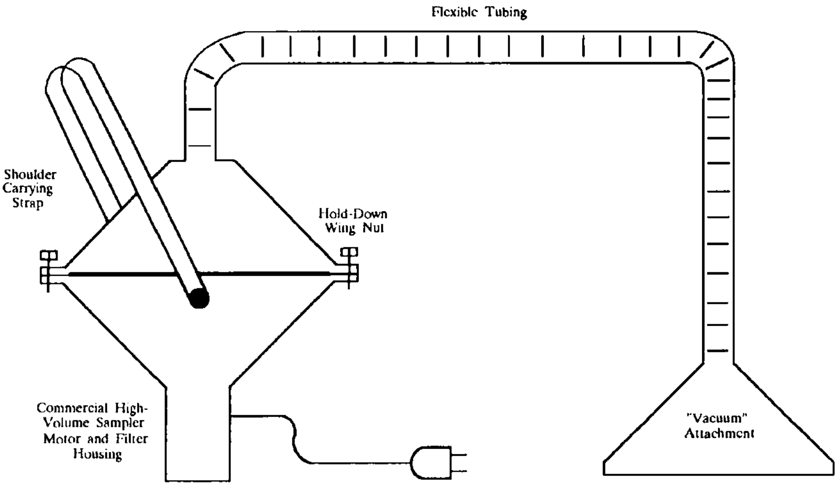

Figure 5. Schematic of a High Volume Road Dust Sampler.

ambient data set. Traditionally, most source profiles have consisted of 35 to 40 elements measured by x-ray fluorescence spectrometry (XRF), common ions measured by ion chromatography (IC) or wet chemical techniques, and total organic carbon, elemental carbon, and carbonate carbon determined by thermal/optical methods. Uncertainties associated with each chemical species are impacted both by analytical uncertainties and by the variability in the source emissions. Variability in source emissions are in turn dependent on both the inherent variability in emissions in a given source and the mathematical grouping or compositing which is conducted for a given modeling program.

Compositing is an important aspect of developing source profiles, as the objective of source profile development is to obtain profiles that are representative of the material that reaches the ambient (receptor) monitoring site. The apportionment of a given category of particles improves when an accurate representation of the chemical composition of that given category is used. The apportionment of other categories of particles also improves because the relative contributions calculated for each source category are not independent of each other in the CMB receptor model. For example, while it may be of no great concern if road dust from a nearby street impacts a receptor, the contribution of an industrial point source which may be of great concern will be more accurately determined if the chemical composition of the road dust is well characterized.

Composite source profiles are also of value for operational reasons. For example, if two coal-fired boilers contribute to a receptor, a single coal-fired boiler contribution, rather than two individual contributions, may be more desirable to calculate. In addition, frequently, two or more source contributions cannot be separated by the CMB model due to the similarity of their profiles (this similarity is termed collinearity) and a composite profile is desirable. The more dissimilar the sources included into the composite, the greater the uncertainties associated with the composite profile.

Figure 6 illustrates the source profile compositing process. The first level in the compositing hierarchy is the averaging of replicates taken from the same source to produce a mean value and standard deviation for each chemical species. The next level is to calculate weighted averages of these specific source profiles to produce the composite source category profile. A final composite, as shown in Figure 6, is an overall general composite. The weighting factors for the individual sub-types are derived from relative emission rates for the different sub-types or from applying parametrically weighted composite profiles in a source-dominated environment.

To provide the source profile data with the greatest degree of credibility, a quality assurance/quality control (QA/QC) program should be followed during the source sampling activities. The QA/QC program should be, to the extent applicable, consistent with established environmental QA/QC guidelines (U.S. Environmental Protection Agency, 1976, 1977, 1980). These guidelines establish criteria for documenting (1) project organization and responsibilities; (2) data precision, accuracy, completeness, representativeness, and comparability; (3) sampling procedures; (4) sample custody; (5) calibration procedures and frequency; (6) analytical procedures; (7) data reduction, validation, and reporting

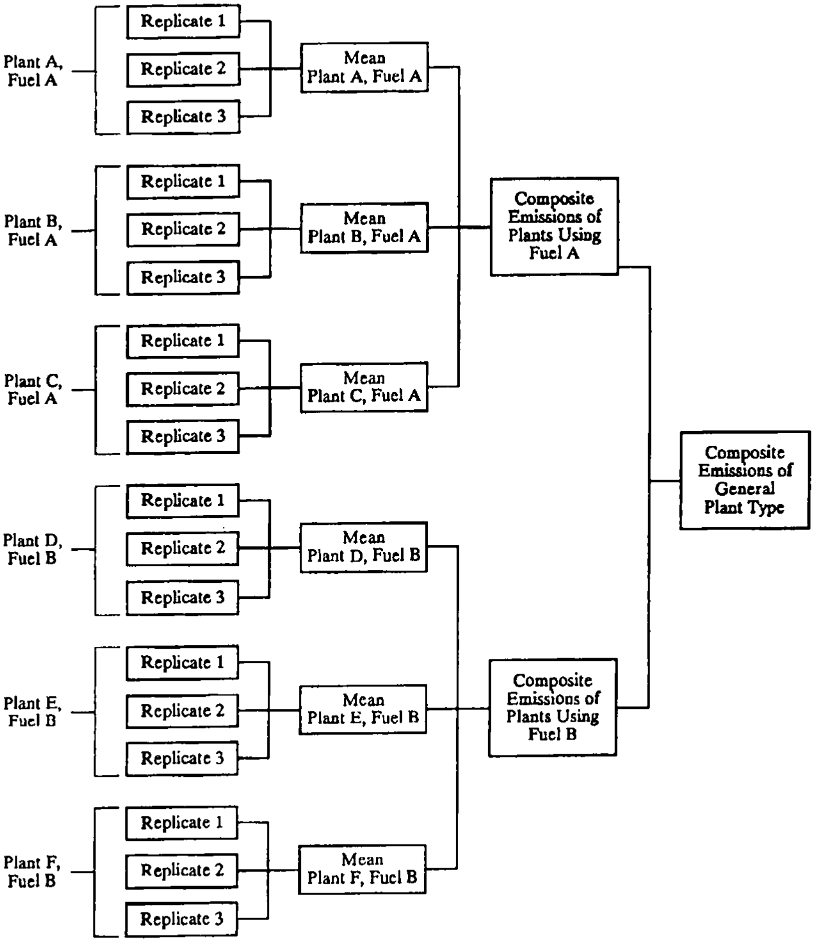

Figure 6. Source Profile Compositing Procedure.

procedures; and (8) a variety of QC checks, audits, maintenance, and corrective action activities.

Once quality source profiles are developed, they need to be formatted in such a fashion as to be useful to the modeler. Example source profiles are illustrated in Table 3 for a municipal waste incinerator, for urban road dust, and for woodstove emissions. As can be seen, the profiles consist of concentrations of chemical species and associated uncertainties, both in units of weight percent. The concentration value for each chemical species in each profile is the mean value of three replicate samples. The uncertainties are either the standard deviation around the mean value or the root mean square propagated analytical uncertainty as determined for each sample. Whichever value was greater is used as the uncertainty and and is used in the CMB model. Generally the source profiles, such as are listed in Table 3, are compiled on magnetic disks for direct entry into CMB receptor modeling programs.

Histograms, both linear and logarithmic, have been used to effectively illustrate source profiles. Figure 7 is an example set of histograms for a wood-fired boiler. As can be seen, the combined use of both linear and logarithmic histograms permits an intuitive understanding of the key features of each source profile, particularly when they are compared to other profiles which may be present in a given airshed. The linear histograms permit the major chemical species characteristic of a given source to be clearly identified, while the logarithmic histogram permits all chemical species which might be useful in the modeling process to be displayed.

In summary, methods for developing source profiles for use in receptor models have been standardized and data generated by the methods have been successfully used in numerous receptor model studies. The sampling methods have been widely published, as have compilations of source profiles. By using appropriate existing literature source profiles for more generic sources combined with source profiles produced by direct measurement for sources with unique, unusual, or unknown emission chemistry, state-of-the-art receptor modeling results can be obtained.

10. ACKNOWLEDGEMENTS

The contributions made by the staff and faculty of the Desert Research Institute over the course of numerous joint source sampling projects are acknowledged. The contributions made by John Watson, Judy Chow, Lyle Pritchett, and Clifton Frazier are especially appreciated.

Table 3. Example Source Profiles

Chemical Species	Municipal Waste Incinerator, <10μ		Urban Road Dust Bakersfield, CA, <2.5μ		Woodstove Profile Mammoth Lakes, CA, <1μ	
	Conc. (%)	Uncert. (%)	Conc. (%)	Uncert. (%)	Conc. (%)	Uncert. (%)
Na^+	1.3959	2.0533	0.2161	0.0857	0.0155	0.0067
Al	0.6420	1.1119	9.2228	1.0496	0.0023	0.0069
Si	2.3597	2.6211	21.5503	2.4520	0.0056	0.0079
P	0.1040	0.1296	0.4161	0.0526	0.0000	0.0041
S	3.2080	0.7365	0.5725	0.0894	0.0955	0.0069
SO_4^{2-}	10.6065	6.6965	0.5037	0.6581	0.1986	0.0443
Cl	20.7463	17.6659	0.1945	0.0222	0.1261	0.0100
K	0.8485	0.1221	1.9560	0.2223	0.2366	0.1032
K^+	1.0019	0.1134	0.3342	0.1279	0.2366	0.1032
Ca	3.4652	0.9417	3.2958	0.3745	0.0293	0.0390
Ti	0.1053	0.0922	0.5191	0.0590	0.0000	0.0071
V	0.0137	0.0516	0.0290	0.0076	0.0000	0.0030
Cr	0.0435	0.0227	0.0346	0.0039	0.0000	0.0007
Mn	0.0348	0.0212	0.1228	0.0140	0.0003	0.0004
Fe	1.5420	0.4428	6.1414	0.6985	0.0050	0.0055
Co	0.0027	0.0252	0.0056	0.0872	0.0002	0.0003
Ni	0.0137	0.0061	0.0126	0.0014	0.0001	0.0003
Cu	0.0368	0.0136	0.0248	0.0020	0.0001	0.0004
Zn	0.4675	0.1961	0.2373	0.0176	0.0344	0.0074
Ga	0.0023	0.0162	0.0001	0.0054	0.0000	0.0010
As	0.0023	0.0342	0.0034	0.0563	0.0000	0.0009
Se	0.0028	0.0068	0.0003	0.0020	0.0000	0.0005

Table 3 (cont). Example Source Profiles

Chemical Species	Municipal Waste Incinerator, <10μ		Urban Road Dust Bakersfield, CA, <2.5μ		Woodstove Profile Mammoth Lakes, CA, <1μ	
	Conc. (%)	Uncert. (%)	Conc. (%)	Uncert. (%)	Conc. (%)	Uncert. (%)
Br	1.0095	0.9717	0.0101	0.0013	0.0009	0.0002
Rb	0.0007	0.0623	0.0125	0.0012	0.0003	0.0005
Sr	0.0273	0.0084	0.0848	0.0087	0.0001	0.0006
Y	0.0047	0.0148	0.0028	0.0041	0.0000	0.0007
Zr	0.0020	0.0143	0.0092	0.0018	0.0000	0.0009
Mo	0.0101	0.0267	0.0007	0.0052	0.0000	0.0014
Pd	0.0106	0.0477	0.0000	0.0107	0.0000	0.0026
Ag	0.0316	0.0597	0.0012	0.0126	0.0000	0.0030
Cd	0.0146	0.0588	0.0045	0.0134	0.0008	0.0033
In	0.0242	0.0711	0.0020	0.0157	0.0000	0.0038
Sn	0.0629	0.0979	0.0239	0.0191	0.0003	0.0051
Sb	0.0174	0.1116	0.0064	0.0240	0.0049	0.0060
Ba	0.0000	0.4014	0.1151	0.0575	0.0099	0.0209
La	0.0266	0.4475	0.0059	0.0935	0.0106	0.0234
Hg	1.1194	1.0169	0.0035	0.0064	0.0000	0.0016
Pb	0.1627	0.0454	0.3571	0.0267	0.0000	0.0013
NO_3^-	4.0169	3.5789	0.0000	0.3622	0.0293	0.0081
NH_4^+	20.0459	15.7942	0.0278	0.0482	0.0262	0.0052
OC[†]	15.4843	2.4766	16.7761	5.0945	63.8978	8.4425
EC[†]	3.5533	2.1192	2.1172	1.2366	11.6959	2.2468
CC[†]	0.0277	0.2167	0.1517	0.2383	0.0000	0.0219
Sum	80.6893	24.5242	64.2676	5.9593	76.2280	8.7372

[†] OC = organic carbon
EC = elemental carbon
CC = carbonate carbon

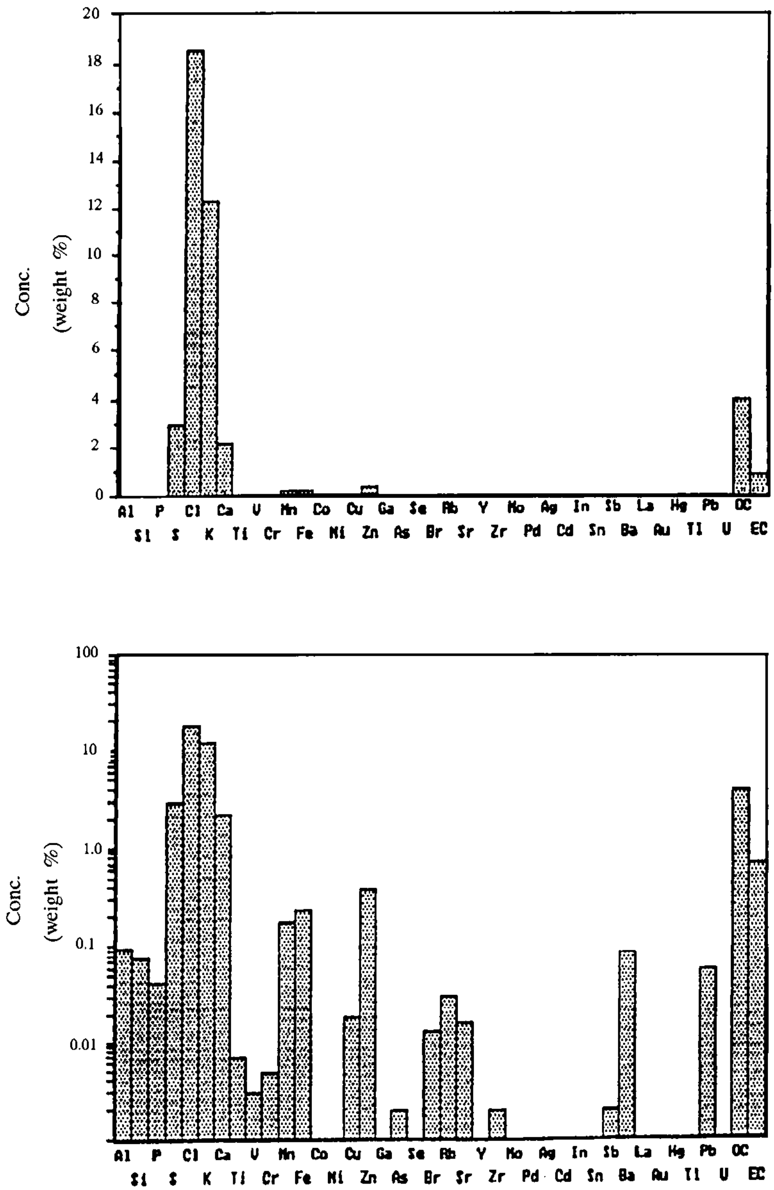

Figure 7. Example Source Profile Histograms, Wood-Fired Boiler, < 10 μm. Top Figure is Linear; Bottom Figure is Logarithmic.

11. REFERENCES

Armstrong, J.A., P.A. Russel, L.E. Sparks, and D.C. Drehmel, Tethered Balloon Sampling Systems for Monitoring Air Pollution, *J. Air Pollut. Contr. Assoc.* 31: 735-743 (1981).

Burnet, P.G., N.G. Edmisten, P.E. Tiegs, J.E. Houck, and R.A. Yoder, Particulate, Carbon Monoxide, and Acid Emission Factors for Residential Wood Burning Stoves; *J. Air Pollut. Contr. Assoc.* 36:1012–1018 (1986).

Cahill, T.A., R.A. Eldred, J. Barans, and L. Ashbaugh, *Ambient Aerosol Sampling with Stacked Filter Units*, Report to Federal Highway Administration, Report No. FHWA-RD-78-178 (1978).

Cooke, W.M., W.H. Piispanen, A.K. Wensky, G.A. Lutz, T.F. Cole, J.S. Ogden, A. Levy, R.H. Barnes, B.W. Cornaby, and K.B. Degner, *Candidate Sampling and Analysis Methods for 21 Suspect Carcinogens in Combustion Emissions*, EPA-600/57-84-078, U.S. Environmental Protection Agency, Research Triangle Park, NC, (1984).

Cooper, J.A., D.C. Redline, J.R. Sherman, L.M. Valdovinos, W.L. Pollard, L.C. Scavone, and C. Badgett-West, *PM_{10} Source Composition Library for the South Coast Air Basin. Volume I: Source Profile Development Documentation Final Report*, South Coast Air Quality Managememt District, El Monte, CA (1987).

Core, J.E., Development of City-Specific Transportation Source Particulate Profiles for Receptor Modeling, in *Receptor Methods for Source Apportionment: Real World Issues and Applications*, T.G. Pace, ed., Transactions No. TR-5, Air & Waste Management Association, Pittsburgh, PA, 251- (1986).

Core, J.E. and J.E. Houck (eds.), *Pacific Northwest Source Profile Library, Sampling and Analytical Protocols, Final Report*, Oregon Department of Environmental Quality, Portland, OR, (1987).

Dzubay, T.G., R.K. Stevens, and L.W. Richards, Composition of Aerosols Over Los Angeles Freeways, *Atmospheric Environ.* 13: 653-659 (1979).

Freeburn, S.A. and B.S. Carhart, *Air Toxic and Volatile Organic Compound Emission Inventory and Program Development for the State of Idaho*, prepared for the State of Idaho Department of Health and Welfare (1985).

Goulet, J.M., J.E. Houck, H.W. Robbins, J.J. Olsen, L.C. Pritchett, and C.A. Frazier, *Plum Creek Timber Company, Inc. Chemical Mass Balance Source Sampling Report*, submitted to Montana Department of Health and Environmental Sciences, 43 pp plus appendices (1989a).

Goulet, J.M., J.E. Houck, H.W. Robbins, J.J. Olsen, L.C. Pritchett, and C.A. Frazier, *Champion International Corporation Chemical Mass Balance Source Sampling Report*, submitted to Montana Department of Health and Environmental Sciences, 38 pp plus appendices (1989b).

Harris, D.B., Source Sampling Systems Used to Develop Source Signatures, In *Receptor Methods for Source Apportionment: Real World Issues and Applications*, T.G. Pace, ed., Transactions No. TR-5, Air & Waste Management Association, Pittsburgh, PA, 46-55 (1986).

Heinsohn, R.J., J.W. Davis, and K.T. Knapp, Dilution Source Sampling Systems, *Environ. Sci. Technol.* 14:1205-1209 (1980).

Houck, J.E., J.A. Cooper, C.A. Frazier, E.R. Larsen, J.F. Mohan, and A.S. Bradeen, *Application of Chemical Mass Balance Methods to the Determination of the Contribution of Potlatch Corporation's Air Particulate Emissions to the Lewiston-Clarkston Airshed*, Report to Potlatch Corporation (1981).

Houck, J.E., J.A. Cooper, and E.R. Larson, Dilution Sampling for Chemical Receptor Source Fingerprinting, Presented at 75th Annual Air Pollution Control Association Meeting, New Orleans, LA, 334-345 (1982a).

Houck, J.E., J.A. Cooper, C.A. Frazier, and R.T. DeCesar, *East Helena Source Apportionment Study*, Report to Montana Department of Health and Environmental Sciences (1982b).

Houck, J.E. and L.C. Pritchett, *Baseline Environmental Quality, Kilauea East Rift, Summary Report*, State of Hawaii Department of Planning and Economic Development report, 130 pp (1985).

Houck, J.E., J.H. Carlson, C.F. Frazier, and R.A. Cary, The Contribution of Residential Wood Combustion to Respirable and Inhalable Particulate Concentrations in Missoula, Montana, *Measurement of Toxic and Related Air Pollutants*, Proceedings No. VIP-8, Air & Waste Management Association, Pittsburgh, PA, 691-696 (1987).

Houck, J.E. and L.C. Pritchett, Particulate Monitoring on the Kilauea East Rift, Hawaii, in *Particulate Matter/Fugitive Dusts, Measurement and Control in Western Arid Regions*, Proceedings of APCA Specialty Conference, Tucson, Arizona, Air & Waste Management Association, Pittsburgh, PA, 128-138,(1988).

Houck, J.E., J.C. Chow, J.G. Watson, C.A. Simons, L.C. Pritchett, J.M. Goulet, and C.A. Frazier, *Determination of Particle Size Distribution and Chemical Composition of Particulate Matter from Selected Sources in California*, report to California Air Resources Board, Agreement No. A6-175-32, NTIS PB89 232805, 3 volumes and executive summary (1989a).

Houck, J.E., J.C. Chow, and M.S. Ahuja, The Chemical and Size Characterization of Particulate Material Originating from Geological Sources in California, *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 322-333 (1989b).

Houck, J.E., J.C. Chow, and M.S. Ahuja, The Chemical Characterization of Agricultural Sources of PM_{10} Particles in the San Joaquin and Imperial Valleys of California, Air & Waste Management Association Meeting Paper 89-145.1 (1989c).

Houck, J.E., J.M. Goulet, J.C. Chow, and J.G. Watson, Chemical Source Characterization of Residential Wood Combustion Emissions in Denver, Colorado, Bakersfield, California, and Mammoth Lakes, California, Air & Waste Management Association Meeting Paper 89-95.1 (1989d).

Houck, J.E., L.C. Pritchett, R.B. Roholt, J.G. Watson, J.C. Chow, J.M. Goulet, and C.A. Frazier, *Determination of Particle Size Distribution and Chemical Composition of Particulate Matter from Selected Sources in the San Joaquin Valley*, report to California Air Resources Board, Contract 88-11, 35 pp plus appendices (1990).

Huynh, C.F., T. Vuduc, C. Schwab, and H. Rollier, In-Stack Dilution Techniques for the Sampling of Polycyclic Organic Compounds, Applications to Effluents of a Domestic Waste Incineration Plant, *Atmospheric Environ.* 18:255-259 (1984).

Javitz, H.S., J.G.Watson, J.P. Guertin, and P.K. Mueller, Results of a Receptor Modeling Feasibility Study, *J. Air Pollut. Contr. Assoc.* 38:661-669 (1988).

McCain, J.D. and A.D. Williamson, *Development and Evaluation of Dilution Probes Used for Sampling to Determine Source Signatures*, EPA 600/3-84-045, U.S. Environmental Protection Agency, Research Triangle Park, NC (1984).

Olmez, I, A.E. Sheffield, G.E. Gordon, J.E. Houck, L.C. Pritchett, J.A. Cooper, T.G. Dzubay, and R.L. Bennett, Compositions of Particles from Selected Sources in Philadelphia for Receptor Modeling Applications, *J. Air Pollut. Contr. Assoc.* 38:1392-1402 (1988).

Pan, Y.S., *Review Summary of Stack Sampling of Fine Particulates*, U.S. Department of Energy Report DOE/PETC/TR-87/2, Pittsburgh Energy Technology Center, Pittsburgh, PA (1986).

Pierson, W.R. and W.W. Brachaczek, Particulate Matter Associated with Vehicles on the Road, Paper No. 760039, Society of Automotive Engineers, Inc., Warrendale, PA, 19 pp. (1976).

Pierson, W.R. and W.W. Brachaczek, Particulate Matter Associated with Vehicles on the Road II, *Aerosol Sci. Technol.* 2:1-40 (1983).

Pierson, W.R., W.W. Brachaczek, R.H. Hammerle, D.E. McKee, and J.W. Butler, Sulfate Emissions from Vehicles on the Road, *J. Air Pollut. Contr. Assoc.* 28:123-132 (1978).

Richards, L.W., G.R. Markowski, and N. Waters, Comparison of Nephelometer Telephotometer and Aerosol Data in the Southwest, Presented at 74th Annual Air Pollution Control Association Meeting, Philadelphia, PA (1981).

Richards, L.W., J.A. Anderson, D.L. Blumenthal, J.A. McDonald, E.S. Macias, S.V. Hering, and W.E. Wilson, Chemical Aerosol and Optical Measurements in the Plumes of Three Midwestern Coal-Fired Power Plants, *Atmospheric Environ.* 19:1685-1704 (1985).

Shah, J.J., R. Johnson, and J.E. Houck, Source Characterization Using Tethered Balloons, in *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No TR-14, Air & Waste Management Association, Pittsburgh, PA, 334-345 (1988).

Shelton, J.W. and L.W. Gay, *Evaluation of Low-Emission Woodstoves*, Shelton Research, Inc., Report 1086 to California Air Resource Board, Sacramento, CA, Contract A3-122-32 (1986).

Shelton, J.W., and L.W. Gay, *Colorado Fireplace Report*, Shelton Research, Inc., Report to Colorado Air Pollution Control Division, Contract no. C375322 (1987).

Small, M., M.S. Germani, A.M. Small, W.H. Zoller, and J.L. Moyers, Airborne Plume Study of Emissions From the Processing of Copper Ores in Southeastern Arizona, *Environ. Sci. Technol.* 15:293-298 (1981).

Stiles, D.C., Evaluation of an S^2 Sampler for Receptor Modeling of Woodsmoke Emissions, Presented at 76th Annual Air Pollution Control Association Meeting, Atlanta, GA (1983).

U.S. Environmental Protection Agency, *IERL-RTP Data Quality Manual*, Environmental Protection Agency Report No. EPA-600/2-76-159, Environmental Protection Agency, Research Triangle Park, NC (1976).

U.S. Environmental Protection Agency, *Quality Assurance Handbook for Air Pollution Measurement Systems, Volumes I, II, and III*, Environmental Protection Agency Report Nos. EPA-600/9-76-005, EPA-600/4-77-027a, and EPA-600/4-77-027b, Environmental Protection Agency, Research Triangle Park, NC (1977).

U.S. Environmental Protection Agency, *Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans*, Environmental Protection Agency Report Nos. QAMS-005/80, EPA-600/4-83-004, Environmental Protection Agency, Research Triangle Park, NC (1983).

U.S. Environmental Protection Agency, *Receptor Model Source Composition Library*, Environmental Protection Agency Report No. EPA-450/4-85-002, Environmental Protection Agency, Research Triangle Park, NC (1984).

U.S. Environmental Protection Agency, Method 5G–Determination of Particulate to Emissions from Wood Heaters from a Dilution Tunnel Sampling Location, *Federal Register*, 52:5024-5033, February 18, 1987.

U.S. Environmental Protection Agency, *Air Emissions Species Manual, Volume II, Particulate Matter Species Profile*, Environmental Protection Agency Report No. EPA-450/2-88-003b, Environmental Protection Agency, Research Triangle Park, NC (1988).

Ward, D.E., Source-Strength Modeling of Particulate Matter Emissions from Forest Fires, presented at the 76th Annual Air Pollution Control Association Meeting, Atlanta, GA (1983).

Ward, D.E. and J. Core, Source Emission Profiles for Douglas Fir and Hemlock Slash Burning, in Proceedings, Annual Meeting Pacific Northwest International Section, Air Pollution Control Association, November 12-14, Portland, OR (1984).

Ward, D.E. and C.C. Hardy, Organic and Elemental Profits for Smoke from Prescribed Fires, in *Receptor Models in Air Resources Management*, J.E. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 299-321 (1988).

Ward, D.E., C.C. Hardy, R.D. Ottmar, and D.V. Sandberg, A Sampling System for Measuring Emissions from West Coast Prescribed Fires, presented at Pacific Northwest International Section, Air Pollution Control Association, Vancouver, B.C., Canada (1982).

Ward, D.E., R.M. Nelson, and D.F. Adams, Forest Fire Smoke Plume Documentation, in Proceedings, 72nd Annual Air Pollution Control Association Meeting, June 24-29, Cincinnati, OH (1979).

Watson, J.G., J.C. Chow, L.W. Richards, W.D. Neff, S.R. Andersen, D.L. Dietrich, J.E. Houck, and I. Olmez, *The 1987-1988 Metro Denver Brown Cloud Study, Final Report*, 1987-1988 Metro Denver Brown Cloud Study, Inc., Denver, CO., 3 volumes (1988).

Watson, J.G., J.C. Chow, L.C. Pritchett, J.E. Houck, R.A. Ragazzi, and S. Burns, Chemical Source Profiles for Particulate Motor Vehicle Exhaust Under Cold and High-Altitude Operating Conditions, *Science Total Environ.* 93:183–190 (1990).

Chapter 4

CHEMICAL MASS BALANCE

John G. Watson and Judith C. Chow
Desert Research Institute
P.O. Box 60220
Reno, NV 89506

and

Thompson G. Pace
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

1. INTRODUCTION

The Chemical Mass Balance (CMB) receptor model uses the chemical and physical characteristics of gases and particles measured at source and receptor to both identify the presence of and to quantify source contributions to pollutants measured at the receptor. These chemical and physical characteristics must be such that: 1) they are present in different proportions in different source emissions; 2) these proportions remain relatively constant for each source type; and 3) changes in these proportions between source and receptor are negligible or can be approximated. The chemical mass balance (CMB) is the fundamental receptor model, and the derivation of the Principal Component Analysis (PCA) and Multiple Linear Regression (MLR) receptor models from fundamental physical principles begins with the CMB (Watson, 1984). The CMB was first proposed by Hidy and Friedlander (1972), Kneip *et al.* (1972), and Winchester and Nifong (1971). Applications of the CMB and other receptor models are summarized by Hopke and Dattner (1982), Pace (1986), Watson (1989), Gordon (1980, 1988), and Hopke (1985).

The objectives of this presentation are:

- To derive the CMB receptor model, to state its assumptions, and to summarize the effects of potential deviations from these assumptions.

- To present a procedure for applying and validating the CMB receptor model for PM_{10} source apportionment.

2. CMB MATHEMATICS

The chemical mass balance consists of a least squares solution to a set of linear equations which expresses each receptor concentration of a chemical species as a linear sum of products of source compositions and source contributions. Source compositions (i.e., the fractional amount of the species in the emissions from each source-type) and the receptor concentrations, with appropriate uncertainty estimates, serve as input data to the CMB model. The output consists of the amount contributed by each source-type to each chemical species. The model calculates values for the contributions from each source and the uncertainties of those values. Input data uncertainties are used both to weight the importance of input data values in the solution and to calculate the uncertainties of the source contributions.

The concentration measured at a receptor during a sampling period of length T due to a source j with constant emission rate E_j is

$$S_j = D_j \cdot E_j \tag{1}$$

where

$$D_j = \int_0^T d[\vec{u}(t),\ \sigma(t),\ \vec{x}_j]\ dt \tag{2}$$

is a dispersion factor depending on wind velocity (u), atmospheric stability (σ), and the location of source j with respect to the receptor (x_j). All parameters in Equation 2 vary with time, so the instantaneous dispersion factor, d, must be an integral over time period T.

Various forms for d have been proposed (e.g. Pasquill, 1974; Seinfeld, 1986), some including provisions for chemical reactions, removal, and specialized topography. None are completely adequate to describe the complicated, random nature of dispersion in the atmosphere. The advantage of receptor models in general, and the CMB in particular, is that an exact knowledge of D_j is unnecessary.

If a number of sources, J, exists and there is no interaction between their emissions to cause mass removal, the total mass measured at the receptor, C, will be a linear sum of the contributions from the individual sources.

$$C = \sum_{j=1}^{J} D_j \cdot E_j = \sum_{j=1}^{J} S_j \tag{3}$$

Similarly, the concentration of elemental component i, C_i will be

$$C_i = \sum_{j=1}^{J} F_{ij} \cdot S_j \tag{4}$$

where F_{ij} is the fraction of source contribution S_j composed of element i. The number of chemical species (I) must be greater than or equal to the number of sources (J) for a unique solution to these equations.

Solutions to the CMB equations consist of: 1) a tracer solution; 2) a linear programming solution; 3) an ordinary weighted least squares solution ; 4) a ridge regression weighted least squares solution; and 5) an effective variance least squares solution. Variations on these methods constrain solutions to positive values and/or add a non-zero constant intercept term. An estimate of the uncertainty associated with the source contributions is an integral part of the weighted least squares methods.

Weighted linear least squares solutions are preferable to the tracer and linear programming solutions because: 1) theoretically they yield the most likely solution to the CMB equations, providing model assumptions are met; 2) they can make use of all available chemical measurements, not just the so-called tracer species; 3) they are capable of analytically estimating the uncertainty of the source contributions; and 4) there is, in practice, no such thing as a "tracer," since most species are found in multiple sources.

CMB software in current use (Watson *et al.*, 1990) applies the effective variance solution developed and tested by Watson *et al.* (1984) because this solution: 1) provides realistic estimates of the uncertainties of the source contributions (owing to its incorporation of both source profile and receptor data uncertainties); and 2) gives greater influence to chemical species with higher precisions in both the source and receptor measurements than to species with lower precisions.

The effective variance solution is derived by minimizing the weighted sums of the squares of the differences between the measured and calculated values of C_i and F_{ij} (Britt and Luecke, 1973). The solution algorithm is an iterative procedure which calculates a new set of S_j based on the S_j estimated from the previous iteration. It is carried out by the following steps expressed in matrix notation. A superscript k is used to designate the value of a variable at the kth iteration.

1. Set initial estimate of the source contributions equal to zero.

$$S_j^{k=0} = 0 \qquad j = 1,\ldots,J \tag{5}$$

2. Calculate the diagonal components of the effective variance matrix, V_e. All off-diagonal components of this matrix are equal to zero.

$$V_{e_{ii}}^{k} = \sigma_{C_i}^{2} + \sigma(S_j^{k})^{2} \cdot \sigma_{f_{ij}} \tag{6}$$

3. Calculate the k+1 value of S_j

$$S^{k+1} = (F^{T}(V_e^{k})^{-1} F^{T})(V_e^{k})^{-1} C \tag{7}$$

4. Test the (k+1)th iteration of the S_j against the kth iteration. If any one differs by more than 1 percent, then perform the next iteration. If all differ by less than 1 percent, then terminate the algorithm.

$$\begin{aligned} &\textit{If}\ \frac{|S_j^{k+1} - S_j^{k}|}{S_j^{k+1}} > 0.01,\ \textit{go to step}\ 2 \\ &\textit{If}\ \frac{|s_j^{k+1} - S_j^{k}|}{S_j^{k+1}} \le 0.01,\ \textit{go to step}\ 5 \end{aligned} \tag{8}$$

5. Assign the (k+1)th iteration to S_j and σ_{S_j}. All other calculations are performed with these final values.

$$\sigma_{s_j} = [F^T(V_e^{k+1})^{-1}F)_{jj}]^{-1/2} \tag{9}$$

where C = $(C_1...C_I)^T$, a column vector with C_i as the *i*th component,

S = $(S_1...S_J)^T$, a column vector with S_j as the *j*th component,

F = An I x J matrix of F_{ij}, the source composition matrix,

σ_{C_i} = One standard deviation precision of the C_i measurement,

$\sigma_{F_{ij}}$ = One standard deviation precision of the F_{ij} measurement,

V_e = Diagonal matrix of effective variances

The effective variance represents a simplification of an exact method derived by Britt and Luecke (1973) which has also been termed "structured regression" (White, 1989), and it is used in the CMB owing to its computational efficiency. Tests by Britt and Luecke (1973) show that under most practical circumstances, the differences between the structured regression and effective variance methods are small. The effective variance solution algorithm is very general, and it reduces to most of the solutions cited above with the following modifications:

- When the $\sigma_{F_{ij}}$ are set equal to zero, the solution reduces to the ordinary weighted least squares solution.
- When the $\sigma_{F_{ij}}$ are set equal to the same constant value, the solution reduces to the unweighted least squares solution.
- When a column is added to the F_{ij} matrix with all values equal to 1, an intercept term is computed for the variable corresponding to this column.

- When the number of source profiles equals the number of species (I=J), and if the selected species are present only in a single, exclusive source profile, the solution reduces to the tracer solution.

- When the matrix

  $$(F^T(V_e^k)^{-1}F) \qquad (10)$$

  is re-written as

  $$(F^T(V_e^k)^{-1}F - \phi I) \qquad (11)$$

  with ϕ equal to some non-zero number, known as the smoothing parameter, and I equal to the identity matrix, the solution becomes the ridge regression solution (Williamson and DuBose, 1983).

Only the effective variance method, or the more exact method of Britt and Luecke (1973), are applicable to solving the CMB equations because the uncertainties of the source compositions are generally much larger than the uncertainties in the ambient concentrations.

3. CMB MODEL ASSUMPTIONS

The CMB model assumptions are:

1. Compositions of source emissions are constant over the period of ambient and source sampling.
2. Chemical species do not react with each other, i.e., they add linearly.
3. All sources with a potential for significantly contributing to the receptor have been identified and have had their emissions characterized.
4. The source compositions are linearly independent of each other.
5. The number of sources or source categories is less than or equal to the number of chemical species.
6. Measurement uncertainties are random, uncorrelated, and normally distributed.

Assumptions 1 through 6 for the CMB model are fairly restrictive and will never be totally complied with in actual practice. Fortunately, the CMB model can tolerate deviations from these assumptions, though these deviations increase the stated uncertainties of the source contribution estimates.

The CMB model has been subjected to a number of tests to determine its abilities to tolerate deviations from model assumptions (Watson, 1979; Gordon *et al.*, 1981; Henry, 1982; Currie *et al.*, 1984; Dzubay *et al.*, 1984; Watson and Robinson, 1984; DeCesar *et al.*, 1985; Javitz *et al.*,1988a&b). These studies all point to the same basic conclusions regarding deviations from the above-stated assumptions.

With regard to Assumption 1, source compositions, as seen at the receptor, are known to vary substantially among sources, and even within a single source over an extended period of time. These variations are both systematic and random and are caused by three phenomena: 1) transformation and deposition between the emissions point and the receptor; 2) differences in fuel type and operating processes between similar sources or the same source in time; and 3) uncertainties or differences between source profile measurement methods. Evaluation studies have generally compared CMB results from several tests using randomly perturbed input data and from substitutions of different source profiles for the same source type.

The general conclusions drawn from these tests are: 1) the error in the estimated source contributions due to biases in all of the elements of a source profile is in direct proportion to the magnitude of the biases; and 2) for random errors, the magnitude of the source contribution errors decreases as the difference between the number of species and sources increases.

The most recent and systematic tests are those of Javitz *et al.* (1988a) which apply to a simple 4-source urban airshed and a complex 10-source urban airshed. These tests with 17 commonly measured chemical species showed that primary mobile, geological, coal-fired power plant, and vegetative burning source-types can be apportioned with uncertainties of approximately 30% when coefficients of variation in the source profiles are as high as 50%. This performance was demonstrated even without the presence of unique "tracer" species such as selenium for coal-fired power plants or soluble potassium for vegetative burning. In a complex urban airshed, which added residual oil combustion, marine aerosol, steel production, lead smelting, municipal incineration, and a continental background aerosol, it was found that the geological, coal-fired power plant, and background source profiles were collinear with the measured species. At coefficients of variation in the source profiles as low as 25%, average absolute errors were on the order of 60%, 50%, and 130% for the geological, coal-burning, and background sources, respectively. All other sources were apportioned with average absolute errors of approximately 30% even when coefficients of variation in the source profiles reached 50%. Once again, these tests were performed with commonly measured chemical species, and results would improve with a greater number of species which are uniquely emitted by the different source types.

With regard to the nonlinear summation of species, Assumption 2, no studies have been performed to evaluate deviations from this assumption. While these deviations are generally assumed to be small, conversion of gases to particles and reactions between particles are not inherently linear processes. This assumption is especially applicable to the end products of photochemical reactions and their apportionment to the sources of the precursors. Further model evaluation is necessary to determine the tolerance of the CMB model to deviations from this assumption. The current practice is to apportion the primary material which has not changed between source and receptor. The remaining quantities of reactive species such as ammonium, nitrate, sulfate, and elemental carbon are then apportioned to chemical compounds rather than directly to sources. While this approach is not as satisfying as a direct apportionment, it at least separates primary from secondary emitters and the types of compounds apportioned give some insight into the chemical pathways which formed them. As chemical reaction mechanisms and rates, deposition velocities, atmospheric equilibrium, and methods to estimate transport and aging time become better developed, it may be possible to produce "fractionated" source profiles which will allow this direct attribution of reactive species to sources. Such apportionment will require measurements of gaseous as well as particulate species at receptor sites.

A major challenge to the application of the CMB is the identification of the primary contributing sources for inclusion in the model, Assumption 3. Watson (1979) systematically increased the number of sources contributing to his simulated data from four to eight contributors while solving the CMB equations assuming only four sources. He also included more sources in the least squares solutions than those which were actually contributors, with the following results:

- Underestimating the number of sources had little effect on the calculated source contributions if the prominent species contributed by the missing sources were excluded from the solution.

- When the number of sources was underestimated, and when prominent species of the omitted sources were included in the calculation of source contributions, the contributions of sources with properties in common with the omitted sources were overestimated.

- When source-types actually present were excluded from the solution, ratios of calculated to measured concentrations were often outside of the 0.5 to 2.0 range, and the sum of the source contributions was much less than the total measured mass. The low calculated/measured ratios indicated which source compositions should be included.

- When the number of sources was overestimated, the sources not actually present yielded contributions less than their standard errors if their source profiles were significantly distinct from those of other sources. The over-specification of sources decreased the standard errors of the source contribution estimates.

The linear independence of source compositions required by Assumption 4 has become a subject of considerable interest since the publication of Henry's (1982) singular value decomposition (SVD) analysis. As previously noted, this analysis provides quantitative measures of collinearity and the sensitivity of CMB results to specific receptor concentrations. These measures can be calculated analytically in each application. Henry (1982) also proposed an optimal linear combination of source contributions that have been determined to be collinear.

Other "regression diagnostics" have been summarized by Belsley *et al.* (1980) and have been applied to the CMB by DeCesar *et al.* (1985). Kim and Henry (1989) show that most of these diagnostics are useless because they are based on the assumption of zero uncertainty in the source profiles. Kim and Henry demonstrate, through the examination of randomly perturbed model input data, that the values for these diagnostics vary substantially with typical random changes in the source profiles.

Tests performed on simulated data with obviously collinear source compositions typically result in positive and negative values for the collinear source types as well as large standard errors in the collinear source contribution estimates. Unless the source compositions are nearly identical, the sum of these large positive and negative values very closely approximates the sum of the true contributions.

With most commonly measured species (e.g., ions, elements, and carbon) and source-types (e.g., motor vehicle, geological, residual oil, sea salt, steel production, wood burning, and various industrial processes), from five to seven sources are linearly independent of each other in most cases (Javitz *et al.*, 1988a).

Gordon *et al.* (1981) found instabilities in the ordinary weighted least square solutions to the CMB equations when species presumed to be "unique" to a certain source type were removed from the solution. Using simulated data with known perturbations ranging from 0 to 20%, Watson (1979) found: "In the presence of likely uncertainties, sources such as urban dust and continental background dust cannot be adequately resolved by least squares fitting, even though their compositions are not identical. Several nearly unique ratios must exist for good separation."

With regard to Assumption 5, the true number of individual sources contributing to receptor concentrations is generally much larger than the number of species that can be measured. It is therefore necessary to group sources into source-types of similar compositions so that this assumption is met. For the most commonly measured species, meeting Assumption 4 practically defines these groupings.

With respect to Assumption 6 (the randomness, normality, and the uncorrelated nature of measurement uncertainties), there are no results available from verification or evaluation studies. Every least squares solution to the CMB equations requires this assumption, as demonstrated by the derivation of Watson *et al.* (1984). In reality, very little is known about the distribution of errors for the source compositions and the ambient concentrations. If anything, the distribution probably follows a log-normal rather than a normal distribution. Ambient concentrations can never be negative, and a normal distribution allows a substantial proportion of negative values, while a log-normal distribution allows no negative values. For small errors (e.g., less than 20%), the actual distribution may not be important, but for large errors it probably is important. A symmetric distribution becomes less probable as the coefficient of variation of the measurement increases. This is one of the most important assumptions of the solution method that requires testing.

4. CMB INPUT AND OUTPUT DATA

The ambient concentrations of these species, C_i, and their fractional amount in each source-type emission (source profiles), F_{ij}, are the measured quantities which serve as CMB model input data. These values require uncertainty estimates, σ_{Ci} and σ_{Fij}, which are also input data. Input data uncertainties are used both to weight the importance of input data values in the solution and to calculate the uncertainties of the source contributions. The output consists of: 1) the source contribution estimates (S_j) of each source-type; 2) the standard errors of these source contribution estimates; and 3) the amount contributed by each source-type to each chemical species.

Figure 1 shows examples of ambient fine particle concentrations and source profiles which serve as input data for CMB modeling. For simplicity, only 23 species are shown, though forty or more species in both the gaseous and particulate phases are commonly used to distinguish one source from another. Many more species are available from typical measurements of volatile and particulate organic species.

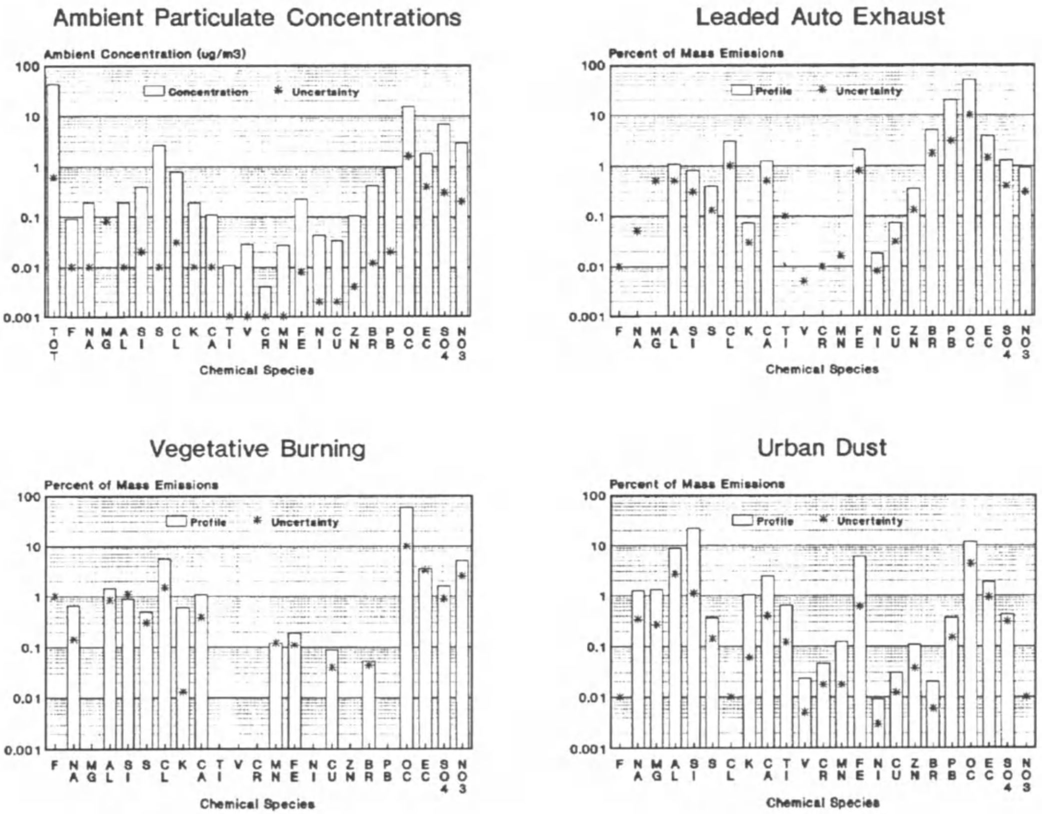

Figure 1. Examples of Ambient Concentration Measurements and Source Profiles from the Potland Aerosol Characterization Study (Watson, 1979).

Emissions inventories are examined to determine the types of sources which are most likely to influence a receptor. Principal components analysis applied to a time series of chemical measurements is also a useful method of determining the number and types of sources. After these sources have been identified, profiles acquired from similar sources are examined to select the chemical species to be measured. Watson (1979) demonstrates that the more species measured, the better the precision of the CMB apportionment.

The ambient values are acquired from chemical measurements of samples from the atmospheres. To quantify particle chemistry, air is drawn through size selective inlets onto filter substrates. These particle deposits can be submitted to a variety of analyses, such as photon or proton induced x-ray fluorescence analysis, instrumental neutron activation analysis, ion chromatography, atomic absorption spectrometry, automated colorimetry, and organic/elemental carbon analysis. Samples from sources are generally acquired by similar sampling through inlets onto substrates by resuspending sieved soil samples, by extracting and diluting with cool air the effluents from an emissions duct, or by sampling ambient air in an area dominated by emissions from a single source.

The ambient sample in the first panel of Figure 1 shows significant concentrations of most chemical species, while each of the source profiles shows a few species which tend to differentiate its composition from that of other sources. Organic (OC) and elemental (EC) carbon dominate the composition of vegetative burning and motor vehicle exhaust, though they are present in different proportions in the two profiles. Carbonaceous material is not so abundant in urban dust, which contains crustal elements such as aluminum, silicon, iron, calcium and titanium in their oxide forms. Lead and bromine stand out in the motor vehicle exhaust profile shown in Figure 1, which was determined by sampling vehicles in use prior to 1978. The abundance of lead and bromine in 1990 motor vehicle exhaust (e.g. Watson *et al.*, 1988 and Cooper *et al.*, 1987) is significantly reduced owing to a larger number of unleaded vehicles and lower lead contents in fuels. The vegetative burning profile contains potassium and chlorine at proportions which are substantially higher than the proportions on other profiles shown in Figure 1.

Differences in chemical composition also exist for particulate matter emitted by sources which are not illustrated in Figure 1. For example, coal combustion contains significant amounts of selenium, copper smelter emissions contain copper, zinc and lead, and steel mill blast furnaces contain large proportions of iron and manganese. It is these differences in the proportions of chemical species which allow the contributions from different sources to be separated by the CMB receptor model.

The ambient data shown in the first panel of Figure 1 also contains significant amounts of sulfate and nitrate which are not found in the source profiles. Sulfates and nitrates (which are usually found in the atmosphere as ammonium sulfate, ammonium bisulfate, sulfuric acid, and ammonium nitrate) are commonly the product of atmospheric reactions which convert gases directly emitted by sources into particles over time periods ranging from hours to days. Particles which are directly emitted by sources and which undergo few changes between source and receptor are termed primary particles. For regulatory purposes, the CMB is currently used to: 1) separate primary contributors from secondary contributors; 2) assign particulate matter of primary origin to source types (e.g. geological, motor vehicle, vegetative burning); and 3) assign secondary particles to their chemical forms (e.g. sulfuric acid, ammonium nitrate). It is an active area of receptor modeling research to develop models which can approximate source profiles after they have aged and interacted with emissions from other sources. When such models have been developed and validated, the CMB model may also be used to apportion secondary particles to the emitters of their precursors.

One of the original objections to receptor modeling for PM_{10} source assessment was that source profile and ambient data were not available. These objections are no longer valid, since a large number of data bases of both source and receptor measurements have been acquired in the United States for use in these models over the past decade. These data bases are widely dispersed, however, and are not generally available for study or evaluation. This situation is unfortunate because: 1) these existing particulate data bases might alleviate the need to acquire new data bases; 2) their potential for receptor model application and testing is untapped; 3) they provide models for success and failure which can enhance the design of new data acquisition projects; and 4) they provide a comprehensive view of particulate levels, concentrations, and source contributions for major parts of the United States. The ideal particulate matter data base for source and receptor measurements has the following characteristics:

- A large number of chemically and size classified concentrations. Mass, elements, ions, and carbon have been found to be the most easily measured and useful species while PM_{10} and $PM_{2.5}$ are the most useful size ranges.

- Comprehensive coverage with respect to time, space, and in the case of source samples, operating conditions. Simultaneous receptor samples taken at locations affected by different source types are useful in the verification of receptor model source apportionments. Similarly, receptor samples taken in different seasons are affected by different emissions sources and meteorological conditions. Source samples need to represent the full range of profiles from a given source category so that uncertainties can be estimated for input to receptor models.

- Documentation of measurement methods, locations, and sampling times. Written records of the entire experimental program which acquired the data base are essential to its extended use. In the case of source characterization, this information should include the fuels, operating cycle, type of facility, location, and time of test.

- Quality control and quality audits. Replicates, field blanks, and independent verifications of field monitoring and laboratory operations are needed to assure that the stated procedures were actually complied with.

- Precision and accuracy estimates. State-of-the-art receptor modeling treats measurement uncertainties as part of the input data and returns uncertainties on source contribution estimates derived from those inputs. The quality control and quality audit data should be processed to quantify these uncertainties.

- Validation summaries or flags. Validation criteria should have been applied to every sample, and the results of that validation should be reported with the data.

- Availability in computerized formats. For research purposes, data cannot be proprietary or secret. If it is not in some computerized and documented form, preferably accessible by personal computers, the expense of putting it into such form usually outweighs the potential benefits derived from examining the data.

The EPA source composition library (Shareef *et al.*, 1989; US EPA, 1988) contains a large number of sources, but its current contents for motor vehicles and residential wood combustion are dated and do not reflect the compositions from modern vehicles, stoves, and fuels. The more recent source libraries (Cooper *et al.*, 1987; Ahuja *et al.*, 1989; Houck *et al.*, 1989; Core *et al.*, 1989; Watson *et al.*, 1988, 1989) contain profiles which are more applicable to data bases acquired today for PM_{10} source assessment. The historical value of source libraries from the Portland Aerosol Characterization Study (Watson, 1979) and the wide range of profiles reported by Hopke (1985) are still of value because of their comprehensiveness and applicability to receptor data taken in an earlier era. Sheffield and Gordon (1986) present the most complete compilation of emissions characteristics from coal- and oil-fired power plants, and this is an excellent resource for studying pollution in areas with these source types.

Most of the collections of source profiles listed in Table 1 contain soil and road dust compositions, and it is unlikely that these profiles change over long periods of time. The Pacific Northwest Source Composition Library (Core, 1989) is one of the first to acquire speciation of the organic carbon fraction of source samples. The measurement of these additional species will allow them to be evaluated in receptor model applications to particulate matter.

Several major ambient particulate studies, and the extent to which they meet the above-stated criteria, are summarized by Lioy *et al.* (1980) and Chow and Watson (1989). The data bases cited in these references are fairly major studies which were initiated for the purpose of applying receptor models, and several of these studies include the results of the modeling. Hopke (1985) identifies a

Table 1. Example of CMB Output

```
SOURCE CONTRIBUTION ESTIMATES -  SITE: PACS1        DATE: 08/13/77  VERSION: 7.0
SAMPLE DURATION        24      START HOUR       0          SIZE:    FINE
       R SQUARE       .98    PERCENT MASS    98.7
     CHI SQUARE      1.12              DF      13

  SOURCE
     * TYPE     SCE(UG/M3)    STD ERR      TSTAT
---------------------------------------------
1      MARIN       12.3889     2.2457     5.5167  Marine Aerosol
3      UDUST        9.5917     1.3876     6.9127  Urban Dust
4      AUTPB       10.0835     1.4942     6.7486  Leaded Motor Vehicle Exhaust
5      RDOIL       11.0603     1.9239     5.7490  Residual Oil Combustion
8      KRAFT        4.6896     5.0467      .9292  Kraft Recovery Boiler
11     ALPRO       10.6023     3.5896     2.9536  Aluminum Pot Line
12     STEEL        8.6729     1.3771     6.2979  Steel Blast Furnace
13     FERMN       11.8754     1.8321     6.4820  Ferromanganese Furnace
---------------------------------------------

MEASURED CONCENTRATION FINE/COARSE/TOTAL:
      80.0+-     8.0/      80.0+-     8.0/     160.0+-     11.3

       UNCERTAINTY/SIMILARITY CLUSTERS  VERSION: 7.0     SUM OF CLUSTER SOURCES
--------------------------------------------------------------------------------

       1      8                                              17.078+-     4.241
       1      5      8                                       28.139+-     3.833
--------------------------------------------------------------------------------

SPECIES CONCENTRATIONS -  SITE: PACS1        DATE: 08/13/77  VERSION: 7.0
SAMPLE DURATION        24      START HOUR       0          SIZE:    FINE
       R SQUARE       .98    PERCENT MASS    98.7
     CHI SQUARE      1.12              DF      13

SPECIES-------I[a]--MEAS[b]----------------CALC[b] -----------RATIO C/M----RATIO R/U
C1      TOT    T  80.00000+-  8.00000  78.96461+-  4.82449     .99+-  .12       -.1
C9      F           .88300+-   .08800    .67644+-   .24792     .77+-  .29       -.8
C11     NA     *   6.93000+-   .69300   6.97025+-   .56446    1.01+-  .13        .0
C12     MG     *    .43000+-   .04300   1.60951+-   .62627    3.74+- 1.50       1.9
C13     AL     *   4.66000+-   .46600   4.02418+-   .88919     .86+-  .21       -.6
C14     SI     *   3.02000+-   .30200   2.92212+-   .13329     .97+-  .11       -.3
C16     S          2.95000+-   .29500   3.02466+-   .31807    1.03+-  .15        .2
C17     CL     *   5.95000+-   .59500   5.69381+-  1.24836     .96+-  .23       -.2
C19     K      *   1.64000+-   .16400   1.73084+-   .46411    1.06+-  .30        .2
C20     CA     *   1.78000+-   .17800   1.43537+-   .11366     .81+-  .10      -1.6
C22     TI     *    .08300+-   .00800    .10088+-   .01630    1.22+-  .23       1.0
C23     V      *    .37200+-   .03700    .39757+-   .08308    1.07+-  .25        .3
C24     CR     *    .31500+-   .03200    .20976+-   .12151     .67+-  .39       -.8
C25     MN     *   2.99000+-   .29900   2.82844+-   .14115     .95+-  .11       -.5
C26     FE     *   4.53000+-   .45300   4.24446+-   .33269     .94+-  .12       -.5
C28     NI     *    .76500+-   .07700    .68246+-   .13428     .89+-  .20       -.5
C29     CU     *    .04400+-   .00400    .05274+-   .00510    1.20+-  .16       1.3
C30     ZN     *    .22500+-   .02300    .26786+-   .03966    1.19+-  .21        .9
C35     BR     *    .41900+-   .04200    .56133+-   .17386    1.34+-  .44        .8
C82     PB     *   2.53000+-   .25300   2.13749+-   .30300     .84+-  .15      -1.0
C201    OC     *   7.54000+-   .75400   8.50978+-  1.35632    1.13+-  .21        .6
C202    EC     *   1.42000+-   .14200   1.33579+-   .34012     .94+-  .26       -.2
C203    SO4    *  10.30000+-  1.03400   9.78819+-  1.47514     .95+-  .17       -.3
C204    NO3    *    .63800+-   .06400    .88402+-   .35938    1.39+-  .58        .7
--------------------------------------------------------------------------------
```

[a] * = species included in calculation of source contribution estimtes

[b] measured (MEAS) and calculated (CALC) concentrations in μg/m^3

number of additional studies which provide data bases of shorter duration. Hopke and Dattner (1982), Pace (1985), Mathai and Stonefield (1988) and Watson (1989) also report the existence of particulate and gaseous data bases suitable for receptor modeling.

5. USING THE CMB MODEL

Every air quality model is a simplification of reality, and the CMB model is no exception. There are some situations under which this simplification is a good approximation, and other situations under which it is not. In is incumbent on the model user to apply the model correctly and to test its validity every time he uses it. To facilitate this process, an applications and validation protocol has been developed for PM_{10} modeling (US EPA, 1987a), and this protocol can be generalized for applications to other types of pollutants. The protocol consists of seven steps which are described in the following sub-sections. The software needed to implement this protocol is available free of charge from the United States Environmental Protection Agency. Familiarity with this software will provide a perspective for the use of this protocol.

5.1 Step 1: Determine the Applicability of the CMB to the Situation Being Studied

The following conditions must be met for the CMB to be applicable:

- A sufficient number of receptor samples have been taken with an accepted sampling method to represent the range of concentrations expected at the receptor.
- Samples are amenable to or have been analyzed for a variety of chemical species.
- The potential source contributors can be identified and grouped into source types of distinct chemical compositions with respect to the species measured at the receptor.
- Profiles for the source types are obtainable which represent the composition of emissions as they would appear at the receptor.

- The number of source types is less than the number of chemical species measured at the receptor.

Unless all five of these requirements are fulfilled, the CMB receptor model is not applicable to the situation under study. These are necessary, but not sufficient requirements. Even if they are met, it may still be found that the precision and validity of CMB results are not adequate to justify pollution control decisions. The remaining steps in the applications and validation protocol must be taken to arrive at this conclusion.

5.2 Step 2: Set Up the Model and Perform an Initial Solution

This step requires the selection of receptor chemical species measurements and source profiles appropriate for the area being studied. CMB modeling software is interactive and allows any subset of chemical species and source profiles to be used in a calculation of source contribution estimates. With most computer memory configurations, more than one hundred source profiles and chemical species can be accommodated.

The chemical species must be present in both the source profiles and in the receptor data sets. A default value of zero with a standard deviation equal to an analytical detection limit may be assigned to a species in a source profile if that species has not been reported but is known to be absent in emissions from that source type. Species with values which are commonly below detection limits in receptor samples, but which are found at detectable levels in source profiles, should not be discarded. Their uncertainty can be set to the analytical detection limit in the receptor data base. When these species are included in the calculations of source contribution estimates, the standard error places an upper limit on the contributions from source types in which these species are present.

Source profiles are selected to represent those which those source types which are known to be operating in a study area. Ubiquitous area sources, such as motor vehicle exhaust, resuspended road dust, and wood combustion should always be included since they are present in nearly all populated areas. Natural sources, such as sea salt and windblown dust, need to be added if the receptor is likely to be affected by such sources. Industrial sources (e.g. steel mills, power plants, smelters) are usually identifiable from emissions inventories. Profiles for secondary particles can be constructed from the stoichiometric ratios of their molecular components. Common secondary profiles include sulfuric acid, ammonium sulfate, ammonium nitrate, and sodium nitrate.

Some profiles may be too similar in composition to allow resolution by the CMB model. If this is the case, these profiles can be combined in different proportions to create a "composite" source profile. This compositing is often needed for leaded, unleaded, and diesel exhaust profiles and for different fugitive dust profiles.

Several profiles for the same source type can be included in the input data files to represent different compositions which might occur for different samples. For example, Watson *et al.* (1988) used composite motor vehicle profiles which contained 50%, 75%, and 85% diesel exhaust to apportion fine particles in wintertime Denver, CO. In general, the 50% diesel profiles best accounted for the receptor concentrations during daytime, while the 75% and 85% diesel profiles best accounted for the receptor concentrations in nighttime samples. This was consistent with the larger proportion of diesel traffic through Denver at night when compared to the daytime traffic.

A few samples from each site are selected for initial CMB testing. A subset of source profiles and chemical species is selected to calculate source contribution estimates and the model outputs, statistics, and diagnostics are examined to determine how well these profiles explain the measured concentrations. This subset is selected after examination of these statistics for several trials in which different profiles are substituted for the same source type.

5.3 Step 3: Examine Outputs, Statistics and Diagnostics

A typical CMB model output is shown in Table 1. The following information is presented in this output:

- Source Contribution Estimate (SCE). This is the contribution of each source type to the pollutant being apportioned, which is mass concentration in the case of particulate matter. Each of the SCE should be greater than zero and none should exceed the total mass concentration.

- Standard Error (STDERR). This is an indicator of the precision or certainty of each SCE. Its magnitude is a function of the uncertainties in the input data and the amount of collinearity among source profiles. It is desirable to have this value be much less than the source contribution estimate.

- t-Statistic (TSTAT). This is the ratio of the source contribution estimate to the standard error. A high value for TSTAT, at least greater than 2.0, shows that the relative precision of the source contribution estimate is high and that the contribution is significant. A low TSTAT value (<2.0) means that a source contribution is not present at a level which exceeds two times the STDERR. Twice the STDERR is a reasonable estimate of the upper limit for a source contribution when TSTAT <2.0.

- R-Square (R SQUARE) and Chi-Square (CHI SQUARE). The R SQUARE statistics measures the variance in the receptor concentrations which is explained by the calculated species concentrations. The CHI SQUARE statistic is the sum of the squares of differences between calculated and measured species concentrations divided by the effective variance and the degrees of freedom (DF). A low R SQUARE (<0.8) indicates that the selected source profiles have not accounted for the variance in the selected receptor concentrations. A large CHI SQUARE (>4.0) means that one or more or the calculated species concentrations differs from the measured concentrations by several uncertainty intervals. The values for these statistics exceed their targets when: 1) contributing sources have been omitted from the CMB calculation; 2) one or more source profiles have been selected which do not represent the contributing source types; 3) precisions of receptor or source profile data are underestimated; and/or 4) source or receptor data are inaccurate.

- Percent of Mass Accounted For (PERCENT MASS). This is the ratio of the sum of the source contributions to the measured mass for particulate samples. For volatile organic compounds it could also be the ratio of the sum of the contributors to total organics measured in a gaseous sample. The target value is 100%, with a reasonable range of 80 to 120%. Percent mass values are outside of this range result when: 1) source profiles have been incorrectly specified; 2) contributing source types have been omitted from the calculation; 3) mass or chemical species measurements are inaccurate; 4) mass measurements are less than 10 $\mu g/m^3$ and within a few precision intervals of the measurements.

- Uncertainty/Similarity Clusters (U/S CLUSTERS). This display shows the results of a singular value decomposition analysis of collinearity. When source profiles are similar in composition, the CMB model does not accurately apportion species to the source types which they represent. This is often the case for geological sources, different types of vegetative burning, and motor vehicle exhaust. These clusters summarize the results of a singular value decomposition analysis of the receptor concentrations, source profiles and their uncertainties. Source identifiers on the same line show the potential for different sources to be collinear. For example, the U/S CLUSTERS in Table 1 show that the marine aerosol and the kraft recovery boiler may be collinear (sources 1 and 8 are on the same line). These sources may also appear in the U/S CLUSTERS when the uncertainties in the source profiles are very high. The SUM OF CLUSTER SOURCES which appears with each cluster can be examined to determine whether collinearity or large input data uncertainties create the cluster.

- Sum of Combined Sources (SUM OF CLUSTER SOURCES). This value is the sum of the source contributions in the cluster coupled with the standard error of that sum. The standard error is not a simple square root of the sum of the squares of the standard errors of the source contributions in the cluster--it contains cross-product terms to account for correlated errors resulting from collinearity. When the relative standard error for this sum is much less than the relative standard errors of any or all of the contributions from sources in the cluster, the standard errors of source contribution estimates can probably be reduced by creating a composite profile for the sources in the cluster. This uncertainty reduction comes at the expense of resolving the collinear source contribution estimates from one another.

- Ratio of Residual to Its Standard Error (RATIO R/U). This column contains the ratio of the signed difference between the calculated and measured concentration (the residual) divided by the uncertainty of that residual (square root of the sum of the squares of the uncertainty in the calculated and measured concentrations). The R/U ratio specifies the number of uncertainty intervals by which the calculated and measured concentrations differ. When the absolute value of the R/U ratio exceeds 2, the residual is significant. If it is positive, then one or more of the profiles is contributing too much to that species. If it is negative, then there is an insufficient contribution to that species and a source may be missing. The sum of the squared R/U for fitting species divided by the degrees of freedom yields the chi square. The highest R/U values for fitting species are the cause of high chi square values.

- Ratio of Calculated to Measured Species (RATIO C/M). The column entitled RATIO C/M shows the ratio of calculated to measured concentration and the standard error of that ratio for every chemical species with measured data. The ratios should be near 1.00 if the model has accurately explained the measured concentrations. Ratios which deviate from unity by more than two uncertainty intervals indicate that an inorrect set of profiles is being used to explain the measured concentrations.

- Source Species Concentrations (SSCONT). Table 2 shows the fractional contribution of each source to each species. This information is used to identify potentially incorrect profiles which might be responsible for a RATIO R/U which exceeds the target range. For example, MG in Table 1 shows a RATIO C/M of 3.74 ± 1.50 and a RATIO R/U of 1.9. Table 2 shows that both MARINE and STEEL profiles contribute more than 100% of the measured magnesium concentrations. Either the ambient or source profile Mg data should be examined for accuracy in this case.

- Modified Pseudo-Inverse Matrix (MPIN). The MPIN is shown in Table 3 for the example of Table 1. This matrix shows the species which are most influential on the source contribution estimates. Absolute values greater than 0.5 show that a species is very influential, while values between 0.3 and 0.5 show that it is moderately influential. For example, Table 3 shows that chloride and sodium are extremely influential in the MARINE source contribution, as expected. The sulfate is also influential for this species, which is not so obvious from a cursory examination of the source profile.

5.4 Step 4: Identify Significant Deviations from Model Assumptions

The source contribution estimates and the diagnostics cited in Section 5.3 can be identified to determine the extent to which the assumptions stated in Section 3 have been complied with. If source profiles have reacted with one another and changed between source and receptor, if contributing sources have been omitted, or if input data have been improperly validated, the CHI SQUARE and R/U RATIOS will not be within target ranges. If the number of sources exceeds the number of fitting species used in the calculation, no solution will be returned and the DF will be less than zero. If profiles are collinear, source contribution estimates will have large standard errors and U/S CLUSTERS will be found.

5.5 Step 5: Adjust Model Inputs to Comply with Assumptions

The ambient data, source data, and combination of species and profiles can be changed to bring the CMB diagnostics into acceptable target ranges. The accuracy and validity of ambient and source data are examined first. If the concentration of one species is suspect for analytical reasons or because it may be contributed by an unknown source type, this species can be removed from the source apportionment calculation. If the suspect species influential in the calculation process, as shown by the MPIN diagnostic, then it a re-analysis of that sample might be desirable.

Table 2. Example of SSCONT Display

```
                                 SPECIES(PER SOURCE)
            INDIVIDUAL RATIO = -------------------------
                               MEAS SPECIES(ALL SOURCES)

                                             SOURCE NAME
SPECIES MARIN  UDUST  AUTPB  RDOIL  KRAFT  ALPRO  STEEL  FERMN
 TOT      .155   .120   .126   .138   .059   .133   .108   .148
 F        .000   .000   .000   .007   .000   .720   .000   .039
 NA       .715   .017   .000   .056   .086   .063   .016   .053
 MG      1.383   .290   .000   .000   .069   .690  1.311   .000
 AL       .000   .182   .024   .013   .003   .614   .012   .016
 SI       .000   .708   .027   .035   .002   .012   .144   .039
 S        .139   .012   .014   .499   .186   .050   .058   .068
 CL       .833   .000   .051   .000   .014   .024   .027   .008
 K        .106   .060   .004   .019   .043   .014   .049   .760
 CA       .097   .131   .071   .098   .000   .020   .302   .087
 TI       .000   .740   .000   .147   .003   .051   .209   .066
 V        .000   .006   .000  1.023   .000   .018   .014   .008
 CR       .000   .014   .000   .017   .042   .000   .578   .016
 MN       .000   .004   .000   .002   .000   .000   .252   .687
 FE       .000   .127   .047   .073   .012   .011   .613   .055
 NI       .000   .001   .002   .775   .008   .026   .079   .000
 CU       .000   .065   .167   .189   .022   .106   .552   .097
 ZN       .000   .047   .157   .197   .014   .007   .463   .306
 BR       .059   .005  1.203   .003   .015   .009   .000   .045
 PB       .000   .014   .797   .005   .000   .001   .026   .002
 OC       .000   .150   .669   .103   .011   .055   .000   .142
 EC       .000   .125   .270   .241   .007   .172   .000   .125
 SO4      .120   .004   .013   .517   .182   .045   .021   .048
 NO3      .000   .000   .144   .113   .000   .068   .000  1.061
```

Large negative residuals indicate that a source is needed which will supply a larger quantity of that species. Source profiles which are not in the solution are examined to determine which ones would supply sufficient quantities of the missing concentrations if they were added to the solution set. The CMB is applied as many times as necessary to determine which source types and profiles best explain the underestimated receptor concentrations. A source type should not be included just because it explains the data. There must be a physical justification for its inclusion. For example, a marine profile may explain the high chloride concentrations observed in wintertime Denver, but the profile is probably more closely associated with road sanding material than it is with sea salt.

Non-contributing source types, or better stated, source types with contributions lower than the standard errors of the source contribution estimates, are identified by a t-statistic less than 2. Table 1 shows that the Kraft Recovery

Table 3. Modified Psuedo Inverse Matrix (MPIN)

TRANSPOSE OF SENSITIVITY MATRIX

					SOURCE NAME			
SPECIES	MARIN	UDUST	AUTPB	RDOIL	KRAFT	ALPRO	STEEL	FERMN
NA	.99	.01	-.03	-.09	.14	.01	-.06	.01
MG	.21	-.05	-.04	-.04	-.07	.13	.18	-.08
AL	-.17	-.03	-.10	-.11	-.05	1.00	-.14	.01
SI	-.02	1.00	-.05	-.05	.04	-.14	-.19	.02
CL	1.00	-.02	.04	.12	-.36	-.04	.01	-.02
K	-.04	.05	-.03	-.05	.11	-.01	-.22	.46
CA	.28	.02	.03	.12	-.19	-.03	.39	-.06
TI	.00	.49	-.07	.08	-.03	-.02	-.07	.02
V	.23	-.04	-.06	1.00	-.40	-.04	-.19	.04
CR	-.10	-.07	-.03	-.12	.13	-.02	.33	-.12
MN	-.02	-.07	-.10	-.04	-.07	-.03	-.05	1.00
FE	-.08	-.12	-.05	-.14	.07	-.06	1.00	-.31
NI	.17	-.06	-.06	.87	-.32	-.02	-.09	-.00
CU	-.12	-.19	.09	.02	.05	.17	.67	-.18
ZN	-.02	-.09	.07	.09	-.02	-.05	.31	.10
BR	.01	-.02	.55	-.05	.03	-.04	-.07	-.01
PB	.00	-.04	1.00	-.06	.00	-.07	-.08	-.05
OC	-.04	.09	.61	.04	.01	.02	-.22	.11
EC	-.01	.03	.16	.17	-.07	.20	-.16	.09
SO4	-.66	.01	-.04	.04	1.00	-.04	-.16	-.01
NO3	-.00	.01	.03	.06	-.04	.04	-.19	.35

Boiler contribution meets this criterion. Such source types may be omitted from the calculation, or they may be retained to provide an upper limit on their contributions if they are not so collinear with other source profiles that they elevate the standard errors of other source contribution estimates.

The most difficult assumption to attain in practice is that of linear independence of the source profiles. As described above, the UC/CLUSTERS can be used to identify potentially collinear source profiles. If this collinearity causes excessive standard errors in source contribution estimates, the following steps can be taken in an attempt to better comply with this CMB assumption: 1) measure additional species which are better markers for the collinear source types; 2) combine profiles for the collinear source types into a composite source profile; 3) use the sum of collinear source contributions presented in the cluster; or 4) reduce the uncertainties in the source profiles by better measurement methods.

5.6 Step 6: Verify the Consistency/Stability of Model Results

There may be more than one solution to the CMB equations which meets the diagnostic criteria equally well. The source contribution estimates may also be very sensitive to small changes in the concentrations of a small number of species. The user must explore these possibilities by performing sensitivity studies on selected samples which are representative of the entire data set.

These sensitivity tests are conducted by: 1) adding and removing species in the calculation; and 2) substituting different profiles for the same source type. For example, lead or bromine can be removed from the solution to test the stability with respect to these species. When elemental and organic carbon are part of the data set, it is often found that the motor vehicle exhaust contribution does not change by more than one standard error, even though the MPIN diagnostic shows that these are very influential species for this source. Watson *et al.* (1988) performed sensitivity tests using source profiles in which the lead content of motor vehicle exhaust varied by a factor of two. They found that the source contribution for this source did not change by more than 25%, which was within one standard error of the source contribution estimates.

5.7 Step 7: Evaluate and Reconcile CMB Source Apportionments

The source contribution estimates calculated by the CMB are not necessarily correct just because all of the diagnostic targets have been achieved. Source contributions for samples from several receptors and for several periods need to be examined to determine whether or not the source contributions make sense. For example, contributions from residential wood combustion are expected to be higher during winter than during summer, and the CMB apportionments should reflect this. Samples taken upwind of a point source should show lower contributions from that source than samples taken downwind. Samples taken during rush hour are expected to show larger motor vehicle contributions than samples taken during off-peak driving periods.

It is especially useful to use the CMB model in conjunction with other models. Chow (1985) describes a composite modeling approach which combines Principal Component Analysis (PCA), CMB, and ISC-ST gaussian plume modeling to apportion suspended particulate matter to coal-fired power plants. Freeman *et al.* (1989) describe a software package which allows this application of multiple models to be easily carried out. US EPA (1987b) provides a procedure for the reconciliation of source apportionments obtained independently from source and receptor models.

6. SUMMARY

The derivation and basic assumptions of the Chemical Mass Balance receptor model have been presented, and a protocol for its application and validation has been described. This protocol is being followed in many states for the justification of state implementation plans for meeting PM_{10} standards. Both the regulating and the regulated communities have used the model and this protocol to come into agreement on cost-effective emissions reductions strategies.

The potential for using the CMB framework for the source apportionment of ambient air pollutants other than suspended particulate matter, indoor and workplace air pollutants, and toxic species found in landfills and water is being actively explored, but has no regulatory status at this time.

Like any other model, the CMB must be used within the confines of its limitations. Both the software and the applications and validation protocol have been constructed to make the user aware of those limitations and to force recognition of them.

The CMB has always been intended for use with other source and receptor models. Using these models together provides a greater ability to apportion pollutants to individual emitters. It also provides a check on the results of each of the modeling systems being used.

7. REFERENCES

Ahuja, M., J. Paskind, J.E. Houck, and J.C. Chow, Design of a Study for the Chemical and Size Characterization of Particulate Matter Emissions from Selected Sources in California, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 145-158 (1989).

Belsley, D.A., E.D. Kuh, and R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley and Sons, New York, NY (1980).

Britt, H.I., and R.H. Luecke, The Estimation of Parameters in Nonlinear, Implicit Models, *Technometrics* 15:233-247 (1973).

Chow, J.C., *Development of a Composite Modeling Approach to Assess Air Pollution Source/Receptor Relationships*, Doctor of Science Dissertation, Harvard University, Cambridge, MA (1985).

Chow, J.C., and J.G. Watson, Summary of Particulate Data Bases for Receptor Modeling in the United States, In *Transactions: Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 108-133 (1989).

Cooper, J.A., D.C. Redline, J.R. Sherman, L.M. Valdovinos, W.L. Pollard, L.C. Scavone, and C. Badgett-West, *PM_{10} Source Composition Library for the South Coast Air Basin. Volume I: Source Profile Development Documentation Final Report*, South Coast Air Quality Management District, El Monte, CA (1987).

Core, J.E., Source Profile Development for PM-10 Receptor Modeling, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 134-144 (1989).

Currie, L.A., R.W. Gerlach, C.W. Lewis, W.D. Balfour, J.A. Cooper, S.L. Dattner, R.T. DeCesar, G.E. Gordon, S.L. Heisler, P.K. Hopke, J.J. Shah, G.D. Thurston, and H.J. Williamson, Interlaboratory Comparison of Source Apportionment Procedures: Results for Simulated Data Sets, *Atmospheric Environ.* 18:1517-1537 (1984).

DeCesar, R.T., S.A. Edgerton, M.A.K. Khalil, and R.A. Rasmussen, Sensitivity Analysis of Mass Balance Receptor Modeling: Methyl Chloride as an Indicator of Wood Smoke, *Chemosphere* 14:1495-1501 (1985).

Dzubay, T.G., R.K. Stevens, W.D. Balfour, H.J. Williamson, J.A. Cooper, J.E. Core, R.T. DeCesar, E.R. Crutcher, S.L. Dattner, B.L. Davis, S.L. Heisler, J.J. Shah, P.K. Hopke, and D.L. Johnson, Interlaboratory Comparison of Receptor Model Results for Houston Aerosol, *Atmospheric Environ.* 18:1555-1566 (1984).

Freeman, D.L., J.C. Chow, R.T. Egami, and J.G. Watson, A Receptor and Dispersion Modeling Software Package, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 243-268 (1989).

Friedlander, S.K., Chemical Element Balances and Identification of Air Pollution Sources, *Environ. Sci. Technol.* 7:235-240 (1973).

Gordon, G.E., Receptor Models, *Environ. Sci. Technol.* 14:792-800 (1980).

Gordon, G.E., W.H. Zoller, G.S. Kowalczyk, and S.H. Rheingrover, Composition of Source Components Needed for Aerosol Receptor Models. In *Atmospheric Aerosol: Source/Air Quality Relationships*, E.S. Macias and P.K. Hopke, eds. Symposium Series No. 167, American Chemical Society Washington, DC, 51-74 (1981).

Gordon, G.E., Receptor Models, *Environ. Sci. Technol.* 22:1132- (1988).

Henry, R.C., Stability Analysis of Receptor Models that Use Least Squares Fitting, In *Receptor Models Applied to Contemporary Air Pollution Problems*, S.L. Dattner and P.K. Hopke, eds.,Proceedings No. SP-48, Air & Waste Management Association, Pittsburgh, PA, 141-157 (1982).

Hidy, G.M., and S.K. Friedlander, The Nature of the Los Angeles Aerosol, 2nd IUAPPA Clean Air Congress, Washington, DC (1972).

Hopke, P.K., and S.L. Dattner, editors, *Receptor Models Applied to Contemporary Pollution Problems*, Proceedings No. SP-48, Air & Waste Management Association, Pittsburgh, PA (1982).

Hopke, P.K., *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons, New York, NY (1985).

Houck, J.E., J.C. Chow, J.G. Watson, C.A. Simons, L.C. Pritchett, J.M. Goulet, and C.A. Frazier, *Determination of Particle Size Distribution and Chemical Composition of Particulate Matter from Selected Sources in California. Volume I.* Prepared for California Air Resources Board, Sacramento, CA, by OMNI Environmental Services, Inc., Beaverton, OR, and Desert Research Institute, Reno, NV (1989).

Javitz, H.S., J.G. Watson, and N.F. Robinson, Performance of the Chemical Mass Balance Model with Simulated Local-Scale Aerosols, *Atmospheric Environ.* 22:2309-2322 (1988a).

Javitz, H.S., J.G. Watson, J.P. Guertin, and P.K. Mueller, Results of a Receptor Modeling Feasibility Study, *J. Air Pollut. Contr. Assoc.* 38:661-669 (1988b).

Kim, B., and R.C. Henry, Analysis of Multicollinearity Indicators and Influential Species for Chemical Mass Balance Receptor Model, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 379-390 (1989).

Kneip, T.J., M.T. Kleinman, and M. Eisenbud, Relative Contribution of Emission Sources to the Total Airborne Particulates in New York City, 3rd IUAPPA Clean Air Congress (1972).

Lioy, P.J., J.G. Watson, and J.D. Spengler, APCA Specialty Conference Workshop on Baseline Data for Inhalable Particulate Matter, *J. Air Pollut. Contr. Assoc.* 30:1126 (1980).

Malm, W.C., M. Pitchford, and H.K. Iyer. Design and Implementation of the Winter Haze Intensive Tracer Experiment - WHITEX, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 432-458 (1989).

Mathai, C.V. and D.H. Stonefield, eds, *PM_{10}: Implementation of Standards*, Transactions No. TR-13, Air & Waste Management Association, Pittsburgh, PA (1988).

McDade, C.E., M.R. McGowen, and J.G. Watson, Identifying Long-Range Tracers of Los Angeles Emissions, In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 407-417 (1989).

Pace, T.G., ed., *Receptor Methods for Source Apportionment: Real World Issues and Applications*, Transactions No. TR-5, Air Pollution Control Association, Pittsburgh, PA (1986).

Pasquill, F., *Atmospheric Diffusion* (2nd Ed.), Ellis Horwood, England (1974).

Seinfeld, J.H., *Air Pollution: Physical and Chemical Fundamentals: Second Edition*, McGraw-Hill, New York, NY (1986).

Shareef, G.S., L.A. Bravo, J.H.E. Stelling, W.B. Kuykendal, and J.D. Mobley, Air Emissions Species Data Base. In *Receptor Models in Air Resources Management*, J.G. Watson, ed., Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA 73-83 (1989).

Sheffield, A.E. and G.E. Gordon, Variability of Particle Composition From Ubiquitous Sources: Results from a New Source-Composition Library, In *Receptor Methods for Source Apportionment: Real World Issues and Applications*, T.G. Pace, ed, Transactions No. TR-5, Air Pollut. Control Assoc., Pittsburgh, PA (1986).

U.S. Environmental Protection Agency, *Protocol for Applying and Validaing the CMB Model*, Environmental Protection Agency Report No. EPA-450/4-86-010, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987a).

U.S. Environmental Protection Agency, *Protocol for Reconciling Differences Among Receptor and Dispersion Models*, Environmental Protection Agency Report No. EPA-450/4-87-008, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987b).

U.S. Environmental Protection Agency, *Air Emissions Species Manual: Volume II Particulate Matter Species Profiles*, Environmental Protection Agency Report No. EPA-450/2-88-003b, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC (1988).

Watson, J.G., *Chemical Element Balance Receptor Model Methodology for Assessing the Sources of Fine and Total Particulate Matter*, Ph.D. Dissertation, Oregon Graduate Center, Beaverton, OR, University Microfilms International, Ann Arbor, MI (1979).

Watson, J.G., Overview of Receptor Model Principles, *J. Air Pollut. Contr. Assoc.* 34:620-623 (1984).

Watson, J.G., J.A. Cooper, and J.J. Huntzicker, The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model. *Atmospheric Environ.* 18:1347-1355 (1984).

Watson, J.G. and N.F. Robinson, A Method to Determine Accuracy and Precision Required of Receptor Model Measurements, In *Quality Assurance in Air Pollution Measurements*, Air Pollution Control Association, Pittsburgh, PA, (1984).

Watson, J.G., J.C. Chow, D.L. Freeman, R.T. Egami, P. Roberts, and R. Countess, *Model and Data Base Description for California's Level I* PM_{10} *Assessment Package*, Document 8066-002.1F1, Final Report, prepared for the California Air Resources Board, Sacramento, by Desert Research Institute, Reno, NV (1987).

Watson, J.G., J.C. Chow, L.W. Richards, S.R. Anderson, J.E. Houck, and D.L. Dietrich, *The 1987-88 Metro Denver Brown Cloud Air Pollution Study, Volume III: Data Interpretation*, DRI Document No. 8810.1F3, prepared for the Greater Denver Chamber of Commerce, Denver, CO, by Desert Research Institute, Reno, NV (1988).

Watson, J.G., editor, *Receptor Models in Air Resources Management*, Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, (1989).

Watson, J.G., N.F. Robinson, J.C. Chow, R.C. Henry, B.M. Kim, T.G. Pace, Q. Nguyen, and E.L. Meyer, EPA/DRI Version 7.0 Chemical Mass Balance Modeling Software, *Environmental Software* 5:38-49 (1990).

White, W.W., MCLS Modeling Software, In *Receptor Models in Air Resources Management*, J.G. Watson, ed, Transactions No. TR-14, Air & Waste Management Association, Pittsburgh, PA, 379-390 (1989).

Williamson, H.J. and D.A. DuBose, *Receptor Model Technical Series, Volume III: User's Manual for Chemical Mass Balance Model*, Environmental Protection Agency Report No. EPA-450/4-83-014, U.S. Environmental Protection Agency, Research Triangle Park, NC (1983).

Winchester, J.W. and G.D. Nifong, Water Pollution in Lake Michigan by Trace Elements from Aerosol Fallout, *Water, Air, and Soil Pollut.* 1:50-64 (1971).

Chapter 5

MULTIVARIATE RECEPTOR MODELS

Ronald C. Henry
Environmental Engineering Program
Civil Engineering Department
University of Southern California
University Park
Los Angeles, Californai 90089-0231

1. INTRODUCTION

1.1 The Need for Multivariate Receptor Models

Chemical mass balance models given by:

$$C_{ik} = \sum_{j=1}^{p} a_{ij} S_{jk} + \mathit{background} + \mathit{errors} \tag{1}$$

require that the composition of all the contributing source types be known. In many cases this information is not easily obtained either because the emissions are difficult to sample or because the source class consists of many small sources with widely varying compositions. Examples of the first problem are fugitive emissions of particulate matter from roadways or coal storage piles and the emissions of hydrocarbon gases from leaking valves in a refinery. On the other hand, vehicle emissions are a good example of many sources with a great variety of emissions profiles.

Another more technical problem which often occurs in CMB modeling is the disparity between the species determined in the sample and the species given in published lists of source compositions. Many times the published source compositions do not contain all the species that were determined in the sample. As a result, these species cannot be used for the CMB fitting. Volatile organic compounds, trace elements, and toxic species often cannot be included in the CMB for this reason.

As regulatory action is increasingly directed towards nontraditional sources of particles, toxics, and volatile organic compounds, there is a need for receptor modeling techniques that can overcome these problems.

Multivariate receptor models seek to surmount these problems of ordinary CMB by using the data to estimate not only the source contributions but the source compositions as well. They do this by using the information contained in the correlations between the observed species. The basic idea is that species from the same source will be correlated and that these correlations can be used to estimate the composition of the source.

1.2 Challenges to Multivariate Receptor Modeling

All multivariate receptor models have two fundamental difficulties which must be appreciated if they are to be applied in an intelligent manner. First, the problem of deducing the source compositions and contributions using only the data and natural, physical constraints can be shown to be mathematically indeterminate, i.e., there is no unique solution. Secondly, observed correlations between species which are assumed to bear source composition information may instead the result of the mutual effects of meteorology and coincident source location.

Concentrations of air pollutants are determined primarily by the ability of the atmosphere to disperse the contaminants. Generally, concentrations of all species tend to rise and fall together with changing meteorological conditions. Also, some sources are coincident in space and time, making it impossible to separate them by multivariate means alone. For example, direct vehicle roadway emissions are virtually perfectly correlated with resuspended road dust. Thus, the basic information utilized by multivariate models may be misleading.

These fundamental difficulties do not have simple solutions. The non-uniqueness problem can only be addressed by adding additional constraints, physical or mathematical, on the possible solutions. Likewise, the problem of coincident sources can only be overcome with additional assumptions or information, such as particle size and shape (Hopke and Casuccio, 1991).

1.3 Outline of the Chapter

This chapter will review the major multivariate methods which have been applied to receptor modeling. The Source Apportionment by Factors with Explicit Restrictions (SAFER) model is discussed in greater detail than other models because all the others suffer from the fundamental mathematical indeterminacy discussed above. The other models continue to be important as semi-quantitative methods to estimate composition of sources from the data alone. Of course, all the multivariate methods can be extremely valuable in identifying the existence of unsuspected sources.

In the next sections the concepts basic to multivariate receptor models are introduced. This is followed by a more detailed description of the SAFER model and its application to Los Angeles PM-10 data.

2. MULTIVARIATE MODEL FUNDAMENTALS

The mathematical fundamental of receptor models in general and multivariate models in particular are reviewed in Henry *et al.* 1984. All these models are based at least implicitly on the singular value decomposition of the data.

Several multivariate statistical models have been applied to air pollution studies: principal component analysis (Gatz, 1978; Henry and Hidy, 1979, 1982), factor analysis (Blifford and Meeker, 1967; Hopke *et al.*, 1976; Gaarenstroom *et al.*, 1977; Henry, 1978; Heidam, 1984; Thurston and Spengler, 1985; Keiding *et al.*, 1986; Keiding and Sorensen, 1987), target transformation factor analysis (Alpert and Hopke, 1980, 1981; Hopke *et al.*, 1980, 1983), factor analysis followed by multiple linear regression (Kleinmann *et al.*, 1980a, 1980b), and self-modeling curve resolution (Henry and Kim, 1989, 1990; Kim, 1989).

The mathematical foundation of all these techniques, whether used explicitly or not, is singular value decomposition.

2.1 Singular Value Decomposition (SVD), Principal Component Analysis (PCA)

Almost all multivariate receptor models begin, implicitly or explicitly, with a PCA of the data, which has usually been transformed in some manner. After the PCA, most models transform the principal components. Once transformed, the principal components are known as factors and the model is a factor analysis model. PCA is best understood in terms of SVD, as described below.

2.1.1 Singular Value Decomposition

The singular value decomposition theorem (Lawson and Hanson, 1974; Dillon and Goldstein, 1984) states that any matrix C can be written uniquely as

$$C = UDV^T. \tag{2}$$

In this application, C is a data matrix of m rows of species and n observations arranged in columns. V is the n x n eigenvectors of C^TC, U is the m x m eigenvectors of CC^T, and D is an m x n diagonal matrix made up of the singular values, i.e, the square roots of the corresponding eigenvalues. Several mathematical properties of the SVD are useful: the eigenvalues of C^TC and CC^T are always equal; the matrices U and V are orthogonal, i.e., $U^TU = UU^T = V^TV = VV^T = I$, the identity matrix.

2.1.2 Principal Component Analysis

The principal component solution can be obtained from the singular value decomposition of the data matrix (Jolliffe, 1986). In equation 2, let L = UD and F = V^T, then,

$$C = LF. \tag{3}$$

In PCA, L is called the component loading matrix, and F is called the matrix of component scores.

It is also possible to define L=U and F=DV^T. This is called Q-mode analysis while the former approach is known as R-mode analysis. Applications of PCA in physical sciences usually use R-mode while Q-mode is the rule in the social sciences. A major advantage of the SVD formalism is that it avoids this pedantic and often unnecessary confusion over terminology. The SVD theorem makes clear that PCA is essentially unique and which mode to use depends mostly on the form of the underlying physical model.

The basic purpose of PCA is to reduce the dimensionality of a data set of interrelated variables so that a minimum number of factors can explain the maximum variance of the interrelated data. The principal components are extracted so that the first component accounts for the largest amount of the total variation in the data, the second principal component accounts for the maximum amount of the remaining total variation not already accounted for by the first principal component, and so on. This is accomplished by orthogonally transforming the correlated data into a new set of uncorrelated variables. Mathematically, this procedure is equivalent to an eigenanalysis, which produces eigenvalues and corresponding eigenvectors.

PCA has been used as a technical tool to reduce the dimensionality of a large highly collinear data set or as a qualitative tool to describe the interrelationships among such data. However, it can form the basis for predictive physical models. These are known as factor analysis models and their derivation is given next.

2.2 Factor Analysis

The classical factor analysis model (Harman, 1976) is expressed as

$$C = LF + U, \tag{4}$$

where C is the data matrix as defined above, L is the n by p matrix of factor loadings, F is the p by m matrix of factor scores, and U is the n by m unique factor matrix. The reduces to the PCA if the unique factor term is not included in the analysis, and L is the matrix of eigenvectors of CC^T and **F** is the matrix of eigenvectors of C^TC multiplied by the singular values.

The main differences between factor analysis and PCA stems from the explicit inclusion of unique factors and the fact that the factors need not be unique or even orthogonal as in the case of PCA. Detailed explanations of the differences between these two methods are made by Jolliffe (1986), and Dillon and Goldstein (1984).

The application of factor analysis models to source apportionment is based on the fact that the factor analysis model as given in Equation 4 has exactly the same mathematical form as the CMB equation,

$$C = AS + E, \tag{5}$$

where C is the n x m data matrix, A is source composition matrix of n species by p sources, S is a p x m source contribution matrix, and E is a n x m matrix of random observational errors. In the following, the effects of errors E is not considered. The reader is referred to the statistical literature on this subject. In particular, the estimation of communalities in factor analysis. Generally, this has little impact on source apportionment calculations.

2.2.1 Transformations and Factor Analysis Models

Eigenvectors V and U contain the information regarding the source composition and source apportionment. However, U and V are not necessarily equal to the actual source composition and apportionment matrices. The abstract mathematical solution of the eigenvectors V has to be transformed to a solution having physical meaning as source composition vectors. Mathematically, this is equivalent to finding the appropriate transformation matrix, T, that satisfies the equation

$$C = (UT^{-1})(TDV^{T}) = AS. \tag{6}$$

The transformation matrix in receptor modeling serves as a mathematical bridge between the eigenvectors and the source compositions. Theoretically, an infinite number of transformation matrices could explain the ambient data equally well. Therefore, the ultimate objective of the model is to determine one unique transformation matrix, T. The question is how to find such a transformation matrix.

There are several different transformation methods, such as orthogonal transformations which preserve the statistical independence of the factors (VARIMAX, QUARTIMAX, PROMAX), and oblique transformations which allow the factors to be dependent (OBLIMIN, target transformation). However, as Henry (1987) and Lowenthal and Rahn (1987) pointed out, these transformation methods cannot be relied upon to produce results consistent with physical reality. They transform the abstract eigenvectors to other abstract solutions that do not guarantee the transformed results are physically meaningful. For example, the transformed results often give negative source compositions which are difficult to interpret physically.

Physical constraints cannot be easily incorporated into factor analysis receptor models through the formalism of transformations as presented above. However, there is a geometrical interpretation of the SVD which is exceptionally well suited to this task. It is introduced in the next section.

2.3 Self-Modeling Curve Resolution

In analytical chemistry, identifying the components in samples that consist of mixtures of unknown amounts of an unknown number of unknown compounds is entirely analogous to the factor analysis receptor model problem. Self Modeling Curve Resolution (SMCR), as proposed by Lawton and Sylvestre (1971), is a way of imposing physical constraints on the mixture problem.

They pointed out that any source composition vector that is consistent with the observed data must be in the space spanned by the columns of the U matrix in equation 2. By similar reasoning, each vector of observed concentrations in a sample can also be written as a linear combination of the columns of U. Mathematically this is expressed as

$$c_{ij} = \sum_{k=1}^{p} \alpha_{ik} u_{kj}, \tag{7}$$

where

$$\alpha_{ik} = C_j \cdot U_j = \sum_{j=1}^{n} c_{ij} u_{jk}. \tag{8}$$

Geometrically, this means that the true source composition vectors and the data vectors can be represented as points in the space spanned by the columns of U. In the above, the ith data point is plotted at coordinates α_{ik}, k=1...p. The geometric relationship between the data points and the source composition points is easy to see. Because the data points are non-negative linear combinations of the source composition points, the data points must lie inside the convex hull formed by the source composition points and the origin. This geometrical interpretation is very useful a way of representing physical constraints as well be discussed in a later section.

2.4 Scaling the Data

Statistical applications of multivariate analysis usually work with scaled data rather than the raw data as in the above discussion. The most common transformations are centering to the mean and scaling to the standard deviation:

$$c_{ij} = \frac{c_{ij} - \bar{c}_i}{\sigma_i}, \tag{9}$$

where $\bar{c}_i$ is the average of species i and σ_i is the standard deviation.

If the data are not centered on the mean, then the first principle component is basically the average. Centering on the average removes this source of variability. If the data are not centered, this is called centering on the origin.

The main reason for scaling to the data to the standard deviation is to compensate for the wide range of concentrations of different species. Since SVD or PCA explains the variability in the data, species with large concentrations will dominate. By scaling as above all the species have a standard deviation of 1. In this way a trace species is given as much weight as a major component. Sometimes this is desirable, however, if the measurement of the trace species has a large error, this scaling could amplify the noise in the measurement.

Scaling is the responsible for the difference between Q and R mode factor analysis in that one may scale according to row or column means and standard deviations which will generally be different.

Statisticians universally recommend that the data be centered and scaled as above. However, statisticians are usually not interested in building a physical model. In the application of multivariate methods to source apportionment, scaling causes several problems. By centering on the mean, the average of all the variables becomes zero. Thus, the average source apportionment is similarly equal to zero. A result of little regulatory interest. Thus, if the data are centered, some means must be found to un-center the results. Thurston and Spengler (1985) presents one method which is straightforward but lacks fundamental theoretical support. A generally acceptable solution to this problem has yet to be developed for models that use centered data. Perhaps the best approach is simply not to center the data.

Undoing the effects of scaling to the standard deviations is also a problem. In this case, the source apportionments (factor scores) must have a standard deviation of one. Henry (1978) proposed a method to unscale a factor analysis model which required that each source have a unique tracer species. Theoretically, the effects of scaling on the model results typically are easily determined so that the results by be presented in unscaled terms which are necessary for physical models.

2.5 Target Transformation Factor Analysis

Target transformation was originally applied to chemeometrics by Malinowski and coworkers (Weiner, 1970; Malinowski, 1980). Hopke and coworkers have further refined and applied it to source apportionment of the ambient aerosol (Hopke, 1980; Alpert, 1980, 1981). The current state-of-the-art in receptor modeling application of TTFA reviewed in Hopke (1988).

The target transformation method transforms the V eigenvectors to minimize the difference between the transformed vector and a target vector. Mathematically the problem is stated as:

$$\underset{t_{kj}}{\mathit{minimize}} \sum_{l=1}^{n} \left(b_l - \sum_{k=1}^{p} v_{lk} t_{kj} \right)^2 \qquad (11)$$

where the parenthetical expression is a component of the transformed vector and b_{ij} is a component of the target composition vector for the jth source. The least squares estimate of the best T_j is

$$T_j = (V^T V)^{-1} V^T b_j \qquad (12)$$

In this form, target transformation begs the question, since it requires a fairly accurate idea of the composition of the source, the lack of which is the main reason for applying factor analysis. In an attempt to get around this problem TTFA can be applied in an iterative manner (Hopke *et al*, 1983). The process starts with a unique test vector in which a tracer element is assigned an initial value of 1.0 and all other elements are assigned a value of zero. Using the unique test vector, a new value of test vector, b_j', is estimated as

$$b_{ij}' = \sum_{k=1}^{p} v_{ik} t_{kj} \qquad (13)$$

This predicted test vector is then used as input for Equation 12 to find a new T_j. Any negative values that occur in the predicted test vector are replaced with small positive ones. This process is repeated until the average percent changes in the values in b' and b are less than 10^{-4} or some other acceptable value. In this way, the test vector is refined iteratively to the source composition profiles and the CMB model is used to estimate the source contributions from each source.

The iterative TTFA described above often converges on a reasonable source composition. However, there is no theoretical reason why it should converge on the true source composition, even in the absence of random error in the data. Thus, TTFA is a useful screening tool to get an idea of possible source compositions, but it cannot be relied upon to produce a physically meaningful model.

2.6 Multiple Linear Regression

The Factor Analysis followed by Multiple Linear Regression (FA/MLR) approach was proposed by Kleinman and coworkers (Kleinmann *et al.* 1980a, 1980b; Kneip *et al.*, 1983, Morandi *et al.*, 1987) to apportion total suspended particulates (TSP) in the New York City area. This approach was developed to avoid the difficulties in transforming the eigenvectors. It starts with a FA with a VARIMAX rotation to identify sources and to determine tracers for each source. Then a multiple linear regression analysis of TSP against the tracers is used to apportion the sources, as follows

$$TSP_k = \sum_{i=1}^{p} b_i T_{ik} + b_0 + \varepsilon \qquad (14)$$

where TSP_k = observed TSP on k-the day

b_i = regression coefficient for tracer i

T_{ik} = observed concentration of tracer i on k-the day

b_o = unexplained background concentration

ϵ = random error

Once b_i is determined, the source contribution to TSP is obtained from the product of b_i and T_{ik}. This approach is appropriate when impacting sources are known but source compositions are not available or not reliable because the source apportionment can be made without knowing source compositions. However, this method requires tracer elements for each source that are highly correlated with the source and statistically independent of other tracers.

2.7 Required Number of Samples and the Effect of Errors

By definition, multivariate analysis requires several observations, the more the better. If there are too few samples, the results of a multivariate model are not likely to be reliable. But how few are too few? Experience has shown that 100 samples or more is generally acceptable and that 20 or 30 is usually too few. A quantitative rule-of-thumb is derived in this section to provide guidance on this issue. A brief discussion of error in multivariate models will also be given.

Much of the power of statistics derives from the fact that certain statistics derived from a large set of data are more reliable than the individual data points themselves. Errors tend to average out, so the average of 100 numbers is more reliable than the average of 10 numbers. For averages, this is proven by the Central Limit Theorem. A similar theorem does not exist for singular value decomposition, yet the eigenvalues and eigenvectors of a 100 observations must be more trustworthy than those from a set of 10 observations.

A simple rule-of-thumb was developed by Henry and reported in Henry *et al.* (1984). The idea is to calculate the degrees of freedom per variable in the analysis. For example, an average of N observation of 1 variables has N-1 degrees of freedom per variable since the average itself constitutes 1 relationship between the observations. To put it another way, all the data can be reconstructed if one knows N-1 data points and the mean. In the same way, the standard deviation of N observations of 1 variable has N-2 degrees of freedom per variable, because, in addition to the mean, one also has a relationship which defines the standard deviation. Each statistic used in the analysis is a relationship between the observations and reduces the degrees of freedom in the analysis by 1.

In a typical multivariate receptor model with N observations of V variables, there will be V means and V(V-1)/2 variances and correlations. The total number of degrees of freedom in the data set is NV. Thus, the degrees of freedom are reduced to NV - V - V(V-1)/2. This is now divided by V to get the number of degrees of freedom per variable in the analysis:

$$N - 1 - \frac{V-1}{2}.$$

Experience has shown that this number should be at least 30 and preferably 60 or greater.

Admittedly, there is no theoretical support for this approach. However, the author (Henry) has found it to be very useful. A systematic evaluation of the effects of sample size on a multivariate receptor model have been reported in Ito *et al.* (1986). They evaluated the rule-of-thumb as given above and verified that their model results were unreliable if the degrees of freedom per variable were less than 30.

Another rule-of-thumb that the author has found to be generally accurate stems from the fact that the eigenvalues of a correlation matrix are essentially the signal-to-noise ratios for the associated eigenvector, or principal component. In determining the number of factors to include in a multivariate receptor model, one can be guided by this fact. Generally, eigenvectors with eigenvalues greater than 1 are more signal than noise and should be kept in the model. Those eigenvectors with eigenvalues less than 0.5 have more than twice as much noise as signal and should be eliminated from the model. For eigenvalues between 1 and 0.5, the associated eigenvectors are more noise than signal, but they may be important enough to keep in the model. Few of this class of eigenvectors should be kept as is consistent with formulating a physically reasonable model.

Finally, estimates of the errors in the source compositions and source apportionments from a multivariate receptor model are needed if the results are to have any credibility. Unfortunately, an error analysis for multivariate models has not been possible since the effects of errors on the singular value decomposition are not yet been characterized mathematically. The only approaches available are computationally intensive methods using resampling or Monte Carlo methods. In the later, simulated error free data are produced so that the "true" source compositions and apportionments are known. The data, and perhaps the source compositions, are corrupted by known amounts of error. The effects of the error on the multivariate methods can then be estimated if the model is applied a large number of simulated data sets are which differ only in random errors. An early example of the simulated data approach is found in Currie *et al.* (1984).

3. Multivariate Factor Analysis as Applied to Los Angeles Particulate Composition Observations

Los Angeles is justly famous for its intractable air pollution problem. Even the strictest emission controls in the nation have been unable to cope with the combination of explosive growth and poor air pollution meteorology found in the Los Angeles area. While ozone levels are the worst violators of health standards, the concentrations of airborne particles less than 10 μm, known as PM-10, also constitute a health risk to the general population much of the time.

In this section describes the application of an advanced multivariate receptor model to an exceptionally complete set of PM-10 composition data from several sites in the L.A. area. The model used is the Source Apportionment by Factors with Explicit Restrictions (SAFER) model developed by Henry and coworkers. A full description of the model can be found in Kim (1989), and Henry and Kim (1990).

The basic concepts of the model will be described next. This is followed by a brief description of the data and the model results.

3.1 The SAFER Model

The SAFER model is based on self-modeling curve resolution as described in Section 2.3. To recapitulate, the method starts with the singular value decomposition of the data expressed as a matrix **C**:

$$C = UDV^t, \tag{16}$$

where the columns of **U** are the eigenvectors of $\mathbf{CC}^t$, the columns of **V** are the eigenvectors of $\mathbf{C}^t\mathbf{C}$, and **D** is a diagonal matrix of the corresponding singular values (square roots of the eigenvalues). Further, assuming that there are p sources and that n is greater than p, only p of the singular values will be different from zero, assuming there are no random errors. The effects of random errors on our method will be considered later.

As pointed out in section 2.3, if a source's composition is written as a p-dimensional vector, then every possible source composition consistent with the observed data can be written as a linear combination of the eigenvectors that make up the columns of the V matrix. The same is true for the n m-dimensional vectors made up of the observations of m elements at the receptor. Thus, each source composition vector and vector of observations can be thought of as a point in the eigenvector space spanned by the columns of **V**. As such, physical constraints on the source compositions and contributions can be expressed as linear equality or inequality constraints in the V-space. The constraints are divided into two groups, natural physical constraints which must be satisfied by all physically meaningful models and additional physical constraints which apply only to a specific source. These two classes of constraints are considered next.

3.1.1 Natural Physical Constraints

The fundamental, natural physical constraints that must be obeyed by any source composition vector are:

1) The orginal data must be reproduced by the model; the model must explain the observations.

2) The predicted source compositions must be non-negative; a source cannot have a negative percentage of an element.

3) The predicted source contributions to the aerosol must all be non-negative; a source cannot emit negative mass.

4) The sum of the predicted elemental mass fractions for each source must be less than or equal to 1; the whole is greater than or equal to the sum of its parts.

These natural physical constraints can be expressed as linear inequality constraints defined by hyperplanes in the eigenvector space spanned by the columns of **V**. The intersection of the half-spaces of all the non-redundant hyperplanes defines the basic feasible region, which must contain the points corresponding to the composition of the true sources. Geometrically, the basic feasible region takes a predictable form, as shown for the three source case in Figure 1. The non-negativity constraints on the source compositions define the interior of an irregular pyramid with apex at the origin with the first principal component as the axis. A second pyramid, interior to the first, is defined by the non-negativity constraint on predicted source contributions. The feasible points must lie outside this pyramid. Finally, the condition that the sum of the source composition vectors be less than or equal to one defines a hyperplane not passing through the origin which forms the base of the pyramids that define the feasible region. The feasible region is the area between this plane and the origin, as shown in Figure 1.

3.1.2 Additional Physical Constraints

Generally, the major source categories in an area are easily identified. While the exact composition of these emissions is not well known, the major species in specific sources can usually be assumed to lie in some fairly narrow range. In a source dominated by crustal material, silicon composition can be safely assumed to lie between 15 and 25%, for example.

Constraints such as these are easily expressed as linear constraints in the V-space. Like the natural physical constraints, these constraints take a predictable geometrical form. Using the example given above, silicon must lie between two hyperplanes neither of passes through the origin. Of course, the true crustal source must also satisfy the natural physical constraints. Thus, it must lie in the intersection of the basic feasible region and the region between the two hyperplanes.

As more additional physical constraints are added, the region in V-space where point corresponding to the true source composition must lie is further restricted. If a sufficient number of tight constraints can be assumed, then the true composition of the source can be estimated as the centroid of the feasible region and the range of possible source compositions calculated from the vertices of the feasible region.

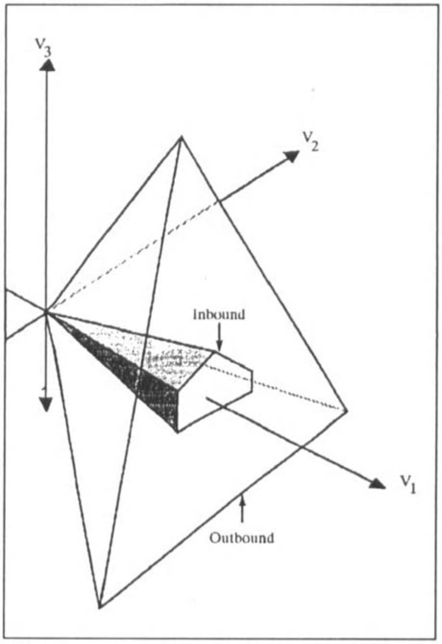

Figure 1. Typical form of the basic feasible region in eigenvector space. Any point inside Outbound and outside Inbound will describe a completely non-negative source composition vector with non-negative source contributions.

Linear programming is the natural method for attacking multidimensional problems with linear constraints. However, linear programming applies only to convex regions, and the feasible region described above may not be convex because of the inner boundary constraints. The problem of determining the region of physically feasible factor models must, therefore, be broken down into a number of separate linear programming problems, as described in Kim (1989).

3.2 Description of the Los Angeles Data

The California Institute of Technology's Environmental Quality Laboratory conducted 24-hr PM10 aerosol sampling every six days from January 1986 to December 1986 in the Los Angeles area of the South Coast Air Basin at nine sampling sites: Burbank, Long Beach, Lennox, Rubidoux, Anaheim, San Nicolas

Island, Downtown L.A., Upland, and Tanbark Flats. Their locations are shown in Figure 2.

PM10 particulate matter was sampled at all the sites and PM-2.5 at Downtown Los Angeles only. Total mass and 41 species (Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Se, Br, Rb, Sr, Y, Zr, Mo, Pd, Ag, Cd, In, Sn, Sb, Ba, La, Hg, Pb, Na, Mg, OC (Organic Carbon), EC (Elemental Carbon), NO_3, SO_4, NH_4) were analyzed. A more detailed description of the monitoring site, sampling schedule, and sample analysis can be found in Solomon *et al.* (1989).

Secondary particles formed by chemical reactions in the atmosphere are a major component of L.A. PM-10. Chemical mass balance models, multivariate or otherwise, cannot directly apportion the secondary particles to sources of precursors. The approach taken in this study was to assume that the chemical reactions in the atmosphere constitute a "secondary source" which appears to be the source of this material at a given site.

In this model application, CO and O_3 gas data have been included with the PM10 data because CO and O_3 should be unique tracers for the motor vehicle source and secondary source, respectively. Therefore, the use of these two gaseous species in additional physical constraints (APCs) should improve the

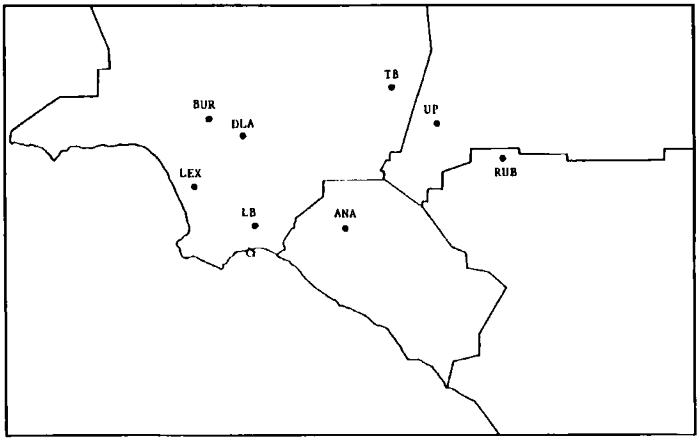

Figure 2. Map of the Los Angeles Basin Showing the Location of the Sampling Sites.

ability of the model to estimate these sources from the data. Mixing particle composition data and gaseous data in this way very difficult for other receptor models, but turns out to be a simple but important aspect of the SAFER model approach.

Data for CO and O_3 are not available from Tanbark Flats, San Nicolas Island, Lennox, and Upland. These data are, however, available from sites at Hawthorne and Ontario, which are within a few miles of Lennox and Upland. Therefore, the Hawthorne and Ontario data are used for the Lennox and Upland sites, respectively. As a result, two coastal sites, Long Beach (LB) and Lennox (LEX), and five inland sites, Downtown Los Angeles (DLA), Burbank (BUR), Anaheim (ANA), Upland (UP) and Rubidoux (RUB), have been selected for SAFER modeling.

3.3 SAFER Modeling of Los Angeles PM-10

The number of contributing sources to each site can be estimated from the number of eigenvalues of the correlation matrix of the ambient data that are greater than 0.5. The number of probable sources identified at each site are: six for RUB; seven for DLA, LB, UP; eight for LEX and ANA; and nine for BUR.

At each site, estimates of the source composition were made for the major source categories: roadway emissions, secondary particle formation, marine particles, and crustal material not associated with roadways.

A unique feature of the SAFER modeling of this data is the use of physical constraints on secondary particles based on the stoichiometric relations among species, as described below.

3.3.1 Additional Physical Constraints for Los Angeles

As indicated above, setting up additional physical constraints is essential to the SAFER model. The starting point for setting up the appropriate additional physical constraints for each source is the selection of species and reasonable ranges for them. Major species or species which are found in only one or two sources are the best candidates for APCs. The acceptable range can be determined either from values in the literature or from source sampling. Both methods were used in this study.

Stoichiometric relations among species are used as APCs to sharpen the feasible ranges of the model predictions of source compositions. Water-soluble ionic species, such as NH_4, NO_3, SO_4, Na, and Cl, are present as a chemical compounds. For example, NH_4, NO_3, and SO_4 may exist as the compounds NH_4NO_3, $(NH_4)_2SO_4$, and NH_4HSO_4. These three compounds, less any primary sulfate, and including secondary organic species, constitute most of the mass of the secondary source. The stoichiometric constraints are developed next.

Let the fractions of each compound in the secondary source be assumed as follows:

OC : f_1,
NH_4NO_3 : f_2,
$(NH_4)_2SO_4$: f_3,
NH_4HSO_4 : f_4.

Then the source compositions of OC (a_{OC}), NH_4 (a_{NH4}), NO_3 (a_{NO3}), and SO_4 (a_{SO4}) in the secondary source can be expressed as the sum of the fraction of each compound f_i,

$$a_{OC} = 0.769\, f_1 \tag{17}$$

$$a_{NH4} = 0.225\, f_2 + 0.273\, f_3 + 0.157\, f_4 \tag{18}$$

$$a_{NO3} = 0.775\, f_2 \tag{19}$$

$$a_{SO4} = 0.727\, f_3 + 0.835\, f_4. \tag{20}$$

These expressions are solved for f_i and the following constraints applied:

$$f_i \geq 0\ , \tag{21}$$

$$0.9 \leq \sum f_i \leq 1.0\ , \qquad \text{for } i = 1, \ldots , 4. \tag{22}$$

where the f_i are expressed as linear combinations in the V-space. These physical constraints, derived from the stoichiometric relation among the species, turn out to be very strict in the sense of restricting the feasible region.

The stoichiometric constraint for the secondary source has been explained above. This concept of using stoichiometry among species as APCs can be extended to the marine and roadway sources. For the marine source, it is well known that NaCl is attacked by acids such as HNO_3 and H_2SO_4 to form $NaNO_3$, Na_2SO_4, and HCl (Hidy, *et al.* 1974; Hitchcock, 1980). As before, fractions of each compound of NaCl, $NaNO_3$, Na_2SO_4, and SO_4 are assumed as f_i, and APCs for f_i are set up. For the roadway source, fractions of each compound of NH_4NO_3, $(NH_4)_2SO_4$, and NH_4HSO_4 are used to set up APCs to estimate the amount of primary sulfate and nitrate that may be associated with roadway emissions.

3.3.2 Additional Physical Constraints for Roadway Source

The motor vehicles re-suspend road dust as they move. Thus, these two source categories are spatially correlated and will almost always affect the receptor site at the same time. Because vehicle exhaust emissions are so closely correlated with resuspended road dust both in time and space, these two sources appear to multivariate factor models as a single composite roadway source.

Setting up APCs for this composite roadway source is not a easy problem because the composite profile depends on how much of the motor vehicle and crustal sources are mixed together and this will differ from site to site. The composite roadway source profiles were obtained at three different sites (DLA, LEX, and RUB) by minimizing the difference between the linear combination of the target source compositions and those predicted. The predicted source compositions were obtained from the linear combination of the eigenvectors of the CC^T of the ambient data matrix, C. The complete methodology of obtaining the roadway composite profiles is discussed in Henry and Kim (1988). Si and Pb were selected for setting up APCs and their ranges were taken from results of the above study. The actual APCs used for each site are listed in Table 1.

Table 1. Additional Physical Constraints Used for Roadway Source for Each Site.

Stoichiometry among NH_4NO_3, $(NH_4)_2SO_4$, and NH_4HSO_4	
DLA	$.04 \le Si \le .07$ $.004 \le Pb \le .008$ $.0 \le O_3 \le .001$
LB	$.048 \le Si \le .088$ $.007 \le Pb \le .018$ $.0 \le O_3 \le .001$
BUR	$.04 \le Si \le .07$ $.004 \le Pb \le .008$
LEX	$.048 \le Si \le .088$ $.007 \le Pb \le .018$ $.0 \le O_3 \le .001$
ANA	$.06 \le Si \le .12$ $.002 \le Pb \le .005$ $.026 \le Al \le .049$
RUB	$.06 \le Si \le .12$ $.035 \le Pb \le .006$
UP	$.06 \le Si \le .12$ $.035 \le Pb \le .006$ $.0 \le O_3 \le .001$

The stoichiometric constraints have also been used as APCs. Although the exact fractions of compounds NH_4NO_3, $(NH_4)_2SO_4$, and NH_4HSO_4 in the roadway source are not known, it is obvious that their sum must be a small fraction. For all sites, the sum of these three compounds is assumed to be less than 10 percent, i.e.,

$$0.0 \le \Sigma f_i \le 0.1 \ . \qquad (23)$$

After APCs for each source category were set up, the model was applied to predict the source compositions of these source categories, and source apportionments were made from the CMB model using these compositions. The model's results for each site are summarized in the next section.

3.4 SAFER Los Angeles Results and Discussion

The estimated roadway source compositions for each site are shown in Table 2. The reasonableness of these profiles is examined below by checking some important elemental ratios. Key elemental ratios, especially for elements not used as APCs, of the predicted source compositions should be consistent from site to site. These elemental ratios are compared to the ratios expected from source compositions found in the literature.

3.4.1 Predicted roadway composite source composition profile

Key species including those that come mainly from the roadway source, have been selected to examine the elemental ratios of the predicted roadway source compositions. The elemental ratios of these species for each site are summarized in Table 3. These elemental ratios are compared with those found by other investigators and source compositions (or profiles) found by source sampling.

Three profiles have been taken from the SoCAB source composition library (Library No. 1-05, 1-39, and 1-43), Cooper *et al.* (1987), and one profile is taken from Cass and McRae (1983). Two profiles taken from SoCAB Library No. 1-05 and Cass are composite profiles of several different types of motor vehicles, tire tread and brake linings. SoCAB Library No. 1-39 and 1-43 are the profile composite of the paved road dust and the soil profile composite, respectively.

The predicted roadway profile is a composite profile of motor vehicles and resuspended road dust. Therefore, predicted roadway source compositions cannot be directly compared with composite profiles of motor vehicles or soil, separately. The ratios of some elements that come primarily from the motor vehicle or soil sources are compared with those of the model estimated roadway profiles.

The Si/Al ratio is similar at all the sites: it ranges from 2.05 at ANA to 2.82 at DLA. This ratio is close to the ratios of 2.72 and 2.10 for the SoCAB road dust and SoCAB soil profiles, respectively. The Si/Al ratio for average rock ranges from 3.41 to 4.02 and for clay material from 1.04 to 2.07 (Rahn, 1976).

Table 2. Predicted roadway source compositions and standard deviations at each site.

	DLA	BUR	LB	LEX
OC	.27987 (.08930)	.30504 (.08448)	.34090 (.03474)	.31869 (.05928)
EC	.15622 (.03290)	.09135 (.06294)	.12835 (.00821)	.13735 (.02081)
NH4	.01532 (.00992)	.01564 (.01007)	.01517 (.01005)	.00000 (.00000)
NO3	.00990 (.02440)	.01107 (.02550)	.01034 (.02536)	.00000 (.00000)
SO4	.04723 (.03527)	.04575 (.03549)	.04566 (.03776)	.00000 (.00000)
Na	.00732 (.01182)	.00710 (.01091)	.00474 (.00886)	.00485 (.00813)
Mg	.00227 (.00116)	.00202 (.00116)	.00182 (.00106)	.00246 (.00075)
Al	.01953 (.00490)	.02045 (.00449)	.02155 (.00424)	.02642 (.00644)
Si	.05498 (.01465)	.05390 (.01447)	.05374 (.01068)	.06518 (.01831)
P	.00093 (.00065)	.00168 (.00083)	.00095 (.00010)	.00135 (.00037)
Cl	.01347 (.01472)	.01095 (.01133)	.02463 (.01354)	.02661 (.01182)
K	.00786 (.00155)	.00862 (.00160)	.01014 (.00117)	.00937 (.00166)
Ca	.02122 (.00442)	.01758 (.00510)	.01987 (.00283)	.01834 (.00425)
Ti	.00352 (.00062)	.00328 (.00103)	.00286 (.00056)	.00334 (.00065)
V	.00034 (.00004)	.00034 (.00004)	.00031 (.00006)	.00038 (.00004)
Cr	.00034 (.00005)	.00043 (.00008)	.00031 (.00004)	.00039 (.00003)
Mn	.00124 (00014)	.00098 (.00018)	.00113 (.00010)	.00139 (.00014)
Fe	.03437 (.00476)	.02399 (.00630)	.02724 (.00414)	.03375 (.00607)
Ni	.00029 (.00005)	.00031 (.00008)	.00023 (.00005)	.00007 (.00002)
Cu	.00939 (.00319)	.00562 (.00629)	.00870 (.00291)	.00526 (.00069)
Zn	.01028 (.00246)	.00525 (.00340)	.01397 (.00144)	.00590 (.00050)
Br	.00101 (.00024)	.00098 (.00024)	.00141 (.00015)	.00195 (.00032)
Sr	.00038 (.00004)	.00024 (.00008)	.00033 (.00003)	.00039 (.00003)
Pb	.00613 (.00079)	.00552 (.00076)	.00741 (.00063)	.00812 (.00133)

The Pb/Br ratio is also quite similar at all the sites. It ranges from 4.16 at LEX to 6.07 at DLA. This ratio has been compared to the ratios of the SoCAB and

Table 2 (cont). Predicted roadway source compositions and standard deviations at each site.

	ANA	RUB	UP
OC	.23750 (.04501)	.26424 (.03736)	.29961 (.04085)
EC	.08866 (.01480)	.08716 (.01098)	.10938 (.01290)
NH4	.01555 (.00563)	.01184 (.01075)	.01563 (.00273)
NO3	.00546 (.01510)	.01871 (.03039)	.00292 (.00834)
SO4	.06417 (.02236)	.02045 (.02661)	.07434 (.01183)
Na	.00358 (.00742)	.00955 (.01034)	.01652 (.00454)
Mg	.00197 (.00079)	.00423 (.00153)	.00351 (.00044)
Al	.03983 (.00938)	.03615 (.00720)	.02652 (.00407)
Si	.08166 (.02469)	.09045 (.02932)	.06570 (.01076)
P	.00115 (.00025)	.00141 (.00114)	.00004 (.00008)
Cl	.00949 (.01245)	.00755 (.00692)	.01076 (.00386)
K	.01525 (.00317)	.01693 (.00437)	.01094 (.00125)
Ca	.02630 (.00401)	.04281 (.04464)	.02275 (.00321)
Ti	.00471 (.00100)	.00414 (.00071)	.00357 (.00039)
V	.00038 (.00006)	.00032 (.00006)	.00027 (.00002)
Cr	.00033 (.00004)	.00023 (.00004)	.00027 (.00002)
Mn	.00125 (.00015)	.00129 (.00018)	.00107 (.00009)
Fe	.04035 (.00869)	.03851 (.00748)	.02986 (.00397)
Ni	.00022 (.00004)	.00015 (.00003)	.00026 (.00002)
Cu	.00650 (.00464)	.00557 (.00220)	.00821 (.00173)
Zn	.00557 (.00337)	.00543 (.00186)	.00762 (.00117)
Br	.00089 (.00015)	.00081 (.00009)	.00074 (.00007)
Sr	.00037 (.00005)	.00034 (.00006)	.00026 (.00003)
Pb	.00450 (.00068)	.00366 (.00037)	.00379 (.00032)

Cass composite motor vehicle profiles and the SoCAB road dust profile. The ratios predicted by the SAFER model are always larger than the ratios of the two

Table 3. Elemental ratios of the predicted roadway source compositions at each site and the comparison with the other profiles.

	Si/Al	Pb/Br	Si/Ca	Si/Pb	Pb/OC (x100)	Si/OC	Si/TC	Pb/TC (x100)
DLA	2.82	6.07	2.59	8.97	2.19	0.196	0.126	1.41
BUR	2.64	5.63	3.07	9.76	1.81	0.177	0.136	1.39
LB	2.49	5.26	2.70	7.25	2.17	0.158	0.115	1.58
LEX	2.47	4.16	3.55	8.03	2.55	0.205	0.143	1.78
ANA	2.05	5.06	3.10	18.15	1.89	0.344	0.250	1.38
RUB	2.50	4.52	2.11	24.71	1.39	0.342	0.257	1.04
UP	2.48	5.12	2.89	17.34	1.26	0.219	0.161	0.93
SoCAB Road[1]	2.72	31.60	4.08	52.97	1.91	1.012	0.953	1.80
SoCAB Auto[2]		3.93			2.95			1.28
SoCAB Soil[3]	2.10		9.65					
Cass[4]		2.57						22.78

[1] SoCAB source composition library no. 1-39 (Cooper *et al.*, 1987)
[2] SoCAB source composition library no. 1-05 (Cooper *et al.*, 1987)
[3] SoCAB source composition library no. 1-43 (Cooper *et al.*, 1987)
[4] Cass and McRae (1983)

composite motor vehicle profiles given by Cass. This fact could be explained by loss of Br.

The roadway composite profiles show spatial variations according to how much of the motor vehicle source is mixed with the resuspended road dust. However, the roadway composite source compositions do fall into certain ranges. Some of them are shown below.

$.238 \leq \text{OC} \leq .341$

$.087 \leq \text{EC} \leq .156$

$.020 \leq \text{Al} \leq .040$

$.054 \leq \text{Si} \leq .090$

$.018 \leq \text{Ca} \leq .043$

$.024 \leq \text{Fe} \leq .040$

$.0007 \leq \text{Br} \leq .0020$

$.0037 \leq \text{Pb} \leq .0081$

Simulation studies have been conducted to test the model's performance and to determine the random error effects both in the source compositions and the ambient concentrations on the predicted source composition profiles. It has been shown that the SAFER model can estimate source compositions with acceptable error and bias. The maximum percentage uncertainties in the predicted compositions of the roadway are less than 10 percent for most of the major species, except EC which is 20 percent (Kim, 1989).

3.4.2 Los Angeles Source Apportionment

A summary of the source apportionment estimated by the CMB model for all the sites is shown in Table 4. The source contribution from the roadway source ranges from 20 $\mu g/m^3$ to 34 $\mu g/m^3$. The minimum contribution of 19.597 $\mu g/m^3$ is found at LEX and the maximum contribution of 34.202 $\mu g/m^3$ at RUB. However, except for these two extreme values, the roadway contribution is about the same for all the sites. The large roadway contribution at the RUB site might be related to the large fraction of the crustal source in the roadway composite profile at this inland, arid site. Similarly, the small roadway contribution at the LEX site could be explained by a large motor vehicle fraction in the composite profile.

Table 4. Summary of Source Apportionment Results.

	DLA	LB	LEX	ANA	RUB	UP
Roadway	26.987	25.106	19.597	28.911	34.202	25.433
Secondary	20.825	17.750	22.540	19.416	31.396	20.351
Marine	1.671	1.513	2.498	2.230	---	---
Crustal	6.001	4.060	---	---	8.127	6.575
Sum	55.483	48.428	44.635	50.557	73.725	52.359
Unexplained	4.774	1.404	0.493	2.109	14.058	5.604
Total Mass	60.257	49.892	45.128	52.666	87.783	57.963

4. CONCLUSIONS

It can be seen from the review of multivariate techniques and the example presented above that a considerable amount of insight into the sources of airborne particles can be obtained directly from the ambient concentration data. These methods provide the ability to screen the data to bigin to determine the possible structure and then to perform detailed source apportionment without initial input of specific source profiles. These techniques are thus a useful complement to the Chemical Mass Balance model.

5. REFERENCES

Alpert, D.J. and P.K. Hopke, A Quantitative Determination of Sources in the Boston Aerosol, *Atmospheric Environ.* 14:1137-1146 (1980).

Alpert, D.J. and P.K. Hopke, A Determination of the Sources of Airborne Particles Collected during the Regional Air Pollution Study, *Atmospheric Environ.* 15:675-687 (1981).

Blifford, I.H. and G.O. Meeker, A Factor Analysis Model of Large Scale Pollution, *Atmospheric Environ.* 1:147-157 (1967).

Cass, G.R. and G.J. McRae, Source-Receptor Reconcillation of Routine Air Monitoring Data for Trace Metals: An Emission Inventory Assisted Approach, *Environ. Sci. Technol.* 17:129-139 (1983).

Cooper, J.A., D.C. Redline, J.R. Sherman, L.M. Valdovinos, W.L. Pollard, L.C. Scavone, and C. Badgett-West, *PM_{10} Source Composition Library for the South Coast Air Basin. Volume I: Source Profile Development Documentation Final Report*, South Coast Air Quality Managememt District, El Monte, CA (1987).

Dillon, W.R. and M. Goldstein, *Multivariate Analysis: Methods and Applications*, John Wiley & Sons, New York (1984).

Garrenstrom, P.D., S.P. Perone, and J.L. Moyers, Application of Pattern Recognition and Factor Analysis for Charaterization of Atmospheric Particulate Composition in Southwest Desert Atmosphere, *Environ. Sci. Technol.* 11:795-800 (1977).

Gatz, D.F., Identification of Aerosol Sources in the St. Louis Area using Factor Analysis, *J. Appl. Met.* 17:600-608 (1977).

Harman, H.H., *Modern Factor Analysis*, University of Chicago Press, Chicago (1976).

Heidam, N.Z., The Components of the Arctic Aerosol, *Atmospheric Environ.* 18:329-343 (1984).

Henry, R.C., *A Factor Model of Urban Aerosol Pollution: A New Method of Source Identification*, Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon (1978).

Henry, R.C. and G.M. Hidy, Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Components-Part I. Annual Data from Los Angeles and New York, *Atmospheric Environ.* 13:1581-1596 (1979).

Henry, R.C. and G.M. Hidy, Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Components-II. Salt Lake City, Utah and St. Louis, Missouri, *Atmospheric Environ.* 16:929-943 (1982).

Henry R.C., C.W. Lewis, P.K. Hopke, and H.J. Williamson, Review of Receptor Model Fundamentals, *Atmospheric Environ.* 18:1507-1515 (1984).

Henry, R.C., Current Factor Analysis Models are Ill-Posed, *Atmospheric Environ.* 21:1815-1820 (1987).

Henry, R.C. and B.M. Kim, *Factor Analysis Screening of Source Profiles for Analysis of Los Angeles PM10 Data*, Draft, Appendix V-I, Air Quality Management Plan 1988 revision, South Coast Air Quality Management District, Los Angeles (1988).

Henry, R.C. and B.M. Kim, Extension of Self-Modeling Curve Resolution to Mixtures of More than Three Components. Part I. Finding the Basic Feasible Region, *Chemometrics and Intelligent Laboratory Systems* 8:205-216 (1990).

Hidy, G.M., P.K. Mueller, H.H. Wang, J. Karney, S. Twiss, M. Imada, and A. Alcocer, Observations of Aerosols over Southern California Coastal Waters, *J. Appl. Met.* 13:96-107 (1974).

Hitchcock, D.R., Sulfuric Acid Aerosols and HCl Release in Coastal Atmospheres: Evidence of Rapid Formation of Sulfuric Acid Particulates, *Atmospheric Environ.* 14:165-182 (1980).

Hopke P.K., E.S. Gladney, G.E. Gorden, W.H. Zoller, and A.G. Jones, The Use of Multivariate Analysis to Identify Sources of Selected Elements in Boston Urban Aerosol, *Atmospheric Environ.* 10:1015-1025 (1976).

Hopke P.K., R.E. Lamb, and D.F.S. Natusch, Multielemental Characterization of Urban Roadway Dust, *Environ. Sci. Technol.* 14:164-172 (1980).

Hopke P.K., D.J. Alpert, and B.A. Roscoe, Fantasia-A Program for Target Transformation Factor Analysis to Apportion Sources in Environmental Samples, *Computers & Chemistry* 7:149-155 (1983).

Hopke P.K., *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons, New York (1985).

Hopke, P.K., Target Transformation Factor Analysis as an Aerosol Mass Apportionment Method: A Review and Sensitivity Study, *Atmospheric Environ.* 22:1777-1792 (1988).

Ito K., T.J. Kniep, and P.J. Lioy, The Effects of Sample Size and Random Error on the Factor Analysis/Multiple Regression (FA/MR) Receptor Modeling Technique, *Atmospheric Environ.* 22:1433-1440 (1986).

Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New York (1986).

Keiding, K., F.P. Jensen, and N.Z. Heidam, Absolute Modeling of Urban Aerosol Elemental Composition by Factor Analysis, *Anal. Chim. Acta* 181:79-85 (1986).

Keiding, K. and M.S. Sorensen, A Receptor Model for Urban Aerosols Based on Oblique Factor Analysis, *Anal. Chim. Acta* 193:295-307 (1987).

Kim, B.M., *Development of a New Multivariate Receptor Model and its Application to Los Angeles Airborne Particle Data*, Ph.D. Thesis, University of Southern California, Los Angeles, CA (1989).

Kleinman M.T., M. Eisenbud, M. Lippmann, and T.J. Kneip, The Use of Tracers to Identify Sources of Airborne Particles, *Environment Inter.* 4:53-62 (1980a).

Kleinman, M.T., B.S. Pasternack, M. Eisenbud, and T.J. Kneip, Identifying and Estimating the Relative Importance of Sources of Airborne Particulates, *Environ. Sci. Technol.* 14:62-65 (1980b).

Kneip T.J., R.P. Mallon, and M.T. Kleinman, The Impact of Changing Air Quality on Multiple Regression Models for Coarse and Fine Particle Fractions, *Atmospheric Environ.* 17:299-304 (1983).

Lawson C. L. and R.J. Hanson, *Solving Least Squares Problems*, Prentice Hall, Englewood Cliffs, NJ (1974).

Lawton W.H. and E.A. Sylvestre, Self Modeling Curve Resolution, *Technometrics* 13, 617-633 (1971).

Lowenthal D.H. and K.A. Rahn, Application of the Factor-Analysis Receptor Model to Simulated Urban- and Regional-Scale Data Sets, *Atmospheric Environ.* 21:2005-2013 (1987).

Malinowski E. R. and D.G. Howery, *Factor Analysis in Chemistry*, John Wiley & Sons, New York (1980).

Morandi M.T., J.M. Daisey, and P.J. Lioy, Development of a Modified Factor Analysis/Multiple Regression Model to Apportion Suspended Particulate Matter in a Complex Urban Airshed, *Atmospheric Environ.* 21:1821-1831 (1987).

Rahn K.A., Silicon and Aluminum in Atmospheric Aerosols: Crust-Air Fractionation?, *Atmospheric Environ.* 10:597-601 (1976).

Solomon P.A., T. Fall, L. Salmon, G.R. Cass, H.A. Gray, and A. Davidson, Chemical Characteristics of PM10 Aerosols Collected in the Los Angeles Area, *J. Air Poll. Control Assoc.* 39:154-163 (1989).

Thurston G.D. and J.D. Spengler, A Quantitative Assessment of Source Contributions to Inhalable Particulate Matter Pollution in Metropolitan Boston, *Atmospheric Environ.* 19:9-25 (1985).

Weiner P.H., E.R. Malinowski, and A.R. Levinstone, Factor Analysis of Solvent Shifts in Proton Magnetic Resonance, *J. Phys. Chem.* 74:4537-4542 (1970).

Chapter 6

SCANNING ELECTRON MICROSCOPY

by

Philip K. Hopke
Department of Chemistry
Clarkson University
Potsdam, NY 13676

and

Gary S. Casuccio
R.J. Lee Group
350 Hochberg Road
Monroeville, PA 15146

1. INTRODUCTION

Chemical receptor methods as described in this volume have employed various statistical techniques to quantitatively apportion the ambient aerosol mass based on bulk sample analyses. However, there are has limitations on apportioning pollutant sources because of limited number of measurable features in source profiles. Because of the combination of collinearity and measurement imprecision, it is generally not possible to resolve more that 6 to 7 sources for any given sample. Alternatively, microscopic methods have been used for specific source identification studies based on particle-by-particle analyses (Crutcher, 1982). However, these methods are laborious and expensive when full quantitative analysis is required. In this chapter, an alternative approach will be present. Computer controlled scanning electron microscopy (CCSEM) coupled with advanced pattern recognition methods, permits improved quantitative apportionment with high source specificity.

Ever since the first scanning electron microscope (SEM) became available commercially in 1965, the SEM has gained increasing acceptance as a powerful research tool. With the development of the energy dispersive detector for x-ray analysis, the SEM has moved from an instrument which could only produce images of small portions of matter (micrographs) to an extremely versatile analytical instrument. The ability of the SEM/XRF system to perform elemental analysis of extremely small volumes of materials is very useful for the characterization of individual particles and the identification and quantitative mass apportionment of those particles to sources. Individual particles can be analyzed for their elemental constituents as well as mophological information that aide in determining the processes by which the particle was formed.

With improvements in computers, the scanning and image analysis can be done in close to real time under computer control so that size, shape, and elemental composition can be obtained for a large number of particles in a reasonable length of time. Recent advances now permit the automatic capture of a high quality image of each particle as well as the results of the image and x-ray analysis. Thus, computer-controlled scanning electron microscopy (CCSEM) represents a powerful tool for the identification and classification of particles. The development of CCSEM has overcome the limitations on the number of particles that can be analyses in a practical manner. Coupled with intensive data analysis methods, computer-controlled scanning electron microscopy has a very substantial capability for resolving a larger number of sources and providing more accurate source apportionments.

2. SCANNING ELECTRON MICROSCOPE

2.1 General Configuration

CCSEM combines several analytical tools including the scanning electron microscope, the energy dispersive x-ray analyzer, and the digital scan generator for image processing. It can quantitatively provide size as well as chemical compositional information in a short analysis time. Free electrons are obtained by thermionic emission from the heated tungsten filament in an evacuated chamber. The electrons are accelerated through a high voltage. The resulting beam of energetic electrons is focussed onto the sample surface by electromagnetic lenses. The lenses are used to demagnify the electron source (~4-6 mils in diameter) to a diameter of about 10.0 nm. The electron beam can be scanned over the specimen in a rectangular raster.

Several forms of emission result from the interaction of these energetic electrons with the particle surface. One type of emission resulting from the complex beam-matter interaction is secondary electrons. These are low energy electrons ejected from the sample by the primary electron beam and are collected by a scintillation counter. The output of the scintillation counter is connected to a cathode ray display. The deflection of the electron beam over the specimen and the deflection in the cathode ray display are synchronized by connecting the two sets of scanning coils to the same X-Y generator. In this way an image of the specimen appears on the display. Image contrast arises from the differences in the efficiency of producing secondary electrons from different regions of the specimen. This efficiency is dependent on the topography of the sample so that the morphology of the specimen is displayed.

Alternatively, electrons from the microscope beam that are directly scattered by the particle can be used for particle imaging. The energy of these backscattered electrons is sensitive to the average atomic number in the particle. This variation in energy helps to distinguish particles from the background or to distinguish among different types of particles. The backscattered electron intensity is also affected by the particle morphology, but does not provide the same high level of detail in the surface as secondary electrons.

The major development that has made the use of SEM practical for quantitative receptor modeling is the coupling of computers to the SEM (Kelly *et al.*, 1980; Lee and Kelly, 1980). The computer is used both to control the electron beam and automatically process the image produced by the sample. The computer directs a digital beam-control for a real-time point by point analysis with up to 4096 by 4096 points.

2.2 Imaging System

In the automated mode, the beam-control system moves the electron beam in a step-wise fashion across the sample. The spacing between the points examined determines the *minimum* particle size that will be observed with certainty. For the system typically used, this minimum particle size is around 0.2 μm. The intensity of the signal, backscattered or secondary electron, is compared to a threshold that is established for the particular sample backing material. If the signal level is below the threshold, it assumes no particle is present at that location. If the signal is above the threshold, then a secondary beam control system drives the beam in a preset pattern to determine the size and shape of the entity or feature that is causing the increased reflectance.

The preset pattern consists of pairs of diagonals across the particle where the lengths are determined by the points where the signal drops below the threshold values. The pattern is repeated twice. The first pattern is used to locate the feature and the second provides the lengths of the diagonals through the centroid. It is also possible to move the beam around the end of the particle storing the locations of its perimeter. The maximum, minimum and average diameter are calculated and stored as well as the centroid of each particle. The measurement of the particle dimensions is made more accurate by using a closer spaced point array once the particle is located. The stored centroid locations are compared to the currently examined particle to insure that particles are not counted twice.

Once the parameters have been stored, the beam is returned to the centroid of the particle and the x-ray spectrum is recorded in a multichannel analyzer. In other approaches, the beam is rastered across the particle within the previously measured perimeter. A density can be estimated on the basis of the chemical composition and the mass calculated assuming that the particle is a spheroid of revolution. The process takes as little as 1.5 seconds per particle depending on the time needed to obtain a statistically acceptable x-ray spectrum. The system then returns to the search mode to find additional particles. Thus, it is possible to measure the characteristics of up to a thousand particles an hour depending on the length of time chosen to accumulate the fluoresced x-ray spectrum.

The image analysis system provides a reasonable estimate of the particle size. However, only a rough idea of the shape of the particle is obtained. The instrument does provide an image of the particle that can be captured photographically, but it has not been available for further computer analytical study. Thus, we are currently not taking full advantage of the information available from the instrument. Major improvements in automated particle imaging by the RJ Lee Group now make it possible to store the particle image directly. Automatic capture of single particle images (256 gray level, 256 pixel by 256 pixel) now permits the accumulation of a large number of particle images for more detailed off-line analyses of shape and particle texture. This large amount of data can be stored on a write-once, read many (WORM) optical disk storage system with a disk capacity of 800 Mbytes of data. These images may provide clearer indications of the particle's origin and what processing it has undergone in transit in the atmosphere.

To illustrate the qualitative information that is available from the imaging process, four single particles are displayed in Figures 1 to 4. These figures show the secondary electron image and the x-ray spectrum taken with the electron beam focussed on the particle center. Figure 1 is a quartz particle whose x-ray spectrum shows only Si and O. The diatom shown in Figure 2 has an identical x-ray spectrum, but displays a very different shape and texture. A similar distinction can be seen between the fly ash particle in Figure 3 and the clay mineral particle in Figure 4 where the spherical shape of the fly ash particle clearly suggests its origins. It can be seen from these figures that both shape and composition can be useful in identifying particle types.

2.3 X-Ray Fluorescence System

The incident electron beam can also cause emission of x-rays which are generated from a sample volume of about 1-3 μm^3. Because the cross-sections for excitation of atomic electrons by electrons are much smaller than those for heavy charged particles or x-rays as described elsewhere in this volume by Dzubay and Stevens, the sensitivity of the x-ray fluorescence analysis within the microscope is much lower than for photon-induced XRF or PIXE. The fluoresced x-rays are collected and analyzed by an x-ray spectrometer system. The x-ray detector is a lithium drifted, silicon diode is similar to those used in other x-ray

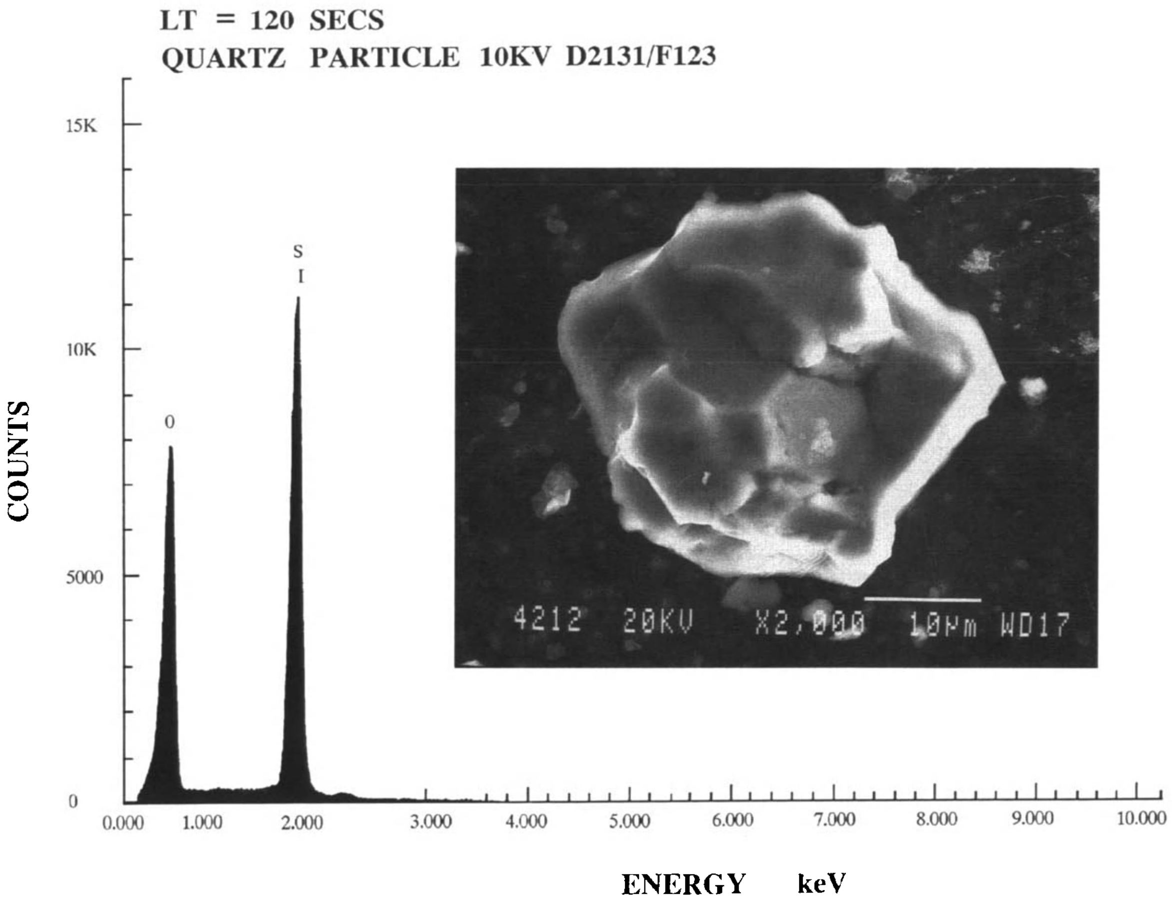

Figure 1. X-Ray Fluorescence Spectrum and Secondary Electron Image of a Quartz Particle.

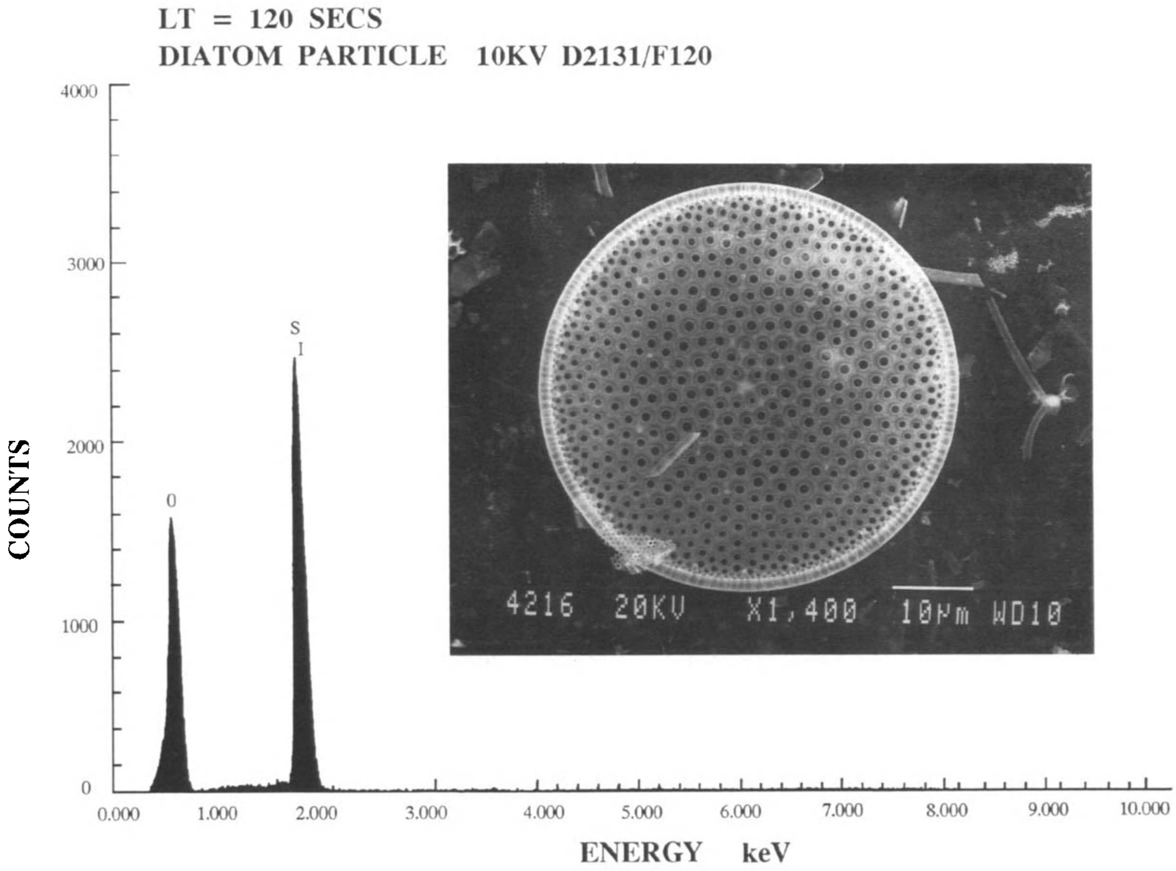

Figure 2. X-Ray Fluorescence Spectrum and Secondary Electron Image of a Diatom Particle.

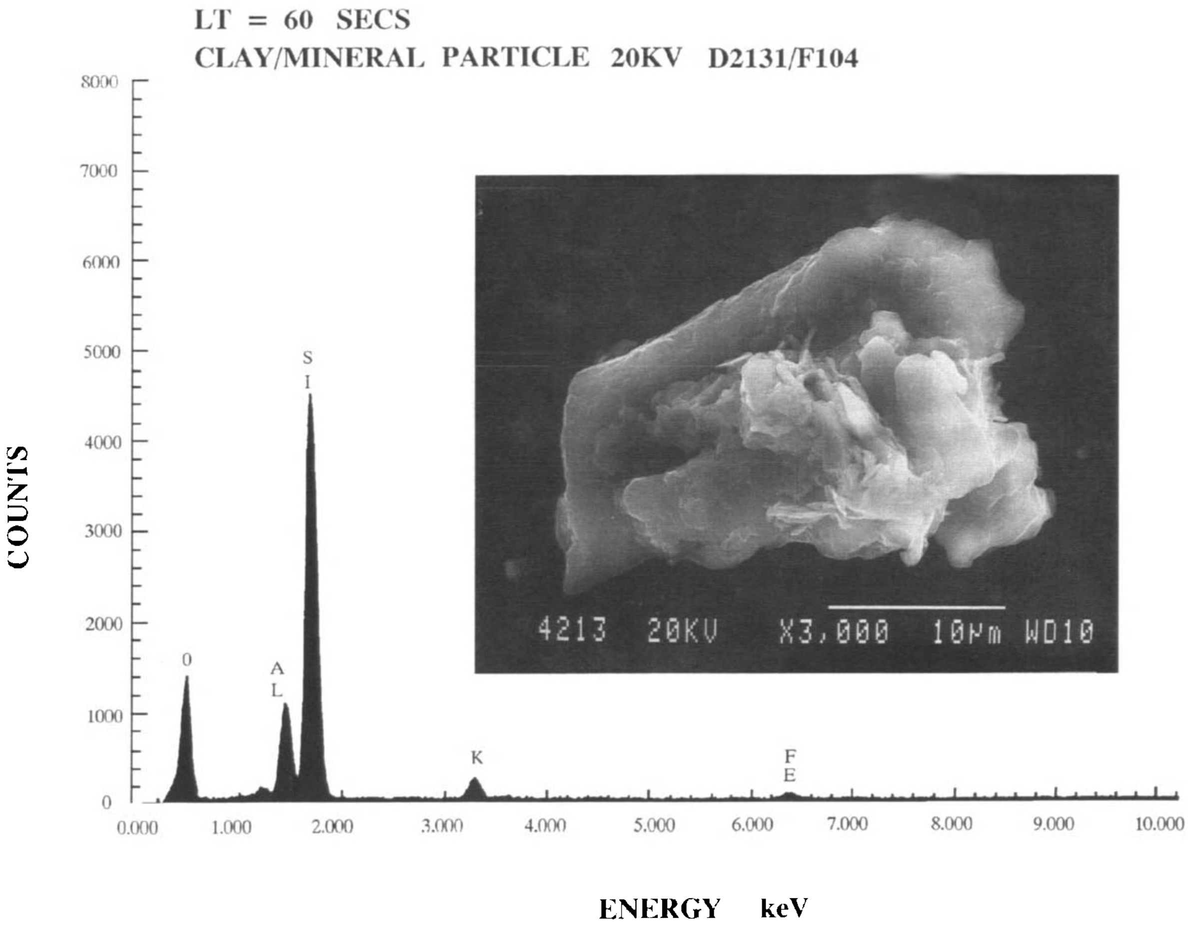

Figure 3.. X-Ray Fluorescence Spectrum and Secondary Electron Image of a Clay Mineral Particle.

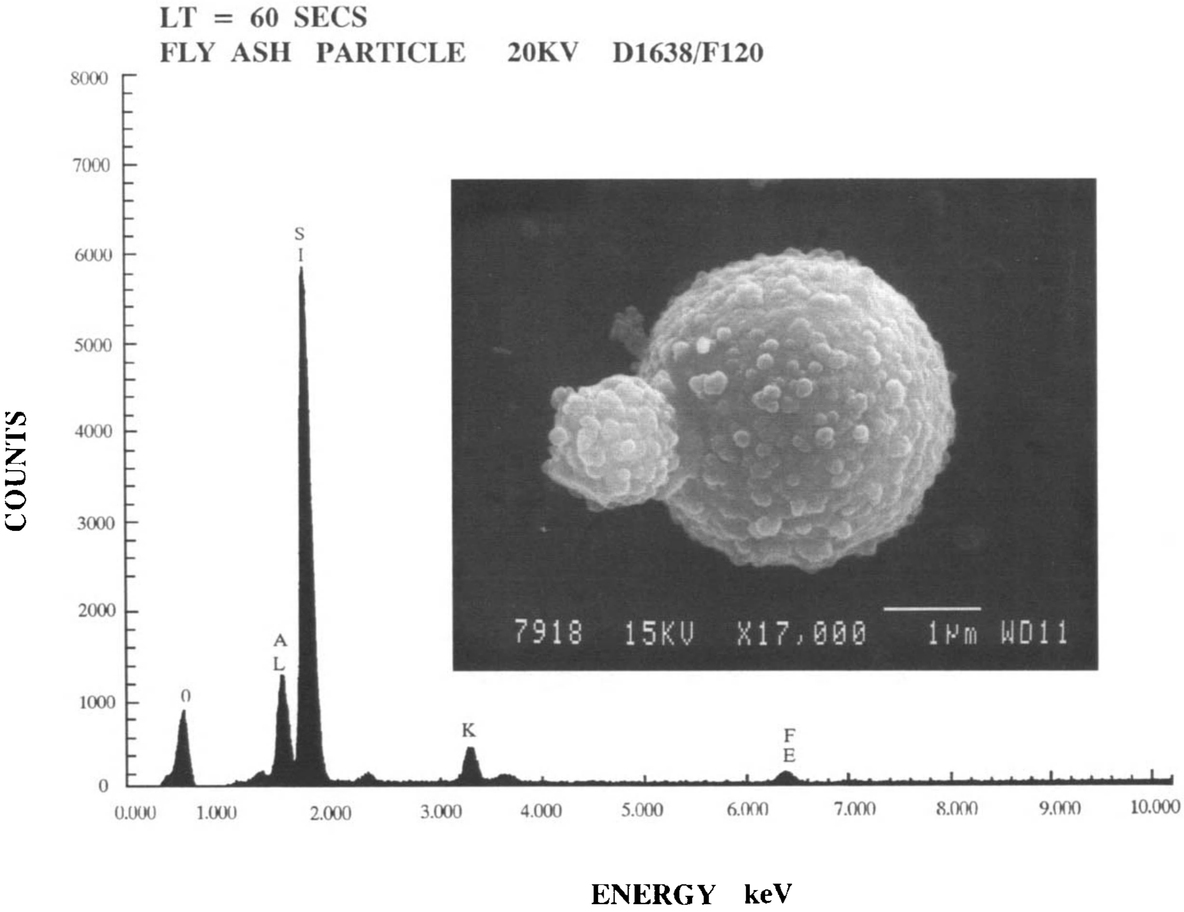

Figure 4.. X-Ray Fluorescence Spectrum and Secondary Electron Image of a Fly Ash Particle.

fluorescence methods. However, because a vacuum is maintained on both sides of the entrance window of the detector, it is possible to make the window sufficiently thin that even low energy x-rays from the lightest elements can penetrate to the detector. These detectors are called "windowless" even though that is not fully true.

It is generally difficult to obtain a precise quantitative analysis of the elemental composition of the particle. However, it is possible to obtain relative concentrations of elements present at $\geq$ 0.1 atomic percent in the particle. Janossy *et al.* (1979) describe a method for obtaining accurate atomic weight ratios for ultrathin samples. The elements that can typically be observed include Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Cu, Ni, Zn, Pb, and Br. Additional elements could be sought but are generally below detection limits except in unusual samples. It is possible to observe the lighter elements because of the fact that the sample is in a vacuum. In many of the earlier applications of CCSEM, a particle that is observed but for which there are not observable x-rays was generally assigned to be a carbonaceous particle. Carbonaceous particles can be imaged and information can be obtained on the possible particle sources from the particle morphologies (Griffin and Goldberg, 1979). The improved "windowless" detectors have become available that permit the detection of x rays down to carbon. Thus, now in addition to the elements listed above, C, N, and O can be observed directly (Kim *et al.*, 1989). However, the analysis time per particle needs to be 15 to 20 seconds in order to obtain useful spectra.

There have been several recent improvements in the assignment of observed x-ray spectral lines to elements and quantitation of elemental concentrations in individual particles. Janssens *et al.* (1988 a & b) have developed an expert system for the automated interpretation of large numbers of single particle x-ray spectra. The x-ray data are often not extensively processed on-line, but are stored for subsequent off-line analysis. Each particle is characterized by a series of peaks in the spectrum for which an energy and a net peak intensity are stored. The expert system uses a data base of x-ray energies and measured relative intensities to determine which elements are present, identify spectral interferences, and determine the net intensity assignable to each specific x-ray line.

The x-ray intensities can then be converted into elemental concentrations. Methods for the quantitative analysis of particles have been recently reviewed by Markowicz *et al.* (1986). However, many of the methods are not applicable to the analysis of single particles. Van Borm (1989) has adapted a standardless ZAF correction method initially developed by Raeymaekers (1986) based on the work of Wernisch (1985, 1986). The program, QUANTA, incorporates theoretical expressions for the efficiency of an Si(Li) detector, an atomic number correction (Z), an absorption correction (A) based on the quadrilateral model of Sewell *et al.* (1985), and the fluorescence correction (F) for the increased fluorescence caused by secondary fluorescence excitation. With this approach, quantitative estimation of elemental concentrations can be calculated for each particle.

Van Borm (1989) has tested this approach with bulk alloys, bulk mineral particles, and microscopic mineral particles. The analyses of the bulk alloys provided results within 5% of the known values. For the bulk minerals, there was a consistent underestimation of low-Z elements such as Na, Mg, and Al, but within the range of 1 to 100 weight % the standard errors were within 10% for all elements. For the microscopic mineral particles a similar result was obtained in that there were significant underestimations for the low-Z elements and more variability because the particles are not spherical. Relative standard errors of up to 40% were found for low-Z elements present in low concentration. Elements present at greater than 10% weight percent were analyzed with much lower errors including oxygen. For $Z > 14$, accuracies of better than 10% can be obtained.

It is possible to utilize the concentration measurements made on individual particles to estimate the average bulk concentration values for the collection of particles. Casuccio *et al.* (1983a) and Energy Technology Consultants (1983) have described a detailed intercomparison study that compares the results of scanning electron microscopy with bulk chemical analysis for several high volume sampler filters taken in El Paso, TX. The chemical methods include x-ray fluorescence, both photon- and proton-induced, atomic absorption spectrometry and ion chromatography. The results generally show good agreement between methods. The lead values were in reasonable agreement although the CCSEM value for some cases is low possibly because of the inability to remove deeply penetrating, fine particles. In general these results from Casuccio *et al.* (1983a) do show that a reasonable representation of the bulk concentration values can be obtained from the microscopic results. However, the value of the SEM analysis is the characterization of particles on an individual basis and the ability to also obtain a reasonable estimate of the bulk average concentration is an additional bonus.

3. SAMPLE PREPARATION

The nature of the sample preparation method depends on the type of sample to be analyzed. Ideally the particles should be uniformly distributed as a monolayer with particles separated from one another on a relatively smooth background with a uniform atomic number lower than that of any particle to be identified. The primary emphasis of SEM receptor studies has been on airborne particulate matter collected on filter media. If the particles have been collected on a fibrous material such as glass fiber, quartz fiber, or paper filters, the particles are distributed within the volume of the filter and cannot be directly analyzed.

Under these conditions, it is necessary to remove the particles from the filter and redeposit them on an appropriate material. There are serious problems associated with any method to transfer particles. It is difficult to insure representative removal of particles since there will be differential penetration of particles into the fibrous filter pad depending on particle size. Thus, more of the large particles may be removed from the surface layer than smaller particles from the deeper regions of the filter. If the primary interest is in particle mass, this loss of some small particles may not be a problem. However, if respirable particles or specific toxic element apportionment is the objective of the study, particle transfer may lead to a biased result.

Another problem with redeposition is that the solvent used to remove the particles from the filter may dissolve some particles or leach materials from others. The results of a study of particles in El Paso, TX (Janocko *et al.*, 1982) indicated that the CCSEM results of the average aerosol composition were similar to those of other methods indicating that it may be possible to transfer a representative sample of TSP samples. A punch of material from the glass fiber filter was sonicated such that acetone was forced through the filter. The acetone flush may be performed in each direction using clean solvent for each. The process was repeated six times and the particulate matter was then redeposited on a 0.2 μm pore Nuclepore filter. For a paper filter, a less rigorous procedure was used. In this case, 50 ml and a punch were ultrasonically agitated for 5 minutes. The piece of filter was removed from the solvent and washed with a stream of filtered acetone. The material was again redeposited on a 0.2 μm pore Nuclepore filter.

A reasonable level of precision is obtained between replicate depositions, and it appears that a reproducible sample can be removed from the original filter. SEM results for samples taken in El Paso, TX were compared with optical microscopy with generally reasonable agreement (Energy Technology Consultants, 1983). In both of these analyses, particles were removed from the original filter medium and it appears that sufficiently representative samples can be obtained for such analyses.

Collection on membrane type filters can directly provide an acceptable sample if the particle loading is not too high. In this case, it is fairly simple to obtain a sample. Often optical microscopic examination is made to find locations on the filter for taking SEM samples. Small sections are then removed from the filter and mounted on an aluminum stub attached with a dispersion of amorphous graphite in butanol. If an impactor is being used, it may be possible to use a metal disk or a SEM stub as the collection surface. In this case, beryllium can be used to provide a very low atomic number background that permits a better identification of carbonaceous particles. Alternatively particles removed from a filter could also be deposited on an SEM stub by drying a droplet of solvent in which the particles are suspended directly on the stub surface. However, this approach can lead to artifacts as described above. In order to be able to easily distinguish carbonaceous particles from the background scattering from a carbon-based filter medium, it can be useful to redeposit a sample onto a beryllium stub.

By examining both the original filter and the particles on the Be stub, it is possible to ascertain if the redeposited material reflects the same types and distribution of particles in the original samples.

In order to prevent the accumulation of charge on the insulating particles that would distort the image, the sample needs to be coated with a conducting material. For the best imaging, gold has been used as the conducting material. However, gold makes the x-ray fluorescence analysis impossible. The usual method is then to evaporatively deposit a thin coating of carbon on the sample. This coating process does subject the sample to a moderate vacuum potentially resulting in evaporative loss of particles (Leong *et al.*, 1983). This evaporative loss will particularly affect nitrogen compounds (HNO_3, NH_4NO_3) because of their high vapor pressures. The sample is then available for analysis.

4. DATA ANALYSIS

4.1 Size Distributions

One of the advantages of scanning electron microscopy is the direct determination of the physical sizes of the particles in the sample. Thus, more detailed information can be obtained on particle sizes than would be available from the aerodynamic sorting provided by a sampling device. Even a cascade impactor only separates particles into a few categories. As discussed by Dzubay and Stevens, there are often problems of particle bounce in cascade impactors or fine particles in the coarse particle mode sample in dichotomous samplers. Thus, there are uncertainties in the actual size distribution measured by indirect means. Automated SEM makes the direct measurement of the size distribution feasible because of the ability to examine a sufficiently large number of particles.

It is possible to estimate the aerodynamic size distribution from the physical size and estimated density (Casuccio *et al.*, 1983b). It is not possible to precisely convert a measured physical diameter to an equivalent aerodynamic value for non-spherical particles. The approximate aerodynamic diameter is given by

$$d_a = \chi d_p \rho^{1/2} \qquad (1)$$

where d_a is the aerodynamic diameter, d_p is the measured physical diameter, χ is a shape factor, and ρ is the density as estimated from the chemical composition. The shape factor is included to correct for the non-spherical shape of the particles. Values of χ are obtained from Dallavalle (1948). An aerodynamic size particle number distribution of an ambient aerosol sample taken near a major highway outside of San Antonio, TX is shown in Figure 5. The number distribution shows two size modes with peaks at 0.8 µm and at 4 µm. A mass weighted distribution would shift these peaks to larger aerodynamic diameters. SEM is the only way that the complete physical size distribution of the actual sample can be determined.

5. PARTICLE CLASSIFICATION

5.1 Empirical Classification

The CCSEM analysis process described above provides a substantial amount of information characterizing each individual particle. The question then arises as to how to utilize these data to provide an understanding of the system under study. There are several approaches that have been applied to utilize these data in receptor models. The general approach that has been employed in the interpretation of the microscopic analytical data has been to assign each particle to one of a number of empirically defined particle categories (Johnson *et al.*, 1981; Casuccio *et al.*, 1982; Lee and Fisher, 1980) with an assigned density. For example, a particle that is rich in iron must be placed in one of the particle type categories containing significant iron concentrations. These types include iron-rich, spherical iron, chlorite, pyrite, calcium ferrites, etc. If there were no other major elements observed besides the iron, the particle would be assigned as iron-rich unless the aspect ratio (length to width) is less than 1.33:1. For such a situation, the particle should be classified as spherical iron. For a particle with approximately twice the sulfur as iron, the classification would be pyrite. The criteria for these category assignments have been developed and verified by analyzing a large number of reference materials with known characteristics (Lee and Fisher, 1980).

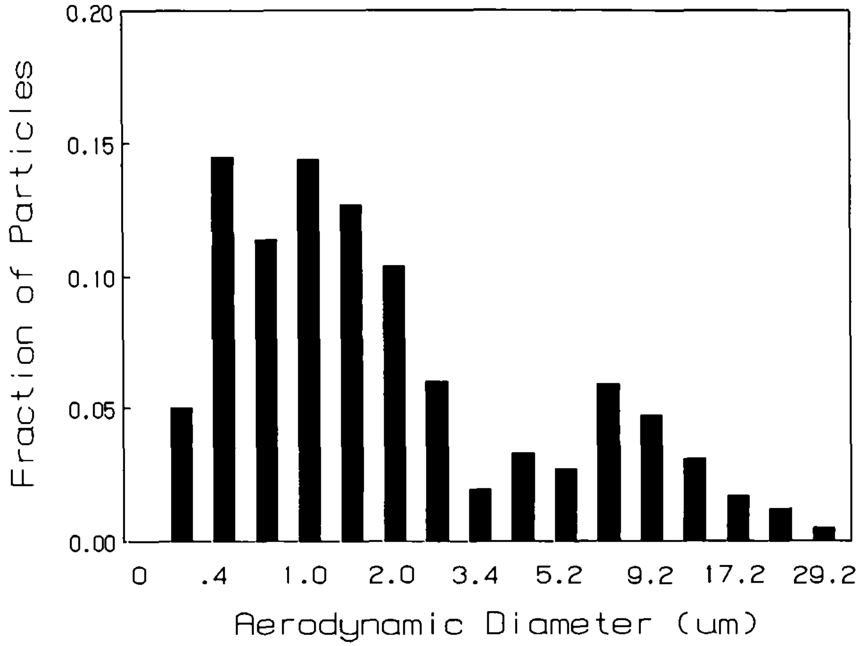

Figure 5. Frequency Distribution of Aerodynamic Diameter for Particles in Mn Smelter Sample T220.

The particles that do not fit into any of the predefined categories are assigned to an "unknown" class. These particles can be examined to determine if there are enough similar characteristic particles to permit the definition of a new particle class. It is possible to determine the number of particles of a given size range within a given particle characteristic class. These number of particles in each of the particle size/chemistry categories become the variables that are used in the further analysis.

Kim *et al.* (1987) tested the quality of the empirical classification scheme using supervised pattern recognition methods. These methods permit quantitative assessment of the classification efficiently. They found that there were a number of empirical categories had substantial overlapping regions where the assignment of a particle to a specific particle class was ambiguous. Thus, it appears that an objective classification scheme could provide better assignment of particles to classes and thus provide the basis for a subsequent mass apportionment.

There is a need to sort the particles into groups of particles that have similar chemical and physical properties. The fractions of mass assigned to each particle class will become the analogous values to the elemental concentrations measured in the bulk sample. Since many particle classes can be identified, a mass balance analysis similar to a chemical mass balance can be performed. This particle class balance (PCB) then is the basis for source apportionment for SEM analysis. The critical step in this analysis is the identification of the types of particles present in source or ambient samples (particle classes) and the accurate assignment of particles to the identified classes. Thus, this section will be devoted to the problem of particle classification.

5.2 Example Data Base

In order to illustrate particle classification and to test alternative methodologies, illustrative examples will be taken in a data sets obtained from a study of El Paso, Tx. The City of El Paso is located at the tip of western Texas, with the state of New Mexico to the west. The Rio Grande River, the boundary between the United States and Mexico, flows south of the city as shown in Figure 6. The city of Ciudad Juarez lies just to the south in Mexico across the river. El Paso also has the Franklin Mountains intruding from the north. Wiersema *et al.* (1984) indicated that based on XRF data, aluminum, arsenic, lead, and manganese levels are significantly higher in El Paso than the statewide average for these elements. Particularly, five-year average lead concentrations in the years from 1978 to 1982 in El Paso were three to four times the five-year average lead levels measured statewide. Therefore, the Texas Air Control Board initiated the El Paso Quantitative Microscopic Study in 1982 to identify the sources of TSP and lead in the air.

Detailed descriptions of the analytical methods and the data in El Paso has been previously reported (Janocko *et al.*, 1982; Casuccio *et al.*, 1983a&b; Dattner *et al.*, 1983; and ETC, 1983). CCSEM was selected as the primary analytical

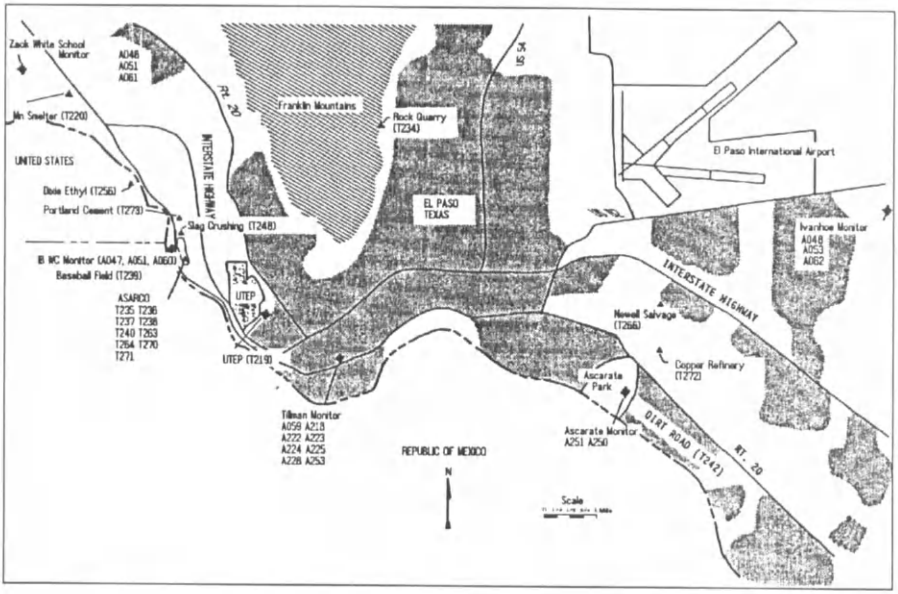

Figure 6. Map of El Paso, Texas showing Sites for Source and Ambient Sampling.

method with several other bulk analysis methods like atomic absorption, ion chromatography, low temperature ashing, and proton induced x-ray emission, used to provide more information and to verify the CCSEM data.

A total of 110 samples including source and ambient samples were collected to analyze a set of TSP filters previously collected by high-volume samplers in El Paso. The hi-vol ambient samples were collected on both cellulose and glass fiber filters. The particles were removed from the filter with acetone using sonication. The particles were then redeposited on a Nuclepore filter for the CCSEM analysis. Although this procedure has the potential for differential removal of particles depending on particle size and for dissolution of some particles into the acetone, the comparison of the CCSEM elemental analysis of a series of test samples showed very good agreement with the bulk analyses performed on the same samples using the conventional methods listed above (Casuccio *et al.*, 1983b).

Source samples were collected from major sources of particulate matter.

For ducted emission, an EPA Method Five sampling protocol was used so the particles were collected on a glass fiber filter. The fugitive emissions samples were collected using a RADER sampler. This sampler consists of an in-line filter holder containing a standard 8" by 10" filter. Both cellulose and glass fiber filters were used. Two inch diameter aluminum pipe were used for the inlet and outlet air flows. Thus, the source sample particles also required removal from fiber filters. This removal was accomplished in the same fashion as the ambient sample. It is hoped any bias in the ambient sample in size and solubility would also apply to the source samples in an identical manner. For studies where it is possible to do so, it is recommended to design a sampling protocol that will collect the particles on a medium that can be used directly in the SEM.

For each sample more than 700 particles are measured. Each particle is characterized by 25 variables including 6 physical descriptive variables (maximum and minimum diameter, ratio of minimum to maximum diameter, mass, aerodynamic diameter, and effective mass) as well as 19 average elemental x-ray intensities (Na, Mg, Al, Si, S, Cl, K, Ca, Ti, Fe, Cr, Mn, Cu, Pb, V, Cu, Zn, Br, and Ni). These samples were analyzed before the ultrathin window x-ray detectors were available and therefore C, N, and O data were not obtained. The effective mass is the particle mass per unit area of the filter surface. The particle mass is estimated from the aerodynamic diameter which in turn is derived from equation 1 using the physical size, shape factor, and estimated density. From the observed elemental x-rays, the particle is assigned a density. From the projected area, the volume is estimated as a spheroid of revolution and the mass is obtained by multiplying this volume by the density. The reproducibility and the precision and accuracy of this method is described by Casuccio *et al.* (1983b). For the classification study, the 21 source samples listed in Table 1 were used to provide the examples of particle classes needed to development of a classification scheme. Ambient samples were then tested and classified using the same scheme.

Table 1. Reference Source Samples Collected in the El Paso and Used for the Classification Study.

Process Stack Samples

Sample ID	TACB ID	Description
T235	#26	ASARCO Zinc Baghouse
T236	#27	ASARCO Copper Reverberatory Furnace
T237	#28	ASARCO Converter Baghouse
T238	#30	ASARCO Zinc Plant Baghouse
T270	#43	ASARCO Sinter Plant
T271	#44	ASARCO Lead Sinter Plant

ASARCO: The American Smelting and Refining Company

Process Fugitive Emission Samples

Sample ID	TACB ID	Description
T220	# 6	American Minerals Sample (Mn Smelter)
T234	# 2	El Paso Rock Quarry
T240	# 9	ASARCO Lead Blast Furnace
T248	#39	Parker Brothers (Slag Crushing Operation)
T263	#32	ASARCO Coke Storage Hoppers
T264	#33	ASARCO Sinter Plant (Front)
T273	#10	Southwest Portland Cement

Fugitive Soil Emission Samples

Sample ID	TACB ID	Description
T219	# 5	UTEP (Soil near Monitor)
T239	# 8	IB & WC (near Monitor)
T242	# 3	Dirt Road, Ascarate/Valencia
T256	#41	Dixie Ethyl Refinery (Old Gas Refinery)
T266	#35	Newell Salvage (Salvage Yard)
T268	#37	IB & WC (near Monitor)
T272	#45	Phelps Dodge (Copper Refinery)

Miscellaneous Samples

Sample ID	TACB ID	Description
T221	# 7	Highway Emission (San Antonio)

5.3 Data Screening and Transformation

To begin the analysis, a simple univariate analysis was made to look at the distribution of values and statistics of the raw data set. As one of the advantages of CCSEM, the frequency distribution for aerodynamic diameter can be directly obtained from the measured physical size and estimated density. The distribution was always positively skewed and bimodal in shape. Figure 7 gives the distribution of manganese x-ray intensities for the manganese smelter source sample T220. It can be seen that a bimodal distribution was observed with a minimum occurring around 3 μm and that the distribution is strongly positively skewed. Because this is a particle number distribution, the accumulation mode contains the larger fraction of the particles although the coarse mode would represent the greater mass.

There were several data treatment steps to begin the CCSEM data analysis such as the selection of variables, the noise reduction of cases, and the transformation of their values. The first step is to select proper variables out of the 25 physical and chemical variables. The 6 size related variables (length, width, its aspect ratio, mass, aerodynamic diameter and effective mass) were strongly correlated with other elemental concentration variables. This type of mixed variables is difficult to use in cluster analysis, especially when Euclidian distance is used. The Euclidian distance is invariant under orthogonal rotation of the pattern space so that variables should be orthogonal or uncorrelated. Otherwise high correlated variables may cause an incorrect classification. Thus, the first technique to screen the raw data was to use only the 19 chemical variables.

The second data treatment to be employed will be noise reduction in the elemental x-ray intensities. The x-ray fluorescence peaks are obtained as a result of a photon counting process having a Poisson distribution. There are observed "peaks" in the spectrum that arise from the statistical fluctuation in the detector background in the particular energy region characteristic of that element. Thus, it is necessary to eliminate those peaks that do not have sufficient intensity to be considered real. If the x-ray count for an element was less than two times the square root of the total number of x-ray counts in the spectrum for this particle, the count was set to zero. It was observed that the cluster pattern gave a much

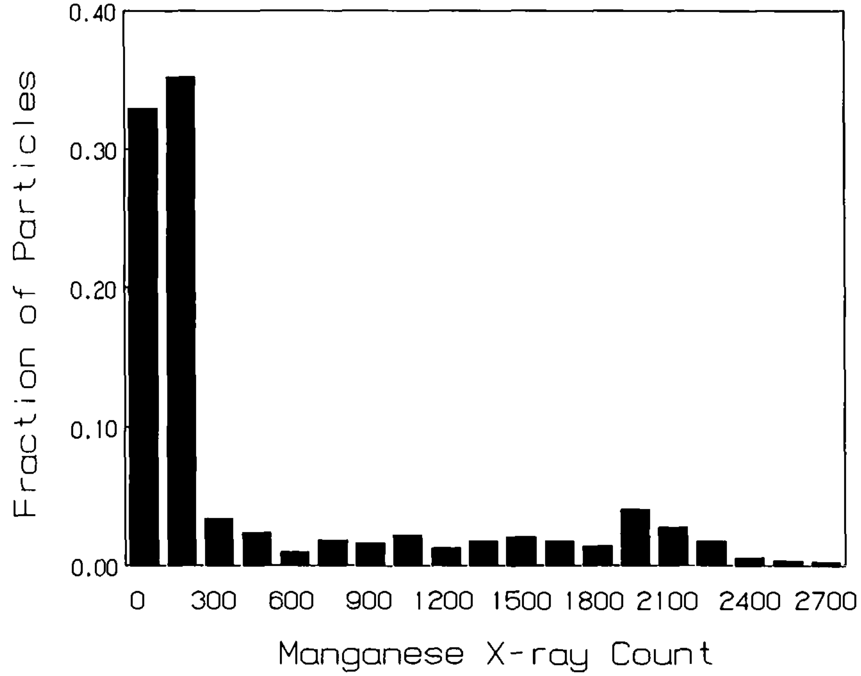

Figure 7. Frequency distribution of the number of manganese x-ray counts in the manganese smelter source sample T220.

more reasonable classification of the particles after the noise reduction scheme.

Finally, a logarithmic transformation of the data was used to compress the distributions. Since many x-ray intensities, x_i, have a value of zero or are set to zero by the noise reduction scheme, a logarithmic transformation can not be made directly. A log $(1+x_i)$ transformation avoids the difficulty with zero values. The addition of a single count makes an insignificant perturbation to the total count values since total x-ray counts were used. The frequency distribution of manganese x-ray intensities for source sample T220 is given again after logarithmic transformation in Figure 8. As can be seen, the distribution has been compressed by the transformation and made more symmetric compared to Figure 7. In the case of the manganese x-ray intensities, there is really a bimodal distribution with peaks occurring around 1.8 and 3.3 (63 and 1,000 of x-ray count)

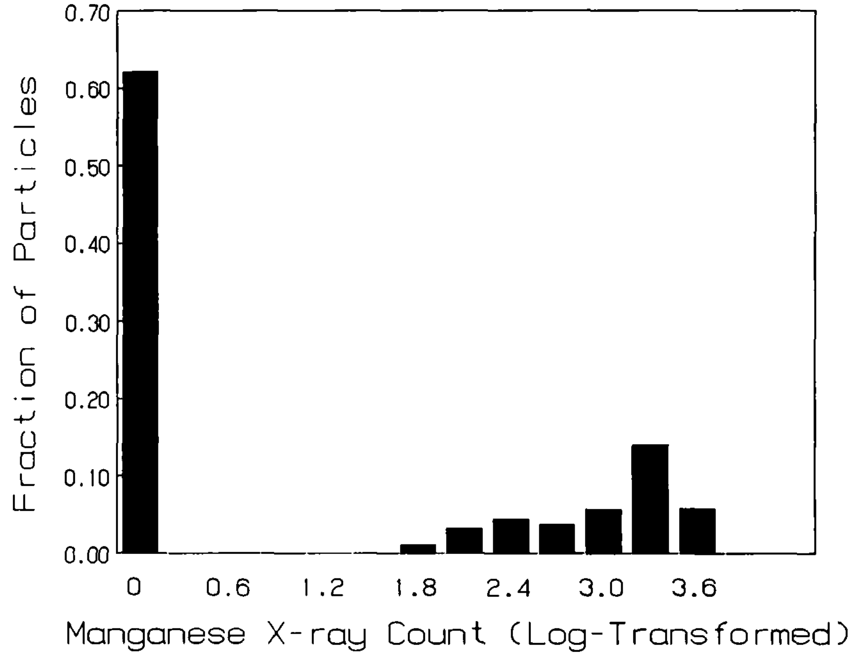

Figure 8. Frequency distribution of the number of manganese x-ray counts after the noise reduction and the logarithmic transformation in the manganese smelter source sample T220.

when excluding the peak of x-ray intensities of zero. After the noise reduction and the logarithmic transformation for the same source sample T220, this bimodal distribution was changed into the reasonable looking symmetric shape with a peak occurring around 3.3, shown in Figure 8, and thus became a more amenable shape for clustering. Finally, the log transformed data were autoscaled to have zero mean and unit variance. This transformation insures that all variables have equal weight in determining the similarity between any given pair of objects (particles).

5.4 Statistical Classification

5.4.1 Cluster Analysis

Cluster analysis is used to find meaningful aggregations (clusters) of objects (individuals, entities, patterns or cases) with little or no *a priori* information and identify useful patterns within a large data set. Mathematically a cluster is a subset of a set of objects.

The basic concept of classification is to identify each object that is similar in a pattern space based on a measure of distance or dissimilarity. Hence, if the distance (dissimilarity) between the two objects is small, they are considered to be in the same cluster. There are many quantitative definitions of dissimilarity and distance measures in the literature to solve cluster problems (Hopke, 1983; 1985).

A commonly used measure of distance or dissimilarity is the squared Euclidean distance. If there are two points in an m dimensional space where m is the number of measured variables (such as x-ray intensities), then the squared Euclidean Distance (SED) is defined as

$$SED_{ij} = \sum_{k=1}^{m} (x_{ik} - x_{jk})^2 \qquad (2)$$

where X_{ik} is the value of the kth variable for the ith datum point (particle) and X_{jk} is the value of the kth variable for the jth datum point (another particle). Alternatively the Euclidean distance can be calculated by taking the square root of equation 2.

Before using various cluster programs, the proper data treatment is needed. Since the cluster analysis is dependent on the data treatment scheme, care must be taken in interpreting the results. It is also important to determine the appropriate measure of dissimilarity or distance. Euclidean distance has been advocated empirically in biological taxonomy since the early 1960's (Sneath and Sokal, 1963). Squared Euclidean distance has been found to be a reasonable measure of distance for environmental data (Hopke, 1983). When Euclidean distance is used as a measure of distance, highly correlated variables should be eliminated prior to the analysis, otherwise it may lead to distorted classifications.

Since the calculation of distance depends on the scale of the data, normalizing variables by dividing by its standard deviation over the total objects is recommended (Massart and Kaufman, 1983).

5.4.2 Applications of Cluster Analysis

Hierarchical and non-hierarchical clustering methods were used to identify the initial classes in the data set. These pattern recognition approaches were applied to extensively explore the auto emission source data set T221, one of the source samples from the El Paso study (Kim *et al.*, 1987). Three clustering methods, including an agglomerative hierarchical program AGCLUS (Oliver, 1973), the k-median program MASLOC (Kaufman and Massart, 1983), and the BMDP (Biomedical Computer Programs) k-means program (1985), were applied to these data to compare the computational efficiency and the ability to isolate clusters.

The agglomerative hierarchical algorithm AGCLUS was first tested for the classification of the CCSEM data. The agglomerative method starts from the individual objects each in a separate cluster and ends with a single cluster containing all of the objects. Larger clusters are constructed by merging smaller clusters. This method requires that once an assignment of objects to a cluster is made, those objects will remain together at higher and higher clustering levels. Thus, errors made early in the clustering cannot be repaired at higher levels when the clusters are larger and better defined.

Hierarchical clustering can be represented in the form of tree-shaped dendrograms such as shown in Figure 9. The horizontal distance between clusters is a representation of their dissimilarity. In this hierarchical method, difficulties exist in choosing significant clusters from the dendrogram. The choice of how dissimilar objects can be while still belonging to the same group is then made by defining the distance from the left side of each segment where a vertical line would separate the figure into possible classes.

Several large classes of very similar particles can be seen in the example shown in Figure 9 as well as some particles that are not likely to be assigned to any well defined particle class for this sample. However, it must be noted that the

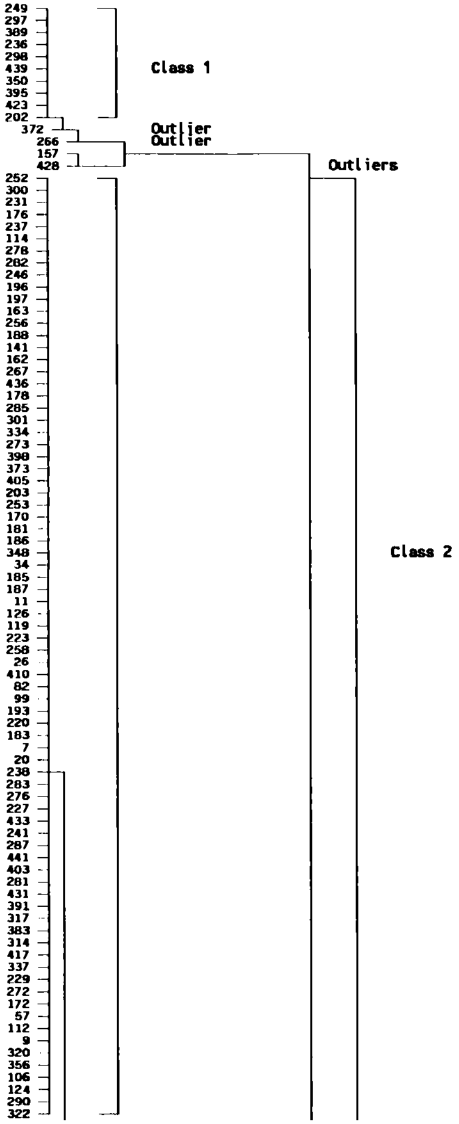

Figure 9. Portion of a Agglomerative, Hierarchical Dendrogram Showing Initial Assignment of Particles to Classes

hierarchical method is simple in terms of computational scheme, and is flexible in terms of decision on the assignment to significant clusters by simply shifting the clustering level in the same dendrogram.

A non-hierarchical method is generally performed in three steps: 1) choosing a specific number of clusters, k; 2) initially allocating objects to these clusters; and 3) reallocating some of the objects to other clusters until an "optimal" solution is obtained. As the initial step, a number of clusters, say k, can be chosen empirically or estimated by the computer simulations (Begovich and Kane, 1982). However, it is generally difficult to decide the specific number of clusters that provides the optimal partitioning for a given data set. There are no comprehensive, generally accepted criteria for choosing the proper number of clusters. Empirical decisions by the investigator are generally needed to solve this problem because the definition of the term cluster and the optimal clustering criterion are dependent on the types of variables and objects in a particular study.

As non-hierarchical methods, MASLOC and BMDP-KM computer programs were applied to the same auto emission source sample T221. The MASLOC program used Euclidean distance as the dissimilarity measure and provided an option for using the algorithm on sequentially increasing values of k. The most significant feature in MASLOC is a determination of significant clusters (robust clusters) that identify the strong relationships that exist in the data. A robust cluster is one in which the objects never aggregate with objects of any other cluster upon increasing the values of k.

In the BMDP-KM program, the Euclidean distance is also used to measure the distance between each object and the center of each cluster, where the center becomes the mean of objects in the cluster. Once a specific number of clusters is chosen, the program places all the data into one cluster and splits one cluster into two clusters step by step until the requested number of clusters is reached. Then objects are reallocated into the cluster whose center is nearest.

The results of MASLOC and BMDP for 60 clusters ($k=60$) clearly identified the isolated clusters where the intracluster distances are small compared to any intercluster value; and the result of AGCLUS at a slightly lower level was similar to them. The computation time to obtain these results differed by orders of magnitude between three algorithms. MASLOC required substantially greater computer time such that it would not be practical to run it for all of the source samples. AGCLUS uses only a short execution time. However, MASLOC performed much better with small clusters with distinctive elemental patterns such as showing high heavy metal concentration groups, in auto emission source sample (Kim *et al.*, 1987) than did AGCLUS.

AGCLUS was run for each of the 21 source samples on a microcomputer coprocessor board that employed a National Semiconductor NS32032 CPU (Hopke, 1986) after the data treatment mentioned above. With the 19 chemical variables for each particle, the dendrograms produced suggest distinct particle classes at the lowest level of the clustering pattern. All the particle groups of more than 4 particles at this level were initially chosen as potential homogeneous classes. Generally as the number of classes created for each source sample increases, the number of particles assigned to each class decreases.

5.4.3 Homogeneous Particle Classes and Outliers

The particles can be characterized by physical and chemical parameters. Although shape may be useful in classification, the other physical parameters such as size, density, and mass are either of not high discriminating power (size) or are strongly dependent on the identification of the chemical composition (density and mass). For example, to assign a density to a particle, the observed x-ray spectrum must be used to estimate the chemical compounds present. Such an assignment will be rather imprecise. Mass and aerodynamic diameter depend on further assumptions regarding the volume as a function of the projected area and the estimated density. Thus, the initial classification efforts were based entirely on the x-ray intensity data.

In the study of classification for particle classes based on their x-ray intensities, an important consideration is the definition of the term "homogeneous particle classes." A homogeneous class is one in which all of the objects have similar characteristics. In this case a similar pattern of x-ray intensities. When implementing various clustering methods to classify the objects into homogeneous particle classes, the clustering methods could only provide tentatively homogeneous classes. A possible group of particles chosen from cluster analysis can often have chemically different particles assigned to it. These particles are considered to be outliers. Thus, a class is determined to be homogeneous only when all particles in a class show only the same elements. A particle in Al-Si-Fe class will show only Al, Si, and Fe x-ray intensities. In the sense of this concept, all particles initially clustered together from cluster analysis can be examined, and thus it is possible to remove outlier particles for each class. Those particles in the class after deleting outliers can be potential "representative" examples for the expert system.

With the 19 chemical compositions as variables, the $2^{19} - 1$ combinations of particle classes can occur without replacement of the elements. For example, when there are only three elements, Al, Si, and Fe, the possible particle classes can be observed as one of the seven particle classes (Al, Si, Fe, Al-Si, Al-Fe, Si-Fe, and Al-Si-Fe). However, because of possible isomorphous substitutions of elements in terms of geochemistry of formation of the material, the sequential order of elements in the class name is not related to the relative x-ray intensities for particle classes having more than two elements. With two elements, it is likely that a change from one being the major x-ray intensity to the other being dominant reflects a change from one particle type to another. Thus, Si-Al and Al-Si classes are different from each other. However, since there is the possibility of composite particles as well as substitutions, the Si-Al-Fe, Al-Si-Fe, and Fe-Si-Al classes are considered to be of the same homogeneous class.

The total number of possible particle classes is then $(2^{19} - 1) + C(19,2)$, where the second term explains the number of combinations with replacement of the element when considering only two elements. The objective is to have a precise enough classification method to obtain sufficient resolution of one source material from another. There are practical limitations to how fine the classification can be. As the class definitions become more specific, more particles must be examined so that the mass fraction that class represents can be determined with reasonable precision. Thus, it is possible to provide more classes if needed at the cost of analyzing enough particles to insure each class mass fraction is well determined.

5.4.4 Expert Systems

Expert systems, a class of high performance computer programs in the area of artificial intelligence, are applied as knowledge-engineering tools in any field to interpret, predict, diagnose, design, plan, monitor, and control systems. An expert system is dependent on obtaining the knowledge in fully concrete terms. Mittal and Dym (1985) warned of potential mismatch because a single domain expert in complex and varied domains like applied science can only provide information on a small subset of the tasks in the domain. Thus, a system attempting to expand beyond that subset may provide poor or misleading guidance. In general, to draw a conclusion about scientific evidence it is necessary to collect all the possible evidence or examples characterized by various variables and then analyze them systematically. Recent trends to reach this task are to analyze data statistically and develop experimental design to collect data easily analyzed by expert systems. Gale and Pregibon (1985) reviewed a project applying artificial intelligence methods in statistics.

5.4.5 EX-TRAN 7

EX-TRAN 7 (Hassan *et al.*, 1985) is a series of programs designed to generate a set of rules in the form of a decision tree based on examples for which various attributes (variables) are known and which have known outcomes (classes). It then produces a self-contained FORTRAN program that can implement these rules. The rule-generator program is called the Analog Concept Learning Translator (ACLTRAN). It can use numerical or logical variables as attributes for the objects in the known classes. ACLTRAN uses an extension of Quinlan's ID3 algorithm (Quinlan, 1983).

The program searches the features one at a time to identify the one for which it can "best" separate one class from the others. The choice of attribute is based on the maximum decrease in the entropy derived from information theory between the undivided and the split classes. The information theoretical approach used to determine the entropy of each state is described by Hunt *et al.* (1966). For a given variable, the entropy of the whole and divided states are calculated, and the separation is made based on the maximum entropy decrease obtained. The decision rule is then formulated for the attribute with the largest entropy change. The rule splits the data such that an attribute value greater than or equal to the value midway between the closest points of two data sets are assigned to one group and those less than that critical value are assigned to the other class. The program systematically divides the data set until all of the objects are separated into single class subsets. Thus, EX-TRAN is a sequential univariate classification system.

The ID3 algorithm is an iterative procedure that forms a succession of decision trees. The trees increase in their accuracy in classifying the objects until each rule has no contradictory examples. It does so by selecting a subset of objects and developing rules for them. It then adds objects that represent exceptions to the existing rules and develops a new decision tree. The process is repeated until all objects can be properly classified. This procedure will always work as long as no two objects have identical attribute values and different assigned classifications. However, the number of rules may become rather large to accommodate overlapping classes (Derde *et al.*, 1987). The algorithm was tested for a data set of 2000 objects and 14 attributes (Quinlan, 1979) and was found that the algorithm converged rapidly; decision rules could be developed using only a small fraction of the total data set; the initial subset size was not a critical factor; and the computation time is a linear function of the product of the number of examples, the number of attributes and the number of nodes in the decision tree.

Although EX-TRAN is a one attribute at a time approach, it does take some of the correlation between variables into consideration since the splitting of the data set based on one of perfectly correlated variables will be identical to that obtained for any of the others. The system always splits the data along directions orthogonal to the variable axes and thus cannot directly determine diagonal class boundaries, but rather would approximate it by a series of orthogonal rules. EX-TRAN does provide the derived decision tree directly as well as its implementation as a FORTRAN code or compilable subroutine. This subroutine can then be linked with the DRIVER program of the EX-TRAN system to provide an executable program to classify other objects based on their attribute values and the existing set of rules.

5.4.6 EX-TRAN Implementation

This rule-building expert system was used to build a decision tree from examples of particle classes following classification studies. To perform this study, it was necessary to find a set of "representative" examples of the "homogeneous" particle classes.

Initially the agglomerative hierarchical clustering algorithm, AGCLUS, was used to obtain the potential number of the homogeneous particle classes. Distinct particle classes at the lowest level of the dendrogram in cluster analysis were selected when the class contained more than four particles. The initial classes were not perfectly homogeneous and, therefore, the outlier particles were removed. Some classes were split into smaller classes to obtain homogeneity. Each data set representing one of twenty-one source samples was split into 23 to 67 homogeneous particle classes. A total of 11,294 (73%) out of 15,499 particles (80% of total mass) were placed into one of the homogeneous classes so that they could be used as examples for expert system implementation. Table 2 shows the total number of particles, the total number of particles placed in homogeneous classes, and the number of homogeneous classes for each source data set.

It is important to note that the remaining particles (27%), including both particles not selected from the cluster analysis and outlier particles removed from the potential homogeneous classes, were placed into a miscellaneous class that was not active in the EX-TRAN 7 algorithm. As described above, since there are many possible combinations for homogeneous classes, in order for the

Table 2. Number of Particles, Number of Classified Particles, and Number of Classes for Each Ambient Sample Identified by the Classification Rule

Sample ID	No. of Particles	No. of Classified Particles	No. of Classes
A218	735	582	115
A222	645	459	90
A223	735	539	136
A224	735	530	129
A225	735	471	122
A228	735	611	104
A250	735	518	84
A251	735	450	98
A253	735	578	99
Total	6,525	4,738	

miscellaneous class to be active in the EX-TRAN 7, a corresponding number of examples would be needed to build a rule for the universal particle classification. This limitation could be overcome by simple set theory.

Let the totality of particles be the universal set, S; the homogeneous classes be a subset, H; and the miscellaneous class be a subset, M. Further, let P be the probability of belonging to any given set. Then P(H) is the probability of the event H, so that the following properties are satisfied:

i) $P(H) > 0$ and $P(M) > 0$

ii) $P(H) + P(M) = P(S) = 1$

iii) $P(H \cap M) = 0$

iv) If $H_1, H_2, \ldots H_k$ are the sub-sub sets of sub set H, then

$$P(H_1 \cup H_2 \cup \ldots \cup H_k) = P(H_1) + P(H_2) + \ldots + P(H_k)$$
$$= P(H)$$

for each positive integer k.

$H_1, H_2, \ldots H_k$ are independent from one another; therefore $H_i \cap H_j = 0$, $i \neq j$. In particular, if $k = (2^{19} - 1) + C(19,2)$, then $P(M) = 0$ and $P(H) = P(S) = 1$. Since H is the complement of M, even though subset M is not active in the expert system study, this problem can be solved after obtaining a perfect subset H.

In order to create a classification rule from the expert system, it is necessary to find a set of "representative" examples. For this study, 283 homogeneous particle classes from the 21 source data sets were created after deleting outlier particles. Table 3 shows a portion of the homogeneous classes, all the sodium-containing particle classes observed in the 21 source samples. As can be seen, some classes were unique, i.e., observed for only one particular source, and other classes were observed in two or more sources. Among all 283 classes, some classes such as Na-Al-Si-Ca, Mg-Al-Si-Ca, Mg-Si-Ca, Al-Si-K, Al-Si-K-Ca, Al-Si-K-Ca-Fe, Al-Si-Ca, Al-Si-Ca-Fe, Si, Si-Al, Si-S-Ca, Ca, and Ca-Si were frequently observed in more than twelve sources. Particularly, the Al-Si-K class was the most commonly observed class being found in 17 of the 21 sources.

It must be noted that all of 11,294 particles assigned to one of the 283 classes cannot be used directly as the representative examples since the maximum number of examples that can be used by EX-TRAN is 300. To select 300 representative particles from the 11,294, various techniques have been tested to provide the best set from which to induce an accurate rule tree. EX-TRAN sets its decision value at the midpoint value between the variable values of the two closest points between the groups of particles being separated (Derde *et al.*, 1987). Thus, an example which has minimum or maximum value for an attribute (variable) in a given class can be considered as a representative example. For instance, the Si-Al-K particle class observed from the 17 sources consisted of 434 particles. A particular particle is then selected as a representative example which has the maximum value of the Si x-ray intensity among all the particles in the class. Next, a particle with the minimum value of Si is selected. The Al and the K x-ray intensities are also considered in the same manner. Then 6 particles can be chosen to represent this class among the 434. If a class consists of 10 attributes, at most 20 particles are needed. For all the 283 classes, the same treatments were used. However, the 283 classes were more than enough to exceed the limitation of the 300 examples. The next technique used to solve this problem was to make subsets such as H'_1, H'_2,....H'_j, and then $S > H > H'_j > H_i$, where i,j = integer. To do this, the first subset H'_1 was defined as a set of all the Na-

Table 3. Homogeneous Particle Classes and Those Number of Particles for Each Source Sample.

Particle Type	ID	Number of Particles for Each Source Sample				
Na-Mg-Al-Si- S-Fe	1	T266(3)				
Na-Mg-Al-Si- K-Ca-Fe	2	T242(2)	T268(5)			
Na-Mg-Al-Si-Ca	3	T239(8)	T242(4)	T268(2)		
Na-Mg-Si- S- K	4	T263(4)				
Na-Mg-Si- K-Ca	5	T234(2)				
Na-Mg-Si-Ca	6	T219(4)	T234(2)	T239(4)	T242(2)	
Na-Mg-Si-Ca-Pb	7	T270(2)				
Na-Mg-Si-Pb	8	T263(4)	T264(7)	T270(8)	T271(11)	
Na-Mg-Ca	9	T234(8)				
Na-Mg-Pb	10	T264(5)	T270(7)	T271(8)		
Na-Al-Si	11	T219(7)	T220(5)	T239(10)	T242(6)	T248(4)
		T256(6)	T268(9)	T272(10)		
Na-Al-Si- S	12	T266(8)				
Na-Al-Si- S- K-Fe	13	T266(4)				
Na-Al-SI- S-Ca	14	T221(7)	T266(2)			
Na-Al-Si- S-Ca-Fe	15	T266(5)				
Na-Al-Si- S-Fe	16	T266(6)				
Na-Al-Si- K	17	T219(4)	T220(5)	T221(2)	T234(9)	T239(11)
		T242(13)	T248(4)	T256(8)	T266(7)	T268(9)
		T272(14)				
Na-Al-Si- K-Ca	18	T219(7)	T239(11)	T242(6)	T256(4)	T268(8)
		T272(8)	T273(5)			
Na-Al-Si- K-Ca-Fe	19	T221(3)	T239(7)	T242(7)	T248(4)	T268(3)
		T272(6)				
Na-Al-Si- K-Fe	20	T221(5)	T239(10)	T242(7)	T256(8)	T268(3)
		T272(3)				
Na-Al-Si- K-Mn	21	T220(2)				
Na-Al-Si-Ca	22	T219(19)	T220(4)	T221(4)	T234(6)	T239(16)
		T242(15)	T248(6)	T256(3)	T266(4)	T268(13)
		T272(15)	T273(8)			
Na-Al-Si-Ca-Fe	23	T219(3)	T239(4)	T242(2)	T248(3)	T256(5)
		T273(4)				
Na-Al-Si-Ca-Mn	24	T220(2)				
Na-Al-Si-Fe	25	T239(6)	T256(5)	T266(2)		
Na-Al-Si-Fe-Mn	26	T220(3)				
Na-Al-Si-Pb	27	T221(3)	T271(5)			
Na-Si- S	28	T236(6)	T263(4)			
Na-Si- S- K-Ca	29	T270(4)				
Na-Si- S-Ca	30	T242(3)	T256(4)	T263(4)	T266(6)	T272(3)
Na-Si- K-Ca	31	T234(2)				
Na-Si- K-Pb	32	T235(5)	T270(3)			
Na-Si-Ca	33	T219(12)	T221(4)	T234(17)	T239(15)	T256(2)
		T272(4)	T273(4)			
Na-Si-Ca-Fe	34	T248(9)				
Na-Si-Ca-Fe-Mn	35	T220(3)				
Na-Si-Fe-Mn	36	T220(5)				
Na-Si-Pb	37	T236(5)	T237(4)	T263(7)	T264(4)	T266(3)
		T270(10)	T271(8)			
Na- S-Ca	38	T263(10)	T266(10)	T271(5)		
Na- S-Fe	39	T263(2)				
Na- K-Pb	40	T235(8)				
Na-Fe-Mn	41	T220(2)				
Na-Fe-Pb	42	T263(2)				

containing classes; the second subset, H'_2 as a set of the Mg-containing classes except objects in H'_1; and the third subset, H'_3 as a set of Al-containing classes except objects in H'_1 and H'_2, and so on.

5.4.7 Building and Testing a Classification Rule

Initially the transformation of log (1+x) was made for the x-ray intensities following the setting of sufficiently small values to zero. Thus, an example file for subset H'_1 was prepared for ACLTRAN. Each example is characterized by the 19 elemental x-ray intensities as attributes and labeled with a homogeneous class type. ACLTRAN generates induced classification rules in several different forms such as a decision tree, a standard FORTRAN-77 code, and a FORTRAN-77 subroutine to be compiled and linked with the Driver program. All the source samples were used to build the rule tree, test the rules and correct examples.

Figure 10 shows an example of a tree from the decision rule, and its corresponding standard FORTRAN-77 code is listed in Figure 11. In Figure 10, the decision tree shows the assignment of particles containing sodium into these classes. For example, if the transformed Ca value is less than 1.0205, the particle will go into classes, C1, C4, C8, or C10 depending on the variables S, Si, and Al. For particles with Ca greater or equal to 1.0205, they will be assigned to classes C2, C3, C5, C6, C7, or C9 depending on the values of Al, Si, and K. The names given to each class are listed in the box below the figure. The FORTRAN routine listed in Figure 11 implements this decision tree where the Ca value is the first point of separation and the same class assignments are made.

The decision rules for all the subsets were determined in an identical manner, and then each FORTRAN code was incorporated into a complete universal classification code. Since the FORTRAN-77 subroutine for the Driver program is designed for the diagnosis of one sample at a time, it will be time consuming to use and thus was not employed. The FORTRAN code was compiled and linked so that the rules could be executed as a program on an IBM or compatible microcomputer.

In the beginning of the test step, a decision rule obtained from the 21 source

```
[Ca     ] :
    < 1.0205 : [S      ] :
           < 1.00 : [Si     ] :
                   < 0.8265 : c10
                   > =0.8265 : c8
           > =1.00 : [Al     ] :
                   < 1.145 : c4
                   > =1.145 : c1
    > =1.0205 : [Al     ] :
           < 1.00 : [Si     ] :
                   < 2.1125 : [Si      ] :
                          < 1.0395 : c9
                          > =1.0395 : c7
                   > =2.1125 : [K       ] :
                          < 1.057 : c6
                          > =1.057 : c5
           > =1.0000 : [K       ] :
                   < 1.0395 : c3
                   > =1.0395 : c2
```

Class c1......Na-Mg-Al-Si- S-Fe	Particle Type
Class c2......Na-Mg-Al-Si- K-Ca-Fe	Particle Type
Class c3......Na-Mg-Al-Si-Ca	Particle Type
Class c4......Na-Mg-Si- S- K	Particle Type
Class c5......Na-Mg-Si- K-Ca	Particle Type
Class c6......Na-Mg-Si-Ca	Particle Type
Class c7......Na-Mg-Si-Ca-Pb	Particle Type
Class c8......Na-Mg-Si-Pb	Particle Type
Class c9......Na-Mg-Ca	Particle Type
Class c10.....Na-Mg-Pb	Particle Type

Figure 10. Example of a Tree from the Decision Rule.

samples was examined sequentially for each source data set for all 15,499 particles. For this test, each raw source data set was screened and transformed as described above and its particles were labeled with one of the 284 classes, including the miscellaneous class. Thus, the test was performed like a supervised pattern recognition analysis. Only two (0.01%) out of the previously assigned particles were misclassified by the final routine. These misclassifications occurred when the number of representative examples in a class was much less than that of the active attributes. Thus, the frequency of the misclassification was small and

```
CHARACTER*8 decisn
REAL na    ,mg    ,al    ,si    ,s     ,cl    ,k
REAL ca    ,ti  ,fe    ,mn    ,zn    ,cr    ,pb
REAL v     ,cu    ,p     ,br    ,ni

IF(ca    .LT.   1.0205)THEN
      IF(s     .LT.     1.00)THEN
            IF(si    .LT.   0.8265)THEN
                 decisn='c10'
            ELSE
                 decisn='c8'
            END IF
      ELSE
            IF(al    .LT.    1.145)THEN
                 decisn='c4'
            ELSE
                 decisn='c1'
            END IF
      END IF
ELSE
      IF(al    .LT.   1.0000)THEN
            IF(si    .LT.   2.1125)THEN
                 IF(si    .LT.   1.0395)THEN
                       decisn='c9'
                 ELSE
                       decisn='c7'
                 END IF
            ELSE
                 IF(k     .LT.    1.057)THEN
                       decisn='c6'
                 ELSE
                       decisn='c5'
                 END IF
            END IF
      ELSE
            IF(k     .LT.   1.0395)THEN
                 decisn='c3'
            ELSE
                 decisn='c2'
            END IF
      END IF
END IF

END
```

Figure 11. Standard FORTRAN-77 Code for the Decision Rule.

acceptable. From the results of testing all particles, it was found that the universal decision rule could also detect the misclassification by the hierarchical cluster analysis. In sample T240 from Table 2, 501 of 735 particles were assigned to one of the identified homogeneous classes. However, three events in the miscellaneous set were found to belong to one of the homogeneous classes. In this manner, a total 48 particles from the 21 source samples were misclassified by the hierarchical cluster analysis. The misclassified examples were reassigned to a homogeneous particle class, and the mass fraction values for each homogeneous class were corrected. It is an important step to correct the mass fractions of each class since they will be used as source profiles in a subsequent source apportionment study.

It is also noted that the decision rule could classify miscellaneous particles in a sample into one of homogeneous classes, where that class does not appear to be in the source sample. Again, in sample T240 from Table 2, 52 of 234 particles in the miscellaneous set were assigned to one of homogeneous classes other than the 33 homogeneous classes originally created for the sample T240. 1,367 (8.8% of the total particles) from the 21 sources could be relabeled in this manner.

6. PARTICLE CLASS BALANCE

To use the individual particle data, the particle class balance (PCB) was first proposed by Johnson and McIntyre (1982). They used this approach to determine the source contributions at the receptor site in Syracuse, N.Y. based on the particle-by-particle information.

In this section, we will examine the use of PCB analysis using carefully classified particles and through this analysis provide the contributions to the ambient aerosol mass in El Paso, Texas. Using the 21 source profiles developed from the homogeneous particle classes described above, the PCB method was used to identify and quantify the major TSP sources in El Paso. By examining samples from 2 sources, 283 particle classes were defined and the fraction of mass in each class were calculated. The full set of source profiles and their estimated uncertainties are available (Kim, 1987).

The starting point of the PCB is the mass conservation principle and the resulting mass balance. The measured particle mass in a given particle class at the receptor is assumed to be the sum of the contributions of the sources to those classes. Mathematically, the model is expressed as

$$x_{ij} = \sum_{k=1}^{p} c_{ik} f_{kj} \qquad \begin{array}{l} i=1,\ m\ classes \\ j=1,\ n\ samples \end{array} \tag{3}$$

where x_{ij} is the mass of particles of ith class for the jth sample, c_{ik} is the mass fraction of ith class from the kth source, and f_{kj} is the mass contribution of kth source for the jth sample. The commonly used chemical mass balance (CMB) is expressed analogously with the PCB except in a CMB analysis, x_{ij} is the ith elemental concentration measured in the jth sample, c_{ik} is the gravimetric concentration of the ith element from the kth source, and f_{kj} is the airborne mass concentration of kth source for the jth sample. The collection of the class mass fractions or gravimetric concentration for all sources, the C matrix, is referred to as the source profile matrix.

Watson (1982) presented the assumptions and the limitations of chemical receptor methods as follows: 1) composition of source emissions are constant, 2) components do not react with each other, and 3) only identified sources contribute to the receptor. Especially for the chemical mass balance approach, he provided two additional assumptions: 4) the number of sources is less than or equal to that of the components, and 5) the compositions of all sources are linearly independent of each other. Among the above assumptions, the problem of collinearity described in item 5) lead to serious degradation of precision intensity in the CMB analysis. At the Mathematical and Empirical Receptor Models Workshop (Quail Roost II), Stevens *et al.* (1984) noted that additional species characterization is needed to resolve similar sources. The PCB is constrained by assumptions 1) to 3) since it is a linear receptor-oriented measurement model, but it can potentially overcome the mathematical limitations of 4) and 5) by greatly increasing the number of variables in the fit.

6.1 Diagnostics for the Source Profile Matrix

Before solving for the unknown mass fractions by multiple linear regression, it is necessary to determine if the C matrix contains collinearity. When some of the explanatory variables (i.e., some of the source profiles) are collinear, the least square solutions for the regression coefficient may yield poor solutions and the solutions obtained have poor precision (Cheng and Hopke, 1989). Weisberg (1985) suggested that deletion of predictor variables (sources) from a regression model could reduce apparent collinearity.

Belsley *et al.* (1980) developed methods for detecting and identifying the presence of collinear relations among explanatory variables. According to them, collinearity has to do with specific characteristics of the source profile matrix C and not the statistical aspects of the linear regression model

$$x = Cf + e \tag{4}$$

where **e** is a vector containing the observed fitting residuals, **x** is the vector of class mass fractions, **f** is the vector of mass contributions, and C is the matrix of source mass fraction profiles.

To test the degree of collinearity for El Paso source profiles obtained from the universal classification rule, a singular value decomposition (SVD) approach was used (Henry, 1982). The source profile matrix C can be decomposed as

$$C = UDV^T \tag{5}$$

where U is an n class by n orthogonal matrix, V is a p source by p orthogonal matrix, and D is an n by p diagonal matrix. In matrix D, its diagonal elements d_k, $k=1,...p$, are known as the singular values of matrix C. The singular-value decomposition is related to the eigenvalue and eigenvector concepts since the diagonal elements of D are the square roots of the eigenvalues of C^TC. Generally small singular values, d_k, demonstrate the existence of collinearity, and lead to problems in the regression analysis. Belsley *et al.* (1980) suggested an empirical criterion to decide the degree of linear dependencies by introducing a condition index, C.I., as the relative ratio of singular values.

$$C.I. = d_{max}/d_k \qquad k=1,...p \tag{6}$$

where p in this case is 21. Weak dependencies are associated with condition indexes around 5 to 10, whereas moderate to strong relations are associated with condition indexes of 30 to 100.

Further, they introduced the variance-decomposition proportion to identify the collinear vectors. The variance-covariance matrix of the least-square estimator, **f**, is

$$\sigma^2(C^T C)^{-1} = \sigma^2 V D^{-2} V^T \tag{7}$$

where σ^2 is the common variance of the components of e in the linear model, **x** = C**f** + **e**. Then, kth component of the variance-covariance matrix becomes

$$Var(f_k) = \sigma^2 \sum_j v_{kj}^2 / d_j^2 \qquad k,j = 1,\ldots,p \tag{8}$$

where d_j and v_{kj} can be obtained from Equation 4. For a small singular value, i.e., a small denominator in the equation, the variance component will be large compared to the others. The variance-decomposition proportions can be calculated as follows.

$$\phi_{kj} = V_{kj}^2 / d_j^2 \qquad and \qquad \phi_k = \sum_{j=1}^{p} \phi_{kj} \tag{9}$$

Then, the variance-decomposition proportions are

$$\pi_{kj} = \phi_{kj} / \phi_k \tag{10}$$

Belsley *et al.* (1980) suggested that this concept be used for the case of near dependencies after using the SVD information. When there is a condition index larger than 5, the existence of a variance-decomposition proportion larger than 0.5 for two or more components of the source matrix indicate collinear vectors.

Condition indexes and variance-decomposition proportions for the source matrix were calculated for the source profile matrix to check for potential collinearity. As recommended by Belsley *et al.* (1980), all source profiles were scaled to unit column length. Table 4 shows the condition indexes and Table 5 gives the values for the pi-matrix for the scaled profiles. Since the matrix decomposed by the singular value decomposition had 21 columns, there are 21 singular values numbered 1 to 21 and 21 values of the condition index. The column on the left is this index of the singular values. All the condition indexes are less than 5, and thus it shows excellent independence among source samples. Although there are some pi values greater than 0.5 (or 5000 in the scaled table), they are not significant because the condition index for that singular value is sufficiently small. Therefore, the C matrix in Equation 3 can be used in a regression model without deletion of any source profile to reduce collinearity and it will not degrade the precision of the estimation for the regression coefficients.

6.2 Source Apportionment of Ambient Aerosols in El Paso

Nineteen ambient samples were chosen from the El Paso airshed and examined for source apportionment. All of the ambient samples were initially screened and transformed as described by Kim and Hopke (1988). After the universal classification rule was executed, each particle was assigned to one of the 283 possible homogeneous classes identified in the source samples. Approximately 100 homogeneous particle classes were observed in each ambient sample as shown in Table 6.

Mass fractions and their uncertainties were calculated in the same manner as those of the source samples. For examples, Table 7 shows mass fractions and their uncertainties for ambient sample A224. From the classification of ambient particles in the 19 samples, it was observed that 76% of the total particles were assigned into the homogeneous classes in terms of number fraction, and 83% of the total mass in terms of mass fraction. Complete results for all of the samples are given by Kim (1987). Thus, the ambient mass fractions, **x**, and the source profiles, C, are available to solve for the source contributions, **f**, in Equation 3.

For the El Paso apportionment study, regression analysis was performed by the program LSCODE (Cheng, 1986). The program was developed for solving CMB source apportionment models and provides options for various weighting

Table 4. Condition Indexes and Pi Matrix (Variance-Decomposition Profiles) for Unscaled Source Profile.

ID	S.I.	T219	T220	T221	T234	T235	T236	T237	T238	T239	T240	T242	T248	T256	T263	T264	T266	T268	T270	T271	T272	T273
T219	1.00	.0086	.0024	.2146	.0376	.0099	.2038	.1497	.0355	.9690	.0001	.0100	.0007	.0023	.0199	.0124	.0767	.1535	.0028	.0023	.0171	.0003
T220	1.08	.0090	.0005	.0001	.0796	.0006	.0022	.0211	.0006	.0001	.0000	.0000	.0001	.0000	.0006	.0004	.0000	.0000	.0108	.0000	.0006	.0000
T221	1.46	.0054	.0043	.0210	.0000	.0104	.0001	.0677	.0091	.0009	.0002	.0000	.0212	.0000	.0000	.0062	.0000	.0098	.0556	.0011	.0011	.0000
T234	1.56	.0181	.0000	.0032	.0000	.0024	.0042	.0018	.0000	.0000	.0000	.0089	.0000	.0000	.0029	.0001	.0025	.0000	.0030	.0000	.0000	.0000
T235	1.64	.0142	.0001	.0032	.0174	.0973	.0027	.0057	.0001	.0002	.0000	.0000	.0077	.0000	.0002	.0002	.0057	.0058	.0097	.0001	.0066	.0000
T236	1.75	.2443	.0004	.0503	.0304	.6301	.0319	.0688	.7168	.0045	.0000	.0012	.0065	.0000	.1087	.0032	.0865	.0871	.0000	.0000	.0003	.0000
T237	1.82	.0003	.0000	.0388	.0000	.0000	.0013	.0017	.0067	.0000	.0001	.0001	.0001	.0000	.0001	.0022	.0026	.0115	.0007	.0004	.0003	.0000
T238	1.85	.0042	.0000	.0148	.0001	.0034	.0005	.0007	.0080	.0000	.0000	.0056	.0002	.0000	.0000	.0000	.0013	.5432	.0000	.0000	.0034	.0000
T239	1.88	.1107	.0002	.0021	.3157	.0001	.0274	.0018	.0003	.0000	.0000	.0000	.0001	.0000	.2090	.0035	.0002	.0008	.0000	.0004	.0002	.0000
T240	1.92	.0000	.0003	.0017	.0125	.0000	.0003	.0384	.0555	.0160	.0000	.0000	.0096	.0000	.0000	.0181	.0063	.0027	.6093	.0089	.0000	.0000
T242	2.06	.0175	.0004	.2259	.0221	.0041	.0015	.0007	.0042	.0000	.0000	.0029	.0028	.0002	.0001	.0019	.0174	.0046	.0012	.0000	.0003	.0000
T248	2.26	.0002	.0000	.1127	.0001	.0025	.0087	.0000	.0000	.0008	.0000	.0046	.0001	.0000	.0000	.0023	.0001	.0000	.0005	.0012	.0008	.0000
T256	2.40	.0001	.0000	.0253	.0063	.2075	.0322	.0016	.0003	.0001	.0000	.0010	.0000	.0000	.0000	.0184	.0000	.0004	.0002	.0196	.0191	.0000
T263	2.63	.0156	.0115	.0489	.0016	.0088	.0000	.0225	.0006	.0002	.0000	.0007	.0016	.0000	.0003	.0001	.0086	.0083	.0005	.0000	.0000	.0000
T264	2.93	.0000	.0001	.0000	.0001	.0001	.0025	.0301	.0015	.0000	.0004	.0000	.0011	.0000	.0001	.0001	.3546	.1213	.0055	.0003	.0076	.0000
T266	5.09	.0335	.0001	.0159	.0061	.0010	.6015	.0000	.0000	.0000	.0000	.6308	.0021	.0000	.2158	.0000	.1586	.0018	.0001	.0007	.0184	.0000
T268	4.55	.0016	.0000	.1883	.0407	.0018	.0001	.0201	.0022	.0000	.0016	.0000	.0002	.0000	.1824	.0032	.0102	.0422	.0176	.9620	.8579	.9993
T270	4.13	.3039	.9796	.0014	.3060	.0169	.0315	.5675	.1584	.0080	.9974	.3341	.9381	.9974	.2512	.9260	.2438	.0033	.0490	.0019	.0003	.0002
T271	3.21	.2039	.0000	.0255	.0318	.0003	.0409	.0000	.0000	.0001	.0000	.0000	.0039	.0000	.0072	.0000	.0168	.0029	.1636	.0009	.0657	.0002
T272	3.32	.0090	.0000	.0064	.0918	.0027	.0069	.0000	.0002	.0001	.0000	.0000	.0040	.0000	.0015	.0017	.0079	.0007	.0007	.0000	.0004	.0000
T273	3.53	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0690	.0000	.0000	.0000

Table 5. Condition Indexes and Pi Matrix (Variance-Decomposition Profiles) for Scaled Source Profile.

ID	S.I.	T219	T220	T221	T234	T235	T236	T237	T238	T239	T240	T242	T248	T256	T263	T264	T266	T268	T270	T271	T272	T273
T219	1.00	.0240	.3579	.1583	.0039	.0270	.1581	.1381	.0506	.4982	.0052	.0353	.0110	.0852	.0717	.1475	.0987	.0729	.0240	.4998	.0612	.4107
T220	1.15	.0014	.0364	.0049	.6931	.0046	.0132	.0092	.0021	.0041	.0129	.0001	.0005	.0000	.0041	.0042	.0000	.0001	.0071	.0000	.0188	.0003
T221	1.18	.0012	.0944	.0087	.0003	.0035	.0000	.1677	.0053	.0009	.0686	.0001	.0618	.0000	.0000	.0152	.0001	.0055	.0335	.0274	.0046	.0002
T234	1.23	.0022	.0000	.0007	.0000	.0000	.0097	.0016	.0020	.0002	.0003	.0048	.0000	.0000	.0024	.0029	.0000	.0000	.0001	.0132	.0001	.0000
T235	2.59	.0103	.0061	.0000	.0024	.2788	.0019	.0048	.0010	.0043	.0074	.0001	.0053	.0000	.0018	.0000	.0011	.0130	.0193	.0036	.1128	.2266
T236	2.44	.0835	.2590	.0087	.2304	.5756	.0223	.2932	.1203	.0793	.1594	.0023	.0033	.0024	.2234	.0009	.0023	.0458	.0007	.0036	.5069	.0001
T237	1.29	.0009	.0001	.1772	.0000	.0000	.0178	.0100	.0520	.0096	.3621	.0001	.0020	.0000	.0003	.0003	.0001	.0011	.0083	.0006	.0112	.0115
T238	2.06	.0010	.0056	.0006	.0134	.0068	.0000	.0023	.0127	.0028	.0009	.0000	.0000	.0000	.0000	.0004	.0002	.5076	.0072	.0063	.0330	.2127
T239	1.95	.0257	.0010	.4728	.0150	.0000	.0268	.0013	.0002	.0000	.0014	.0055	.0003	.0001	.2504	.0148	.0001	.0004	.0007	.0896	.0134	.0169
T240	1.69	.0123	.0612	.0722	.0007	.0028	.0396	.0951	.0340	.3127	.0002	.0000	.0112	.0186	.0000	.0825	.0105	.0015	.2708	.0724	.0517	.0152
T242	1.41	.2053	.0145	.0051	.0002	.0010	.0000	.0042	.0005	.0002	.0000	.0004	.0000	.0010	.0002	.0002	.0002	.0022	.0004	.0030	.0669	.0020
T248	1.48	.0004	.0298	.0255	.0078	.0150	.0008	.0000	.0005	.0012	.0009	.0069	.0022	.0026	.0002	.0001	.0032	.0018	.0002	.0000	.0245	.0001
T256	1.45	.0007	.0019	.0290	.0005	.0109	.0261	.0045	.0011	.0000	.0200	.0620	.0002	.0027	.0005	.0065	.0061	.0040	.0004	.0819	.0014	.0248
T263	1.44	.0000	.1173	.0074	.0149	.0022	.0001	.0141	.0000	.0048	.0266	.0000	.0013	.0001	.0047	.0000	.0003	.0000	.0000	.0006	.0000	.0001
T264	1.53	.0001	.0050	.0001	.0003	.0093	.0030	.0035	.0000	.0093	.0076	.0001	.0002	.0000	.0002	.0076	.1428	.0490	.0004	.0069	.0039	.0000
T266	1.79	.0110	.0032	.0000	.0030	.0006	.1728	.0000	.0002	.0001	.0013	.0441	.0047	.0000	.0055	.0003	.1091	.0006	.0001	.0356	.0216	.0042
T268	1.65	.0001	.0061	.0112	.0000	.0001	.0001	.0003	.0000	.0169	.0014	.5043	.0122	.2400	.4203	.6148	.5677	.2879	.4947	.0181	.0108	.0268
T270	1.69	.5174	.0002	.0056	.0001	.0599	.4708	.2501	.7158	.0548	.3236	.3334	.8797	.6455	.0119	.0910	.0396	.0043	.1238	.0087	.0104	.0298
T271	1.71	.0996	.0002	.0008	.0059	.0002	.0112	.0000	.0010	.0009	.0004	.0002	.0026	.0014	.0004	.0061	.0168	.0021	.0007	.1252	.0271	.0179
T272	1.59	.0029	.0002	.0111	.0080	.0018	.0256	.0000	.0006	.0000	.0000	.0001	.0015	.0004	.0018	.0048	.0011	.0003	.0012	.0035	.0199	.0003
T273	1.60	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0063	.0000	.0000	.0000

Table 6. Mass Fractions and their Uncertainties for Ambient Sample A218.

Particle Type	Class ID	Mass Fraction	Uncer-tainty	Particle Type	Class ID	Mass Fraction	Uncer-tainty
Na-Mg-Al-Si- K-Ca-Fe	C 2	.00253	.00253	Al-Si-Ca-Pb	C228	.00110	.00110
Na-Mg-Al-Si-Ca	C 3	.00605	.00129	Al-Si-Fe	C229	.00359	.00179
Na-Mg-Si- S- K	C 4	.00208	.00208	Al-Si-Br	C240	.00068	.00068
Na-Mg-Si-Ca	C 6	.00002	.00002	Al-Pb-Br	C247	.00004	.00004
Na-Mg-Si-Pb	C 8	.00009	.00009	Si	C301	.02500	.00465
Na-Mg-Ca	C 9	.00007	.00007	Si-Na	C302	.00775	.00184
Na-Mg-Pb	C 10	.00028	.00028	Si-Mg	C303	.00891	.00483
Na-Al-Si	C 11	.00803	.00285	Si-Al	C305	.00873	.00248
Na-Al-Si- S	C 12	.00083	.00083	Si- S-Ca	C313	.01293	.00283
Na-Al-Si- K	C 17	.01391	.00213	Si- S-Ca-Fe	C314	.00136	.00136
Na-Al-Si- K-Ca	C 18	.00802	.00114	SI- S-Zn	C322	.00082	.00041
Na-Al-Si- K-Ca-Fe	C 19	.00461	.00117	Si- K-Ca	C327	.00552	.00276
Na-Al-Si- K-Fe	C 20	.00446	.00206	Si- K-Ca-Fe	C328	.00117	.00117
Na-Al-Si-Ca	C 22	.02715	.00473	Si-Ca	C334	.00665	.00218
Na-Al-Si-Ca-Fe	C 23	.00118	.00067	Si-Ca-Fe	C335	.00839	.00216
Na-Al-Si-Fe	C 25	.00094	.00047	Si-Ca-Fe-Pb	C338	.00011	.00011
Na-Si- S	C 28	.00042	.00042	Si-Ca-Zn-Pb	C340	.00003	.00003
Na-Si- S-Ca	C 30	.00009	.00005	Si-Ca-Pb	C342	.01255	.00699
Na-Si- K-Pb	C 32	.00009	.00009	Si-Fe	C344	.00185	.00185
Na-Si-Ca	C 33	.01661	.00348	Si-Fe-Zn-Pb	C347	.00155	.00155
Na-Si-Ca-Fe	C 34	.00117	.00117	Si-Fe-Pb	C348	.00238	.00238
Na-Si-Pb	C 37	.00124	.00062	Si-Zn-Pb	C350	.00055	.00004
Na- S-Ca	C 38	.00258	.00129	Si-Zn-Pb-Br	C352	.00061	.00061
Mg-Al-Si- S-Ca	C104	.00238	.00529	S	C401	.00102	.00102
Mg-Al-Si- K	C108	.00420	.00037	S-Na	C402	.00008	.00008
Mg-Al-Si- K-Ca	C109	.01272	.00210	S-Ca-Fe	C409	.00076	.00076
Mg-Al-Si- K-Ca-Fe	C110	.02525	.00315	S-Fe-Zn	C414	.00299	.00299
Mg-Al-Si- K-Fe	C111	.02110	.00462	S-Fe-Cu	C415	.00083	.00083
Mg-Al-Si-Ca	C112	.02875	.00372	S-Zn	C417	.00193	.00096
Mg-Al-Si-Ca-Fe	C113	.02526	.00549	S-Zn- V	C419	.00019	.00007
Mg-Al-Si-Fe	C116	.00187	.01523	K-Ca-Zn-Pb	C501	.00042	.00042
Mg-Si- S-Ca	C126	.00673	.00055	Ca	C520	.06858	.00721
Mg-Si-Ca	C136	.02361	.00121	Ca-Na	C521	.00472	.00207
Mg-Si-Ca-Fe	C137	.00061	.00435	Ca-Mg	C522	.02564	.00414
Mg-Si-Ca-Fe-Mn	C138	.00005	.00061	Ca-Al	C523	.00020	.00010
Mg-Si-Ca-Mn	C139	.00005	.00005	Ca-Si	C524	.06316	.00765
Mg-Si-Ca-Pb	C141	.00492	.00005	Ca- S	C525	.01078	.00312
Mg-Si-Fe	C142	.00267	.00492	Ca-Cl	C526	.00006	.00006
Mg-Si-Fe-Zn-Pb	C144	.00529	.00127	Ca-Br	C527	.00007	.00003
Mg-Si-Zn-Pb	C147	.00028	.00529	Ca-Zn-Pb	C528	.00027	.00027
Mg-Si-Pb	C149	.00014	.00014	Ca-Zn-Pb-Br	C529	.00045	.00045
Mg- S- K	C152	.00004	.00004	Fe	C570	.00216	.00216
Mg- S-Ca	C154	.00455	.00004	Fe-Si	C571	.00454	.00454
Mg- K-Zn-Pb	C162	.00003	.00233	Fe- S	C572	.00416	.00416
Mg-Ca	C164	.00004	.00003	Zn-Ca	C656	.00529	.00529
Mg-Fe-Pb	C166	.00529	.00004	Zn-Fe	C657	.00234	.00234
Al-Si- S	C203	.00110	.00110	Zn-Cr-Pb	C660	.00012	.00012
Al-Si- S-Ca	C206	.00662	.00148	Zn-Pb	C662	.00019	.00005
Al-Si- S-Ca-Fe	C207	.00964	.00183	Zn-Pb- V	C663	.00004	.00004
Al-Si- S-Zn	C210	.00238	.00238	Pb	C750	.03624	.01274
Al-Si- K	C212	.03381	.00591	Pb-Na	C751	.00138	.00096
Al-Si- K-Ca	C213	.02340	.00415	Pb-Mg	C752	.00106	.00106
Al-SI- K-Ca-Fe	C214	.03320	.00495	Pb-Si	C753	.00189	.00058
Al-Si- K-Fe	C217	.02095	.00282	Pb- K	C754	.00014	.00014
Al-Si- K-Mn	C219	.00003	.00003	Pb-Ca	C755	.01134	.01134
Al-Si-Ca	C223	.06201	.00527	Pb-Zn	C757	.00333	.00158
Al-Si-Ca-Fe	C224	.02398	.00460	Pb-Br	C759	.00038	.00019
I-Si-Ca-Mn	C226	.00003	.00003	Others	C999	.13557	.01354

methods such as unweighted, ordinary weighting, and effective-variance weighting techniques. The program has an iterative feature for subset analysis by deleting a source at a time from matrix C. Sources are eliminated until both the values of all of the regression coefficients are non-negative and the smallest t-statistic is greater than a predetermined value. The t-statistic is defined as the ratio of the regression coefficient (estimated source contribution) to its estimated standard error.

6.2.1 Regression Analysis for Source Apportionment

It was decided to remove the source sample of salvage yard soil from the source profile matrix because the material was collected after extensive disturbance of the oiled soil layer. Thus, only 20 source profiles were retained in the source matrix. The number of variables used in the regression analysis of a sample was the number of homogeneous classes observed in that ambient sample. Thus, of the 283 classes, only those classes observed in the ambient sample were used. For instance, in ambient sample A218 shown in Table 6, 129 homogeneous classes were identified. The other 154 classes were removed from the original 20 x 283 source matrix.

Since uncertainties are available for both the ambient concentrations and the source profiles, weighted least squares methods could be employed and it must be determined which weighting mode, if any, is appropriate for estimating the source contributions. Although ordinary weighting and effective-variance weighting provide generally acceptable results for source apportionment using the CMB model, these weighting schemes not be ideal for the PCB model. For CMB, the average elemental concentrations may vary over many orders of magnitude from major to trace elements. Thus, weighting is essential to bring the elemental concentration variables to a similar metric. However, in PCB, all of the variables are on the same metric. Thus, weighting is not as theoretically important as in a CMB analysis. Also the uncertainty-based weights may actually distort the analysis. The uncertainty was estimated by the jackknife method, and it increases with decreasing numbers of particles in a homogeneous class. There are many classes in the ambient samples that contain only one or two particles. For those classes, 100% of the mass fraction has been arbitrarily assigned as their uncertainties. Kim (1987) has examined ordinary weighted least squares fitting and found that it provided poorer quality fits compared to an unweighted analysis. Thus, PCB model source apportionment study will use unweighted least-squares for the 19 ambient samples.

Ambient sample A218 is used as an example. Initially, Studentized residuals (Belsely *et al.*, 1980) for each variable (class mass fraction) were calculated. The Studentized residual is defined as

$$E_i^* = \frac{e_i}{s(i)\sqrt{1-h_i}} \tag{11}$$

where e_i is the ordinary residual from Equation 4, and $s(i)^2$ is the variance in the fit with class mass fraction i eliminated from the fit, and h_i are the diagonal elements of the $C(C^TC)^{-1}C^T$ matrix (Cheng and Hopke, 1989). The Studentized residuals provide a good way to examine the data since they have equal variances and should be distributed as t-distribution with n-p-1 degrees of freedom. The residuals can be plotted against each class number. Figure 12 shows the Studentized residuals plot for ambient sample A218. The points are generally symmetrically distributed near the zero line.

Table 7 shows the final results obtained from the multiple regression analyses for the 19 ambient samples collected at the 5 different monitor sites. The table includes source identification, its corresponding contributions, the TSP collected, and statistical information such as the correlation coefficient, the degrees of freedom, and the residual sum of squares for the individual ambient samples. When a t-statistic was negative, the source profile was discarded from the source matrix. The regression analysis was then repeated. In addition, when a t-statistic was less than a preset value, its corresponding source profile was deleted and the sample was reanalyzed in order to obtain a statistically significant solution. For the regression of the El Paso data, using the PCB model, 0.5 was

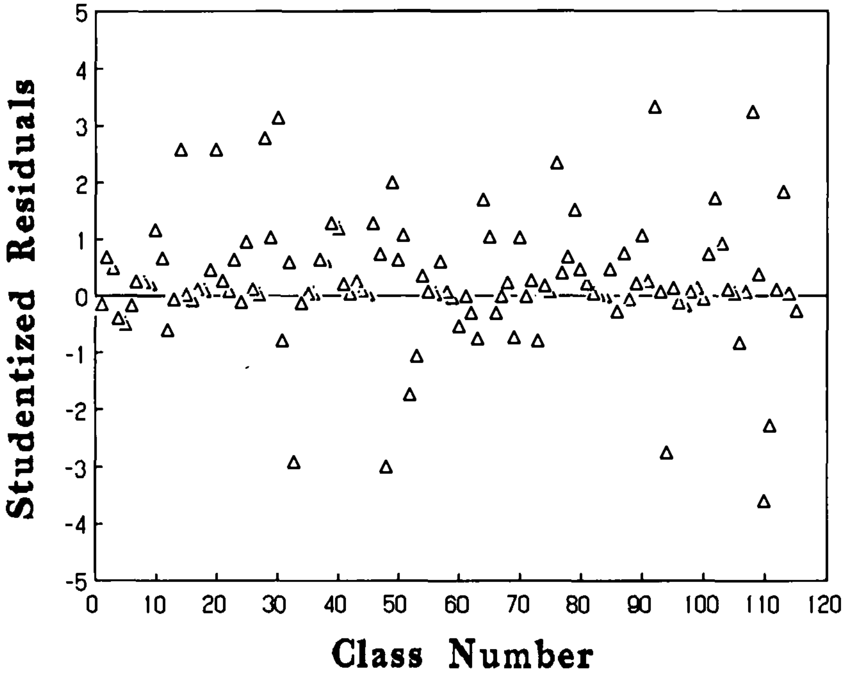

Figure 12. Plot of Studentized residuals against class number for ambient sample A218.

empirically used as a cut-off point of t-value since at this level, samples from co-located samplers yielded very similar results (Kim, 1987). The 70th percentile value of the t-distribution with 100 degrees of freedom is approximately 0.5. The results of the source apportionment in El Paso are discussed in the following section.

6.2.2 Results and Discussion for Apportionment Study

One of the limitation of receptor modeling is that source samples are assumed to be representative particulate emissions during all ambient sampling periods. To understand the El Paso source apportionment better, it is useful to briefly review the source sampling methods. First of all, highway emissions were considered to incorporate both the coarse particle road dust emissions generated

Table 7. Results Obtained from the Particle Class Balance Analysis for the 19 Ambient Samples Collected at the 5 Different Sites.

Source ID	Tillman Monitor								Ascarate		Ivanhoe Stat.			Zack White			IB&WC Monitor		
	A225	A222	A223	A224	A218	A059	A228	A253	A251	A250	A048	A053	A062	A046	A052	A061	A047	A051	A060
T219 (Soil: UTEP)	.081	.066	-	.036	.079	-	.073	.101	-	-	-	-	-	-	-	-	-	-	-
T220 (Mn Smelter)	.038	-	-	.156	-	-	.270	-	.296	-	.428	-	-	-	.220	.144	.045	-	.715
T221 (Auto Emission)	.115	.183	.261	.172	.109	-	.182	.378	.079	-	.236	-	-	-	-	.073	-	-	-
T234 (Rock Quarry)	.077	.083	.175	.177	.143	.122	.186	.229	-	.112	-	-	.163	-	-	.210	-	-	-
T235 (Zn Baghouse)	.025	.048	-	.034	-	-	-	-	-	-	-	-	-	-	-	-	.427	-	-
T236 (Cu Rever. Furnace)	-	-	-	-	-	-	-	-	-	-	-	-	.349	.112	-	-	.299	-	-
T237 (Converter Baghouse)	-	-	-	.057	.041	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T238 (Zn Plant Baghouse)	.071	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T239 (Soil: IB & WC)	.167	-	.038	.083	.147	.342	.135	.099	-	-	-	-	.070	-	-	-	-	-	-
T240 (Pb Blast Furnace)	-	-	.061	.024	-	-	-	.029	-	.335	-	-	-	-	-	-	-	-	-
T242 (Soil: Dirt Road)	.039	.224	.250	.139	.089	-	-	.044	.022	.114	.336	-	-	.130	-	.235	.193	-	-
T248 (Slag Crushing)	.085	.070	.031	.039	.031	.141	.049	.056	-	.178	-	.057	.185	-	.029	.193	.007	.258	.048
T256 (Soil: Oil Gas Refin.)	-	-	-	.027	-	-	-	-	.269	.262	-	.515	-	.449	.089	-	.029	.568	.194
T263 (Coke Storage Hoppers)	.207	.160	-	.057	-	.395	.072	.064	-	-	-	.106	-	.310	-	.147	-	-	-
T264 (Sinter Plant)	-	-	.102	-	.185	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T268 (Soil: IB & WC)	-	-	-	-	.176	-	-	-	-	-	-	.280	-	-	.063	-	-	-	-
T270 (Sinter Plant)	-	-	-	-	-	-	.035	-	-	-	-	-	-	-	.599	-	-	-	-
T271 (Pb Sinter Plant)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.174	.043
T272 (Soil: Cu Refinery)	-	-	-	-	-	-	-	-	-	-	-	-	.234	-	-	-	-	-	-
T273 (Portland Cement)	.096	.166	.080	-	-	-	-	-	-	-	-	.042	-	-	-	-	-	-	-
Estimated Unknown Fraction	.183	.171	.175	.192	.188	.200	.204	.182	.231	.173	.215	.150	.136	.185	.282	.185	.151	.174	.244
Observed Unknown Fraction	.192	.215	.182	.138	.136	.269	.111	.143	.231	.180	.328	.096	.232	.195	.060	.155	.233	.025	.088
TSP ($\mu g/m^3$)	111	156	100	90	170	69	316	226	36	34	67	92	46	42	99	49	84	469	.341
Correlation Coefficent (r)	.927	.944	.962	.964	.952	.801	.916	.950	.784	.767	.808	.894	.873	.616	.825	.800	.901	.859	.894
Degree of Freedom (DF)	111	82	128	117	106	79	96	91	92	79	69	57	65	82	57	68	66	58	87
Residual Sum of Square	.003	.003	.002	.002	.002	.009	.003	.003	.006	.008	.001	.010	.007	.028	.018	.016	.005	.019	.011

by motor vehicles as well as the direct fine particle exhaust since the source sample (T221) was collected near a highway, not from the direct auto exhaust (Kim *et al.*, 1987).

Six urban soil samples were used in this source apportionment study. Each soil sample was collected over a period of ten minutes by a portable sampler using an electric powered air pump. The soil was slightly disturbed by a sampling effort to simulate the reentrainment by the wind. The six soil samples were collected near the TSP monitors, near the smelter that was suspected of substantially contributing to the measured TSP, and from a dirt road. Among these samples, the dirt road sample (T242) collected in the Ascarate Park area may be the most representative surface-soil sample along the river valley in the city of El Paso since it will have the least contamination from industrial activities. The fugitive sampling procedure was similar to the soil sampling except that the inlet was held horizontally near the center of the fugitive source. Thus, the industrial fugitive emission samples include a large portion of the crustal classes. The process fugitive emission samples were quite different from the process stack samples. The stack samples were collected isokinetically using EPA Method 5.

As shown in Table 6, 10 to 20% of the mass fraction in each ambient sample was due to unclassified particles. The table also includes the unknown mass fraction estimated from the sum of mass fraction times the unclassified fractions of corresponding source samples. In the ideal case, the observed and the estimated unclassified fractions should be the same. In general the estimated values agree closely with the observed values. However, samples A051, A052, and A060 were highly overestimated, and sample A048 was underestimated. Although there may be potential TSP contributors to El Paso, no source samples were collected in Mexico. TSP from unidentified sources in Mexico and elsewhere in the airshed may be the cause of the poor results for some samples in the source apportionment study.

The Tillman monitor is located in the southern part of El Paso near the Mexican border. The identified sources contributed ambient aerosol mass to the receptor with multiple correlation coefficients (R) greater than 0.9. Figure 13 shows source apportionment pie diagrams for Tillman monitor samples A222 and A223 taken on two different days. Samples from co-located samplers were useful not only to assure the sampling quality, but also to assure the statistical reliability of the analyses. Each pair of samples show relatively good agreement with each other. For example, Figure 14 shows bar diagrams for the samples A222 and A225 with similar major peaks. The Tillman monitor was impacted by highway emissions, rock quarrying, and various urban soil sources including T219 (soil from UTEP), T239 (IB & WC soil source), and T242 (Ascarate Park soil). Industrial fugitive samples, T234 and T248, impacted persistently on this monitor, and T263 (fugitive emission source.from a coke storage) was frequently observed.

Two ambient samples were from the Ascarate Park monitor located in the southeast of El Paso. The TSP concentration at this monitor was always lower than at the other monitors. The multiple correlation coefficient (R) for the PCB estimations was also lower at this site. The site was impacted mainly by highway emissions and the dirt road. No industrial stack sources were detected.

The Zack White School monitor was located northwest of the main area of industrial activity. Ambient sample A048 was apportioned mainly by T256 (soil source from old gas refinery) and by T263 (fugitive emission source from coke storage hoppers). The other samples, A061 and A062, were affected by mostly industrial stack and fugitive emission sources. Compared to the other sites, the residual sums of squares in the regression analysis were relatively high. This sampling site may be affected by sources other than those included in this study, and high residuals suggest this possibility. Without local meteorological data in this complex valley system, a definitive analysis of these results is not possible.

The IB & WC monitor was also located near the ASARCO plant. As expected, the monitor was affected by industrial sources. Sample A051 (Table 6) had the highest TSP concentration and was principally soil. For the ambient sample A060, 71.5% was estimated to be from the manganese smelter (source T220).

Three samples from the Ivanhoe Fire Station monitor were examined. The monitor was located east of El Paso near the International Airport. Physically

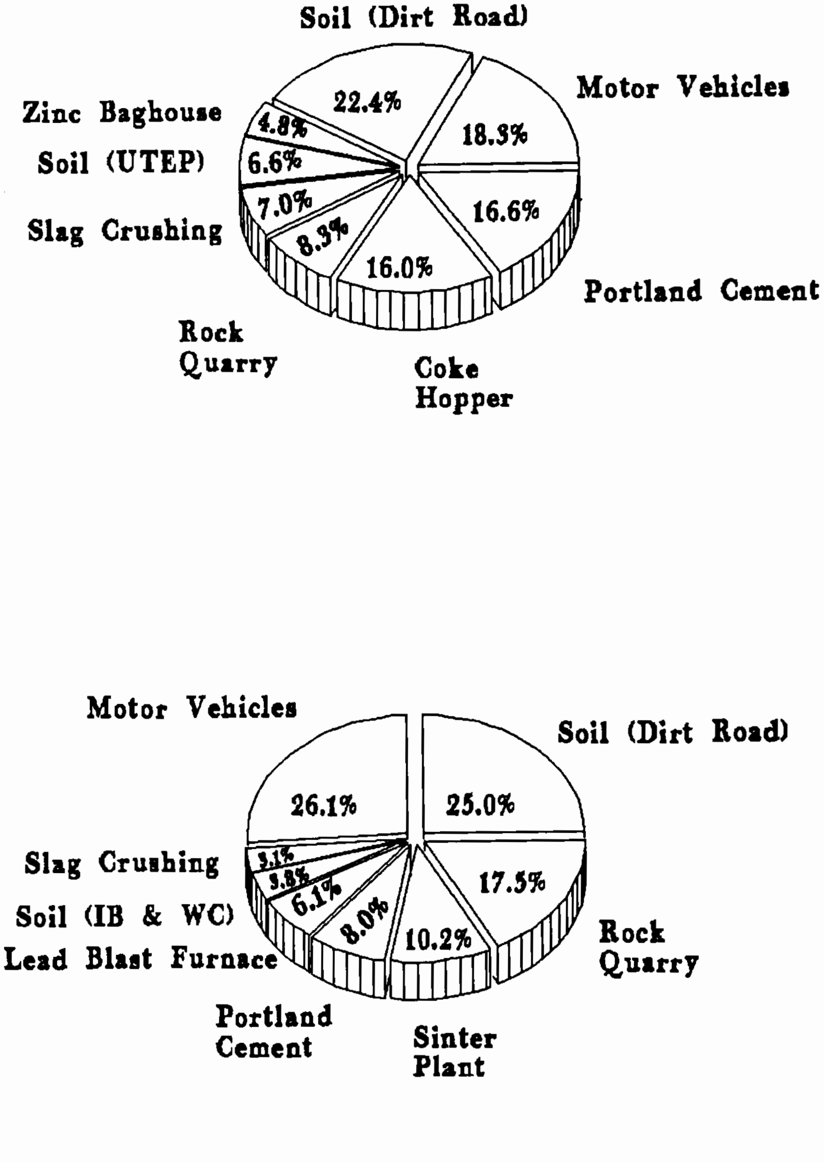

Figure 13. Pie charts for source apportionment of ambient samples A222 (top) and A223 (bottom).

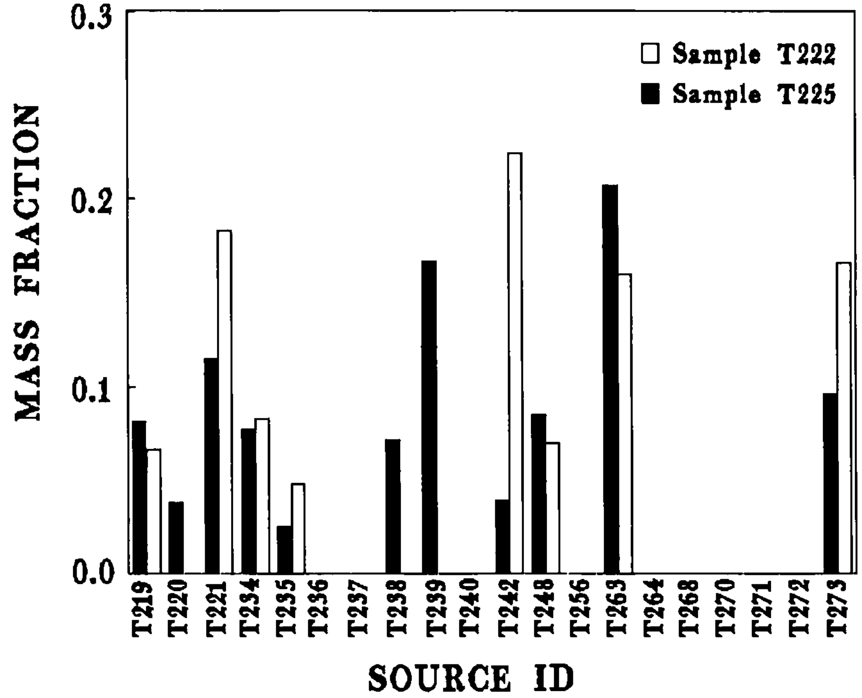

Figure 14. Bar charts for source apportionment of co-located ambient samples A222 and A225.

unreasonable results were obtained for all of the samples. The soil sources, like T242 (dirt road), and the copper refinery (T272) near the monitor, were expected to be the major contributors. In sample A048, more than 40% of the mass fraction was apportioned to the fugitive emissions from the manganese smelter (T220) located in far west El Paso. In sample A053, approximately 80% of the mass fraction was attributed to soil sources T256 and T268 collected on the opposite side of the city. Thus, previously discarded soil sample T266 was reintroduced in the source profile matrix, and the regression analysis repeated. However, there was very little change in the results. Once a small amount of T266 was included, sources T220 (fugitive emission sample from a Mn smelter) and T256 (soil sample from an old gas refinery) were observed as the main contributors to ambient samples A048 and A053, respectively. It appears that there are unidentified sources that are being aliased by the included sources. In a retrospective reanalysis such as this study, it is not possible to re-sample and improve the analysis.

As a whole, the slag crushing source (T248) was most frequently observed throughout the city for all seasons, even though it was not a major mass contributor. Highway emissions, dirt road emissions, emissions from the rock quarry, and from the coke storage hoppers were the principal contributors of TSP in El Paso. Generally, the stack emission sources were not observed frequently since its portion of mass near the source is lower than that from the other sources. The stack emissions are presumed to pass over the nearby ground level monitors. When TSP concentration in an ambient sample is lower than 50 $\mu g/m^3$ or higher than 400 $\mu g/m^3$, the multiple correlation coefficients in the regression analyses were relatively low. Finally, samples from co-located samplers provided a useful basis for assuring statistical reliability.

6.3 Conclusions of the El Paso Study

The particle class balance (PCB) model has been shown to be a useful receptor model. In this study, 20 identified sources were fitted to ambient samples taken in El Paso, Texas. The large number of homogeneous classes provide potentially better resolution in a PCB analysis than the limited number of elemental concentrations can yield in a CMB analysis. Less than 20% of mass was unclassified using the PCB model. Thirdly, similar sources like various soil sources can be used in PCB source profiles without needing to eliminate any source due to collinearity. The increased number of variables breaks the collinearity problem in regression analyses.

Several potential problems do exist with the particle class balance analysis method. First, lead-containing particles have been used as one of the major tracers of motor vehicle emissions. With the phase-out of lead in gasoline, this tracer is eliminated. However, improvements in x-ray detectors now permit detection of carbon and oxygen. Thus, the carbonaceous particles emitted by motor vehicles can be detected. These particle are coemitted with reintrained road dust again giving a suite of particle classes that permit resolution of motor vehicles in the ambient samples.

Second, there is the problem of secondary sulfate particles. As in any mass balance problem, there is generally no simple manner to assign this mass to sources. Other information is required to make this assignment whether using a chemical mass balance or a particle class balance. Without other tracers that have not yet been found, it will be difficult for the PCB analysis to do any more than a CMB analysis.

Finally, there is a problem of volatility of material in the high vacuum of a scanning electron microscope. Various particle materials including organic carbon, nitric acid and ammonium nitrate will be readily lost and thus unaccounted for in both source and ambient samples. If such materials represent a large fraction of the actual particle mass, a biased analysis would be obtained using these techniques.

This methodology has been applied to several other problems. Kim *et al.* (1989) have used CCSEM data to compare upwind and downwind samples taken in the vicinity of a large coal-fired power plant. A helicoptor was used to collect samples that were then sectioned and analyzed. Distinct differences in the types of particles and the mass fractions in the identified particle classes were observed between the upwind and in-plume samples.

Subsequently a much more extensive sampling campaign was made in the vicinity of the same power plant. In-plume samples were obtained using a specially built high volume sampler mounted under a helicopter. Upwind samples were obtained using a fixed wing aircraft while stack samples were obtained with a dilution sampler similar to the one described in the chapter in this volume by Houck and with an EPA Method 5 particle sampling train modified by Battelle Columbus Laboratories (Battelle, 1988). The samples were again characterized using CCSEM with a light element detector that provides C, N, and O x-ray intensities and analyzed using the methods described above. Hopke and Mi (1990) provide the detailed description of the results. The general conclusions were that the samples from the dilution sampler could be distinguished from those obtained by the in-plume sampler. Thus, the dilution sampler was not able to fully mimic the processes that occur in the ambient atmosphere after the plume has left the stack and begun to dilute and cool. The samples from the same sampler were found to be quite similar from sampling interval to sampling interval, but the samples from each sampler were distinguishable from one another. With the small size of the samples that could be obtained using aircraft, CCSEM coupled with advanced data analysis methods provide a powerful means for characterizing an aerosol and comparing it with the aerosol found at other times and in other locations.

7. OTHER APPROACHES TO DATA ANALYSIS

The approach described above is one that we have found to be effective in the cases to which it has been applied. However, other studies have been performed by other investigators to make use of the information available from electron microscopic analysis of particle samples. Dzubay and Mamane have used manual microscopy in a number of cases to identify and quantify the impact of specific source types. Mamane and Dzubay (1988) were able to distinguish the presence of coal fly ash in samples from Philadelphia, PA where the samples also contain substantial amounts of clay minerals. As can be observed in Figures 3 and 4, these two particle types have very similar chemistry and it is the visual morphology of the particles that permit them to distinguish one type of particle from the other. Coarse fraction crustal matter estimated from x-ray fluorescence elemental data agreed well with that based on electron microscopy. Mamane (1988) has also estimated the contribution of municipal refuse incinerators to the airborne particle mass.

In a more extensive investigation of the same Philadelphia samples, Dzubay and Mamane (1989) have apportioned the mass concentrations attributed to botanical, coal fly ash, mineral (soil), and incinerator sources. Again these results are in good agreement with those obtained using a chemical element balance analysis obtained with the bulk sample XRF data. These results show that very intensive manual analysis can succeed, but the analyses required considerable time and effort.

Saucy *et al.* (1987) have also found cluster analysis to be a useful technique for the identification and characterization of atmospheric aerosol particles. They have developed methods to analyze a large number of samples in order to intercompare one sample with another. They began using Euclidean distance and a modified chi-square measure and found that the cluster analysis did not provide a good separation as judged by a manual inspection of the cluster results. They then use a new dissimilarity measure, the angle between the vectors drawn from the origin to the points representing the particles in the multivariate space. They find that this measure provided a better cluster pattern.

In addition to the work that has been done at the University of Antwerp on the chemical analysis of particles described earlier, they have also examined data analysis methods. Van Borm and Adams (1988) and Van Borm (1989) have used cluster analysis to examine the resuspension of soil particles in the vicinity of a zinc smelter. They also use hierarchical clustering to identify groups of similar particles. They examined a number of clustering criteria similar to those described by Hopke (1983) and found that Ward's error sum method performed the best.

8. CONCLUSIONS

The scanning electron microscope under computer control and with automated image analysis represents a powerful method for obtaining a great deal of information on individual particles. It is possible to combine these individual particle results to obtain a reasonable approximation to the bulk concentration of the collection of particles. The advantages and disadvantages of this methodology have been summarized by Casuccio *et al.* (1983b) as follows:

Advantages

Physically measures particles within a broad size range, from 0.2 to 300 μm.

Elemental chemistry is obtained from every particle.

Each particle is classified by size and composition.

Distributions are obtained for each particle class as a function of size.

Both geometric (optical) and aerodynamic size distributions are calculated.

Size and x-ray data are stored for future retrieval.

Analysis time averages less than 1.5 seconds per particle, including sizing and chemical analysis.

Data acquisition is designed so that size and weight distributions have an absolute uncertainty, independent of the size range selected.

The analysis is compatible with most sampling methods.

Effects of operator bias, fatigue and subjectivity, inherent in manual microscopic techniques, have been minimized.

Results are reproducible.

Disadvantages

All samples must be prepared for the SEM. Most samples must be redeposited.

The nature of dichotomous fine fraction samples presents special sample preparation problems.

Species with an average atomic number close to that of the substrate are difficult to detect.

Particle volume is inferred from projected area.

Calculated particle mass assumes that the density is known for each particle type.

Chemical inhomogeneities within a particle may not be recognized.

The software sorting algorithm must be modified if a significant fraction of undefined particles is encountered.

These advantages and disadvantages must be given strong consideration when formulating a receptor modeling study. As with other analytical methods, CCSEM may not, by itself, provide all the desired information. However, it seems likely that there will be an increasing role for such a methodology in future receptor modeling studies.

9. REFERENCES

Battelle, *Biological Effect of Plume Fly Ash*, Final Report of EPRI Project PR2482-5, Battelle Columbus Laboratories, Columbus, OH (1988).

Begovich, C.L. and V.E. Kane, Estimating the Number of Groups and Group Membership Using Simulation Cluster Analysis, *Pattern Recog.* 15:335-342 (1982).

Belsley, D.A., E. Kuh, and R.E. Welsch, *Regression Diagnostics*, John Wiley & Sons, Inc., New York (1980).

BMDP, BMDP Statistical Software, University of California Press, Berkeley, CA (1985).

Casuccio, G.S., P.B. Janocko, R.J. Lee, and J.F. Kelly, The Role of Computer Controlled Scanning Electron Microscopy in Receptor Modeling, Paper No. 82-21.4, Air Pollution Control Association, Pittsburgh, PA (1982).

Casuccio, G.S., P.B. Janocko, S.L. Dattner, J.S. Mgebroff, and J.L. Zalar, Measurement of Ambient Hi-Vol Filters by Microscopic and Bulk Chemistry Methods, Paper No. 83-17.1, Air Pollution Control Association, Pittsburgh, PA (1983a).

Casuccio, G.S., P.B. Janocko, R.J. Lee, J.F. Kelly, S.L. Dattner, and J.S. Mgebroff, The Use of Computer Controlled Scanning Electron Microscopy in Environmental Studies, *J. Air Pollut. Control Assoc.* 33:937-943 (1983b).

Cheng, M.D., *A Unified Theory for Aerosol Source Apportionment Models*, Ph.D Thesis, University of Illinois at Urbana-Champaign (1986).

Cheng, M.D., and P.K. Hopke, Identification of Markers for Chemical Mass Balance Receptor Model, Atmospheric Environment 23:1373-1384 (1989).

Crutcher, E.R., Light Microscopy as an Analytical Approach to receptor Modeling, in *Receptor Models Applied to Contemporary Pollution Problems,* S.L. Dattner and P.K. Hopke, eds, Proceedings No SP-48, Air Pollution Control Association, Pittsburgh, PA, 266-284 (1982).

Dallavalle, J.M., *Micromeritics, the Technology of Fine Particles*, 2nd Edition, Pitman Publishing Corporation, New York (1948).

Dattner, S., S. Mgebroff, G. Casuccio, and P. Janocko, Identifying the Sources of TSP and Lead in El Paso Using Microscopy and Receptor Models, Paper No. 83-49.3, Air Pollution Control Association, Pittsburgh, PA (1983).

Derde, M.P., L. Buydens, C. Guns, D.L. Massart, and P.K. Hopke, Comparison of Rule-Building Expert Systems with Pattern Recognition for the Classification of Analytical Data, *Anal. Chem.* 59:1868-1871 (1987).

Dzubay, T.G. and Y. Mamane, Use of ELectron Microscopy Data in Receptor Models for PM-10, *Atmospheric Environ.* 23:467-476 (1989)

ETC, Identification of the Sources of TSP and Particulate Lead in the El Paso Area by Quantitative Microscopic Analysis, Vol. I and Vol. II, Final Report Prepared for the Texas Air Control Board, Austin, TX, Energy Technology Consultants, Monroeville, PA (1983).

Gale, W.A. and D. Pregibon, Artificial Intelligence Research in Statistics, *AI Magazine*, Winter:72-75 (1985).

Griffin, J.J. and E.D. Goldberg, Morphologies and Origin of Elemental Carbon in the Environment, *Science* 206:563-565 (1979).

Hassan, T., M.A. Razzak, D. Michie, and R. Pettipher, EX-TRAN-7: A Different Approach for an Expert System Generator, presented at the 5th International Workshop for Expert Systems and Their Applications, Avignon, France (1985).

Henry, R.C., Stability Analysis of Receptor Models that Use Least Squares Fitting, in *Receptor Models Applied to Contemporary Pollution Problems,* S.L. Dattner and P.K. Hopke, eds, Proceedings No SP-48, Air Pollution Control Association, Pittsburgh, PA, 141-157 (1982).

Hopke, P.K., An Introduction to Multivariate Analysis of Environmental Data, in *Analytical Aspects of Environmental Chemistry*, D.F.S. Natusch and P.K. Hopke, eds., John Wiley & Sons, Inc., New York, 219-262 (1983).

Hopke, P.K., *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons, Inc., New York (1985).

Hopke, P.K., Super Micro Computing with the DSI-32 Board, *Trends in Anal. Chem.* 5:204-205 (1986).

Hopke, P.K. and Y. Mi, The Use of a Rule Building Expert System for Analyzing Electron Microscopy Data, in *Data Handling in Science and Technology*, E.

Karjalainen, ed., Elsevier Science Publishers, Amsterdam, (in press 1990).

Hunt, E.B., J. Marin, and P.T. Stone, *Experiments in Induction*, Academic Press, Inc., New York (1966).

Janocko, P.B., G.S. Casuccio, S.L. Dattner, D.L. Johnson, and E.R. Crutcher, The El Paso Airshed: Source Apportionment Using Complementary Analyses and Receptor Models, in *Receptor Models Applied to Contemporary Pollution Problems,* S.L. Dattner and P.K. Hopke, eds, Proceedings No SP-48, Air Pollution Control Association, Pittsburgh, PA, 249-265 (1982).

Janossy, A.G.S., K. Kovacs, and I. Toth, Parameters for the Ratio Method by X-ray Microanalysis, *Anal. Chem.* 51:491-495 (1979).

Janssens, K., W. Dorrine; and P. Van Espen, The Development Process of an Expert System for the Automated Interpretation of Large EPMA Data Sets, *Chemometrics and Intelligent Laboratory Systems* 4:147-161 (1988a).

Janssens, K., W. Van Borm, and P. Van Espen, Increased Accuracy in the Automated Interpretation of Large EPMA Data Sets by the User of an Expert System, *J. Research N.B.S.* 93:260-264 (1988b).

Johnson, D.L., B. McIntyre, R. Fortmann, R.K. Stevens, and R.B. Hanna, A Chemical Element Comparison of Individual Particle Analysis and Bulk Analysis Methods, *Scanning Electron Microscopy* 1:469-476 (1981).

Johnson, D.L., and B.L. McIntyre, A Particle Class Balance Receptor Model for Aerosol Apportionment in Syracuse, N.Y., *Receptor Models Applied to Contemporary Pollution Problems*, S.L. Dattner and P.K. Hopke, eds, Proceedings No SP-48, Air Pollution Control Association, PA, 238-247 (1982).

Kaufman, L. and D.L. Massart, *MASLOC User's Guide*, Version 1.2, Vrije Universiteit Brussel, Belgium (1983).

Kelly, J.F., R.J. Lee, and S. Lentz, Automated Characteristics of Fine Particulates, *Scanning Electron Microscopy* 1:311-322 (1980).

Kim, D.S., *Particle Class Balance for Apportioning Aerosol Mass*, Ph.D Thesis, University of Illinois at Urbana-Champaign (1987).

Kim, D.S., P.K. Hopke, D.L. Massart, L. Kaufman, and G.S. Casuccio, Multivariate Analysis of CCSEM Auto Emission Data, *Sci. Total Environ.* 59:141-155 (1987).

Kim, D.S. and P.K. Hopke, The Classification of Individual Particles Based on Computer-Controlled Scanning Electron Microscopy Data, *Aerosol Sci. Technol.* 9:133-151 (1988).

Kim, D.S. and P.K. Hopke, Source Apportionment of the El Paso Aerosol by Particle Class Balance Analysis, *Aerosol Sci. Technol.* 9:221-235 (1988).

Kim, D.S., P.K. Hopke, G.S. Casuccio, R.J. Lee, S.E. Miller, G.M. Sverdrup, and R.W. Garber, Comparison of Particles Taken from the ESP and Plume of a Coal-Fired Power Plant with Background Aerosol Particles, *Atmospheric Environ.* 23:81-84 (1989).

Lee, R.J. and R.M. Fisher, *Quantitative Characterization of Particulates by Scanning and High Voltage Electron Microscopy*, National Bureau of Standards Publication 533, pp. 63-83 (1980).

Lee, R.J. and J.F. Kelly, Overview of SEM-Based Automated Image Analysis, *Scanning Electron Microscopy* 1:303-310 (1980).

Leong, K.H., P.K. Hopke, and J.J. Stukel, Evaporative Mass Losses From Particle Samples, *J. Aerosol Sci.* 14:611-613 (1983).

Mamane, Y., Estimate of Municipal Refuse Incinerator Contribution to Philadelphia Aerosol, *Atmospheric Environ.* 22:2411-2418 (1988).

Mamane, Y. and T.G. Dzubay, Fly Ash Concentrations in Philadelphia Aerosol Determined by Electron Microscopy, *Water, Air, and Soil Pollution* 37:389-405 (1988).

Markowicz, A., B. Raeymaekers, R. Van Grieken, and F. Adams, *Physical and Chemical Characterization of Individual Particles*, Ellis Horwood Ltd., Chichester, U.K (1986).

Massart, D.L. and L. Kaufman, *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, John Wiley & Sons, Inc., New York (1983).

Mittal, S. and C.L. Dym, Knowledge Acquisition from Multiple Expert, *AI Magazine,* Summer:32-36 (1985).

Oliver, D.C., Aggregative Hierarchical Clustering Program Write-Up, Preliminary Version, National Bureau of Economic Research, Cambridge, MA (1973).

Quinlan, J.R., Induction over Large Data Bases, Technical Report No. HPP-79-14, Heuristic Programming Project, Stanford University, Palo Alto, CA (1979).

Quinlan, J.R., Learning Efficient Classification Procedures and their Application to Chess End Games, in *Machine Learning, An Artificial Intelligence Approach*, R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, eds., Tioga Publishing Company, Palo Alto, CA, 463-482 (1983).

Raeymaekers, B., *Characterization of Particles by Automated Electron Probe Microanalysis*, Ph.D. Thesis, University of Antwerp (UIA), Belgium (1986).

Saucy, D.A., J.R. Anderson, and P.R. Buseck, Cluster Analysis Applied to Atmospheric Aerosol Samples from the Norwegian Arctic, *Atmospheric Environ.* 21:1649-1657 (1987).

Sewell, D.A., G. Love, and V.D. Scott, Universal Correction Procedure for Electron-Probe Microanalysis: II. The Absorption Correction, *J. Phys. D.* 18:1245-1267 (1985).

Sneath, P.H.A. and R.R. Sokal, *Principles of Numerical Taxonomy*, W.H. Freeman & Co., San Francisco (1963).

Stevens, R.K., and T.G. Pace, Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II), *Atmospheric Environ.* 18:1499-1506 (1984).

Van Borm, W.A., and F.C. Adams, Cluster Analysis of Electron Microscopy Analysis Data of Individual Particles for the Source Apportionment of Air Particulate Matter, *Atmospheric Environ.* 22:2297-2307 (1988).

Van Borm, W., *Source Apportionment of Atmospheric Particles by Electron Probe X-ray Microanalysis and Receptor Models*, Ph.D. Thesis, University of Antwerp (UIA), Belgium (1989).

Watson, J.G. (1982). Overview of Receptor Model Principles, in *Receptor Models Applied to Contemporary Pollution Problems*, S.L. Dattner and P.K. Hopke, eds, Proceedings No SP-48, Air Pollution Control Association, PA, 6-17 (1982).

Weisberg, S., *Applied Linear Regression*, 2nd Ed., John Wiley & sons, Inc., New York (1985).

Wernisch, J., Quantitative Electron Microprobe Analysis without Standard Samples, *X-Ray Spectrometry* 14:109-119 (1985).

Wernisch, J., Application of Different $\phi(\rho z)$-Distributions in Quantitative Electron Probe Microanalysis with Standard Samples, *Radex-Rundschau* 2/3:110-113 (1986).

Wiersema, J.M., L. Wright, B. Rogers, R. Barta, L. Haeuser, and J.H. Price, Human Exposure to Potentially Toxic Elements through Ambient Air in Texas, Paper No. 84-1.2, Air Pollution Control Association, Pittsburgh, PA (1984).

Chapter 7

RECEPTOR MODELING FOR VOLATILE ORGANIC COMPOUNDS

Peter A. Scheff and Richard A. Wadden
Environmental and Occupational Health Sciences
University of Illinois at Chicago
Box 6998
Chicago, IL 60680

1. INTRODUCTION

The Chemical Mass Balance (CMB) source reconciliation model has been found to be a useful tool for the evaluation of ambient speciated volatile organic compound (VOC) concentration data. The method has been applied in a simplified form to evaluate total hydrocarbon concentrations (THC) in Los Angeles, California (Mayrshon and Crabtree, 1976; and Feigley and Jeffries, 1979) and Sidney, Australia (Nelson *et al.*, 1983). (THC was defined as the sum of specified "unreactive" hydrocarbons in the atmosphere.) Mayrshon and Crabtree used ethane, acetylene, propane, i-butane, n-butane, i-pentane, and n-pentane as tracer compounds for automobile exhaust, gasoline and gas vapor, commercial natural gas, and geogenic natural gas. Feigley and Jeffries added methane and hexane to these 7 tracers and concluded that 93 % of the total NMHC in the air came from emissions related to vehicles. The Australian researchers expanded the list of tracer compounds to include hexane, toluene, ethylbenzene and the xylene isomers in an attempt to distinguish the area solvent contributions.

We have evaluated source contributions to the total non-methane hydrocarbon concentration (NMHC) in Tokyo, Japan (Wadden *et al.*, 1986), Newark and Linden, New Jersey (Scheff and Klevs, 1987), and Chicago, Illinois (O'Shea and Scheff, 1988) and to the total non-methane organic concentration (NMOC) in Chicago, Illinois (Aronian *et al.*, 1988; Hegberg *et al.*, 1989). The Tokyo evaluation used 17 hydrocarbons to estimate ambient contributions to 192 samples collected aloft from four sources (vehicles, gas vapor, petroleum refineries and paint solvents). The study in New Jersey used 24 hydrocarbons to quantify contributions from five sources (vehicles, gas vapor, petroleum refineries, paint solvents and petrochemical industries). The most recent Chicago evaluations used a group of 23 organics and NO_x to estimate contributions from eight source categories (vehicles, gas vapor, petroleum refineries, printing solvents, architectural coatings, vapor degreasing, wastewater treatment and drycleaning). These studies demonstrate the practical application of the CMB method in quantifying source contributions to ambient measurements. Details on the development of fingerprints for VOC modeling applications and descriptions of the Tokyo and Chicago CMB studies are presented in the following sections on Source Fingerprints and Applications.

A number of unique aspects of CMB modeling of VOCs are worth noting. For our studies in Tokyo, New Jersey and Chicago, the tracer compounds comprise the majority of the mass of the categorical pollutant (NMHC or NMOC) modeled. This is in contrast to studies of particulate matter where the elemental tracers used in the CMB calculations comprise a small fraction of the total mass evaluated (e.g. see Miller *et al.*, 1972; Kowalczyk *et al.*, 1982; Scheff *et al.*, 1984; and Dzubay *et al.*, 1988).

The modes of generation of VOC's are also significantly different from sources of particulate matter. While the elemental signatures of particulate matter sources are directly related to raw material composition (and this is frequently related to the earth's crust), the composition of VOC sources is based on the physical and chemical processes that consume or modify the raw material. For example, all petroleum refineries process hydrocarbons with similar unit operations, the emissions of which are not strongly a function of crude-oil composition. As a result, we have observed remarkable similarity in the composition of emissions from this source from studies across the U.S. and Japan (Scheff *et al.*, 1989a).

As with elemental source compositions, collinearities between specific VOC categories cause statistical problems and limit source resolution by the CMB method. Our experience to date suggests that the sources we have examined fall into three major groups with respect to collinearity: 1) vehicles, gas vapor and refineries (all major sources of the C2 to C6 hydrocarbons); 2) solvent and coating sources including architectural coatings, printing inks, and auto painting (all major sources of toluene and other aromatic compounds); and 3) sources of chlorinated organics including vapor degreasing, drycleaning and wastewater treatment (Aronian *et al.*, 1988; Hegberg *et al.*, 1989). The approaches to handling the collinearity problem are the same as used for particulate matter models. Singular value decomposition and associated methods of regression diagnostics can be used to evaluate the causes of collinearity (DeCesar *et al.*, 1986; Belsley *et al.*, 1980; and Cheng and Hopke, 1989). Sources should only be included in the CMB calculation for a specific sample if they are known to contribute. For example, vehicle exhaust and gas vapor is always present in urban air and these sources will always need to be included in the model. In contrast, petroleum refineries (if present in the air shed) will only contribute when they are up-wind of the receptor. Wind direction or trajectory information can be used to determine if this source should be included in the model (O'Shea and Scheff, 1988). Collinear categories can also be estimated as a lumped category as is done with the USEPA CMB software (USEPA, 1987).

In general, uncertainties in emissions inventories are greater for VOCs than the criteria pollutants (e.g. particulate matter, CO, SO_2, and NO_x). These uncertainties have proved to be a difficult obstacle to accurate photochemical modeling. There is, therefore, a clear need for a source reconciliation technique to evaluate and validate VOC emission inventories developed by traditional survey methods. The CMB model can help to fill this need. In addition to accurate VOC inputs, photochemical models need information on the total reactivity of the VOC source inventory. The source fingerprints combined with CMB source calculations provides the first step in developing source/reactivity contributions. And by extending the fingerprints to include highly reactive compounds (materials that may be too reactive to be used as tracer species), source contributions to total reactivity can also be calculated.

The reactivity of the organic tracer fitting compounds may also affect the model. One underlying assumption of the CMB receptor model is the conservation of mass of the tracer species relative to the total emission. While this is not completely true for any categorical pollutant, the conservation assumption poses special problems when the tracers are all reactive hydrocarbons (in varying degrees). This problem, along with discussions of emission inventory, validation, modeling un-identified sources and the use of extended fingerprints are all examined in the Evaluation section of the chapter.

2. SOURCE FINGERPRINTS

Source fingerprints are an essential input for CMB receptor model calculations. We have found that ambient measurements, downwind plume characterization studies, data from source tests, product usage information, as well as product composition information are all useful for the development of fingerprints for sources of VOC. Examples showing how these types of information are evaluated is presented in this section.

We have identified a set of 23 compounds that can be used as fitting compounds for CMB mass balance calculations. These compounds are listed in Tables 1 through 4 and were selected for a variety of reasons including: 1) they are ubiquitous and, because they are usually above minimum detectable levels in urban environments, relatively easy to measure; 2) a number have been identified as toxic organics (benzene, ethylbenzene, the xylenes, 1,1,1-trichloroethane, trichloroethylene, tetrachloroethylene, carbon tetrachloride, and chloroform); 3) they make up the majority of the NMOC mass emissions from most of the sources studied; 4) the emission data for these materials are generally consistent from study to study; 5) except for propylene (which is highly reactive) and ethane, acetylene and benzene (which have low reactivities), the hydrocarbons have similar hydroxyl radical reaction rate coefficients (k_{OH}) (all within an order of magnitude of each other); and 6) the non-chlorinated hydrocarbon fitting compounds have been applied in CMB modeling studies with reasonable success in Japan (Wadden *et al.*, 1986), New Jersey (Scheff and Klevs, 1987) and Chicago (O'Shea and Scheff, 1988; Aronian *et al.*, 1988).

Table 1. Vehicle Emissions Source Fingerprint (Scheff *et al.*, 1989a).

	Dynamometer[a]		Roadway[b]		Tunnel[c]		Japan[d]	VOC Species Manual[e]
Compound	(g/mile)	(wt%)	($\mu g/m^3$)	(wt%)	(ppbC)[f]	(wt%)	(wt%)	(wt%)
Ethane	0.0272	1.68	9.24	1.77			1.61	1.57
Ethylene	0.1592	9.84	36.72	7.03	408.7	9.39	6.22	4.37
Propane	0.0589	3.64	4.41	0.84			3.16	
Propylene			13.12	2.51	122.4	2.80	2.91	1.62
Acetylene	0.0683	4.22	29.88	5.72	160.7	3.43	2.85	0.83
i-Butane	0.0104	0.64			76.4	1.81	2.38	1.89
n-Butane	0.0790	4.88	16.42	3.14	198.1	4.69	7.77	10.57
i-Pentane	0.0628	3.88	31.55	6.04	305.8	7.18	5.39	7.28
n-Pentane	0.0282	1.74	15.11	2.89	145.9	3.44	3.11	3.21
2-Methylpentane	0.0253	1.56					1.97	1.95
3-Methylpentane	0.0165	1.02			73.8	1.73	1.19	
n-Hexane	0.0120	0.74			74.8	1.76	2.28	0.97
2,4-Dimethylpentane	0.0101	0.62			68.9	1.61		0.64
Benzene	0.0580	3.58	15.49	2.96	202.2	4.29	2.90	3.11
Toluene	0.1238	7.65	31.76	6.08	303.4	6.50	5.44	5.72
Ethylbenzene	0.0176	1.09	6.43	1.23			0.83	0.85
p-Xylene/m-Xylene	0.0568	3.51	21.65	4.14	162.7	3.53	1.24	2.84
o-Xylene	0.0343	2.12	6.88	1.32	74.3	1.61	0.57	1.73
Sub-total	0.8484	52.43	238.67	45.67	2378.1	53.77	51.81	50.54
Total paraffins	0.6910	42.70	214.25	41.02	1678.1	39.6		
Total olefins	0.3686	22.78	95.37	18.26	940.3	21.6		
Total aromatics	0.4867	30.08	146.30	28.01	1516.4	32.8		
Unknown	0.0719	4.44	36.93	7.07				23.30
Compounds			29.46	5.64	161.0	6.0		11.15
Unexplained	1.6182	100.00	522.30	100.00	4295.8	100.0		100.00
Total NMOC	0.0425	2.6264	9.66	1.85				
Total Aldehydes	16.28	1006.06	4211.0	806.7	15600.	715.2		
Carbon Monoxide	2.18	134.72	343.0	65.7	2416.	118.6		
Oxides of Nitrogen								

[a] Based on average emissions from FTP driving cycle test for 46 in-service vehicles (Sigsby *et al.*, 1987).

[b] From upwind-downwind sampling (Zweidinger *et al.*, 1988).

[c] From NMOC measurements in the Lincoln Tunnel (Lonneman and Sella, 1986).

[d] From ambient monitoring in Kanagawa Prefecture (Wadden *et al.*, 1986; Wakamatsu *et al.*, 1984). Wt % of the 17 measured components adjusted to total % for same components in dynamometer average.

[e] Composite (USEPA, 1988).

[f] parts per billion carbon for hydrocarbons, parts per billion for CO and NO$_x$.

Table 2. Vapor Degreasing and Drycleaning Source Fingerprints (Scheff *et al.*, 1989a).

	Vapor degreasing[a]		Drycleaning[b]
Compound	(10^3tons)	(wt%)	(Mg/yr)
Chloroform		0.00	0.00
1,1,1-Trichloroethane	113.6	49.65	0.00
Carbon Tetrachloride		0.00	0.00
Trichloroethylene	68.8	30.07	0.00
Perchloroethylene	23.9	10.45	55,000
Methylene Chloride	22.5	9.83	0.00
Sub-total	228.8	100.00	63. %
Other C8 to C12 Hydrocarbons		0.00	31,700
Total		100.00	86,700

[a] Derived from 1985 total usage data of 228,800 metric tons of chlorinated solvent used for vapor degreasing (Storck, 1987).

[b] 55,000 Mg/yr of perchloroethylene used for drycleaning in 1984 (estimated) (USEPA, 1980). 31,000 Mg/yr of petroleum solvents used for drycleaning (USEPA, 1982). The solvents are similar to kerosene and are primarily used by large drycleaning operations for industrial, commercial, and institutional customers.

Tables 1 through 4 summarize selected information on source fingerprints. Included in the tables are data expressed in the original units of measurement and as a weight % of the total fitting compounds or total non-methane organic compounds (NMOC). An entry of zero represents a quantitative estimate of source composition for a specific compound. In contrast, a blank entry represents the case when the compound was not measured and the actual value in unknown. For modeling purposes, unknown values should be taken as zero.

Table 1 shows a summary of the most recent and complete data sets on the hydrocarbon emissions from motor vehicles. The table lists data collected using a wide variety of methods. For example, the dynamometer data represent the average tail-pipe emissions (source test data) from the FTP driving cycle test for 46 in-service vehicles collected under laboratory conditions (note that the vehicles

Table 3. Architectural Coatings Source Fingerprint (Scheff *et al*, 1989a).

Compounds	Solvent Coating[a]	Thinning Solvents[b]	Water Coating[c]	Coating Composite[d]
Percentage of VOC emissions contributing to total[d]	64%	11%	25%	100%
2-Methylpentane				
3-Methylpentane				
n-Hexane				
2,4-Dimethylpentane	1.70			1.09
Benzene			0.36	0.09
Toluene	37.86	15.21		25.90
Ethylbenzene	0.54	0.93		0.45
p-Xylene/m-Xylene	3.70	2.72		2.67
o-Xylene	4.47			2.86
Sub-total	48.27	18.86	0.36	33.06
n-Heptane	2.94	3.07		
n-Butylalcohol			20.09	
Acetone	1.27			
Methyl-ethyl-ketone	0.54	4.06		
Methyl-isobutyl-ketone	0.36			
Other Paraffins	22.28	39.21	1.33	
Other Olefins	0.15	9.44		
Other Aromatics	0.31			
Other Alcohols		4.91	4.03	
Other Ketones	4.58	1.99	1.04	
Other Esters	19.28	9.27	56.03	
Glycols			2.01	
Chlorinated Organic		1.22	9.52	
Sub-total	51.71	73.17	94.05	
Other Compounds	0.02	7.97	5.59	
Total NMOC	100.00	100.00	100.00	

[a] Composite of profiles for lacquer, primer and enamel in proportion to usage in southern California (USEPA, 1988).

[b] Composite based on sales volume from nine solvents used with architectural coatings (USEPA, 1988).

[c] Composite of seven coatings in proportion to 1980 California sales figures (USEPA, 1988).

[d] Percentage of VOC emissions from 1984 survey of New York major metropolitan area and entire state of New Jersey (Leone *et al.*, 1987). Composite calculated by applying these percentages to compositions in each category and summing contributions from all categories.

Table 4. Source Profiles Normalized to Fitting Compounds (wt %)[a].

Fitting Compounds	Vehicle	Gas Vapor		Refinery	Arch. Coatings	Waste-water	Graphic Arts	De-greasing	Dry Cleaning	Auto Paint	Petro-Chem.
		Winter	Summer								
Ethane	3.12	0.00	0.00	4.81	0.00	0.00	0.00	0.00	0.00		3.0
Ethylene	18.18	0.00	0.00	0.73	0.00	0.00	0.00	0.00	0.00		50.0
Propane	6.72	1.66	0.39	21.29	0.00	0.00	0.00	0.00	0.00		4.3
Propylene	3.16	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.00		3.9
Acetylene	7.77	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00		1.8
i-Butane	1.17	18.46	13.39	4.73	0.00	0.00	0.00	0.00	0.00		2.6
n-Butane	9.02	41.68	30.24	17.60	0.00	0.00	0.00	0.00	0.00		5.5
i-Pentane	7.19	20.22	31.43	16.85	0.00	0.00	0.00	0.00	0.00		4.0
n-Pentane	3.20	8.51	13.21	7.32	0.00	0.00	0.00	0.00	0.00		2.8
2-Methylpentane	2.89	2.89	4.95	7.22				0.00	0.00	0.99	1.5
3-Methylpentane	1.87	1.48	2.54	4.32				0.00	0.00	0.95	1.1
n-Hexane	1.37	1.15	1.97	3.60				0.00	0.00	1.77	4.3
2,4-Dimethylpentane	1.15	0.36	0.31	1.66	3.30			0.00	0.00		
Benzene	6.64	1.28	0.54	1.39	0.27	2.33		0.00	0.00		3.4
Toluene	14.16	1.32	0.93	4.68	78.34	6.17	93.09	0.00	0.00	67.42	7.7
Ethylbenzene	2.01	0.36	0.04	0.58	1.36	1.26		0.00	0.00	6.24	1.6
p-Xylene/m-Xylene	6.64	0.37	0.04	1.44	8.08		6.91	0.00	0.00	17.10	1.4
o-Xylene	3.90	0.19	0.04	0.84	8.65			0.00	0.00	5.54	1.1
Chloroform						12.98		0.00	0.00		
1,1,1-Trichloroethane						28.43		55.06	0.00		
Carbon Tetrachloride						12.79		0.00	0.00		
Trichloroethylene						16.35		33.35	0.00		
Perchloroethylene						19.69		11.59	100.00		
Total fitting compounds	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
THC[b] as % of NMOC	52.43	78.22	91.26	80.47	33.06	62.50	11.90	90.17	63.00		
NO_x (wt % of THC)	257.29	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00		

[a] From Scheff *et al.*, 1989a; Hegberg *et al.*, 1989.

[b] THC = sum of 23 organic fitting compounds.

were tested without any modifications or engine tuning) (Sigsby *et al.*, 1987); the roadway data represent ambient measurements collected down-wind and up-wind of a busy intersection (Zweidinger *et al.*, 1988); the tunnel data are derived from NMOC measurements in the Lincoln Tunnel (Lonneman and Sella, 1986); the Japanese fingerprint is based on ambient roadside monitoring over a 36 month period in Kanagawa Prefecture (Wadden *et al.*, 1986; Wakamatsu *et al.*, 1984); and the VOC Species Manual values are a composite from the recent update of that document (USEPA, 1988). Despite the different sampling methodologies, the data in Table 1 are remarkably consistent. For example, the weight fractions for benzene range from 2.90 to 4.29 %, from 1.57 to 1.77 % for ethane, and from 45.67 to 53.77 % for the fitting compound fraction of the total NMOC . Given the close agreement between the laboratory dynamometer data and the ambient roadway, tunnel and Japanese data, we believe that the dynamometer data are most representative of the actual motor vehicle tail-pipe contributions and suggest that these be used as a fingerprint for CMB modeling. Also included in Table 1 are data on total paraffins, olefins, aromatics, aldehydes, carbon monoxide, and oxides of nitrogen. NO_x (i.e. NO + NO_2) can also be used as a fitting compound.

The architectural coatings fingerprint (Table 2) is based on an extensive survey of product-type consumption in the New York City-State of New Jersey region (Leone *et al.*, 1987). The consumption data are shown as the % of VOC emissions contributing to the total. These data are in excellent agreement with previously reported nation-wide per capita NMOC emissions from architectural coatings (USEPA, 1985). Profiles for the three product types shown are composites based on sales in California (derived from product composition information). For example, the solvent-based coatings include contributions from lacquer, primer and enamel in proportion to use in Southern California, the thinning and cleanup solvent fingerprint is based on sales volume from 9 solvents used with architectural coatings, and the water-based coating is a combination of 7 coatings in proportion to 1980 California sales figures (USEPA, 1988). The composite is then calculated by applying these percentages to compositions in each category and summing the contributions from all categories. Note that the fitting compounds only represent 33.06 % of the total emission for this source. Non-fitting compounds make up most of the NMOC emissions from this category and include alcohols, ketones, esters, glycols, and other chlorinated organics.

The fingerprint for vapor degreasing (Table 3) is based on a 1985 nation-wide chlorinated solvent consumption estimate of 228,800 metric tons for this application; and the individual fractions attributed to trichloroethylene, 1,1,1-trichloroethane, perchloroethylene and methylene chloride (product usage information from the Halogenated Solvents Industry Alliance) (Storck, 1987). The total vapor degreasing usage for these four solvents was about the same as that reported from a 1974 survey (USEPA, 1979), although the more recent data reflect significant reductions in trichloroethylene and perchloroethylene, and corresponding increases in 1,1,1-trichloroethane. Note that this fingerprint is only representative of vapor degreasing since no contributions were included from maintenance and manufacturing cold cleaning.

The data for drycleaners (Table 3) is based on total solvent usage by perchloroethylene cleaners (55,000 Mg/yr estimated for 1984) (USEPA, 1980a) and petroleum based cleaners (31,000 Mg/yr) (USEPA, 1982). The petroleum solvents used are composed of C8 to C12 hydrocarbons and do not significantly contribute to the fitting compounds.

Table 4 is a summary of 10 source profiles normalized to 100 % of the fitting compounds (Scheff *et al.*, 1989a). Included in the table is the fitting compounds as a percent of the total NMOC emission. The vehicle emission fingerprint is based on the dynamometer test from Table 1, gasoline vapor from vapor-space analyses of unleaded winter and summer blends, refinery emissions from a plume study downwind of a large modern petroleum refinery, architectural coatings and printing inks are composite fingerprints combining product consumption and composition information, wastewater processing is from the average of two ambient measurement and two mass transfer modeling studies, auto painting and petrochemical from downwind ambient measurement studies, and vapor degreasing and drycleaning from solvent usage data. When used with the CMB model, the profiles in Table 4 generate source contributions to ambient concentrations of the 23 fitting compounds. Leaving the profiles normalized to the total NMOC emissions would produce estimates of source contributions to the total of volatile organics.

The consistency in the data for vehicles, gasoline vapor and petroleum refineries suggests that the profiles for these sources can be applied to a variety of settings without modification (Scheff *et al.*, 1989a). This is in contrast to our experience with profiles for sources of particulate matter (Scheff *et al.*, 1984). The reason for the difference is in the physical basis of the profile. For the particulate matter case, profiles (typically based on elemental composition) are a function of the composition of the raw materials used in the process. For example, the profile for a coal-fired power plant is largely controlled by the elemental composition of the coal burned and the profile of a soil dust is based on an area's geology. In contrast, the profiles for sources of NMOC are controlled by the physical and chemical processes that consume or modify the raw material. For example, all petroleum refineries process hydrocarbons with similar unit operations, the emissions of which are not strongly a function of crude-oil composition. It is, therefore, not surprising that different profiles which represent a wide variety of refineries around the world are all so similar.

3. APPLICATIONS

3.1. Tokyo, Japan

A data set consisting of 192 samples of ambient air, speciated for 17 hydrocarbons, was collected over Tokyo, July 16-17, 1981. Weighted least-squares fitting of the hydrocarbons was used to estimate ambient contributions from four source categories: petroleum refineries, vehicle exhaust, gasoline storage, and paint solvents (Wadden *et al.*, 1986). The samples were obtained as part of a series of summertime aircraft surveys of photochemical pollutants carried out by the National Institute for Environmental Studies (NIES) to develop a better understanding of photochemical smog formation over Tokyo (Wakamatsu *et al.*, 1983; Uno *et al.*, 1984; Uno *et al.*, 1985a).

Six runs (flights of 1-1½ h), using two instrumented aircraft, were spaced throughout each day starting with sun-up and continuing through about 9:30 p.m. The number of samples/run varied between 11 and 23. In general, most runs followed nearly the same flight pattern, and consequently run-averaged compositions constituted a chemical "snapshot" of the air over the Tokyo Metropolitan Area. Most of the observation data were collected between altitudes of 350 and 700 m although occasional samples were collected up to 1500 m. The study-average O_3 concentration aloft over all runs was 118 ppb, and the maximum run-averaged value was 188 ppb. There were numerous hourly O_3 excursions above 120 ppb at ground level stations in the study area during the two days of monitoring.

The vertical wind profile was also monitored at about 20 sites by using pilot-balloon observations. Wind data were collected at each 100-m interval up to 3000 m. These were subsequently integrated to determine the areal wind profile for subsequent trajectory estimation. Back-trajectories at an altitude of 350 m are shown for three of the runs in Figures 1 to 3. For both days, trajectories entering the sampling area in the forenoon came from the northeast and passed over Tokyo Bay. In addition on both days, parcels entering the sampling area later in the day passed over Tokyo Bay in early morning but then were advected over Sagami Bay for a number of hours before being returned by a southwest ocean breeze. The 13:00 trajectory in Figure 2 shows the beginning of this pattern which became more pronounced through the afternoon and evening (i.e., more hours over water as displayed in Figure 3).

Hydrocarbons were analyzed with a 99% confidence interval corresponding to a precision of ± 5% of the mean value of repeated analyses for C_2-C_5 and ± 10% of the mean for C_6 and heavier. These values corresponded to $\sigma/\overline{x}$ of 0.019 and 0.039 for C_2-C_5 and C_6^+, respectively, which were used to determine the weighting function in the least-squares fit. Other details of the chemical and data-reduction procedures are given in Wadden *et al.* (1986).

Coefficients for each of the four source categories were determined for each sample using the weighted least squares technique and 17 of the 18

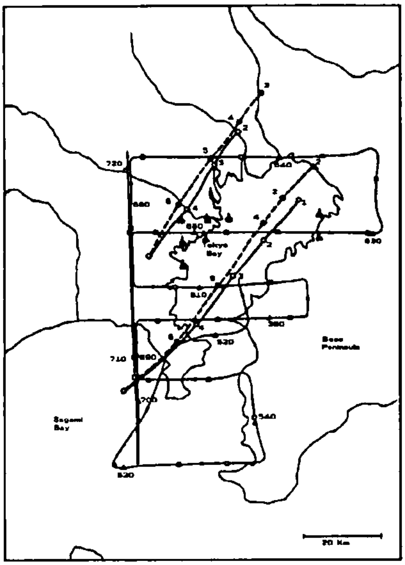

Figure 1. Wind trajectories and flight path for run 11, 5:29-7:16 JST, July 16, 1981 (Wadden *et al.*, 1986).

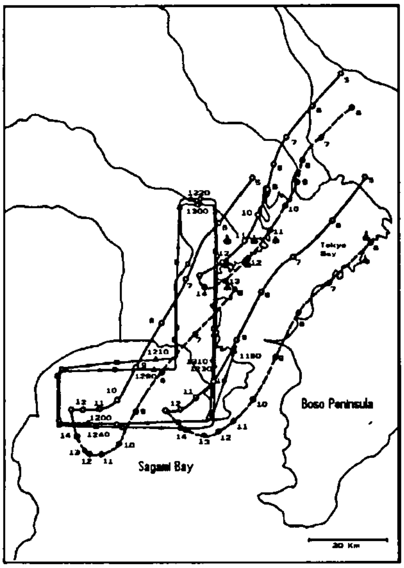

Figure 2. Wind trajectories and flight path for run 13, 11:55-13:08 JST, July 16, 1981 (Wadden *et al.*, 1986).

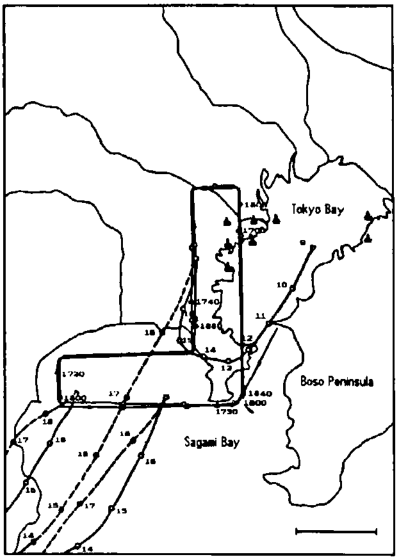

Figure 3. Wind trajectories and flight path for run 15, 17:21-18:31 JST, July 16, 1981 (Wadden *et al.*, 1986).

hydrocarbons listed in Tables 1 and 4 (excluding 2,4-dimethylpentane). The source fingerprints for refineries, gas vapor, and vehicles were determined from monitoring data in the Tokyo area. The vehicle exhaust composition is given in Table 1, and the refinery and gas vapor fingerprints were quite similar to those listed in Table 4. The paint solvents fingerprint consisted of 25.7% toluene, 32.5% ethylbenzene, 30.3% p,m-xylene and 11.5% o-xylene. This composition is a weighted value based on the usage amounts for all of Japan and the specific composition of each of the following categories: varnish; enamel; and, "veneer, epoxy, urethane, polyester, anticorrosive and other paints."

The average coefficients are given for each run and for all 192 samples in Table 5. The unexplained source contribution is determined by difference between the total hydrocarbon concentration for the 17 components and the amounts allocated by the model to the other four sources. This category contains contributions from uncharacterized sources, such as printing solvents, and experimental error in source and receptor compositions. While the other

Table 5. Mean Source Coefficients for Tokyo Application (Wadden *et al.*, 1986).

		Coefficients, $\mu g/m^3$					
Flight Run Period	Number of samples	Vehicle Exhaust	Gasoline Vapor	Petroleum Refinery	Paint Solvent	Unexplained	Total Hydrocarbon
July 16, 1981							
11 5:29- 7:16	23	1.9	0.6	2.7	14.1	20.4	39.7
12 8:52-10:37	18	2.6	4.7	16.4	22.2	31.8	77.7
13 11:55-13.08	19	7.1	12.9	31.7	34.5	67.4	153.6
14 14:35-15:43	12	4.4	12.6	20.7	31.9	31.9	101.5
15 17:21-18:31	11	2.0	5.6	17.9	12.5	32.0	70.0
16 20:10-21:27	12	0.8	4.2	9.6	25.8	5.5	45.9
July 17, 1981							
21 5:20- 6:53	23	8.8	5.2	29.8	26.1	53.6	123.5
22 8:24- 9:55	20	4.6	4.7	18.1	15.5	33.7	76.6
23 11:31-12:41	20	7.2	20.2	41.5	31.3	62.6	162.8
24 14:20-15:31	12	4.3	9.5	28.4	21.5	23.5	87.2
25 17:18-18:18	11	1.1	2.2	16.3	7.5	20.1	47.2
26 20:08-21:07	11	1.4	2.4	11.5	7.0	29.7	52.0
Overall	192	4.0	6.4	20.3	23.3	37.3	91.3

coefficients are constrained to be ≥ 0, this restriction was not placed on the unexplained contribution. However, as a measure of the consistency of the data and the modeling approach, negative contributions for this category were calculated for fewer than 10% of all the sample compositions.

The vehicle exhaust coefficient varies from 2 to 7% (overall 4%) and refinery sources contributed from 7 to 35% of the total hydrocarbons (overall 22%). On the basis of an emission inventory for the Tokyo Metropolitan area (consisting of Tokyo and the six surrounding prefectures), vehicle emissions would be expected to contribute about 27% and refineries 5% of the 17 fitting hydrocarbon compounds. The emission inventory reflects an evaluation of annual use and emission data, but is not necessarily representative of short-term emission patterns. This apparent inconsistency between inventory and source-reconciliation values can be at least partly resolved when wind trajectory and traffic frequency are included in the evaluation.

From the trajectory analysis it became clear that most of the air sampled during any run had advected over the Tokyo Bay industrial areas in early morning but during its daytime history had not passed over other land areas with substantial hydrocarbon emissions. As indicated in Figures 1-3, the major refinery sources are located on the east and west shores of Tokyo Bay. This trajectory-source interaction suggests the reason for the relatively high refinery source coefficients. In addition, the vehicle use pattern in early morning is significantly different from the 24-hr average (which is reflected in the emission inventory). The period from 21:00 to 7:00 only contributes about 20% of the total vehicular traffic. So if the air sampled during each run actually represents only early morning conditions (because of the changeover from land breeze to ocean breeze), a reduced vehicular contribution would be expected. The average paint solvent contribution is less likely to be strongly affected by wind direction, and the average coefficient of 26% was in reasonable agreement with the inventory usage value of 32%. The gasoline vapor contribution for major urban areas is also not usually dependent on direction and the average contribution of 7% compares well with the annual emission value of 4%. In addition since maximum ground-level temperatures on both days were above 30° C, the gasoline vaporization from storage would be expected to be greater than the average.

3.2 Chicago, Illinois

We have tested the CMB model for ambient measurements of volatile organic compounds (VOC) in Chicago (Aronian *et al.*, 1988; Hegberg *et al.*, 1989). The study included a four-month winter sampling program and a three-month summer sampling program at three sites in the Chicago Metropolitan Area; a suburban background location approximately 50 miles north of downtown Chicago (SUB), an inner-city urban site at the University of Illinois at Chicago located one mile west of downtown Chicago (URB), and an industrial location on Chicago's southeast side (IND). The sites are roughly located on a north-south line and are 21 km and 55 km apart, from IND to URB to SUB, respectively. Four-hour samples (8:00 am to 12:00 pm) were collected simultaneously at all three sites. A twenty-hour sample (12:00 pm to 8:00 am) was also collected at the central city location for each sampling session. This sampling strategy, therefore, allowed for the evaluation of the impact of major point sources of Non-Methane Organic Compounds (NMOC) at three widely varying receptor locations as well as the evaluation of the long-term average impact at the central location. For example, since all four refineries in the Chicago area are located in the southern portion of the study area, each receptor location will view the refineries with a different emission-to-collection transport time when the wind is from the south.

A total of twenty-six winter-time, and 55 summer-time whole air samples were collected at the three sites and analyzed for the compounds listed in Table 4. In addition to the 23 organics, NO_x was measured at the URB receptor location during the winter, and at all three locations during the summer monitoring program and included as a fitting compound. The average temperatures ranged between -15.0°C and 4.4°C for the winter, and between 16.8°C and 28.8°C for the summer samples. Each ambient sample was evaluated for the contribution from eight sources of VOCs in the Chicago region; vehicles, gasoline vapor, petroleum refineries, architectural coatings, graphic arts, vapor degreasing, drycleaning, and wastewater treatment. These sources were selected because they are known to emit significant quantities of NMOC in the Chicago area and are sources for which reasonable fingerprints could be developed (IEPA, 1987; Scheff *et al.*, 1989). Details of the environmental sampling and analytical procedures have been presented (Aronian *et al.*, 1988).

Weighted least squares solutions of the CMB model were calculated for each of the 81 ambient samples. This procedure weights the regression analysis by the ambient measurement error. The calculations were performed on a personal computer using the SYSTAT statistical package (SYSTAT Inc., Evanston, IL). The standard deviations of the measurements (measurement error for the weighted least squares calculations) for the alkanes and alkenes were developed from an analysis of ten duplicate samples. The measurement error was calculated as the maximum difference between the replicate analyses divided by the student's t with 9 degrees of freedom at the 95 % confidence level. (Note that this calculation assumes that the maximum difference between the 10 replicate analyses represents the 95 % confidence interval.) The measurement errors for the aromatic and chlorinated compounds were calculated as the standard deviation of repeated analyses of a laboratory calibration standard.

The source profiles used in the CMB model are shown in Table 4. The profiles are normalized to 100 % of the 23 organics studied. Therefore, the predicted source coefficients represent contributions at the receptor sites for the specific 23 fitting compounds studied. Note that the profile for gasoline vapor reflects the change in fuel volatility for the summer and winter periods (USEPA, 1980b; USEPA, 1988). Both profiles for gasoline vapor are based only on unleaded fuels which comprise more than 80 % of gasoline usage in the Chicago area.

CMB source coefficients for each complete sample were developed. The 81 sets of coefficients averaged by receptor location, sample duration and season are shown in Table 6. The coefficients in this table were calculated from source profiles normalized to the sum of the 23 organic fitting compounds. The

Table 6. Mean Source Coefficients for Chicago Application (Hegberg *et al.*, 1989).

				Coefficients, µg/m^3								
Location	Duration (hrs)	No. samples	VEH	GV	REF	AC	GA	WAS	VDG	DCL	Other	THC
WINTER												
Industrial	4	5	45.34	20.09	33.28	2.47	2.00	3.23	3.39	0.74	46.6	157.1
Urban	4	7	70.25	4.79	27.76	3.54	4.78	4.76	4.78	1.05	32.9	153.2
Urban	20	7	56.96	16.01	39.51	3.23	2.34	4.30	4.31	0.90	64.0	191.5
Suburban	4	7	43.43	10.53	10.28	0.24	0.32	1.75	2.17	0.29	43.2	112.1
SUMMER												
Industrial	4	14	44.69	22.99	16.89	2.31	1.81	8.42	8.10	0.35	120.5	226.1
Urban	4	15	37.40	12.34	11.48	2.54	1.81	6.48	5.14	0.84	101.1	179.2
Urban	20	13	44.96	5.98	14.86	1.91	1.52	6.92	6.12	0.49	105.4	188.1
Suburban	4	13	22.42	9.66	9.94	1.95	1.57	5.03	4.76	0.19	67.2	122.7

GV = Gasoline Vapor, REF = petroleum refinery, AC = architectural coatings, WAS = wastewater treatment, VEH = vehicles, GA = graphic arts, VDG = vapor degreasing, DCL = drycleaning, Other = sum of residual organic, THC = fitting compounds.

unexplained fraction shown in the table is defined as the difference between measured organics and the sum of the source coefficients. In this way, the "other" category represents contributions to the organics measured from all sources which were not included in the model. Note that the other category was positive for 24 of 26 winter samples, and positive for all 55 summer samples.

Regression diagnostics of the CMB model with all eight sources in the equation identified two problems with collinearity. For a typical run, two of the eight eigenvectors had condition indices greater than 10. (Condition index is a measure of collinearity and a value greater than 10 is an indication of a major collinearity (Belsley *et al.*, 1980)). An examination of the source profile matrix shows the cause of the collinearities. The largest condition index (representing the smallest eigenvalue) is associated with large variance proportions for architectural coatings and graphic arts. This collinearity can be seen in Table 4 as these two sources are primarily composed of toluene. Because of this collinearity, it is generally not possible to simultaneously solve for both of the toluene sources. This problem was handled by solving the CMB model with either architectural coatings or graphic arts in the model and averaging the two solutions. Although the results for these sources are shown separately in Table 6, we interpret the sum of the two as the combined impact of printing and architectural coating solvents. The exact split between the two sources is less certain.

The second highest condition index is associated with a very large variance proportions for wastewater treatment, vapor degreasing and drycleaning. This collinearity can be seen in Table 4 as these are the three sources of chlorinated organics. As was the case for the solvent sources, the sources of chlorinated organic were estimated separately and the three solutions averaged to give the result in Table 6; and the sum of the three sources represents their combined impact. The model was usually able to resolve the vehicle, gas vapor, and refinery sources without problems of negative coefficients.

As shown in Table 6, contributions from vehicles and gasoline vapor were approximately equal for summer and winter samples at the industrial and urban sites. In contrast, the contribution from vehicles at the suburban site was lower in the summer than winter. Two of the sources of chlorinated organic, vapor degreasing and wastewater treatment were higher during the summer at all sites suggesting increased releases for chlorinated solvents during warm conditions. Also note that the mass concentration of unexplained hydrocarbon is higher in the summer compared to winter results.

The average refinery source coefficients are highest for the industrial, and lowest for the suburban receptor during both winter and summer sampling periods. This is also consistent with source-receptor geometry as all of the region's refinery point sources are located in the southern suburbs closest to the industrial site and furthest from the study's suburban site which is located 55 km north of the city. Table 6 further shows that the suburban location had the smallest impacts from architectural coatings, wastewater treatment, graphic arts solvents, vapor degreasing and drycleaning. Since most of the major printing, industrial, and wastewater sources are located far south of the site (and much closer to the inner-city and industrial sites), lower values for these sources are expected.

4. EVALUATION

4.1 Estimation of Emission Inventories from CMB Predictions

One of the most promising applications of the CMB model for ambient organic concentration data is the evaluation of the VOC emission inventory. A comparison of CMB source allocations to emission inventory data for Chicago is shown in Table 7 (IEPA, 1987). The emission inventory data in the table is expressed as kg day^{-1} and weight percent of the total VOC emission and is representative of a typical summer day. The inventory includes 61 VOC emission source types representing 800,509 kg day^{-1}. Note that the inventory does not include specific halogenated compounds (including 1,1,1-trichloroethane and freon) and this will affect the comparison for degreasers as this source contains a significant mass fraction of 1,1,1-trichloroethane. The CMB data is taken from the Chicago application described in the previous section.

Table 7. Comparison of Emission Inventory Data to CMB Source Coefficients (Hegberg *et al.*, 1989).

Source	Emission Inventory		CMB Winter			CMB Summer		
			THC[a]		NMOC[b]	THC[a]		NMOC[b]
	(kg day^{-1})	(%)	($\mu g\ m^{-3}$)	(%)	(%)	($\mu g\ m^{-3}$)	(%)	(%)
Vehicles	286,155	33.5	54.0	35.2	34.6	37.4	20.9	33.8
Gasoline Vapor	61,179	7.6	12.8	8.4	5.4	12.7	7.1	6.4
Petroleum Refineries	10,749	1.3	27.7	18.0	11.3	13.3	7.4	7.6
Architectural Coatings	44,230	5.5	2.4	1.5	2.3	2.2	1.2	3.1
Graphic Arts	78,268	9.8	2.0	1.3	5.6	1.7	1.0	6.6
Vapor Degreasing	25,078	3.1	3.7	2.4	1.3	6.0	3.4	3.1
Dry Cleaning	758	0.1	0.7	0.5	0.4	0.5	0.3	0.4
Other	312,092	39.0	50.2	32.7		105.2	58.7	
Total Hydrocarbons	800,509	100.0	153.5	100.0		179.0	100.0	

[a] Average of the CMB coefficients from the four combinations of site and sampling duration normalized to the 23 fitting compounds.

[b] Weight %, determined from the average CMB coefficients, normalized to NMOC. The sum of the seven source categories is assumed = 61 %, which is the sum of the contributions from these categories to the emission inventory.

Generally, the emission inventory results are in reasonable agreement with the average winter and summer CMB coefficients. When the CMB coefficients are normalized to the fitting compounds, the inventory estimates of vehicle emission, gasoline vapor and vapor degreasers are very close to the CMB estimates. In contrast, the average graphic arts contribution of 1.3 % and 1.0 % (winter and summer, respectively) is much smaller than predicted by the emission inventory (9.8%). Note, however, that because the CMB estimate for graphic arts is based on a small fraction of the total emission (11.9 %) it is under-represented in the results normalized to the fitting compounds. This was also true for architectural coatings. Normalizing the CMB coefficients in terms of the total NMOC emissions by using extended fingerprints, brings these predictions much closer together for both winter and summer average coefficients (Scheff *et al.*, 1989a).

Table 7 also shows the average CMB coefficients normalized to the total NMOC. The NMOC coefficients were calculated by dividing the average CMB coefficients by the THC as a fraction of NMOC (see Table 4). Since we do not have a % of NMOC factor for the "other" sources category, the coefficients are expressed as a % of NMOC such that the emission inventory percent of the seven categories in Table 7 (61%) equals the sum of the normalized CMB estimates. The comparison between the emission inventory as weight % and NMOC normalized CMB estimates for vehicles, gasoline vapor, architectural coatings, graphic arts, vapor degreasing, and dry cleaning is very reasonable for both summer and winter samples. It is interesting to note that these similarities exist despite the fact that average daily temperatures for the winter-time samples varied between -15.0 and 4.4°C while those in the summer were between 16.8 and 28.8°C. This agreement, despite considerable temperature variation suggests in a general way that reactivity in the atmosphere does not greatly distort the fingerprints nor interfere with the CMB analysis.

The largest difference between the emission inventory and CMB predictions is seen with refinery contributions. We suspect this results from under-estimation of the emission inventory values. A previous study in the Chicago area indicated good agreement between calculated refinery coefficients and refinery contributions determined from wind trajectories and refinery capacity (with refinery capacity used as a surrogate for emission); but poor agreement when estimates of emissions were substituted for capacity (O'Shea and Scheff, 1988). The inventory values also do not include the contributions from Northeast Indiana refineries and other sources.

It is encouraging to note that the emission inventory and CMB model's estimate of "other" sources are very close. This suggests that the calculations do not over-predict the modeled sources nor under-predict the sources not included in the model.

4.2 Validation

Calculated source coefficients are most useful if they can be validated, or at least are consistent, with other independently observed source and meteorological characteristics. As a simple example, if a receptor location is directly downwind of a large auto paint plant, the paint solvent source coefficient should be higher than when the wind is blowing in the opposite direction. Consideration of prevailing meteorological patterns is particularly important for understanding coefficients calculated from single short-term samples, (i.e, 1-4 hours). As discussed in the previous section, comparison of emission inventory values with the average coefficients, determined from many samples, constitutes another type of validation over longer time intervals.

Two examples of short-term validation come from studies in Tokyo (Wadden *et al.*, 1986) and Chicago (O'Shea and Scheff, 1988). Both are for refinery contributions to ambient samples. In the Tokyo study (Figure 4), back-trajectories were determined for each sample from the point in the flight path at which sampling began. If the trajectory for a particular sample intersected a refinery location, the sample was given a score equal to the annual emission rate. The average refinery emission score per sample was computed for each run and compared with the petroleum refinery coefficient calculated from the run-averaged composition. Use of the run-averaging approach was appropriate because each sample consisted of air collected over 3-4 km and some imprecision was

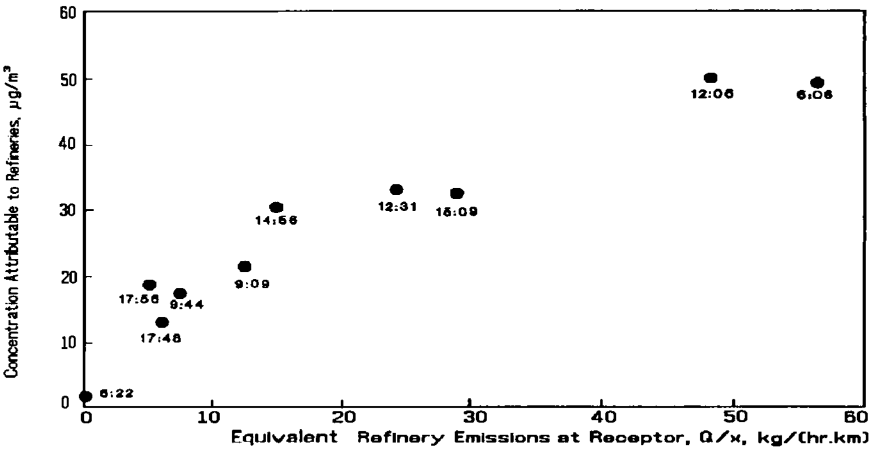

Figure 4. Refinery contributions to ambient concentrations of total hydrocarbons vs refinery emissions at the receptor point (Wadden *et al.*, 1986). (Average time of sampling flight given on plot.)

associated with the trajectories. Comparison of average refinery sample score and average refinery coefficient for the 10 runs for which trajectories could be developed gave an r^2 = 0.70.

What was not reflected in this relationship was the effect of travel time on the contribution of the petroleum refinery sources to the run-averaged composition. Usually receptor and source are separated by a fixed distance. In this case there is a fixed sampling location because the flight path was essentially repeated and most samples were collected within altitudes of 350-700 m, and the locations of major refineries were known. However, because of the trajectory pattern, some refinery emissions in early morning subsequently entered the sampling zone in late afternoon and were aloft for much longer periods than those that intersected the zone earlier in the day. Both dilution and reaction can have an effect on the amount and composition coming from a specific source which is ultimately measured. At this point we will limit our discussion to the effect of dilution and return to reactivity in a later section.

When the Gaussian plume model is used, the receptor concentration, C, due to a particular elevated source is

$$C_j = \left[\frac{Q_j}{2\pi\sigma_y\sigma_z u}\right]\left[exp\left(-\frac{y^2}{\sigma_y^2}\right)\right]\left[exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right) + exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right)\right] \tag{1}$$

where Q_j is the emission rate from a particular source, u is the wind velocity, σ_y and σ_z are the dispersion coefficients in the vertical (z) and horizontal (y) direction, and H is the source plume rise. For large downwind distances (typically 10-70 km from the trajectories) and elevated sampling ($z \approx 500$ m), the plume rise will no longer be a factor. This is particularly the case for refineries, where leaks are major emission sources. Back trajectories from each sample were based on plume centerlines (i.e, intersection with the refinery site) so $y = 0$.

For many conditions the dispersion coefficients can be approximated by $\sigma_y = Ax^{\alpha}$ and $\sigma_z = Gx^{\gamma}$, where x is the downwind distance (or travel time) and A, G, α, and γ are empirically determined. For long downwind distances, the exponential terms in Equation 1 will become ~1.0. If in addition we assume that u is relatively constant, then we find from Equation 1 that

$$C_j \propto \frac{Q_j}{x^{\alpha+\gamma}} \tag{2}$$

where $1 \leq \alpha + \gamma \leq 2$. Consequently, if $\alpha + \gamma = 1$ the concentration should be proportional to the emission rate divided by the downwind distance from source to receptor (or alternatively, the travel time). And the total concentration allocated to all sources in a particular category will be the sum of each of the emission contributions adjusted for its upwind distance.

Figure 4 shows the relationship between the total run-averaged concentration of 17 fitting components at the receptor point predicted from the CMB model to be due to petroleum refinery emissions and the corresponding value of emission rate adjusted for travel distance from refinery sources. For these points the refinery coefficient was determined from the run-averaged composition. The Q/x values were determined by drawing back-trajectories from each sample location in a run and averaging the resulting scores. Points in Figure 4 with low values of Q/x represent runs during which the air sampled followed trajectories that infrequently passed over refinery sites or were far removed in time from such sites. The point at Q/x = 0 represents a run during which none of the sampled air had trajectories intersecting refinery locations. The positive predicted refinery contribution to ambient concentration for this run is consistent with the fact that part of the refinery emissions are contributed by storage of petroleum materials. Storage depots are located throughout the Tokyo area, but only those at refineries were included in the analysis. Therefore, some refinery emissions are expected for samples from air parcels not advected over refinery sites.

The value of r^2 = 0.90. When $\alpha + \gamma = 2$, r^2 = 0.52. While we have no way of further discriminating individual values of $\alpha + \gamma$, the results shown in Figure 4 are consistent with limits on σ_z due to inversions over the ocean surface and aloft and with the observation that σ_y may be proportional to the square root of travel time for long downwind distances (Gifford, 1982). The association reemphasizes that the petroleum refinery coefficients do specifically reflect the impact of refinery emissions.

A similar approach was used to evaluate refinery contributions in a Chicago study (O'Shea and Scheff, 1988). Twenty 1-hour samples were collected from noon to 1:00 pm during September and October, 1985. Nine hydrocarbon components were used as fitting compounds to determine contributions from three source categories: vehicle exhaust, gasoline vapor, and petroleum refineries. Figure 5 shows typical back-trajectories to the refinery locations. Pasquill-Gifford definitions of σ_y were used and were corrected to hourly values for downwind distances greater than 10 km (Mueller and Reisinger, 1986). On the particular day shown, since the receptor location was within 2σ of the centerlines from refineries 16 and 25, these sources, but not refineries 13 and 20, were expected to affect the ambient measurement. Figure 6 compares the CMB-

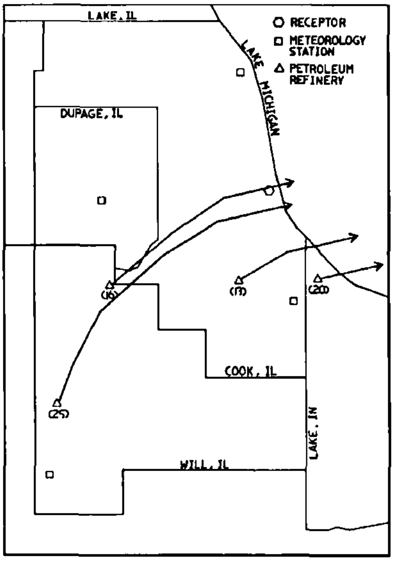

Figure 5. Location of the receptor, refineries and meteorological stations (O'Shea and Scheff, 1988). (Trajectory of surface wind shown for September 24, 1985.)

predicted refinery contributions to the sum of the nine measured hydrocarbons with an emissions score similar to Equation 2. Since the measurements were made on different days a windspeed adjustment was also incorporated into the emission estimate, i.e.,

$$C_j \propto \frac{Q_j}{ux} \tag{3}$$

As indicated in Table 8 the association was not strong (r^2 = 0.20). However, since there were some inconsistencies in the reported refinery inventory emissions, refinery capacity was substituted as a surrogate for Q. Figure 7 shows this comparison and displays a much stronger association with refinery score (r^2 = 0.64). In addition we would expect the correlation of the CMB prediction to progressively increase when compared with Q, Q/x, and Q/ux, since each of these forms of scoring incorporates an additional physical aspect. As shown in

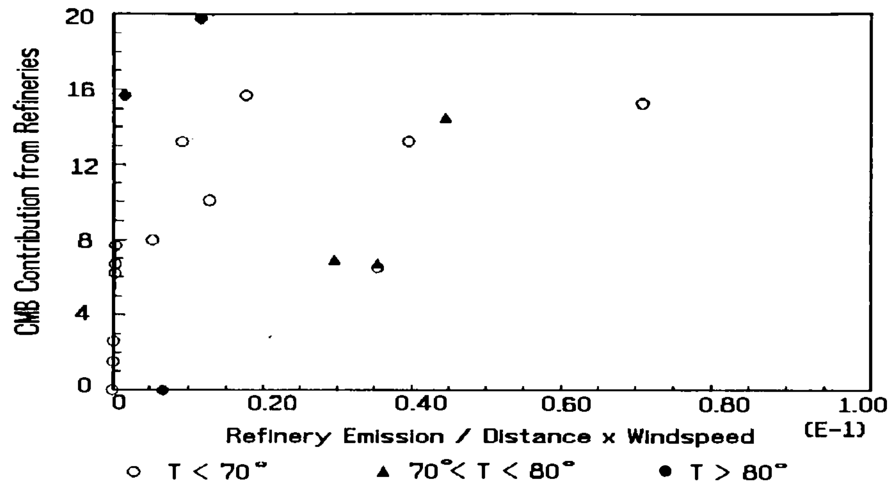

Figure 6. Refinery contributions to ambient concentrations of total hydrocarbons vs refinery emissions adjusted for distance and wind speed (O'Shea and Scheff, 1988). (Ambient temperature during sampling periods indicated on the plot.)

Table 8, the strength of the correlation does improve when using refinery capacity, but not with emissions. These observations indicate that the CMB correctly allocated mass to the refinery source category, and suggest caution in interpreting the emission inventory values.

4.3 Non-Identified Sources

At this stage of CMB organic modeling, not all major NMOC source categories have been included in the models. Consequently, a fraction of the sum of the fitting compounds (THC) in the ambient sample(s) is not predicted. We have interpreted this fraction as being primarily from sources not included in the model rather than from experimental error or variability in source fingerprints or ambient composition. For the Chicago study approximately 33% (winter) and 59% (summer) of the average THC was not explained by the 8 specified sources (Hegberg *et al.*, 1989). About 40% of the ambient THC was not explained by the

Table 8. Comparison of CMB and Dispersion Model Predictions for Refineries (O'Shea and Scheff, 1988).

Components	r	r^2
CMB vs. emissions	0.45	0.20
CMB vs. capacity	0.74	0.54
CMB vs. emissions/x	0.34	0.11
CMB vs. capacity/x	0.79	0.62
CMB vs. emissions/$u\,x$	0.45	0.20
CMB vs. capacity/$u\,x$	0.80	0.64

4 source types in the Tokyo study (Wadden *et al.*, 1986). The earlier Chicago study which only included 3 source categories resulted in 18% unexplained of the 9 hydrocarbons measured. In all of these cases the unexplained difference for individual samples was usually positive although this was not a constraint of the model. This fact suggests that the CMB method does not overestimate emissions from modeled sources by mis-allocating residual mass to the estimated coefficients. The seasonal consistency of source contributions to NMOC (Table 7) despite considerably different unexplained source contributions also supports this premise. We believe that the residual mass source category represents actual sources of VOC and has physical meaning. This observation is also consistent with previous work with particulate matter (Scheff and Wadden, 1986). Additionally, the residual composition is a first estimate of the speciated contribution from a mix of non-modeled sources. Such data provide information for further CMB or tracer evaluation, which has been found to be useful for some recent evaluations of ambient particulate (Scheff, 1989; Scheff *et al.*, 1989b; Dzubay *et al.*, 1988).

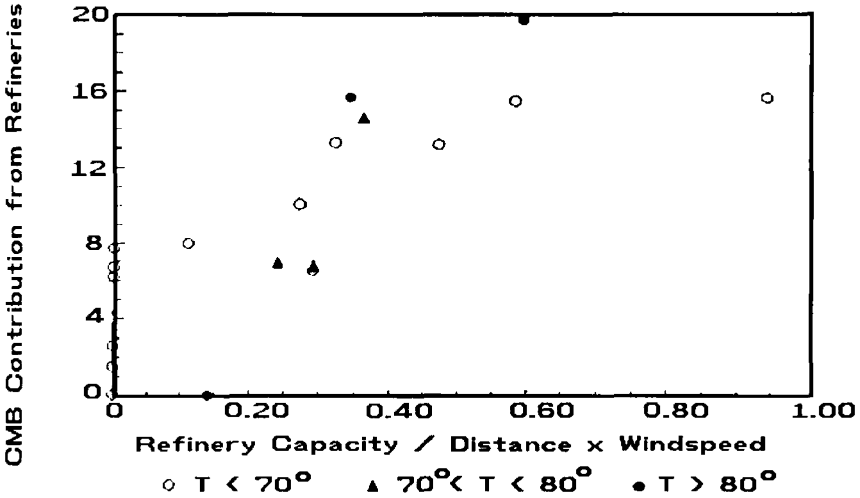

Figure 7. Refinery contributions to ambient concentrations of total hydrocarbons vs refinery capacity adjusted for distance and wind speed (O'Shea and Scheff, 1988). (Ambient temperature during sampling periods indicated on the plot.)

4.4 Extended Fingerprints

Table 4 summarizes the organic composition for 23 fitting components for 10 source fingerprints. The choice of these fitting compounds has been discussed under Source Fingerprints. However, as is evident in Tables 1 to 3, the fitting organics only comprise a fraction of the total NMOC. These fractions are also given in Table 4. Table 9 shows a typical component distribution for non-fitting materials for the graphic arts source category (Scheff *et al.*, 1989a). Many of these materials are too reactive to be measured at ambient concentration levels representative of their emission contribution. However, the information is useful in several other ways. These fingerprints can be used directly, without reference to the CMB evaluation, to speciate source emissions required as input for photochemical smog modeling. Alternatively, since we have shown earlier that source coefficients averaged over a representative number of ambient samples reflect the source distribution of emissions, these non-fitting component data can be used in conjunction with average source coefficients to estimate actual

Table 9. Composite Non-fitting Component Fingerprint for Graphic Arts (Scheff *et al.*, 1989a).

Component	Weight %
Fitting compounds	
Toluene	11.05
p-Xylene/m-Xylene	0.82
Non-fitting compounds	
VM&P Naphtha	16.19
C_6-C_7 paraffins	
C_7-C_8 aromatics	
93-127°C cut	
Acetone	3.18
Methyl-ethyl-ketone	7.06
Methyl-isobutyl-ketone	1.92
Ethyl Alcohol	8.04
Ethyl Acetate	2.50
Isopropyl Acetate	3.24
Petroleum Distillate	46.00
C_{12}-C_{50}	
243-370°C cut	
Total	100.00

speciated emissions. If we let Wf_j be the average fraction, over a distribution of samples, of the sum of the fitting compounds allocated to source category j, and R_j be the ratio of the sum of the fitting compounds to total NMOC in source j emissions, then the fractional contribution, $w_{i,j}$ of non-fitting component i from source j to the total NMOC for the whole airshed will be:

$$w_{i,j} = \left[\frac{Wf_j}{\sum R_i} \right]_j x_{i,j} \tag{4}$$

where $x_{i,j}$ is the weight fraction of i in the source j fingerprint.

4.5 Reactivity

One of the basic assumptions of the CMB approach is that the concentration of each component from a particular source remains constant with respect to the other components from that source. This does not mean that there can be no decay or deposition between source and receptor but only that there be no compositionally selective removal. In addition, any significant changes in relative composition due to reaction effects must take place in a time frame that is shorter than the sampling time and the travel time between source and receptor. Table 10 shows the reaction rates of typical fitting compounds with the OH radical (k_{OH}), a relationship which is frequently used as a measure of reactivity. While the values of k_{OH} do vary from compound to compound, most are within a factor of 10, ethane being the significant exception. However, because a statistical fitting technique is used, an outlying point ordinarily will not have significant effect on the predicted coefficients. (We have found this to be the case by selectively removing fitting components with high or low values of k_{OH} from the fitting procedure (Wadden *et al.*, 1986; Scheff and Klevs, 1987)).

Figures 8-10 partly demonstrate the typical changes in composition which can take place in an air mass under reactive conditions (Uno *et al.*, 1985b). Figure 8 shows a series of horizontal sections of a sampling flight over Tokyo at about 2:00 in the afternoon on August 6, 1980. Sample 602 was collected over the Miura Peninsula but samples 603, 605, and 606 were all taken over Sagami Bay. At a first estimate these samples are representative of the same steady-state sources, since each sample falls approximately within the prevailing wind trajectory, but are separated by advection time. The pattern of concentration change (relative to benzene) with advection time based on wind trajectories is displayed in Figures 9 and 10. Least-square fitting lines are shown for each component. Because the samples were taken over open water there was no reinforcement from additional sources, and because of the advective inversion at the ocean's surface, there was little or no loss due to deposition or surface reaction. As indicated on the Figures, the ozone concentration increased from 0.095 ppm to 0.148 ppm, so there was removal of hydrocarbons by reaction with

Table 10. k_{OH} for Typical Fitting Compounds (Uno *et al.*, 1985a).

Component	k_{OH}^* (ppm^{-1} min^{-1})	Component	k_{OH}^* (ppm^{-1} min^{-1})
Ethane	4.1×10^2	Benzene	2.1×10^3
Ethylene	1.2×10^4	M-cyclo-Hexane	1.2×10^4
Propane	2.8×10^3	2,2-DM-Hexane	6.3×10^3
Propylene	3.7×10^4	2,3,4-TM-Pentane	1.1×10^4
iso-Butane	4.0×10^3	3,M-Heptane	1.1×10^4
n-Butane	4.8×10^3	2,M-Heptane	1.1×10^4
Acetylene	1.0×10^3	3,E-Hexane	1.1×10^4
1-Butene	5.0×10^4	cis-1,3-DM-c-Hexane	1.4×10^4
iso-Butene	8.3×10^4	n-Octane	1.2×10^4
iso-Pentane	5.1×10^3	trans-1,3-DM-c-Hexane	1.4×10^4
1-Pentene	5.3×10^4	Toluene	9.0×10^3
n-Pentane	7.4×10^3	E-cyclo-Hexane	1.4×10^4
cis, trans-2-Pent	1.3×10^5	3,3-DM-Heptane	8.0×10^3
3,3-DM-1-Butene	4.0×10^4	2,3-DM-Heptane	1.2×10^4
2,2-DM-Butane	2.9×10^3	c-cyclo-Octene	1.0×10^4
2-M-Pentane	7.4×10^3	1,3,5-TM-c-Hexane	1.5×10^4
3-M-Pentane	1.0×10^4	t-cyclo-Octane	1.0×10^4
n-Hexane	8.7×10^3	cyclo-Octane	1.4×10^4
3,3-DM-Pentane	4.6×10^3	E-Benzene	1.2×10^4
cyclo-Hexane	1.0×10^4	p-Xylene	1.9×10^4
3-M-Hexane	9.1×10^3	m-Xylene	3.3×10^4
3-E-Pentane	9.1×10^3	o-Xylene	2.0×10^4
n-Heptane	9.1×10^3	Styrene	7.8×10^4
2,2,4-TM-Pentane	5.5×10^3		

*Most of values were taken from Atkinson *et al.*, 1979; styrene from Bigozzi *et al.*, 1981. When not experimentally available for paraffinic compounds, k_{OH} was estimated from an empirical equation

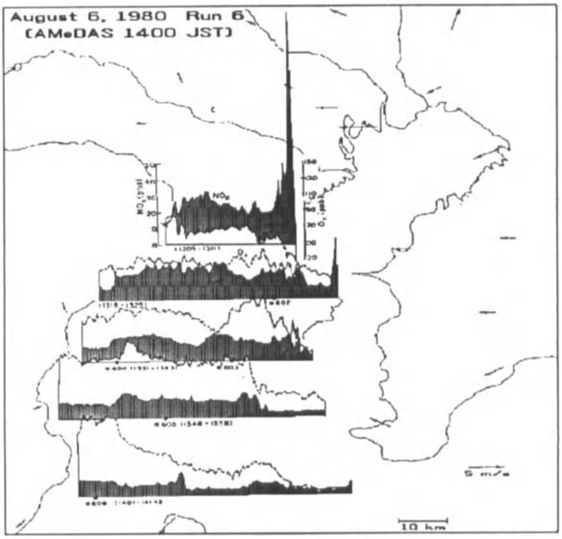

Figure 8. August 6, 1980 flight path over Tokyo for organic sampling. Arrows indicate wind vectors; # sampling points; NO_2 and O_3 concentrations are at 350 m altitude (Uno *et al.*, 1985b).

O_3 as well as with the OH-radical. Nonetheless, over 2 or 3 hours there was relatively little change in the composition ratios except for propylene and possibly ethylene. It is also important to recognize that the multi-variable fitting technique is based on the ratio of each component to each of the others, and so will not be affected even as much as might be suggested by Figures 9 and 10 where ratios are only shown with respect to benzene. The 2 or 3 hour advective time of passage from source to receptor is typical of many urban settings.

Figures 4 and 7 also indicate that reactivity was not a problem for comparing refinery emissions with allocated contributions to ambient concentration. In Figure 4, the mid-flight time is given for each point. However, the data show no groupings which reflect time of day. For example, the two points at Q/x = 48 and Q/x = 56 kg/h·km are for early morning and noon. If reactivity were very important for the refinery fingerprint, this would be reflected in the coefficient but not in Q/x, and the noon point should not agree with the rest of the points, and particularly not with a point taken around sun-rise. In Figure 7, ambient

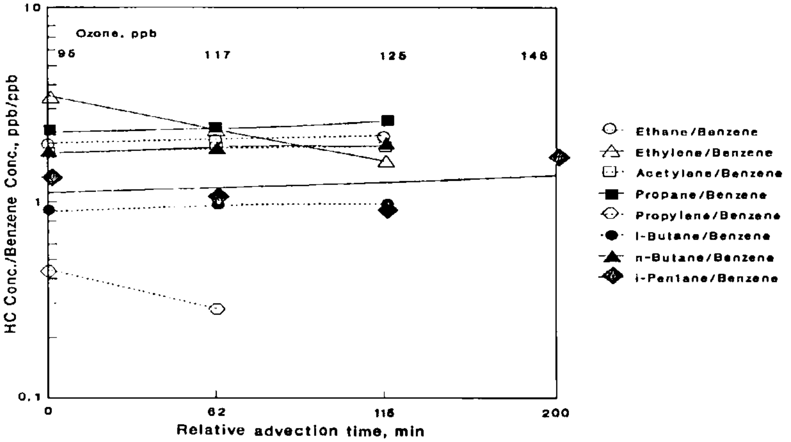

Figure 9. Hydrocarbon concentration/benzene concentration (C_2 - *i*-C_5) vs. relative advection time. Run 6, August 6, 1980; Advection times from back trajectories (Uno *et al.*, 1985b).

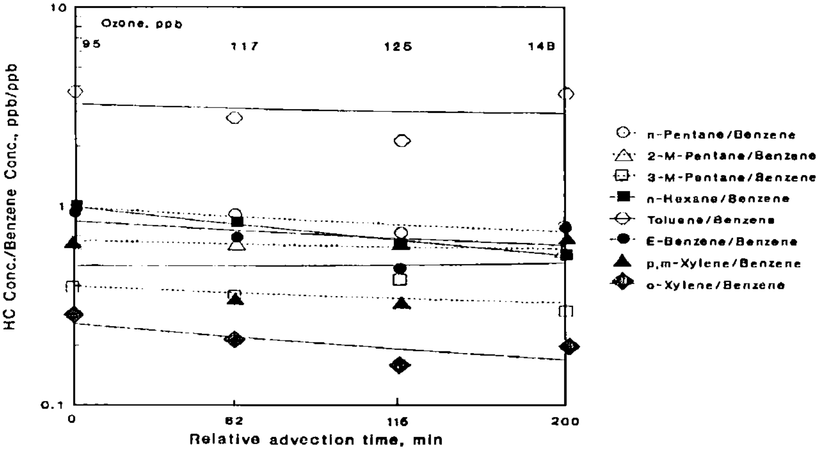

Figure 10. Hydrocarbon concentration/benzene concentration (*n*-C_5 - *o*-xylene) vs. relative advection time. Run 6, August 6, 1980; Advection time from back-trajectories (Uno *et al.*, 1985b).

temperature ranges are given for each point. Temperature is a useful surrogate for ozone-forming potential since ozone concentrations above 0.08 ppm have notbeen observed at maximum temperatures ⟨ 77°F (25°C) for Chicago (Wadden *et al.*, 1979). The data do not show the temperature-dependent patterns which would be expected if reactivity were an important factor in the CMB prediction. It is also notable that all of the samples in this study were collected from noon to 1:00 p.m., a period when temperature effects and ozone production would have had plenty of time to occur.

The Chicago winter-summer NMOC comparison with the emission inventory data in Table 7 also indicates that reactivity effects were not major. With the exception of refineries (discussed above), the summer-time CMB predictions were consistent with the emission inventory which is intended to be representative of warm-weather conditions. In addition, the CMB-estimated contributions from gasoline vapor, architectural coatings, graphic arts, and vapor degreasing all increased from winter to summer, as would be expected for such temperature-dependent emissions.

As a general comment, in the time/space intervals for which the model is applied in urban settings, (1-4 hr, 50-80 km), reactivity does not significantly effect the accuracy of prediction. This means that the various source fingerprints are not distorted by reaction effects to the extent that the respective source categories cannot be separated in the ambient sample. However, application of the approach over much longer distances or extended times, or use of fingerprints with more reactive materials, may require a more specific accommodation of the reactivity effect.

5. REFERENCES

Atkinson R., K.R. Darnall, A.M. Winer, A.C. Lloyd, and J.N. Pitts, Kinetics and Mechanism of the Reactions of the Hydroxyl Radical with Organic Compounds in the Gas Phase, In *Advances in Photochemistry*, J.N. Pitts, G.S. Hammond, K. Gollnick, and D. Grosjean, eds, Vol. 11, J. Wiley & Sons,Inc, New York, NY, 375-488 (1979).

Aronian, P., P.A. Scheff, and R.A. Wadden, Wintertime Source-Reconciliation of Ambient Organics, *Atmospheric Environ.* 23:911-920 (1988).

Belsley, D.A., E. Kuh, and R.E. Welsch, *Regression Diagnostics*, John Wiley & Sons, New York, NY (1980).

Bigozzi, C.A., A. Maldotti, C. Chiorboli, C., Bartocci, and V. Carassiti, Kinetics and Mechanism Reactions between Aromatic Olefins and the Hydroxyl Radical, *Int. J. Chem. Kinetics* 13:1235-1242 (1981).

Cheng, M.D. and P.K. Hopke, Identification of Markers for Chemical Mass Balance Receptor Model, *Atmospheric Environ.* 23:1373-1384 (1989).

Darnall, K. R., R. Atkinson, and J.N. Pitts, Rate Constants for the Reaction of the OH Radical with Selected Alkanes at 300°K, *J. Phys. Chem.* 82:1581-1584 (1978).

DeCesar, R.T., S.A. Edgerton, M.A.K. Khalil, and R.A. Rasmussen, A Tool for Designing Receptor Model Studies to Apportion Source Impacts with Specified Precisions, In *Receptor Methods for Source Apportionment*, T.G. Pace, Ed., Transactions No. TR-5, Air Pollution Control Association, Pittsburgh, PA, 56-67 (1986).

Dzubay, T.G., R.K. Stevens, G.E. Gordon, I. Olmez, A.E. Sheffield, and W.J. Courtney, A Composite Receptor Method Applied to Philadelphia Aerosol, *Environ. Sci. Technol.* 22:46-52 (1988).

Feigley, C.E. and J.H. Jeffries, Analysis of Processes Affecting Oxidant and Precursors in the Los Angeles Reactive Pollutant Program, *Atmospheric Environ.* 13:1369-1384 (1979).

Gifford, R.A., Long-Range Plume Dispersion: Comparisons of the Mt. ISA Data with Theoretical and Empirical Formulas. *Atmospheric Environ.* 16:883-886 (1982).

Hegberg, B.A., P.A. Scheff, R.A. Wadden, B.A. Bates, P.F. Aronian, and Z. Qi, Discrimination of Source Contributions to Ambient Organics, Paper 89-104.4, 82nd Annual Meeting, Air & Waste Management Association, Pittsburgh, PA (1989).

Illinois Environmental Protection Agency, *Illinois Reasonable Further Progress Report for 1986: Ozone and Carbon Monoxide*, Division of Air Pollution Control IEPA/APC/87-015 (1987).

Kowalczyk, G.S., G.E. Gordon, and S.W. Rheingrover, Identification of Atmospheric Particulate Sources in Washington D.C., Using Chemical Element Balances, *Environ. Sci Technol.* 16:79-90 (1982).

Leone, R. M., E.W. Davis, and A.D. Jones, Updating Nontraditional VOC Source Inventories, Paper 87-58.2, 80th Annual Meeting, Air Pollution Control Association, Pittsburgh, PA (1987).

Lonneman, W.A., R.L. Sella, and S.A. Meeks, Non-Methane Organic Composition in the Lincoln Tunnel, *Environ. Sci. Technol.* 20:790-796 (1986).

Mayrshon, H. and J.H. Crabtree, Source Reconciliation of Atmospheric Hydrocarbons, *Atmospheric Environ.* 10:137-143 (1976).

Miller, M.S., S.K. Friedlander, and G.M. Hidy, A Chemical Element Balance for the Pasadena Aerosol, *J. Colloid Interface Sci.* 39:165-176 (1972).

Mueller, S.F. and L.M. Reisinger, Measured Plume Width Versus Sampling Time: A Look Beyond 10 Kilometers, *Atmospheric Environ.* 20:895-900 (1986).

Nelson, P.F., S.M. Quigley, and M.Y. Smith, Sources of Atmospheric Hydrocarbons in Sydney: A Quantitative Determination Using a Source Reconciliation Technique, *Atmospheric Environ.* 17:439-449 (1983).

O'Shea, W.J. and P.A. Scheff, Validation of a Source Apportionment Model for Volatile Hydrocarbons, *J. Air Pollut. Control. Assoc.* 38:1020-1026 (1988).

Scheff, P.A., R.A. Wadden, and R.J. Allen, Development and Validation of a Chemical Element Mass Balance for Chicago, *Environ. Sci. Technol.* 18:923-931 (1984).

Scheff, P.A. and R.A. Wadden, Predicting Unidentified and Secondary Sources with Chemical Mass Balance Receptor Models, In, *Receptor Methods for Source Apportionment, Real World Issues*, T.G. Pace, ed., Transaction No. TR-5, Air Pollution Control Association, Pittsburgh, PA (1986).

Scheff, P.A. and M. Klevs, Source-Receptor Analysis of Volatile Hydrocarbons, *J. Environ. Eng.* 113:994-1005 (1987).

Scheff, P.A., Development of a Two Step Chemical Mass Balance Model for Inhalable Particulate Matter in Chicago, Paper 89-103.8, Annual Meeting, Air Pollution Control Association, Pittsburgh, PA (1989).

Scheff, P.A., R.A. Wadden, B.A. Bates, and P.F. Aronian, Source Fingerprints for Receptor Modeling of Volatile Organics, *J. Air Pollut. Control Assoc.* 39:469-478 (1989a).

Scheff, P.A., G. McKee, and R.A. Wadden, Woodsmoke Contribution to PM_{10} in a Complex Airshed, Paper 89-104.7, Annual Meeting, Air Pollution Control Association, Pittsburgh, PA (1989b).

Sigsby, J.E., Jr., S. Tejada, W. Ray, J.M Lang, and J.W. Duncan, Volatile Organic Compound Emissions from 46 In-Use Passenger Cars, *Environ. Sci. Technol.* 21:466-475 (1987).

Storck, W., Chlorinated Solvent Use Hurt by Federal Rules, *Chem. & Eng. News* p. 11, Sept. 28, 1987.

Uno, I., S. Wakamatsu, M. Suzuki, and Y. Ogawa, Three-Dimensional Behavior of Photochemical Pollutants Covering the Tokyo Metropolitan Area, *Atmospheric Environ.* 18:751-761 (1984).

Uno, I., S. Wakamatsu, R.A. Wadden, S. Konno, and H. Koshio, Evaluation of Hydrocarbon Reactivity in Urban Air, *Atmospheric Environ.* 19:1283-1293 (1985a).

Uno, I., S. Wakamatsu, R.A. Wadden, S. Konno, and H. Koshio, Hydrocarbon Composition in the Environmental Atmosphere and its Source Reconciliation Model, In, *Studies on Photochemical Reactions of Hydrocarbon-Nitrogen Oxides-Sulfur Oxides System, Research on the Photochemical Secondary Pollutants Formation Mechanism in the Environmental Atmosphere (Part 2)*, National Institute Research Report No. 72, pp. 89-128, (in Japanese with English abstract), Tsukuba, Japan (1985b).

U.S. Environmental Protection Agency, *Organic Solvent Cleaners - Background Information for Proposed Standards*, Environmental Protection Agency Report No. EPA-450/2-78-045a, U.S. Environmental Protection Agency, Research Triangle Park, NC (1979).

U.S. Environmental Protection Agency, *Perchloroethylene Dry Cleaners - Background Information for Proposed Standards*, Environmental Protection Agency Report No. EPA-450/3-79-029a, U.S. Environmental Protection Agency, Research Triangle Park, NC (1980a).

U.S. Environmental Protection Agency, *Volatile Organic Compound (VOC) Species Data Manual (2nd Ed.)*, Environmental Protection Agency Report No. EPA 450/4-80-015, U.S. Environmental Protection Agency, Research Triangle Park, NC (1980b).

U.S. Environmental Protection Agency, *Control of Volatile Organic Compound Emissions from Large Petroleum Dry Cleaners*, Environmental Protection Agency Report No. EPA-450/3-82-009, U.S. Environmental Protection Agency, Research Triangle Park, NC (1982).

U.S. Environmental Protection Agency, 4.2.1 Nonindustrial Surface Coating. (April 1981), in *Compilation of Air Pollutant Emission Factors. (Vol. 1)*, Environmental Protection Agency Report No. AP-42, 4th ed., U.S. Environmental Protection Agency, Research Triangle Park, NC (1985).

U.S. Environmental Protection Agency, *Receptor Model Technical Series Volume III (Revised), Chemical Mass Balance Receptor Model User's Manual (Version 6.0)*, Environmental Protection Agency Report No. EPA-450/4-83-014R, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987).

U.S. Environmental Protection Agency, *Air Emissions Species Manual: Volume I. Volatile Organic Compound Species Profiles*, Environmental Protection Agency Report No. EPA-450/2-88-003a, U.S. Environmental Protection Agency, Research Triangle Park, NC, (1988).

Wadden, R.A., E.D. Ross, and J.E. Quon, Characterization of Ozone Episodes in Urban Air, *J. Environ. Eng.* 105:621-628 (1979).

Wadden, R.A., I. Uno, and S. Wakamatsu, Source Discrimination of Short-Term Hydrocarbon Samples Measured Aloft, *Environ. Sci. Technol.* 20:473-483 (1986).

Wakamatsu, S., Y. Ogawa, K. Murano, K. Goi, and Y. Aburamoto, Aircraft Survey of the Secondary Photochemical Pollutants Covering the Tokyo Metropolitan Area, *Atmospheric Environ.* 17:827-835 (1983).

Wakamatsu, S., I. Uno, and R.A. Wadden, A Study of the NMHC Source Fingerprint and its Photochemical Reactivity, In, Proceedings of the Eighth U.S.-Japan Conference on Photochemical Air Pollution, U.S. Environmental Protection Agency, Research Triangle Park, NC, 45-76 (1984).

Zweidinger, R.B., J.E. Sigsby, S.B. Tejada, F.D. Stump, D.L. Dropkin, W.D. Ray, and J.W. Duncan, Detailed Hydrocarbon and Aldehyde Mobile Source Emissions from Roadway Studies. *Environ. Sci. Technol.* 22:956-962 (1988).

Chapter 8

RECEPTOR MODELING IN THE CONTEXT OF AMBIENT AIR QUALITY STANDARD FOR PARTICULATE MATTER

Thompson G. Pace
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

1. INTRODUCTION

The focus of this chapter is a detailed discussion of the source attribution process and the model reconciliation process, including a detailed example of model reconciliation. It begins with a brief discussion of the National Ambient Air Quality Standards (NAAQS) which are established in the United States under the Clean Air Act. The chapter compares the air quality management and emission limitation approaches of source control and briefly overviews the control strategy development process. This chapter is not intended as a reference on the technical aspects of the receptor methods, as are described elsewhere in this text. Receptor methods, briefly stated, are a family of techniques which examine the ambient sample and the conditions of its collection to infer the types and relative mix of sources of pollution impacting the sample. Dispersion models, in contrast, use information on source emission rates and meteorology to estimate the contribution of these sources to daily or annual average conditions. The terms "model" and "method" are used interchangeably, even though techniques such as scanning electron or optical microscopy are methods, not models. The selection of appropriate receptor or dispersion models depends upon several considerations discussed below.

2. AIR QUALITY MANAGEMENT FOR PARTICULATE MATTER

Air Quality Management is the process of 1) determining what atmospheric concentration of a particular pollutant is acceptable in order to protect public health and welfare, 2) determining what level of source emissions of that pollutant in a particular airshed can be allowed if the specified concentration is not to be exceeded and 3) developing regulations to ensure that emissions are reduced to and maintained at or below that level.

2.1 National Ambient Air Quality Standards

Section 108 of the Clean Air Act (CAA-42 U.S.C. 7408) directs the administrator of the U.S. Environmental Protection Agency (EPA) to identify pollutants which may be reasonably expected to endanger public health or welfare and to issue air quality criteria for them. Section 109 directs the Administrator to propose and promulgate primary and secondary NAAQS to protect public health and welfare, respectively for these "criteria" pollutants. Currently, NAAQS have been established for lead, sulfur dioxide, ozone, oxides of nitrogen, carbon monoxide and particulate matter.

In 1971, EPA promulgated NAAQS for particulate matter measured as total suspended particulate matter (TSP). The primary standards were set at 260 ug/m^3, 24-hour average not to be exceeded more than once per year, and 75 ug/m^3, annual geometric mean. Also, a secondary TSP NAAQS was set at 150 ug/m^3, not to be exceeded more than once per year. On July 1, 1987, EPA promulgated revisions to the particulate matter ambient standards which include a replacement to the indicator such that only particles smaller than 10 um aerodynamic diameter are measured (i.e. PM-10). The 24-hour primary PM-10 NAAQS was set at 150 ug/m^3, expected annual arithmetic mean. The secondary NAAQS was set to be identical to the primary standard.

2.2 The State Implementation Planning (SIP) Process

The states are responsible for regulating the level of source emissions of PM-10 in each airshed through the SIP process. Section 110 of the CAA requires the states to submit, for EPA approval, SIPs that provide for attainment and maintenance of the PM-10 NAAQS. This is accomplished through the development of a SIP control strategy to reduce source emissions in those areas which exceed the PM-10 NAAQS and, in all areas, to ensure that the NAAQS will not be exceeded in the future due to growth.

States that have areas which exceed the PM-10 NAAQS are required to adopt and submit a SIP within 9 months after an exceedance of an NAAQS. Section 110(a)(2) lists the requirements for such a SIP. EPA has promulgated regulations under 40 CFP 51.112 requiring that states demonstrate through modeling or an adequate alternative that this control strategy will provide for attainment and maintenance. This "demonstration of attainment" requires the identification of the sources of that pollutant in the airshed and determination of each source's impacts on PM-10 ambient concentrations. This process is often referred to as "source attribution". The sources may be discrete "point" sources (e.g. steel mill refinery) or ubiquitous area sources (e.g. windblown dust). Control strategies must be developed which establish emission limits for both types of sources. The emission limits are designed to attain the NAAQS and ensure that attainment will be maintained for at least ten years.

In contrast to this "air quality management approach" used in the United States for the criteria pollutants, some countries use an emission limitation approach that uniformly regulates the emissions from each specific source category without particular regard to that pollutant's air quality or ambient pollutant concentrations in the airshed in which the source is located. The United States applies this concept of uniform emission limits for sources regulated under the National Emission Standards for Hazardous Air Pollutants (NESHAPS) program, but even then, the ambient concentration estimated for representative sources of the hazardous pollutant of concern is considered when the uniform emission limit is established. The air quality management approach used for PM-10 necessitates a clear understanding of the nature of the PM-10 problem in each airshed of concern, including the attribution of the ambient concentration to its component sources. Receptor modeling is an important tool in the source attribution process.

3. PM-10 SOURCE ATTRIBUTION

As noted earlier, establishment of emission limits in a control strategy requires the attribution of the ambient concentration in an airshed to its component sources. Accurate determination of emission limits is dependent on this source attribution study.

Two methods, receptor and dispersion modeling, are the linchpins of the control strategy development process. Either used alone can usually provide an adequate source attribution and basis for establishing emission limits. However, the synergy created by the combined use of receptor and dispersion modeling often results in a much improved source attribution study. This section contrasts and compares the receptor and dispersion approaches and section 4 will discuss the combined use of these models in detail.

3.1 Receptor Model Selection

Several considerations related to the nature of sources emitting PM-10 influence selection of the receptor model(s) for SIP purposes. These include: 1) the availability of particle size data for the emissions from predominant sources, 2) prior knowledge of the sources, 3) chemical similarity of the sources, 4) the need to identify individual sources vis-a-vis source categories, and 5) the time scale of interest. The factors affecting the selection of receptor models are summarized in Table 1 and discussed below. The chemical mass balance (CMB) is considered the most advanced of the receptor methods listed in Table 1. The other methods include factor analysis (FA), automated scanning electron microscopy (ASEM), and optical microscopy (OM). The FA, ASEM, and OM are not generally considered quantitative and FA requires at least 40 samples to complete an analysis. The reader is referred to the Receptor Model Technical Series (U.S. EPA, 1981a,b; 1983; 1984; 1985a; 1987a; Pace and Watson, 1987; Watson and Pace, 1987) for technical and cost information and U.S. EPA (1987b) for applicability information).

Table 1. Selecting Feasible Source Apportionment Methods Based on Data Availability and Source Characteristics

Method	Sources Known	Sources Unknown	Chemical Similarity	Fingerprints Disimilar	Isolated Single Source	Air Shed (Source Categories)	Air Shed (Specific Sources Within Categories)
Chemical Mass Balance (CMB)	Y		a,b,Y	Y	Y	Y	b,Y
Factor Analysis (FA)	X	X	b,X	X	X	X	
Automated Scanning Electron Microscopy (ASWM)	X	X	b,X	X	X	X	b,X
Optical Microscopy (OM)	X	X	b,X	X	X	X	b,X
Dispersion Model (DM)	YY		YY	YY	YY	YY	YY

a Initial use of factor analysis may be helpful.

b Useful if a source can be isolated from other similar sources by wind direction. Method usually cannot otherwise distinguish between sources in same category.

c Useful for fine particles larger than 1.0 um.

X Appropriate to use in conjunction with CMB or DM.

Y Appropriate to use with DM or FA, ASEM or OM.

YY Appropriate to use.

3.1.1 Particle Size

Many researchers have discussed the bimodal distribution of particulate matter which results from the tendency of some source categories to emit predominantly fine (<2.5 um) or coarse (>2.5 um) aerodynamic diameter particles. Sources which emit predominantly fine particles include those emitted by the combustion of fuels (motor vehicles, boilers, field and forest burning, wood stoves, etc.) and industrial processes involving combustion, chemical reaction, or condensation of vapors. Sources of coarse particles in the atmosphere include windblown dust from storage piles, agricultural fields, etc., vehicle re-suspended road dust, pollens, and fugitive emissions from industrial process sources. Various receptor models such as chemical mass balance and factor analysis are well suited to analyze sources of either fine or coarse particles. Optical and scanning electron microscopy are suitable for coarse and fine particles down to about 1 micrometer in size but work better on coarse particles. Most receptor models generally work best when the sample is segregated by size range (e.g., fine and coarse) because their size ranges are associated with different types of sources.

3.1.2 Prior Knowledge of Sources and Emissions

In many instances, the sources suspected of contributing to ambient PM-10 concentrations at a particular site are apparent. However, the relative contribution of each source is needed. Any of the methods discussed in this text could be used to give source apportionment information if the sources are identifiable, and the methods are properly applied. However, the mass balance requires knowledge of sources and their emission characteristics. If some sources are unknown, FA, ASEM, or OM might prove useful, especially if they are followed by a CMB analysis.

3.1.3 Chemical Similarity

The availability of "fingerprints" or source profiles for the sources of interest will often determine the optimum receptor model to use. A source profile is the characteristic chemical or morphological pattern of the emissions from a source that is used to distinguish it from other sources (U.S. EPA, 1985b). Some sources have fairly distinct fingerprints, while others do not. Since combustion source emissions are predominantly composed of carbon, there is very limited chemical information upon which to differentiate among the various types of combustion sources. Some insights might be gained by examining the optical properties or by using radiocarbon techniques ($^{14}C/^{12}C$ ratios) which can distinguish between modern or fossil carbon (e.g., wood smoke versus fuel oil). Another common source profile problem is distinguishing among sources whose emissions are comprised of soil or distinguishing between flyash and soil. Optical properties can be useful for some situations where fingerprints are similar. Also, X-ray diffraction (XRD) is very useful to identify various minerals by examining their crystalline structure.

3.1.4 Sources Versus Source Categories

Any of the receptor models listed in Table 1 can be useful in identifying the impact of an isolated specific source unless it's profile is similar to that of the background air. Likewise, the techniques can be used to identify many of the source categories within an airshed, recognizing the limitations identified in the preceding discussion. However, the impact of specific individual sources within an airshed which contains multiple sources of the same category may not be reliably identifiable except by using wind directional data or by using a dispersion model. In some cases, a receptor model can be used to estimate the sum total impact at a receptor due to a number of chemically similar sources; then, an emission inventory can be used to help estimate the individual source contributions.

3.1.5 Advantages and Disadvantages of Receptor Models

Receptor models offer a powerful advantage to the source attribution process because their results are based on the interpretation of actual measured ambient data. However, receptor models are not a panacea for all situations. They are limited spatially and temporally to the particular set of samples being analyzed. Obviously, they cannot be used to predict the impact of future sources. Each type of receptor model has specific advantages and limitations as listed in Table 2. More detailed discussion on these models can be found elsewhere in this text.

Table 2. Source Attribution - Receptor Models Advantages and Disadvantages

Source Apportionment Tools	Advantages	Disadvantages
RECEPTOR MODELS		
Microscopy		
Optical	Use of color, surface texture and optical properties for particle identification.	Limited to particles >2um, semi-quantitative, highly dependent on operator skill.
SEM	Can be used with particles >1um.	Costly to use on large numbers of particles, not quantitative if sample contains large variation in particle size.
Automated SEM-XRF	Classifies particles by size, shape and elmental composition. Analytical speed, ability to count large numbers of particles.	Only a few labarotories have ASEM capability and the necessary experience to interpret the results.
Chemical		
Enrichment Factors	Provides evidence of a source category's impact from elemental ratios. Simple.	Usually semi-quantitative; requires source composition data.
Time Series Analysis	Shows temporal variability. Simple, inexpensive.	Generally does not provide specific source impact information.
Chemical Mass Balance	Provides quantitative estimates based on measured data. Impact uncertainties provided.	Source composition data required. Chemically similar sources cannot be independently identified by CMB alone.
Multivariate Analysis	No prior knowledge of sources needed to resolve element patterns. Source composition required to interpret results.	Generally semi-quantitative. Large data sets required.
Radioisotope Analysis	Direct vs. "modern" quantitative measure of fossil carbon (e.g., coal or oil versus wood).	Costly. Limited to fossil-"modern" carbon apportionment. Few labs offer analytical service.
Physical		
X-Ray Diffraction	Direct identification of crystalline composition.	Coarse particles only, not useful for amorphous aerosols.
Trajectory Analysis	Helps identify approximate source location.	Cannot quantitatively estimate specific source impacts.

3.2 Special Considerations for PM-10 Dispersion Modeling

Technical documentation of the various alternatives for dispersion modeling are well established in the U.S. EPA's *Modeling Guideline* and will not be discussed here (U.S. EPA, 1986).

The *Modeling Guideline* provides a detailed discussion on 1) selection of appropriate source and meteorological data for use with dispersion models, 2) location of receptor sites, 3) selection of model options, 4) determination of urban/rural classification and 5) determination of background air quality.

For those cases where the techniques in the *Modeling Guideline* either are not available or not applicable, alternative modeling approaches must be considered on a case-by-case basis.

3.2.1 Input Requirements for Dispersion Models

Dispersion Models are based on knowledge of each source's emission release conditions (e.g., stack height, exit velocity, temperature) and dispersion characteristics (e.g., wind speed, direction, mixing height, atmospheric stability) to estimate it's concentration at a set of locations (receptors) in the airshed. Dispersion models can estimate source contributions reliably if the model is technically applicable to the conditions being modeled and if the source and atmospheric conditions are well characterized. A list of dispersion models applicable to PM-10 analysis is given in Table 3.

3.2.2 Advantages and Limitations of Dispersion Models

Dispersion models are more reliable for estimating longer time-averaged concentrations (e.g., annual average) than for estimating short-term concentrations (e.g., 24-hour) at specific locations (Core and Pace, 1981) Point source models are reasonably reliable in estimating the magnitude of the highest concentrations occurring sometime, somewhere within an area. Errors in estimating the highest concentrations are typically ± 10 to 40 percent for sources that can be adequately characterized. The multiple source urban model RAM showed no significant bias in estimating 1-hour ground level concentrations for a

Table 3. Dispersion Models Applicable to PM-10 Analyses*.

1 TO 24-Hour Average	Annual Average	Screening Techniques**
CRSTER	CRSTER	PTPLU-2
MPTER	MPTER	COMPLEX I
RAM	RAM	VALLEY
ISCST	ISCLT	
	CDM 2.0	

* For more information concerning the applicability of these models, consult the *Guideline on Air Quality Models (Revised)* (USEPA,1986). As noted in this document these models may also be used for TSP modeling analyses in conjunction with a suitable TSP emission inventory, as a surrogate, where PM-10 data bases are inadequate.

** These models are considered to be screening techniques for use prior to a more refined analysis as outlined in the *Guideline on Air Quality Models (Revised)* (USEPA,1986).

13-station monitoring network in St. Louis. The average network cumulative frequency distributions of hourly estimated and observed concentrations differed by no more than + or - 30 percent over the entire concentration range. However, estimates of concentrations that occur at a specific time and site tend to be poorly correlated with observed concentrations and are less reliable, should this performance attribute be important in a regulatory application (U.S. EPA, 1987b)

3.3 Model Selection

Table 4 contains general recommendations for source apportionment based on the time scale (annual or 24-hour) of the non-attainment problem and the data base available. The choices in the table are in order of preference, with the preferred approach listed first.

Table 4. Order of Preference for Approaches for PM-10 Source Apportionment (USEPA, 1987b).

Applicable dispersion and receptor model
Applicable dispersion model
Receptor methods (at least 2) CMB w. corroborating method

The CMB is recommended as the primary method to be used in regulatory applications of receptor models to PM-10 data (U.S. EPA, 1981b; 1983; 1987a). However, there is uncertainty in any source apportionment approach. Therefore, if CMB is used for source apportionment (without combining with a DM), it is required that at least one other receptor modeling approach be used as a corroborating analysis (U.S. EPA, 1987b). This may be FA, OM, ASEM, microinventory, trajectory analysis, XRD, or other corroborating approach as selected from those discussed in Volume I of the *Receptor Model Technical Series* (U.S. EPA, 1981a). Other receptor models are discussed in the *Digest of Ambient Particulate Analysis and Assessment Methods* (Throgmorton and Axetell, 1978). For regulatory applications involving PM-10, it is strongly urged that either optical microscopy or ASEM be used to corroborate CMB, along with intensive chemical analysis (sulfate, carbon and other elements) of the samples (U.S. EPA, 1987b). The OM or ASEM should be used instead of CMB if only TSP data are available. It is also strongly urged that the CMB be performed on size fractionated PM-10 samples, (i.e. fine and coarse fractions, below and above 2.5 um). This greatly increases the resolution of the techniques.

3.4 Using Receptor and Dispersion Models in Combination

Several demonstrations have been made where receptor models were used to help evaluate the results of dispersion modeling (Core *et al.*, 1982). This is the recommended approach for control strategy development. It is especially useful when the emission inventory used in a dispersion model is determined to be marginally adequate. The results of the receptor model can be used to carefully scrutinize the inventory assumed in the dispersion model to deduce whether emissions from certain source categories appear to have been adequately characterized. The use of a receptor model, such as CMB, in conjunction with dispersion modeling, is highly recommended in such situations. However, the use of dispersion and receptor models together often produces different results and these results must be reconciled.

4.0 RECONCILING RECEPTOR AND DISPERSION MODEL RESULTS

There are at least three generic cases which require model differences to be reconciled (Trijonis, 1985). The *Protocol for Reconciling Receptors and Dispersion Model Results* discusses this process (U.S. EPA, 1987c). The first case is when the results strongly suggest that at least one of the models has a significant error. The second case is when the model estimates are not disparate, but the overall uncertainty is judged to be large. Resources devoted to further refinement of the models and data bases might significantly reduce the uncertainties. The third case, a very common one, includes situations where one model is better for some source types, and another is better for other source types. Refinement of the final source contribution estimates should be cognizant of the suitability of each model to the suspect sources. Trijonis (1985) reports that the literature assumes that CMB models are more reliable than dispersion models for quantifying the impact of source categories. In contrast, the DM can deal explicitly with emissions from single, identifiable sources within the same source category. However, in order to do so, the DM must rely on several potentially uncertain inputs (emissions data, meteorological data, and the transport-diffusion-transformation-deposition mechanisms). Because of their fundamentally different approaches, the CMB and DM are complementary in their approach to source apportionment and using both can reduce the limitations of each alone.

4.1 CMB-DM Reconciliation

The CMB-DM reconciliation process is based on the premise that the receptor and dispersion models provide independent estimates of source contributions, rely on different data bases, have different underlying assumptions, and have different strengths and weaknesses. Generally, other receptor methods are used first, then the CMB is performed and, finally, the DM is run. Figure 1 depicts the flow of information if this order is maintained. At each step, the results of the other analyses are used to corroborate the ongoing analysis (U.S. EPA, 1987d). Other receptor methods may be used to corroborate the CMB; These methods are discussed elsewhere in this volume.

The reconciliation of CMB-DM results should follow an eight-step protocol involving comparison, reverification of input data, refining the inputs to both models, and rerunning the model(s) if necessary. The reconciliation protocol described in the remainder of this document assumes that both the dispersion model and CMB have been determined to be applicable and that preliminary results have been obtained for both models. The reconciliation process ends with control strategy development using the dispersion model unless the dispersion model is found to be inconsistent with the majority of the physical data. In such case, the CMB along with other receptor models would be used with a proportional model to develop a control strategy (U.S. EPA, 1987b).

4.1.1 Protocol Step 1: Compare CMB and DM Results

There are several issues that must be resolved in order to compare CMB and DM results: 1) consideration of suitable periods for sampling and analysis, with particular emphasis on source emission variability and meteorological conditions; 2) the issues of secondary particulate matter and background; and 3) grouping of the sources. These issues are discussed in the following paragraphs.

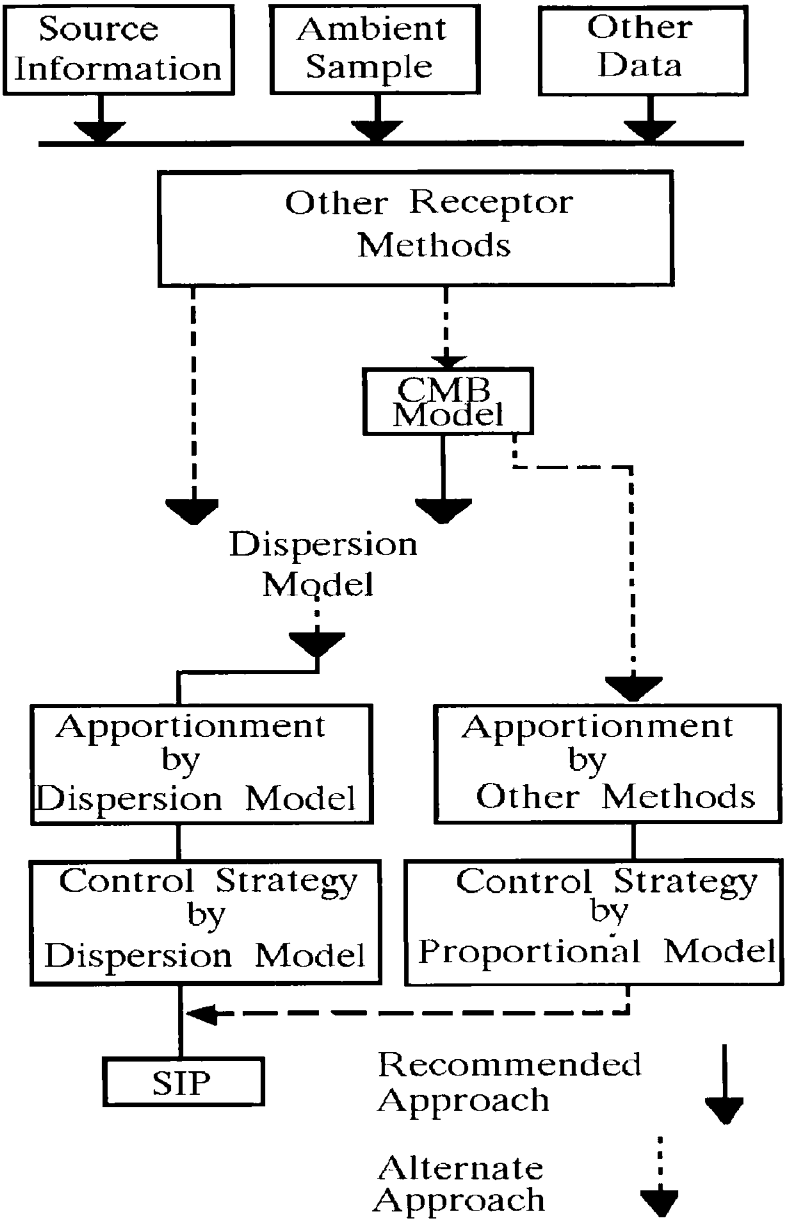

Figure 1. Information Flow for Source Apportionment and Control Strategy Development.

4.1.1.1 Selection of Sampling Periods for Analysis and Comparison.

Attainment of the NAAQS is generally calculated based on ambient data. Ideally, for comparison purposes, the DM would be run for the actual monitoring period and CMB analyses would be available for each day during that same period. However, this ideal is virtually never achieved for numerous reasons of cost and practicality. For example, in assessing potential causes for not meeting the annual NAAQS, the DM may have to be run using meteorological data representing different years than those on which non-attainment is based. In that case, it would be necessary to perform CMB on a subset of days which are representative of the measured concentrations during the time period covered by the DM but also representative of the overall meteorology during that time period. In addition, the number of samples to be compared must reflect the seasonal differences in measured concentrations. In CMB applications relating to the annual NAAQS, a minimum of five samples in each quarter are considered necessary to obtain representative.

For violations of the 24-hour NAAQS, it is preferable to apply the CMB model on the days on which "exceedances" of the NAAQS were observed. In this case, one would analyze and compare all observations greater than the level of the NAAQS. In order to obtain representative results, a minimum of five samples (e.g., the five highest values) should be compared. If there are fewer than five observed "exceedances" of the NAAQS, the five highest values overall should be used so that the analysis is based upon a representative number of days. If a receptor model study is undertaken for a short time period, it must be shown that the period covered is generally representative of the types of source activity and meteorology that are associated with any exceedances which were not measured during the receptor model study period. The source operating levels must be similar and dispersion conditions and wind deviation during the shorter period must be such that the same sources would be expected to contribute to measured concentrations. The receptor modeling is most useful if exceedances are measured during the receptor study.

Unfortunately, a 24-hour emissions inventory appropriate for specific days may not be available or obtainable. Inventories usually contain annual average emissions. Thus, associating "dates" with the dispersion model results, which are based on annual emissions, may give unreliable results. In such cases, an "aggregate" comparison of the analysis results is recommended. To make an "aggregate" comparison, select a subset of dispersion model 24-hour estimates that has dispersion meteorology similar to that which occurred on the CMB analysis days (i.e. days with similar wind velocities and precipitation/cloud cover). Then, compare the source contributions for each CMB analysis with the average source contributions for the corresponding subset of dispersion modeled days.

4.1.1.2 Background and Secondary Particulate Matter.

In many cases, the receptor model is used for source apportionment at urban or source oriented sampling sites, and the apportionment will include not only the "urban" contribution but also the background component which has been transported to the urban area from elsewhere. In such cases, it is usually necessary to perform an analysis on a nearby background receptor so that the background or regional contributions to ambient PM-10 observed at urban or source-oriented sites can be distinguished. This receptor analysis at a background site is necessary, for example, to distinguish between locally generated soil dust and that transported into the area. Receptor analyses from the background and other ambient monitors are then compared. Source categories that were identified on both urban and background monitors should be noted and estimated contributions of these sources at the study (urban) monitor should be reduced by the amount of their impact at the background site. In this process, one must carefully select the background days for receptor analysis to make sure they are "upwind" and are not impacted by local sources, and adjustments must be made to the CMB results prior to CMB-DM comparisons. Also, this reduction in urban contributions must be made carefully when background sources are chemically similar to sources in the urban area (e.g., soil and re-suspended street dust), especially if the background site has considerable impact from localized sources.

4.1.1.3 Grouping of Sources.

CMB is usually only useful for identifying source categories contributing to ambient PM-10. In contrast, DM can identify contributions from individual sources. Thus, source contributions from the CMB and the DM must be grouped in such a way that contributions from chemically similar sources can be intercompared. To accomplish this, the DM results are usually regrouped into larger "categories" or source groups with similar emission chemistries.

For example, perhaps a monitored sample was obtained from a coking operation. From a literature review or from experience, it is determined that the following operations at a steel making facility have emissions chemically similar to the coke sample:

Quenching
Pushing
Charging
Finished Coke Storage Pile

The "coke" source contribution calculated by the CMB would conceivably represent all of these specific sources. The DM output should group all of these sources together into a "coking" group; alternatively, the DM estimates for these sources could be manually grouped in a separate tabulation after the model is run.

For purposes of CMB-DM model reconciliation, a little judicious planning can make the comparison of CMB with DM results fairly straightforward. This comparison can enhance confidence in the models themselves if good agreement is found. Grouping of DM sources is performed as follows:

(a) Review the source profile and CMB source contribution categories to determine what sources they really include or represent. For example, a "sea salt" source might include salt from ocean spray and also salt which was spread for ice anti-skid control (U.S. EPA, 1987a).

(b) Review the Emission Inventory (EI) and decide which DM sources collectively comprise each CMB source category.
Also, identify any un-modeled sources (e.g., continental dust, sea salt, etc.) that might be represented by the CMB source category.

(c) Combine the DM sources into groups that are consistent with CMB source categories.

(d) Identify any sources which were significant contributors in the CMB analysis but which were not in the DM emission inventory.

(e) Add these sources to the EI as appropriate.

(f) Rerun the DM.

(g) Make sure that the CMB has considered all nontrivial sources which were identified in the DM inventory. Rerun the CMB if necessary.

4.1.1.4 Comparison of Results.

The grouping of sources described above allows a comparison of the DM and CMB estimates of the fractional or percentage contributions to the total calculated PM-10 mass. First, establish an interval of + or - 30 percent around the DM results and + or - one standard error (from the CMB output) around the CMB results. If these intervals overlap, no further refinement is generally warranted. However, if the DM results imply a source group is a major contributor and the CMB results imply that it is not (or vice versa), the results must be reconciled, even if the comparison criterion (above) is met. If further refinement of the models or inputs are needed to meet the criterion, the procedure for doing so is described in the following section. Note that the + or - 30 percent interval about the DM is only for use in this intercomparison; it is not intended as a confidence interval for the DM.

4.1.2 Protocol Step 2: Verify Input Data in Both Models and Rerun if Necessary

This step ensures that the difference in CMB and DM results is not due to trivial or inadvertent problems with data entry. It should focus on the data entry process to verify that the models were run using the intended data. The data review should include the ambient and source profile data, emissions data and meteorology. If all data are correct, skip to Step 4.

4.1.3 Protocol Step 3: Recompare Results

Any differences found in model estimates due to data entry should not be eliminated. Revised model results can now be compared, as discussed in Protocol Step 1. If the comparison criteria are not met, the next step is to reexamine model inputs, focusing initially on the CMB model.

4.1.4 Protocol Step 4: Refine CMB Model Inputs

This is an in-depth review of the CMB model inputs, focused on those source groups where large differences occurred between CMB and DM source category contribution estimates. Because of the validation that the CMB application should have received, it is unlikely that there are any substantive CMB model input problems which were not apparent from the preliminary analyses.

However, insights available through the DM analysis may highlight new problems or confirm suspicions that the modeler may have already had about weaknesses in the CMB that could not be previously substantiated. Also, it is appropriate to review the results of other receptor analyses at this point. This review may require modifying the model inputs or addressing collinearity (U.S. EPA, 1987d)

All measurements have some imprecision or uncertainty about them, and there will be some inherent underlying "noise" level in the data set which cannot be reduced by a reasonable allocation of time and resources. The user should review the CMB validation in light of the DM results (U.S. EPA, 1987d) If any weaknesses become apparent, address them to the extent time and resources permit, focusing on the source groups with disparate estimates. Then rerun the CMB model, if any refinements were made. It is not appropriate to be satisfied that a revision to a parameter will "fix" a discrepancy technical justification is required.

4.1.5 Protocol Step 5: Recompare Results

Any remaining differences found between CMB and DM model estimates are best addressed by closer examination of the dispersion model. If the comparison criteria in Step 1 are met, skip to Step 7.

4.1.6 Protocol Step 6: Refine Dispersion Model Inputs and Rerun

This is an in-depth review of the dispersion model, focusing on those source groups for which large differences occur between DM and CMB. This review should include: 1) the appropriateness of emissions data, which are based on the identification of sources and their locations, activity rates, emission factors and release heights; 2) meteorological data; and 3) underlying assumptions in the model itself and its dispersion, transformation, and removal algorithms. U.S. EPA (1984) gives examples of factors that primarily affect individual source-to-monitor impacts:

- Erroneous emission data caused by such items as the omission of unknown sources from the modeling or the use of inaccurate information.
- The use of inappropriate emission rates, such as the use of total particulate matter emission rates which include particle sizes larger than PM-10 sample.
- Incorrect information concerning daily source operating parameters. For example, which a source may operate at 45 percent capacity on an annual basis, its actual mode of operation may be a 90 percent for a large number of days a year.
- Neglect or incorrect consideration of downwash from tall stack sources.
- Neglect or incorrect consideration of particles re-suspended by the wind or mechanical action.
- Building interference causing source-to-receptor (i.e., source to monitor) geometry to be incompatible with Gaussian dispersion assumptions.
- Local meteorology differing from that modeled. A typical problem is wind direction shift or channeling caused by buildings or topographic features.

Anderson *et al.* (1986) suggests that these emissions and meteorological data problems can often be satisfactorily resolved by:

(a) Identifying significant sources of source groups from the CMB results;

(b) Assigning a "level of confidence" to emission and meteorological parameter estimates for each significant source or source group, based on engineering estimates. The "confidence level" information needed for this analysis may be found in Volume V of the Receptor Model Technical Series (U.S. EPA, 1984) Ideally this analysis should be completed before the CMB and DM are run; and

(c) Revising the emission and meteorology parameter estimates, as appropriate. It is not sufficient to know just that a revision to a parameter will "fix" a discrepancy. Technical justification is required to document all changes that are made. The pattern of discrepancies at different receptors may also aid in identifying potential inventory errors. For example, large under- or over-prediction by the DM at a receptor very close to a source may suggest possible errors in source receptor geometry. Large over-prediction at receptors far from a source could suggest an erroneous particle size distribution in emission factors. Experience with dispersion models and emission inventories will play an important role in diagnosing potential inventory problems.

The dispersion model is then rerun, using the revised input parameters.

4.1.7 Protocol Step 7: Recompare Results and Evaluate the Dispersion Model

The revised dispersion model results are again compared with the CMB results to determine whether further modifications to the dispersion model or meteorological inputs are warranted.

Anderson *et al.* (1986) provides examples of factors that are not typically addressed by the dispersion model that may cause systematic biases in model results. These factors are addressed in various nonguideline models that are currently available.

- Heat island effects which cause the actual near-source dispersion of elevated emissions to be greater than that modeled.
- Sea/land breeze and convergence zone effects.
- Effects caused by the fumigation of tall stack emissions.
- Effects caused by the development of thermal internal boundary layers (TIBL) over areas with varying surface heating characteristics.

Remedies to these problems might include acquisition of additional meteorological data, choice of different model options or modification of dispersion algorithms to meet local dispersion conditions and source characteristics. Such problem specific model modifications may prove a time and cost effective alternative to additional sampling and analysis. These remedies should be applied to the extent permitted by time and resource constraints.

4.1.8 Protocol Step 8: Final Model Estimates as Basis for Control Strategy

The reconciliation process should help improve both the initial DM and CMB results so that the differences between models are resolved. If this is the case, it is recommended that the DM model be used for control strategy development. If, however, it is clearly evident that the dispersion model results are inconsistent with the majority of the physical data and cannot be made consistent through justifiable modifications to the input data, the CMB estimates should be used as the basis for control strategy development. It is better to use the results of the method believed to provide the best results, rather than average the results of two or more methods.

4.2 Example Application of Protocol in an Industrial Area

This section describes an example application of the protocol for a nonattainment site located within an industrial area. The example was adapted from an actual environmental case study by Anderson *et al.* (1986). It was modified where necessary to illustrate the use of the protocol. All modifications were cognizant of what could and would have actually been done in the reconciliation had time and resources permitted (DeCesar, 1987). In this example, the causes of the elevated concentrations of PM-10 are not apparent due to the large number of point and area sources located relatively close to the monitoring site. Based strictly on the emission inventory, stack and fugitive dust emissions from a large steel mill as well as traffic generated resuspended road dust appear to be potentially large contributors to the particulate matter observed at the monitoring site. An investigation, which includes ambient and source data collection, OM, pollution rose (a form of trajectory analysis), dispersion modeling (DM) and chemical mass balance (CMB) modeling efforts, was conducted in order to apportion the source impacts with the level of confidence necessary for making control strategy decisions. The results are reconciled using the eight-step process previously described.

The industrial site PM-10 source apportionment example presented in this section is compiled primarily from experience gained from assessing source impact at a single monitoring site. The purpose of this example is to illustrate use of the eight-step reconciliation protocol described in this document. Therefore, the identity of the site and details of the OM, CMB, and DM analyses are unimportant to the objective of the example and are not presented in detail.

A combined receptor and dispersion modeling study was implemented with the objective of identifying and quantifying impacts of the sources contributing to violations of the PM NAAQS recorded at a monitoring site. To supplement the emission inventory, pollution concentration roses were prepared using data for eight TSP monitoring sites located near the PM-10 monitoring site. The pollution roses clearly indicate that the major sources of TSP are located in the quadrant south of the PM-10 monitor. That is, a preponderance of the high 24-hour average TSP measurements correspond with days having resultant winds from the south. This preliminary receptor analysis suggests that additional emphasis should be placed on characterizing steel mill related contributions because the steel mill is the largest source of particulate emissions located nearby and south of the monitoring site.

4.2.1 Receptor Modeling Data Collection Tasks

The ambient data base consisted primarily of: 1) ambient PM-10 samples collected for 1 year on an every-6th-day schedule; and 2) hourly measurements of wind direction, wind speed, and temperature at the PM-10 monitoring station. The PM-10 samples were collected on quartz fiber filters using a size-selective high volume (SSHV) sampler. In addition, for 1 month a dichotomous sampler was in operation at the monitoring site, and it collected several 24-hour coarse and fine fraction samples on Teflon filters. The results of the PM-10 monitoring program provided an annual average PM-10 concentration of 65 ug/m^3 as well as four exceedances of the 24-hour average PM-10 NAAQS of 150 ug/m^3.

A site visit was made to compile a microinventory of emission sources. The microinventory procedure consisted of determining the nature, location and spatial extent of all fugitive dust sources within one-quarter mile of the PM-10 monitoring station. The information was then used to calculate the emission rates for input to a dispersion model. Source types included unpaved roads, paved streets, railroads, and two coal storage piles. Emission rates were determined using published emission factors together with the source area and traffic volumes.

As part of the site visit, samples of emitted materials were collected from six nearby sources that had been identified as potentially important contributors to PM-10. The samples were subsequently analyzed to provide source profiles for use in the CMB model. The sources which were sampled included: a steel mill (blast furnace, coke oven, and basic oxygen furnace); two road shoulders (one near the monitoring site and one near an inoperative lead plant); and an agricultural processing plant (potash, corn gluten pellets, and soybeans).

Two subsets of the ambient PM-10 filters were selected for analysis as discussed above. Subset A contained the filters from 20 sampling days which were selected to be representative of the annual average conditions at the monitoring station. The average PM-10 concentration calculated from the filters in subset A was 67 ug/m^3, which compares well with the annual average calculated from all the filters (65 ug/m^3). Subset B contained the samples from the 4 days on which exceedances of the 24-hour average NAAQS were recorded plus the sample from the highest nonexceeding day. This subset was chosen to guide control strategy development related to the 24-hour standard.

Two multi-elemental characterization techniques were applied to the source and ambient filters. X-ray fluorescence (XRF) was performed on the Teflon filters, and Plasma Emission Spectroscopy (PES) was employed for the quartz filters. The following species were analyzed: Na, Mg, Al, Si, K, Ca, Ti, V, Mn, Fe, Zn, Br, Pb, EC and SO_4. The elemental carbon (EC) content of the samples was determined by optical attenuation analysis. In addition to the chemical characterization procedures, optical microscopy was applied to several of the ambient samples. Optical microscopy was used as a check on the receptor modeling source identifications and to aid in the receptor/dispersion model reconciliation process.

Receptor modeling was performed for this study using the Chemical Mass Balance (CMB) model (U.S. EPA, 1987a,d). The CMB model requires an input file containing the measured ambient concentrations of the elements for which the samples were analyzed. This requirement was fulfilled by inputting the results of the ambient filter analysis using the format specified by the CMB model. In addition, the CMB model requires a file containing the source compositions reported as the elemental mass fractions. A source composition file in CMB specified format was compiled which contained the elemental composition of the nearby sources as well as a number of source profiles from EPA's Source Composition Library (U.S. EPA, 1985b) The emission inventory and the microinventory were used to select sources for inclusion in the source composition file.

4.2.2 Dispersion Modeling Data Preparation

An emission inventory was compiled for all the point sources in the two counties surrounding the PM-10 monitoring site. A total of 140 point sources were included in the inventory. The point source inventory was modified for use in modeling PM-10 impacts by applying source specific PM-10/TSP size fraction ratios to the TSP emission factors which were originally contained in the inventory.

The Industrial Source Complex Short-Term (ISCST) model was used for this investigation. ISCST was used because it is applicable to the dispersion conditions found in the study area, and it contains several features that improve the source apportionment process. Two potentially advantageous features of ISCST are the ability to: (1) model microinventoried fugitive dust sources as volume sources; and (2) calculate the combined impact for selected groups of sources. This latter feature greatly decreases the manipulations necessary to transform dispersion and receptor modeling results into comparable formats for intercomparison.

Two receptor oriented approaches are utilized in this example: the CMB model and the optical microscopy technique. CMB modeling analyses were performed on the data obtained from filters contained in subsets A and B described in Section 4.2.1. The modeling procedure consists of obtaining a solution using the CMB procedures found in reference (Anderson *et al.*, 1986).

Two SSHV samples, two coarse fraction dichotomous samples, and one fine fraction dichotomous sample (ambient samples) underwent optical microscopic analysis to confirm the CMB source identifications. In addition, the results of the particle identification and counting performed as part of the microscopic analysis are used to calculate semi-quantitative source contribution estimates.

As part of the CMB modeling procedure, efforts are made to develop information which could be used to confirm or refute the validity of the CMB results. This information can be classified into four categories: (1) applicability of the model to the situation; (2) evidence of adherence or deviation from model assumptions; (3) stability of source impact estimates with respect to minor changes in the CMB model fitting parameters; and (4) comparisons with preliminary analyses (e.g., pollution rose, OM and microinventory analyses). By evaluating the CMB results on the basis of the above four categories, the CMB analyses are determined to be valid.

Background concentrations are defined as that portion of the measured ambient levels that is not attributable to emissions within the study area. To estimate the appropriate background concentrations for the dispersion model, an eclectic pollution rose[1] was developed using data from TSP monitoring stations located outside the study area. Data were used only from the days having persistent winds blowing into the study area from the direction of the background TSP stations. An average TSP concentration of 26 ug/m^3 was determined to be attributed to sources outside the study area. Application of the site-specific PM-10/TSP ratio (0.57) yields an average background PM-10 concentration of 15 ugm^3. Background concentrations were also estimated for each wind sector using this method.

4.3 The Eight-Step Reconciliation Process

4.3.1 Step 1 - Compare DM and CMB Results

The dispersion model simulation results for time periods corresponding to the PM-10 sampling schedule at the monitoring station were compared with measured PM-10 air quality data. This comparison suggests that the ISCST is underpredicting PM-10 concentrations at the monitoring site. Comparison of the predicted PM-10 with the measured PM-10 minus estimated PM-10 background concentrations provided an indication that the model was underpredicting.

The source contributions from the DM and CMB were grouped in such a way that they could be compared. This was actually done to some extent before the ISCST was run by examining the emission inventory. The emission inventory contained the name of each plant and a description of each emission source. This information was used to develop a preliminary tabulation of the types of

---

[1] The term "eclectic pollution rose" is used to describe a rose developed by compositing the data from several monitors surrounding the urban area such that the eclectic rose reflects concentrations only when the wind is blowing into the urban area.

materials (e.g., specific source chemistry) emitted by each source. The many types of emitted materials were then associated with 1 of 17 general categories which were determined to be consistent with the source profiles used in the CMB analysis. Each of the 140 point sources and 25 volume sources were assigned a code corresponding to one of the categories. Some overlap and "juggling" of categories was inevitable but a reasonably good preliminary grouping was made. The results are presented in Table 5. Adjustment to this preliminary grouping may be necessary in some cases.

The source contribution estimates provided by the CMB and ISCST models for subset A (annual average) and subset B (samples violating the 24-hour NAAQS) are compared in Table 6. The basic oxygen furnace and steel blast furnace groups were combined for the same reason. The results of the two models display reasonable agreement in two categories: (1) the background estimates used by ISCST are approximately equal to the secondary sulfate values estimated by CMB; and (2) resuspended road dust is listed as a major source of PM-10 by both methods. In addition, the similarities and differences between CMB and ISCST are consistent between subsets A and B. Many differences between the CMB and ISCST results were evident. For example, ISCST under predicted the measured mass and CMB over predicted the measured mass. Furthermore, ISCST assigned contributions to several source categories which were not identified by the CMB method. Finally, CMB and ISCST estimated substantially different impacts for wood-fired boilers, coal combustion, the road dust and motor vehicle exhaust combination and the steel mill related emissions from coking operations, blast furnace, and basic oxygen furnace. Unfortunately, Table 6 cannot reflect conclusions drawn from any OM analysis, because the OM analysis is available for only 2 days.

As a result of this comparison, four cases were identified (based on comparison model results as discussed in Step 1 of the Protocol) where additional model refinements may be necessary:

Table 5. Types of Emissions Assigned to Each Source Group in the Dispersion Modeling Analyses

Group Number	Type of Emission
1	Oil and Gas Combustion
2	Coal Combination
3	Wood-fired Boilers
4	Agricultural Products Handling
5	Coking Operations
6	Coal Handling
7	Steel Blast Furnace and BOF
8	Paint Production
9	Cement and Limestone
10	Motor Vehicle Exhaust and Lead Processes
11	Sand and Bentonite
12	Aluminum Production
13	Zinc Processing
14	Tire Production
15	Oil Refining
16	Fertilizer
17	Road Dust, Soil, and Motor Vehicle Exhaust

Table 6. Comparison of Initial Results by ISCST and CMB for Subsets A and B ($\mu g/m^3$).

	(Representing Annual Average) Subset A		(Average of 5 Highest 24-hr Exceedances) Subset B	
Source Categories	ISCST	CMB	ISCST	CMB
Oil & Gas Combustion	1.6	--	2.0	--
Wood-fired Boilers	0.2	3.2 (1.4)	0.8	5.6 (2.1)
Coal Combustion	0.6	2.8 (1.1)	1.2	8.0 (3.2)
Coking Operations	1.4	8.6 (4.2)	4.0	23.2 (8.7)
Blast Furnace &	a	12.0 (3.5)	a	33.6 (8.9)
Basic Oxygen Furnace	0.6	3.8 (1.5)	2.0	10.4 (2.7)
Coal Handling	0.3	--	0.6	--
Agricultural Prod. Handling	1.1	--	1.6	--
Paint Production	--	--	0.8	--
Cement & Limestone	1.9	--	1.9	--
Sand & Bentonite	1.2	--	1.7	--
Aluminum Production	--	--	0.4	--
Zinc Processing	0.8	--	0.6	--
Oil Refining	0.6	--	1.6	--
Fertilizer	--	--	0.8	--
Tire Production	--	--	2.2	--
Motor Vehicle Exhaust	15.1	2.4 (0.8)	2.6	2.4 (0.9)
Road Dust & Soil		30.0 (7.2)	31.2	79.2 (20.3)
Secondary $(NH_4)_2SO_4$	--	15.0 (2.5)	--	27.4 (5.0)
Background	15.0	--	32.0	--
Total	45.4	77.8	89.0	189.8
Measured	67.0	67.0	162.0	162.0

a. Estimated as a composite by ISCST.

Note: Numbers in parentheses indicate one standard error in the CMB analysis.

Case 1. The combined contributions of motor vehicle exhaust and resuspend road dust, as estimated by CMB and ISCST, were in major disagreement.

Case 2. A significant disagreement existed between the CMB and ISCST estimates of the coal combustion impacts.

Case 3. CMB estimated wood-fired boilers to be a much larger contributor to PM-10 than did ISCST.

Case 4. There was a major disagreement between CMB and ISCST for steel mill related impacts (i.e., coking operations, blast furnace and basic oxygen furnace source categories).

Each case is discussed below under the appropriate step in the reconciliation process.

4.3.2 Steps 2 and 3 - Verify Input Data, Rerun Models and Recompare Results

A review of the input data for both models indicated that they were indeed using the data that were intended. None of the four cases were related to inadvertent errors in data input. Therefore, the models were not rerun at this point.

4.3.3 Step 4 - Refine CMB Model Specification

Operationally, all four cases identified in Step 1 were reviewed at this point. Only case 1 was found to be related to the CMB inputs. Thus, discussion of the other three cases is included under step 6. Tables 7 and 8 compare the source impacts estimated by ISCST, CMB and optical microscopy for the 2 days, October 1, 1983 and October 25, 1983, when OM analysis was performed.

Case 1. The combined contributions of motor vehicle exhaust and resuspended road dust, as estimated by CMB and ISCST, were in major disagreement.

Table 7. Comparison of the Receptor and Dispersion Model Results for the October 1, 1983 PM_{10} Samples.

			Modeled Concentrations ($\mu g/m^3$)						
Dispersion Modeling Source Groups	Optical Microscopy Source Categories	CMB Modeling Source Types	Dispersion Modeled Impacts	SSHV Optical Macros. (A)	Coarse Dichot Optical Macros. (B)	Fine Dichot by Diff. (A-B)	SSHV CMB	Fine Dichot CMB	Coarse Dichot CMB
Oil & Gas Combustion	Fuel Comb., Carbon	---	5.8	-.-	1.0	1.0	-.-	-.-	-.-
Wood Combustion	----	Wood Boiler	1.4	-.-	-.-	-.-	-.-	8.6	3.0
Coal Combustion	Fuel Comb., Part. Pyr. Coal	Coal-fired Power Plant	0.8	8.6	3.8	4.8	2.0	-.-	6.6
----	St. Mill, Part. Pyr. Coal	----	-.-	5.8	0.4	5.4	-.-	-.-	-.-
Coke	Steel Mill, Coke	Coke Ash	0.4	7.0	4.4	2.6	12.4	39.6	-.-
Steel	Steel Mill	Blast Furnace	2.0	34.6	12.4	22.2	16.4	-.-	-.-
----	----	Basic Oxygen Furnace	-.-	-.-	-.-	-.-	5.6	13.2	8.4
Coal	Coal Handling	----	0.8	4.6	2.8	1.8	-.-	-.-	-.-
Agricultural Products	Agricultural Products Handling	----	5.4	6.0	-.-	6.0	-.-	-.-	-.-
Paint Products	Spray Painting	----	1.2	-.-	-.-	-.-	-.-	-.-	-.-
Cement, Limestone	Cement Handling	----	2.0	-.-	-.-	-.-	-.-	-.-	-.-
Lead	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Sand, Bentonite	Clay Handling	----	-.-	-.-	1,4	1.4	-.-	-.-	-.-
Aluminum	Alumina Handling	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Zinc	----	----	2.6	-.-	-.-	-.-	-.-	-.-	-.-
Misc. Chemical	----	----	4.4	-.-	-.-	-.-	-.-	-.-	-.-

Table 7. (Continued)

			Modeled Concentrations ($\mu g/m^3$)						
Dispersion Modeling Source Groups	Optical Microscopy Source Categories	CMB Modeling Source Types	Dispersion Modeled Impacts	SSHV Optical Macros. (A)	Coarse Dichot Optical Macros. (B)	Fine Dichot by Diff. (A-B)	SSHV CMB	Fine Dichot CMB	Coarse Dichot CMB
Refinery Catalysts	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Misc. Refinery	Oil Refining	----	3.2	-.-	-.-	-.-	-.-	-.-	-.-
Fertilizer	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Tire Products	----	----	8.6	-.-	-.-	-.-	-.-	-.-	-.-
----	Traffic, Tire Fragments	----	-.-	3.2	2.2	1.0	-.-	-.-	-.-
----	Trafic, < 1um Carbon	Automobile Exhaust	-.-	5.0	2.8	2.2	4.4	1.0	0.8
Road Dust	Traffic, Wind Erosion	Near Street 1	32.2	41.2	17.4	23.8	25.2	2.0	75.2
----	Biologicals & Misc.	----	-.-	-.-	5.2	5.2	-.-	-.-	-.-
----	Secondary Aerosols	Secondary $(NH_4)_2SO_4$	-.-	27.6	5.3	22.3	27.5	24.4	3.0
Background	----	----	31.0	-.-	-.-	-.-	-.-	-.-	-.-
Total	----	----	101.8	143.6	59.1	99.7	93.5	88.8	97.1
Measured $PM_{2.5}$	----	----	99.8	-.-	-.-	99.8	-.-	99.8	-.-
Measured $PM_{2.5\text{-}10}$	----	----	65.4	-.-	65.4	-.-	-.-	-.-	-.-
Measured Dichot. PM_{10}	----	----	165.2	-.-	-.-	-.-	-.-	-.-	-.-
Measured SSHV PM_{10}	----	----	163.2	163.2	-.-	-.-	163.2	-.-	-.-

Table 8. Comparison of the Receptor and Dispersion Model Results for the October 25, 1983 PM_{10} Samples

Dispersion Modeling Source Groups	Optical Microscopy Source Categories	CMB Modeling Source Types	Modeled Concentrations ($\mu g/m^3$)						
			Dispersion Modeled Impacts	Optical Microscopy			SSHV CMB	Fine Dichot CMB	Coarse Dichot CMB
				SSHV	Fine Dichot	Coarse Dichot			
Oil & Gas Combustion	Fuel Comb. Carbon	----	0.4	-.-	1.0	0.4	-.-	-.-	-.-
Wood Combustion	----	Wood Boiler	-.-	-.-	-.-	-.-	-.-	6.8	0.6
Coal Combustion	Fuel Comb., Part. Pyr. Coal	Coal-fired Power Plant	-.-	1.8	1.8	1.2	-.-	-.-	-.-
----	St. Mill, Part. Pyr. Coal	----	-.-	2.4	0.6	0.1	-.-	-.-	-.-
Coke	Steel Mill, Coke	Coke Ash	-.-	3.0	2.8	1.6	11.8	-.-	-.-
Steel	Steel Mill	Blast Furnace	-.-	5.4	5.8	1.0	10.6	-.-	5.0
----	----	Basic Oxygen Furnace	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Coal	Coal Handling	----	-.-	3.0	2.2	1.0	-.-	-.-	-.-
Agricultural Product	Agr. Prod. Handling	----	1.0	3.4	2.6	2.6	-.-	-.-	-.-
Paint Products	Spray Painting	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Cement, Limestone	Cement Handling	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Lead	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Sand, Bentonite	Clay Handling	----	0.6	-.-	-.-	-.-	-.-	-.-	-.-
Aluminum	Alumina Processing	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Zinc	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Misc. Chemicals	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-

Table 8. (Continued)

Dispersion Modeling Source Groups	Optical Microscopy Source Categories	CMB Modeling Source Types	Modeled Concentrations ($\mu g/m^3$)						
			Dispersion Modeled Impacts	Optical Microscopy			SSHV CMB	Fine Dichot CMB	Coarse Dichot CMB
				SSHV	Fine Dichot	Coarse Dichot			
Refinery Catalysts	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Misc. Refinery	Oil Refining	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Fertilizer	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Tire Products	----	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
----	Traffic, Tire Fragments	----	-.-	1.2	1.0	0.6	-.-	-.-	-.-
----	Traffic, <1um Carbon	Automobile Exhaust	-.-	2.4	1.2	1.6	2.8	4.6	2.0
Road Dust	Traffic, Wind Erosion	Near Street 1	41.2	40.2	10.2	5.4	7.4	-.-	20.6
----	Biologicals & Misc.	----	-.-	-.-	-.-	-.-	-.-	-.-	-.-
----	Secondary Aerosols	Secondary $(NH_4)_2SO_4$	-.-	8.0	1.0	4.4	5.4	4.4	0.8
Background	----	----	17.0	-.-	-.-	-.-	-.-	-.-	-.-
Total	----	----	60.2	70.8	30.2	19.9	38.1	15.8	29.6
Measured $PM_{2.5}$	----	----	19.9	-.-	-.-	19.9	-.-	19.9	-.-
Measured $PM_{2.5\text{-}10}$	----	----	30.2	-.-	30.2	-.-	-.-	-.-	30.2
Measured Dichot. PM_{10}	----	----	50.1	-.-	-.-	-.-	-.-	-.-	-.-
Measured SSHV PM_{10}	----	----	70.8	70.8	-.-	-.-	70.8	-.-	-.-

The first step taken to reconcile the disagreement between CMB and ISCST with respect to the combined impact of vehicle exhaust and road dust was to re-examine the comparisons between OM, CMB, and ISCST. This is shown in Tables 7 and 8 for the 2 days for which OM data are available. In this case, OM was found to agree very well with ISCST. The CMB input data were then evaluated for potential errors related to estimating road dust and vehicle exhaust. The coarse fraction filter on which the resuspended road dust source sample was collected appeared to have lost a substantial fraction of the sample between the time it was weighed and the time it was analyzed. Further evaluation of the sample indicated that this was indeed true. Thus, the correct mass was used to recompute the road dust profile.

The CMB model was rerun with the revised road dust profile and good agreement was now found between CMB and ISCST with respect to the combined impacts of road dust and vehicle exhaust. The CMB estimated impacts of the other sources remained unchanged by the revised road dust profile. Finally, the other source samples were re-weighed and no other cases of significant mass loss were identified.

4.3.4 Step 5 - Recompare Results After Rerunning CMB

The CMB model was rerun, the results were recompared and no significant difference in model estimates remained.

4.3.5 Step 6 - Refine Dispersion Model Inputs and Rerun

Case 2. A significant disagreement exists between the CMB and ISCST estimates of the coal combustion impacts.

The first step taken in reconciling the coal combustion impact estimates was to review the ambient and source composition data for errors which could potentially invalidate the CMB results. This review uncovered no obvious errors in the CMB input data related to coal combustion. The next step was to re-examine the comparisons between OM, CMB, and ISCST which are shown in Tables 7 and 8. These tables show that OM predicts coal combustion impacts in reasonable agreement with CMB but predicts larger impacts than ISCST. The emission inventory was then reviewed with respect to coal combustion sources. The review identified four major coal combustion facilities. For the sources closest to the receptor, the emission inventory contained a very high value for the efficiency of the emission controls. Discussions with the plant personnel revealed that the inventory contained an erroneous value and a correct value was obtained. ISCST was then rerun with the corrected inventory and much closer agreement was found between CMB and ISCST with respect to coal combustion impacts.

Case 3. CMB estimates for wood-fired boilers were much larger than the ISCST estimates.

The first step in reconciling the CMB and ISCST estimates for the wood-fired boiler impacts was to review the CMB input data. The review did not identify any obvious errors in the data. Next, the OM results were examined. In this case, OM was of little assistance because OM did not specifically identify any impact from wood-fired boilers while CMB and ISCST both did. The emission inventory was then evaluated in terms of wood-fired boiler sources in the inventory, and a site visit was conducted at each of the two plants. The site visits identified the possibility that one of the wood-fired boilers, located at a lumber mill, could be emitting particulate matter at a much higher rate than was listed in the inventory. Therefore, emission testing was conducted at the mill. The emission testing demonstrated that the emission inventory had severely underestimated the emission rate of the wood-fired boiler located at the lumber mill. ISCST was then rerun with the revised emission rate for the lumber mill boiler and good agreement was found between CMB and ISCST with respect to the wood-fired boiler source category.

Case 4. There was a big disagreement between CMB and ISCST for steel mill related impacts (i.e., coking operations, blast furnace, and basic oxygen furnace source categories).

The steel mill source was of particular concern because there was a big disagreement between CMB and ISCST estimates for each of the three main PM-10 emitting activities within the steel mill. This disagreement resulted in the CMB predicting that the steel mill is the largest industrial source of the PM-10 levels observed at the receptor site while ISCST predicts that the steel mill is a relatively minor source. This discrepancy had a potentially big impact on control strategy development and therefore it was reconciled very carefully.

The first step in the reconciliation process was to review the CMB input data. The review did not identify any obvious anomalies in the CMB input data. In fact, the steel mill related source profiles were judged to be of very good quality due to the close agreement which was found between the plasma emission spectroscopy analysis of the resuspended samples which were collected on quartz fiber and the XRF analysis of those collected on Teflon. The next step was to examine the OM estimated impacts for the steel mill. In this case, OM was found to agree much better with CMB than with ISCST. The emission inventory was then reviewed with respect to steel mill emission rates for similar activities at other steel mills. However, the review revealed that the emission inventory did not contain any fugitive emissions from the steel plant. This omission was viewed as a potentially serious deficiency; therefore, a site visit was conducted to reassess the stack and fugitive emission rates at the steel mill. The results of the stack testing were in good agreement with the original stack emission rates. However, substantial fugitive emissions were found to be associated with the coking operations, blast furnace, and basic oxygen furnace. ISCST was then rerun with the new fugitive emissions for the steel plant included in the inventory. This resulted in very good agreement between CMB and ISCST with respect to the steel mill contribution.

4.3.6 Step 7 - Recompare Results and Evaluate the Dispersion Model

The previous section presented the reconciliation of the CMB and ISCST source impact estimates and resulted in a number of revisions to the input data used by the models. Following Step 4, the CMB was rerun and after Step 6, the ISCST was rerun for subsets A and B using the revised data. The results obtained using the revised data are shown in Table 9. Good agreement now exists between the source impacts estimated by CMB and ISCST for subsets A and B. Comparison of the ISCST results for subset A with the "observed minus background" concentrations was also significantly improved. ISCST was then run using 5 years of meteorological data to obtain estimates of the annual average PM-10 and the six highest values. The results provided by ISCST are shown in Table 10. ISCST predicts that both the annual average and 24-hour PM-10 NAAQS will be exceeded at the monitoring station. Steel mill related activities are the major contributors to PM-10 at the receptor site. The second largest contributor is re-suspended road dust. The highest predicted PM-10 concentrations all occur with persistent south winds, which further substantiates the steel mill contributions.

4.3.7 Step 8 - Final Model Estimates

The dispersion model results have been improved after extensive intercomparison with the CMB results. Thus, the dispersion model results in Table 10 can be used for control strategy development.

Table 9. Comparison of Final Validation Results by ISCST and CMB for Subsets A and B ($\mu g/m^3$)

	(Representing Annual) Subset A		(Avg. of 5 24-hr runs) Subset B	
Source Categories	ISCST	CMB	ISCST	CMB
Oil & Gas Combustion	1.6	--	2.0	--
Wood-fired Boilers	3.5	3.4 (1.6)	6.8	5.5 (2.2)
Coal Combustion	2.3	2.6 (1.4)	9.2	8.2 (3.0)
Coking Operations	10.1	8.4 (3.7)	19.0	22.0 (8.5)
Blast Furnace &	a	11.6 (2.5)	a	31.6 (6.9)
Blast Oxygen Furnace	14.9	3.8 (1.7)	41.0	10.4 (2.6)
Coal Handling	0.3	--	0.6	--
Agricultural Prod. Handling	1.1	--	1.6	--
Paint Production	--	--	0.8	--
Cement & Limestone	1.9	--	1.9	--
Sand & Bentonite	1.2	--	1.7	--
Aluminum Production	--	--	0.4	--
Zinc Processing	0.8	--	0.6	--
Oil Refining	0.6	--	1.6	--
Fertilizer	--	--	0.8	--
Tire Production	--	--	2.2	--
Motor Vehicle Exhaust	15.1	2.2 (0.6)	2.6	2.6 (0.8)
Road Dust & Soil	--	15.1 (3.4)	31.2	39.2 (8.1)
Secondary $(NH_4)_2SO_4$	--	15.5 (2.7)	--	28.6 (4.4)
Background	15.0		32.0	--
Total	68.4	62.6	157.0	148.1
Measured	67.0	67.0	162.0	162.0

a. Estimated as a composite by the ISCST.

NOTE: Numbers in parentheses indicate one standard error in the CMB analysis.

Table 10. Source Contributions Estimated by ISCST using 5 Years of Meteorological Sata (μg/m^3).

Source Categories	Annual Average	24-hr 6th Highest Concentration
Oil & Gas Combustion	1.4	1.7
Wood-fired Boilers	3.0	6.0
Coal Combustion	2.0	7.5
Coking Operations	8.7	27.6
Blast Furnace	10.1	38.0
Basic Oxygen Furnace	2.7	14.8
Coal Handling	0.3	0.4
Agricultural Prod. Handling	0.9	0.4
Paint Production	0.1	1.1
Cement & Limestone	1.6	2.1
Sand & Bentonite	1.0	0.7
Aluminum Production	0.1	0.2
Zinc Processing	0.6	0.6
Oil Refining	0.3	2.1
Fertilizer	0.1	0.9
Tire Production	0.4	1.5
Motor Vehicle Exhaust	12.7	45.4
Road Dust & Soil	--	--
Background	15.0	32.0
Total	61.0	183.0

5. REFERENCES

Anderson, M.K., R. DeCesar, R. Longergan, and E. Brookman, *Example Modeling to Illustrate SIP Development For the New Particulate Matter NAAQS*, Prepared by TRC Environmental for the U.S. EPA under contract 68-02-3886, work assignments 37, 48 and 52, (1986).

Core, J.E. and T. G. Pace, "Receptor Models - How Great Thou Art?!!", APCA Paper No. 81-58.6, Air Pollution Control Association, Pittsburgh, PA (1981).

Core, J.E., J.A. Cooper, P.L. Hanrahan, and W.M. Cox, Particle Dispersion Model Evaluation: A New Approach Using Receptor Models, *J. Air Pollut. Contr. Assoc.* 32:1142-1147 (1982).

DeCesar, R., personal communication to T. Pace, February 1987.

Pace, T.G. and J. G. Watson, *Protocol For Reconciling Differences Among Receptor and Dispersion Models*, Environmental Protection Agency Report No. EPA 450/4/87-008, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987).

Throgmorton, J. and K. Axetell, *Digest of Ambient Particulate Analysis and Assessment Methods*, Environmental Protection Agency Report No. EPA-450/3-78-113, U.S. Environmental Protection Agency, Research Triangle Park, NC (1978).

Trijonis, J., *Model Reconciliation*, Special Report prepared for U.S. EPA and TRC Environmental Consultants under Contract 68-02-3886, Work Assignment No. 13, Research Triangle Park, NC, (1985).

U.S. EPA, *Receptor Model Technical Series, Volume I: Overview Of Receptor Model Application to Particulate Source Apportionment*, Environmental Protection Agency Report No. EPA-450/4-81-016a, PB82-139429, U.S. Environmental Protection Agency, Research Triangle Park, NC (1981a).

U.S. EPA, *Receptor Model Technical Series, Volume II: Chemical Mass Balance*, Environmental Protection Agency Report No. EPA-450/81-016b, PB82-187345, U.S. Environmental Protection Agency, Research Triangle Park, NC (1981b).

U.S. EPA, *Receptor Model Technical Series, Volume IV: Technical Considerations in Source Apportionment By Particle Identification*, Environmental Protection Agency Report No. EPA 450/4-83-018, PB84-103340, U.S. Environmental Protection Agency, Research Triangle Park, NC (1983).

U.S. EPA, *Receptor Model Technical Series, Volume V: Source Apportionment Techniques and Considerations in Combining Their Use*, Environmental Protection Agency Report No. EPA 450/4-84-020, PB85-111524, U.S. Environmental Protection Agency, Research Triangle Park, NC (1984).

U.S. EPA, *Receptor Model Technical Series, Volume VI: A Guide to The Use of Factor Analysis and Multiple Regression (FA/MR) Techniques in Source Apportionment*, Environmental Protection Agency Report No. EPA 450/4-85-007, PB86-107638, U.S. Environmental Protection Agency, Research Triangle Park, NC (1985a).

U.S. EPA, *Receptor Model Source Composition Library*, Environmental Protection Agency Report No. EPA 450/4-85-002, PB85-228823, U.S. Environmental Protection Agency, Research Triangle Park, NC (1985b).

U.S. EPA, *Guideline On Air Quality Models (Revised)*, Environmental Protection Agency Report No. EPA-450/2-78-027R, U.S. Environmental Protection Agency, Research Triangle Park, NC (1986).

U.S. EPA, *Receptor Model Technical Series, Volume III (Revised): User's Manual For Species Mass Balance Computer Program*, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987a).

U.S. EPA, *PM-10 SIP Development Guideline*, Environmental Protection Agency Report No. EPA 450/2-86-001, Research Triangle Park, NC 27711. February 1987b.

U.S. EPA, *Protocol for Reconciling Differences Among Receptor and Dispersion Models*, Environmental Protection Agency Report No. EPA 450/4/87-008, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987c).

U.S. EPA, *Protocol for Applying and Validating the CMB Model*, Environmental Protection Agency Report No. EPA 450/4-87-010, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987d).

J. G. Watson, and T. G. Pace, *Protocol For Applying and Validating The CMB*, Environmental Protection Agency Report No. EPA 450/4/87-010, U.S. Environmental Protection Agency, Research Triangle Park, NC (1987).

Chapter 9

APPLICATION OF RECEPTOR MODELING TO SOLVING LOCAL AIR QUALITY PROBLEMS

John E. Core
Air Quality Division
Oregon Department of Environmental Quality
P.O. Box 1760
Portland, OR 97207

Credibility is a most important attribute in solving local air quality problems. Credibility is not only important in dealing with those that are the facing potential regulation, it is also an essential element of the technical work that forms the basis for control programs. In the absence of technical integrity, control strategy issues becomes difficult to resolve and subject to increasing degrees of skepticism resulting in an absence of community support. Air quality managers must be able to convincingly demonstrate their understanding of the relative contributions of airshed sources not only as a mean of resolving control strategy issues but, more importantly, to assure themselves that their programs make good sense. Unless an effective case, demonstrating the magnitude and identity of source impacts can be made, considerable doubt as to the likely success of new control programs will exist.

This Chapter describes how Receptor Models have been applied in real-world settings to resolve local air quality problems and improve confidence in the effectiveness of emission reduction programs. It focuses on case studies where receptor modeling techniques have played a major role in reshaping air manager's understanding of local airshed problems.

1. INTRODUCTION

Within the past ten years, receptor models have played an increasingly important role in air quality management. Receptor modeling application which initially focused on particulate matter (TSP and PM_{10}), have recently expanded to include several new fields. Receptor models have been applied to apportion sources causing visibility impairment; to toxic air pollutants, ozone precursors (volatile organic compounds) and regional-scale acid deposition aerosols and are just beginning to be applied to groundwater pollution studies. The technology has been applied in China, Japan, Brazil, Mexico and other countries. Within the United States, receptor modeling studies that in 1978 were limited to no more than two or three cities are now common place.

In the field of acid deposition and long range transport research, receptor modeling studies conducted at the University of Rhone Island (Rahn and Lowenthal, 1984) suggested that the ratio of indium to noncrustal manganese could be used as a simple receptor model to indicate the presence of sulfate enriched air masses originating from Ohio River Valley. At the same time, other researchers at the University of Texas were applying chemical mass balance methods to identify sources of acidic deposition in Southern California (Liljestrand, 1982). More recently, the feasibility of applying receptor modeling technology to identify the origin of air masses in the Western States (Watson *et al.*, 1987) has been evaluated as a means of identifying the region of origin of acidic sulfates transported over long distances. This work has concluded that the principal limitation to be overcome in applying the technology are those related to development of the correct experimental design; that the modeling technology was sufficiently advanced and that given the proper regional source profile and ambient aerosol data sets, it may be feasible to determine the origin of an air mass associated with an acid deposition event within reasonable limits of error.

Although the application of receptor modeling technology has diversified, the fundamental framework of the modeling has remained largely unchanged. Receptor modeling theory, computer software, source profile development, ambient air monitoring and laboratory methods have, however, been refined as the understanding of the science has grown.

Three key events have caused the rapid growth and acceptance of receptor modeling within the regulatory community:

- The source "profile" or "fingerprint" libraries have become readily available to receptor modelers; the libraries have grown substantially both in quantity and quality such that national databases have now been established by the Environmental Protection Agency;

- Receptor modeling software that was once available to only a small handful of researchers is now widely available to users. EPA's implementation of the Chemical Mass Balance model on microcomputer systems, documentation and regulatory agency training transformed the technology from a research to an air quality management tool and;

- Advancements in the development of automated, multi-element x-ray fluorescence and ion chromatography techniques brought the sensitivity and cost of laboratory analytical techniques within the reach of regulatory agency budgets.

The advancement of the technology was spurred by EPA's plans to revise the national ambient air quality standards for particulate matter and by the need for new, more advance control strategy development methods. Following nearly ten years of study, EPA in July, 1987 acted to revise the National Ambient Air Quality Standard (NAAQS) for particulate matter to a standard defined by the collection of aerosols with a 50 % collection efficiency at 10 micrometers, thereby triggering the three year timeline specified in the Clean Air Act for submission of SIPs.

Referred to as PM_{10}, the new NAAQS require that state and local air pollution regulatory agencies revise their State Implementation Plans (SIPS) to assure that the PM_{10} standard is attained and maintained throughout the nation. The PM_{10}

SIPs are required to include a technical assessments of PM_{10} air quality, demonstrate the adequacy of proposed control strategies and the legally enforceable means through which emission reductions will be made.

Development of credible emission control strategies requires an in-depth and accurate assessment of the relative contributions of airshed sources. The assessment must not only be technically defensible, it must also be credible in the eyes of elected officials, industry representatives and the public. As noted above, credibility is critical. Because of the direct economic and life-style impacts that control strategies often have on the community, it is extremely important that corroborating evidence be developed to provide independent verification of the accuracy of the source contribution assessment. It is within this setting that receptor modeling has become an established, accepted science.

Until quite recently air quality regulatory professionals relied on emission inventory estimates of source strengths and dispersion modeling source impacts as their sole sources of information. Although the science of urban dispersion modeling has grown in sophistication, scope of application and importance, dispersion models remain severely limited in their ability to simulate the extreme complexity of atmospheric dynamics in mountainous terrain or during the extended calm wind episodes that typify PM_{10} nonattainment area of the Western United States. It is exactly for these reasons that receptor modeling has been so appealing. Recent advances in receptor modeling methods provide new opportunities to advance the accuracy of urban dispersion models by providing an <u>independent</u> means though which dispersion modeling results can be verified (Core and Cooper, 1982). This ability to independently confirm source contribution estimates greatly strengthens one's ability to demonstrate the adequacy of a strategy. The joint application of both forms of modeling to air resource management and policy development has important strategy development advantages that are explored in this Chapter.

2. THE ROLE OF RECEPTOR MODELS IN AIR QUALITY MANAGEMENT

The adoption of the PM_{10} National Ambient Air Quality Standard in July of 1987 ushered in the development of yet another series of control strategies for many airshed in the nation. When air quality agencies began adopting Total Suspended Particulate (TSP) strategies in the early 1970's, the technical basis of programs was founded on emission inventory proportional rollback models. In the late 1970's, when it became clear that the strategies were not effective in attaining the TSP air quality standard and major efforts were made in many communities to revise State Implementation Plans (SIPs) based on emission inventory and dispersion modeling analysis of source contributions.

Unfortunately, the effectiveness of particulate control programs was limited by failure of the emission inventories to accurately estimate emissions of fugitive dust and other area sources, sources now known to be of critical importance in many nonattainment areas. In addition, the inability of dispersion models to simulate formation of secondary aerosols in the atmosphere was a major problem in areas where as much as one-half of the aerosol mass was of secondary origin. Thus, in spite of the fact that cumulative air pollution abatement costs have exceeded $300 billion over the decade ending in 1987, hundreds of counties remained in noncompliance with the TSP NAAQS at the time the standard was repealed (Cass and McRae, 1981).

In addition to emission inventory limitations, many of the control strategies that have not incorporated receptor modeling analysis have been flawed because of limitations inherent in dispersion models. Principal among these are the inability of source oriented models to quantify source impacts during short term episodes, assess impacts in complex terrain or apportion particulate mass in complex urban airsheds. The inability of the models to quantify source impacts to the "background" aerosol being transported into an airshed - aerosol that typically accounts for 50 % or more of the urban PM_{10} mass - is another serious limitation.

Many of these constraints are inexorably linked to the inability of emission inventories to accurately reflect hourly and day-to-day emission variations typical of many urban settings. Improvements to currently available dispersion models will require additional research to cope with the complex and often random nature of atmospheric dispersion under calm wind conditions, of particle deposition, variations in emission variability and secondary aerosol formation pathways.

As state and local agencies begin development of yet a third series of particulate control strategies (this time for PM_{10}), it has become clear that air quality managers must better understand source-receptor relationships if their programs are to remain credible. Local governmental officials, industry representatives and the public are all aware of the limited success of earlier strategies.

It is also important to understand that those that are affected by source emission control strategies have developed a much more sophisticated understanding of the science of air quality source impact assessment than was found during earlier rounds of control strategies. Members of the regulated community as well as concerned public groups are very interested in the technical details of source impact assessments conducted by regulatory agencies. In Oregon, PM_{10} advisory groups have actively questioned emission inventory, dispersion and receptor modeling assumptions as well as the quality of ambient air quality data. In some nonattainment area, interested persons have developed their own, independent assessments of source contributions. In working with these groups, the common-sense simplicity of receptor modeling fundamentals has been well received because the public can understand the technology, the results (hopefully) make sense and can be easily validated through comparisons with other, independent, information.

3. CONTROL STRATEGY DEVELOPMENT USING RECEPTOR MODELS

The task of designing control programs is to determine a set of emission limits for sources spread throughout an airshed that will reduce air pollutant concentrations measured at specific monitoring sites to levels at or below the National Ambient Air Quality Standards. The control strategies must be economically cost-effective, legally enforceable, equitable in the eyes of community leaders (and the public) and achievable within available resources and the time frames provided in the Clean Air Act. Perhaps most importantly, the strategy must be effective in achieving healthful air quality.

The solution to the control strategy design problem requires data on emission strengths, meteorology, ambient air quality and applicable emission control technology. This information is used with air quality dispersion and/or receptor models to quantify source contributions which, in turn, are used in association with information on emission control technology and cost of control to determine the relative air quality benefits of the strategy. In the case of particulate matter strategies, information on the relative health risks associated with various sources can then be used to narrow the number of options.

Receptor models can play an important part in this process from apportioning the sources contributing to nonattainment to tracking the effectiveness of a strategy once it has been implemented.

The first step in strategy development is determination of the 24-hour level from which air quality must be improved to meet ambient air quality standards. This concentration is referred to as the 24-hour worst-case "design value" which represents (based on the past three years of PM_{10} data) the 1/365th percentile of the PM_{10} distribution. This is the PM_{10} concentration that will not be exceeded more than once per year over a three year period. Once this value is determined, air managers must then apportion the design value PM_{10} mass among the contributing sources. Emission growth factors estimating the percentage change in emissions from each source category between the current year and the date on which standard attainment must be reached (defined by the Clean Air Act as three years) are then applied for each category. The end product is an attainment year 24-hour worst case PM_{10} design value and source contribution estimates which form the basis of the control strategy. A similar process is followed to arrive at the attainment year, annual PM_{10} design value and source apportionment.

The ability of receptor modeling methods (especially chemical mass balance, electron microscopy and analytical methods that can be applied to a single, 24-hour sample) to apportion the source contributions based on measurements from actual samples with PM_{10} mass loadings at the current year design value is critically important. First, experience has consistently shown that available dispersion models cannot reliably estimate measured PM_{10} mass concentrations at design value levels because they occur during prolonged periods of calm winds. Secondly, even if adjustments to the model meteorological data base, emission inventories and dispersion assumption are made, they are unlikely to properly apportion the aerosol mass among the contribution sources. Recent efforts to dispersion model one Oregon PM_{10} nonattainment area required application of three different models (a valley stagnation model, RAM and ISC) over a period of many months. This effort underscored the need to validate dispersion modeling results against receptor model derived source contributions to assure that the dispersion modeling results are reasonable. In several airsheds, Chemical Mass Balance receptor modeling pointed out serious area source fugitive dust emission inventories errors that had to be corrected before dispersion modeling efforts could begin.

3.1 Oregon's Receptor Modeling Program

The State of Oregon Department of Environmental Quality's receptor modeling program began in 1976 when it was realized that the Total Suspended Particulate SIP control strategy adopted for Portland in 1970 had failed to attain the National Ambient Air Quality Standards and that the airshed's nonattainment status was hindering community economic growth. The Department believed that the control strategies in force at the time had minimized emissions from all significant sources based on the airshed emission inventory. This included a 60,000 ton per year reduction in industrial emissions. The Department held the firm belief that economic development must be allowed to occur within the constraints of healthful air quality. As a result, a major effort was undertaken to determine the sources of Portland TSP aerosol. The Portland Aerosol Characterization Study (PACS), the first airshed aerosol study to specifically designed for receptor modeling applications, was to become the cornerstone of this effort (Watson, 1979).

3.1.1 Portland Aerosol Characterization Program (PACS)

A new approach to identifying contributing sources was badly needed if a new round of emission controls was to be successful. A turning point in development of the PACS program was reached when the concept of Chemical Element Balance Receptor Modeling was reviewed (Friedlander, 1973). The strength of the approach was it was based on actual analysis of aerosols captured on filters which were also being used to determine compliance with the NAAQS, the approach was conceptually simple and could be applied to single samples to characterize short term air quality episodes. Following technical review of a number of alternatives, a comprehensive plan based on Receptor Modeling was adopted.

In Portland, certain sources were easily identified using "tracer" elements associated with the sources. Automotive exhaust, for example, was by far the largest source of lead while a ferromanganese alloy manufacturing facility was easily identified through measurements of ambient manganese concentrations on TSP filters.

The overall PACS program consisted of a five step process; (1) identification of source contributions to TSP and the fine particle fraction (< 2.0 μm) using the Chemical Mass Balance method, (2) dispersion modeling of source contributions using emission inventories and meteorological data collected during the one year period of ambient sample collection, (3) comparison of the CMB and dispersion model predicted source contributions, (4) reconciliation of the CMB and dispersion modeling results to validate the dispersion model and (5) application of the dispersion model in developing control strategy alternatives.

After one year of staff work to design the field study, raise funding and review meteorological, emission inventory and dispersion modeling adequacy, the first step of the PACS program began. Participation by local governments, industry and community leaders was solicited to assure that all sectors of the community could have confidence in the study results. A public advisory committee was formed to help guide the project and review early drafts of the project results.

PACS was the first field study designed specifically around receptor modeling. It was a three year effort that culminated in development of the following technologies:

1. A source testing train designed to collect fine (<2.0 µm) and coarse (2.0 - 20 µm) aerosol samples suitable for elemental, ion and carbon analysis thereby providing source profiles of all of Portland's major emission sources;

2. The first comprehensive, size resolved "Source Library" of particulate emission profiles, many of which are in wide use today;

3. An expanded, effective variance least squares Chemical Mass Balance Model which formed the basis for the EPA CMB microcomputer software that is widely used today;

4. An aerosol chemistry database consisting of measurements of 27 chemical species on over 2,000 TSP and fine fraction aerosol samples collected at six sampling locations over one calendar year. The data base has been used by a number of researchers to advance receptor modeling technology;

5. A medium-volume, sequential filter aerosol sampling system capable of collecting up to 12, 47 mm filters unattended over any preprogrammed period. The sampler pulls ambient air at 4 cfm flow rate through an appropriate inlet system to obtain the desired upper cut point. A duel port filter system capable of simultaneously collecting aerosol on both Teflon and quartz filter substrates is used to accommodate both thermal-optical carbon and X-ray fluorescence/ion chromatography analytical methods. Because of the excellent agreement between the sampler and EPA's reference methods for

PM_{10}, the sampler (when equipped with a Sierra-Anderson 254 inlet) has been recently been designated by EPA as an acceptable equivalent method.

6. Development of a conservation of mass, advection-diffusion dispersion model (GRID) suitable for use in the complex terrain that typifies the Portland area. The model is still in use within the Department.

Perhaps more importantly, the research conducted during the PACS program (and the many researchers that participated in the program) became the genesis of much of the receptor modeling technology in use today. The concept of source/receptor model reconciliation also came out of the PACS program (Core, 1982). Key results from the program included:

1. Recognition that soil and road dust were major contributors to the particulate loading, identifying a 19,400 ton per year deficiency in the paved road dust emission inventory.

2. The first understanding that vegetative burning (principally residential wood burning) was a major source of particulate accounting for as much as 40% of the fine particulate mass on worst-case winter days. This work led to the addition of 6,500 ton per year TSP to the Portland airshed emission inventory.

3. The finding that industrial sources (previously thought to be the largest contributor to the airshed's nonattainment problem) accounted for only 5% of the TSP mass loading.

The above findings dramatically altered air quality manager's understanding of the relative importance of industrial, road dust and residential wood burning in

Portland and led to adoption of a new series of control strategies to minimize emissions from winter road sanding. The "discovery" of the magnitude of emissions from residential wood stoves led to the 1985 adoption of the Oregon Wood stove Certification Program which, in turn, was used as the basis for the EPA's national wood stove certification program.

4. RECEPTOR MODEL CASE STUDIES

Immediately following completion of the Portland Aerosol Characterization Study, the technology was applied in two new airshed:

4.1 Willamette Valley Airshed Study

The Willamette Valley TSP nonattainment problem had long been believed, on the basis of emission inventories and public perception, to be caused by smoke generated from agricultural field burning, a practice that has long been the subject of thousands of public complaints each year. About 250,000 acres of annual and perennial rye, fescue and other varieties of grass seed are burned by growers each year with as much as 65,000 acres burned per day. It is not uncommon for street lights to be turned on at mid day as the skies are darkened by the smoke. To help resolve culpability for TSP air quality standard exceedances in the Valley, a receptor modeling study was funded to better define field burning's contribution to the problem.

The findings of the receptor modeling study largely exonerated field burning smoke and pointed to fugitive dust as a major contributor to TSP air quality (Core, 1982). Field burning smoke was, however, shown to be a major source of fine fraction aerosol that severely impaired Willamette Valley visibility. As a result of the Willamette Valley study, air quality regulators shifted their TSP standard attainment strategies to reduction of unpaved road dust emissions, pressures on the agricultural community were lessened and air quality manager's perceptions as to the source of the Valley's TSP nonattainment problem were dramatically changed.

Today, pressures to phase out or ban agricultural field burning continue as the public's tolerance of the smoke grows. The issue, however, is not based on TSP or PM_{10} nonattainment but on visibility impairment, nuisance and public health effects.

4.2 Medford Aerosol Characterization Study

The Medford Aerosol Characterization Study (MACS) was conducted concurrent with the Willamette Valley Study to again determine the sources contributing to TSP air quality nonattainment. Unlike Portland, relative few of the major industrial sources in the Medford airshed had at the time been well controlled. This led to the supposition (supported by emission inventory studies) that wood products industry emissions were a major contributor to the TSP nonattainment problem. Since Medford's air quality was consistently worst on winter days characterized by long periods of stagnation, the usefulness of dispersion modeling analysis was questionable without some independent means of validating the modeling effort.

A one year emission inventory, meteorological and receptor - dispersion modeling study was undertaken to develop the database necessary to (a) apportion TSP and fine fraction source impacts on a winter worst case and annual average basis; (b) support development and application of dispersion models and (c) reconcile results from the two modeling studies.

Results from the MACS study were highly useful to the Department in that they resolved the major wood products industry emission sources (hogged fuel boilers) from other sources with a satisfactory margin of uncertainty; identified residential wood stove smoke and fugitive dust sources as significant contributors and provided a basis for dispersion model validation (DeCesar and Cooper, 1980).

Had the Medford Aerosol Characterization Study not been funded, control strategies for wood stoves and fugitive dust emission would not have been adopted and public perceptions that the highly visible plumes from industrial emissions were the major cause of TSP nonattainment would have continued.

In each case where receptor models have been applied in Oregon, the results have consistently resulted in changes in air quality manager and public perceptions of the problem. In every case, results have led to revisions of airshed emission inventories, most commonly for fugitive dust sources.

5. BARRIERS TO PRACTICAL APPLICATIONS OF RECEPTOR MODELS

Prior to 1983, the greatest barrier to widespread application of receptor models by state and local air pollution agencies was the availability of computer software, source profile information and technical guidance on the use of receptor models. With completion of the EPA-Desert Research Institute Chemical Mass Balance program for the IBM PC, distribution of the EPA Source Profile Library (USEPA, 1984) and completion of the six volumes of the EPA Receptor Model Technical Series, the technology fell into widespread use.

Although major advancement have been made, the current modeling techniques and source profile information still falls short of resolving impacts of sources with similar emission profiles (collinear sources). Receptor models still cannot be applied to apportionment of secondary aerosols which, in many airsheds, makes up the majority of the aerosol mass loading.

Within the Pacific Northwest, most of the PM_{10} nonattainment areas are impacted by smoke generated by biomass burning including residential wood stoves, forestry slash, agricultural burning and land clearing. Apportionment of impacts form these sources is very important in developing Northwest control strategies yet receptor models are unable to resolve smoke impacts from these sources with an adequate level of certainty.

New efforts to improve the ability of the Chemical Mass Balance model to resolve collinear sources have focused on (a) expanding source profiles to include more chemical species that may help differentiate the sources, (b) tightening the uncertainties associated with measurements of both the source profile and ambient aerosol species and (c) modification of the CMB software to permit inclusion of an unlimited number of fitting species in the calculation. Important advancements have been made in all three areas:

> Expanded Source Profiles increasing the number of species from (typically) 35 or 40 species to 75 species are now available fifteen point and 60 area sources common to the Pacific Northwest States. The Pacific Northwest Source Profile Library (Core, 1989), was the first program of it's kind to include GC/MS measurement of 25 organic (including 11 PAH) compounds in the profiles in an attempt to improve the ability of the CMB model to revolve impacts from biomass combustion. As both inorganic and organic compound analytical technology advances, more species will be included in future emission profile studies thereby improving the ability of Receptor Models to resolve sources and improving the relative standard errors of the source contribution estimates.
>
> Improvements in Measurement Uncertainties are essential to advancing the ability of receptor models to resolve sources. Synthetic data set studies conducted as part of the Pacific Northwest Source Profile Library Project have shown that inclusion of organic compounds in CMB modeling is only useful if the source profile and ambient aerosol measurement uncertainties can be reduced from the current relative uncertainty level of 80-100 % to less than ten percent. If the organic compound uncertainties are markedly greater than those associated with the inorganic species, the organics are not influential in fitting process and were therefore not helpful in resolving the sources. Lowering the relative uncertainties of organic compounds will require advancements in analytical techniques to minimize extraction/desorption variability and improve system calibration reproducability.

Modification of Receptor Modeling Software to permit inclusion of an unlimited number of fitting species was accomplished with distribution of EPA-DRI Version 7.0 CMB software. This improvement and the advancement of personal computer speed have eliminated a barrier that has long restricted regulatory agency applications.

An additional barrier has been the lack of availability of commercially available air sampling instrumentation suitable for collection of aerosol samples used in receptor modeling studies. Since these studies typically require analysis of fine (< 2.5 µm) and coarse (2.5 - 10 µm) fraction aerosol samples concurrently on quartz and Teflon filter substrate, the only option that most air regulatory agencies have is to operate two collocated dichotomous samplers, one equipped with quartz and the other with Teflon filters. Unfortunately, these samplers have some distinct disadvantages:

- Samples are often lost during heavy filter loading conditions (heavy concentrations of wood smoke with fog) that are of critical importance to control strategy development;

- Since daily sampling is required in Group I PM_{10} nonattainment areas, at least four dichotomous samplers are actually required to meet scheduling needs, increasing the cost of the monitoring program greatly.

For these reasons, few regulatory agencies have designed their PM_{10} monitoring programs to accommodate Receptor Modeling sample analysis needs. Those agencies that do, rely on equipment that is not commercially available. The backbone of the Oregon PM_{10} monitoring program, for example, relies on a dual-port, programmable sequential filter sampler (discussed above) that has been designated by EPA as an equivalent PM_{10} monitoring method. The monitor concurrently collects samples on quartz and filter substrate and sequences to a new set of filters if the pressure drop across the filter drops below a preset point. One week of daily samples can be collected automatically, greatly reducing servicing costs. The instruments are built by Oregon Department of Environmental Quality staff for use in the state. Since they are used on a routine basis throughout the monitoring network, samples suitable for organic carbon and elemental/ion chromatography analysis are readily available. Sample analysis is conducted on a routine basis of all samples that exceed the 24 hour National Ambient Air Quality Standard (NAAQS) of 150 $\mu g/m^3$, assuring that analytical data will be available to allow apportionment of all samples exceeding the 24 hour NAAQS.

Because few regulatory agencies have the facilities, staff and experience needed to design and construct their own sampling equipment, development of a commercially available PM_{10} reference or equivalent instrument would eliminate an major barrier to Receptor Model application.

The final barrier to applying Receptor Models is one of cost. Many agencies do not have the funds needed to collect, analyze and compile aerosol chemistry and source profile data needed to support Receptor Modeling activities.

6. FUTURE APPLICATIONS

Although particulate matter Receptor Modeling applications will continue to dominate the technology, applications to, volatile organic compound, air toxics, groundwater pollution, acid deposition and gaseous pollutants apportionment show promise. Funding will be needed to expanded source profile libraries for toxics, VOCs and gaseous pollutants; analytical and air sampling methods, technical guidance and funding. All of these applications parallel the development of particulate Receptor Modeling in that major advancements in new applications at the regulatory level were not made until all of the above elements were in place.

Applications to apportioning the sources of volatile organic compounds (VOCs) are especially important since they are the principal precursors being regulated under state and local ozone control strategies. The success or failure of the strategies is largely dependent on the strength of local VOC emission inventories and, therefore, on the reliability of VOC emission factors. Failure of many of the nation's largest metropolitan areas to attain the ozone NAAQS has prompted a hard look at the validity of these emission inventories and spurred interest in using Receptor Models as an independent means of judging the validity of the inventories. Given the serious and widespread nature of the ozone nonattainment problem and the potential application of Receptor Models to the problem, it is highly likely that future applications will expand in this area. Source profile compilations to support Receptor Modeling efforts are being developed (Scheff *et al.*, 1988) and a national databases of concentrations of 320 ambient VOC compounds totalling 19 megabytes of data is now available for use (Shah, 1988).

For similar reasons, air toxics applications will also expand in areas where there is serious question as to the origin of toxic compounds. Typical research studies have focused on apportionment of source contributions of dioxins and dibenzofurans (Edgerton, 1988) and toxic organic compounds including benzene, trichlorethylene and carbon tetrachloride in urban atmospheres (Scheff *et al.*, 1988).

An exciting new application area is that of groundwater pollution. New rare earth elemental source tracers have been used to identify the source and extent of water pollution in an area containing oil and coal ash deposits downstream of a large landfill area (Olmez, 1988).

As the issues of long range acid deposition and regional haze become more volatile, new application for Receptor Models will be seen, expanding relatively simple elemental tracer approaches (Rahn and Lowenthal, 1984) to regional, multi-element multivariate and mass balance approaches (Gorse *et al.*, 1988) while inclusion of stable perfluorocarbon isomers emitted for large metropolitan areas (McDade and McGown, 1988) will enhance future applications long range transport applications.

7. CONCLUSIONS

The chemical and physical characteristics of air pollutants, whether in particulate or gaseous form, hold a wealth of information on their source. Determining the characteristic "profiles" of industrial, area sources and urban airsheds is the key to unlocking this information. As more information describing the nature of these emissions becomes available, applications of Receptor Models will grow in scope and breadth beyond particulate matter source apportionment to identifying sources of regional haze, acid deposition, air toxics and ozone precursors.

The strength of Receptor Models and their application to air resource management lies in the fact that source contribution estimates are based on the chemical and morphological properties of pollutants in the atmosphere and how these pollutants vary in time and space. These are concepts that can be easily understood by air resource managers and the public and can be independently verified by other physical measurements. As a result, air quality managers can convincingly demonstrate their understanding of the relative contributions of airshed sources not only as a mean of resolving control strategy issues but, more importantly, to assure themselves that their programs make good sense. It is for these reasons that Receptor Modeling has, and will continue to play, an important role in air resource management.

8. REFERENCES

Cass, G.R., and G.J. McRae, Minimizing the Cost of Air Pollution Control, *Environ. Sci. Techol.* 155:748-757 (1981).

Core, J.E., *Willamette Valley Field and Slash Burning Impact Study*, Final Project Report to the State of Oregon Department of Environmental Quality (1982).

Core, J.E., J.A. Cooper, P.L. Hanrahan and W.M. Cox, Particulate Dispersion Model Evaluation: A New Approach Using Receptor Models, *J. Air Pol. Control Assoc.* 32:1142-1147 (1982).

Core, J.E., J.A Houck, J.A. Cooper, *Pacific Northwest Source Profile Library Project*, Final Report (3 Volumes) to US Environmental Protection Agency Region X (1989).

DeCesar, R.T. and J.A. Cooper, *Medford Aerosol Characterization Study (MACS)*, Final Report to the State of Oregon Department of Environmental Quality. 1980.

Edgerton, S.A., Source Apportionment of Dioxins and Dibenzofurans in Ambient Air in Ohio, in *Receptor Modeling in Air Resource Management*, J.Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 59-70 (1988).

Friedlander, S.K., Chemical Element Balance and Identification of Air Pollution Sources, *Environ. Sci. Technol.* 7:235-245 (1973).

Gorse, R.A., W.W. Brachaczek, S.M. Japar, G.L. Keeler and W.R. Pierson, Source Apportionment of Rural Elemental Carbon Aerosol Using Several Multivariate Statistical Analysis Approaches to Receptor Modeling, in *Receptor Modeling in Air Resource Management*, J.Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 418-431 (1988).

Liljestrand, H.M., Acidic Precipitation Source Identification by Chemical Mass Balance Methods Employing Fractionation Factors, in *Receptor Models Applied to Contemporary Pollution Problems*, S.L. Dattner and P.K. Hopke, eds, Publication No. SP-48, Air Pollution Control Association, Pittsburgh, PA, 212-213 (1982).

McDade, C.E., M. R. McGown and J.G. Watson, Identifying Long-Range Tracers of Los Angeles Emissions, in *Receptor Modeling in Air Resource Management*, J. Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 407-416 (1988).

Olmez, I., Trace Element Signatures in Groundwater Pollution, in *Receptor Modeling in Air Resource Management*, J.Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 3-11 (1988).

Rahn, K.A., and D.H. Lowenthal, Elemental Tracers of Distant Regional Pollution Aerosols, *Science* 223:132-139 (1984).

Scheff,P.A., P.F. Aronian, R.A. Wadden and B.A. Bates, Development of Volatile Organic Carbon Source Fingerprints For Receptor Modeling, in *Receptor Modeling in Air Resource Management*, J.Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 84-107 (1988).

Shah, J.J., VOC Receptor Modeling: Ambient Data Base and a Case Study, in *Receptor Modeling in Air Resource Management*, J.Watson, ed., Transactions No. 14, Air & Waste Management Association, Pittsburgh, PA, 20-38 (1988).

U.S. EPA, *Receptor Model Source Composition Library,* Environmental Protection Agency Report No. EPA 450\4-85-002, U.S. Environmental Protection Agency, Research Triangle Park, NC (1984).

Watson, J.G., *Chemical Element Balance Receptor Model Methodology for Assessing the Sources of Fine and Total Suspended Particulate Matter in Portland*, Oregon, Ph.D. Dissertation. Oregon Graduate Center, Beaverton, Oregon (1979).

Watson, J.G., J. C. Chow and N.F. Robinson, *Western States Acid Deposition Project Phase I: Volume 4..An Evaluation of Ambient Aerosol Chemistry in the Western United States*, Final Report to the Western States Acid Deposition Project, Report No. SYSAPP-87/064, Systems Applications Inc. San Rafael, CA (1987).

INDEX