CONTRACT NO. A832-131
FINAL REPORT
JANUARY 1999

Multivariate Receptor Modeling of SCAQS VOC and Airborne Particle Composition Data

CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY

AIR RESOURCES BOARD
Research Division

Multivariate Receptor Modeling of SCAQS VOC and Airborne Particle Composition Data

Final Report
Contract No. A832-131

Prepared for:

California Air Resources Board
Research Division
2020 L Street
Sacramento, CA 95814

Prepared by:

Ronald C. Henry
Yi Mi
William Moran

University of Southern California
Department of Civil Engineering
Environmental Engineering Program
3620 South Vermont Avenue
Los Angeles, CA 90089-2531

January 1999

For more information about the ARB's Research Division,
its research and activities, please visit our Web site:

http://www.arb.ca.gov/rd/rd.htm

University of Southern California

Department of Civil Engineering

Environmental Engineering Program

3620 South Vermont Avenue

Los Angeles, CA 90089-2531

Summary

Innovative multivariate receptor modeling techniques were applied to several types of data collected during the Southern California Air Quality Study (SCAQS). The purpose of the modeling was to deduce as much information about the sources as possible from the data alone and with the aid of some basic physical constraints. The data examined were the Volatile Organic Compound (VOC) gas phase data, the airborne particle composition data, and routine gas monitoring data. The models applied were the Source Apportionment by Factors with Explicit Restrictions (SAFER) multivariate receptor model, and the Source Identification Through Empirical Orthogonal Functions (SITEOF) hybrid source – receptor model.

The VOC data from all sites for the fall intensive were combined into a singe data set that had a sufficient number of data points for multivariate analysis. The summer VOC data differed too much from site to site to be consolidated into a single set and the number of data points at each site was insufficient to support a multivariate analysis. Three factors were found to explain almost all the variability of the fall data. A graphical, three-dimensional version of the SAFER model determined the composition of these three sources, which could be identified as roadway (direct tailpipe plus running evaporative) emissions, whole gasoline, and gasoline vapor. On average, 60 percent of the Total Non-Methane Hydrocarbons (TNMHC) were apportioned to roadway emissions, 19 percent to whole gasoline, 15 percent to gasoline vapor, and 6 percent were unexplained or to background levels of TNMHC. There was a significant difference between the morning and afternoon source apportionment. In the morning, 71 percent of the TNMHC were apportioned to roadway emission, 14 percent to whole gasoline, 11 percent to gasoline vapor, and 4 percent were unexplained. In the afternoon, 44 percent of the TNMHC was apportioned to roadway emission, 27 percent to whole gasoline, 20 percent to gasoline vapor, and 10 percent were unexplained. This is consistent with increased evaporative emissions from parked vehicles at higher afternoon temperatures.

Multivariate receptor modeling of the particle composition data was restricted to summer PM-10 from five sites: Anaheim, Azusa, Burbank, Hawthorne, and Long Beach. The summer PM-2.5 and all the fall particle composition data sets had too little data for a valid multivariate analysis. Two source types, roadway and soil, account for most of the non-secondary PM-10. The composition and contributions of these sources were estimated by a two-dimensional version of the SAFER model. Roadway is defined as all PM-10 emissions from roadways, including direct tailpipe emissions and reentrained road dust. Soil is defined as all crustal material not directly associated with roadway emissions. The composition of roadway and soil sources varies little between the five sites. The average PM-10 source apportionment allocates 30 percent of the total PM-10 to roadway, 18 percent to soil, 35 percent to inorganic secondary species, and the remaining 17 percent represents, by process of elimination, an upper limit on the amount of secondary organic particulate matter.

The hybrid receptor modeling was unsuccessful because of the limitations in the wind fields.

Table of Contents

SUMMARY II

TABLE OF CONTENTS II

LIST OF TABLES IV

LIST OF FIGURES V

1. INTRODUCTION 1-1

2. AN OVERVIEW OF RECEPTOR MODELING 2-1

2.1 Receptor Model Fundamentals 2-1

2.2 CMB Modeling Methods 2-2

2.3 Multivariate Analysis Methods 2-2

2.4 Hybrid Mathematical Modeling-SITEOF 2-4

2.5 Source Apportionment by Factors with Explicit Restrictions-SAFER 2-5

3. VOLATILE ORGANIC GAS DATA 3-1

3.1 Introduction 3-1

3.2 Data 3-1

3.3 Methods 3-1
3.3.1 Source Compositions and Contributions from SAFER 3-1
3.3.2 Source Compositions from Regression 3-2

3.4 Results And Discussion 3-2
3.4.1 SAFER 3-2
3.4.2 Source Contributions 3-2
3.4.3 Source Compositions 3-3
3.4.4 Other Modeling Results 3-5

3.5 Conclusions 3-3

3.6 Tables 3-5

3.7 Figures 3-13

4. PARTICLE COMPOSITION RECEPTOR MODELING 4-1

4.1 Data Screening 4-1
4.1.1 PM-10 Summer Data 4-1
4.1.2 PM-10 Fall Data 4-2
4.1.3 Summer and Fall Combined PM-10 Data 4-2
4.1.4 PM-2.5 Summer and Fall Data 4-2
4.1.5 Data Screening Conclusions 4-2

4.2 Methods and Results 4-3
4.2.1 Principle Component Analysis 4-3
4.2.2 Determination of the Feasible Region 4-3
4.2.3 Determination of the Source Compositions 4-4

4.3 Results of Source Apportionment 4-4
4.3.1 Other Modeling Results 4-34

4.4 Conclusions 4-4

4.5 Tables 4-6

4.6 Figures 4-11

5. HYBRID SOURCE – RECEPTOR MODELING 5-1

5.1 Introduction 5-1

5.2 Data 5-1
5.2.1 Air Quality Data 5-1
5.2.2 Meteorological Data 5-1

5.3 Methods 5-1
5.3.1 Mixing height fields 5-1
5.3.2 Emission Rates and Locations Estimated from SITEOF 5-2

5.4 Results and Discussion 5-2

5.5 Tables 5-3

6. CONCLUSIONS 6-1

7. REFERENCES 7-1

List of Tables

Table 3-I	VOC Species Included in the Multivariate Analysis	3-5
Table 3-II	Regression of TNMHC on Source Contributions Normalized to Mean 1.	3-6
Table 3-III	Source Contributions by Site and Time (ppbC)	3-6
Table 3-IV	Source Contributions as a Percentage of TNMHC	3-7
Table 3-V	Number of Observations by Site and Time	3-7
Table 3-VI	Source Compositions Determined by Multivariate Receptor Modeling	3-8
Table 3-VII	Comparison of Roadway Composition to Other Studies. (Percent of TNMHC)	3-9
Table 3-VIII	Comparison of Gasoline Composition with Other Studies (Percent of TNMHC)	3-10
Table 3-IX	Comparison of Gasoline Vapor Composition to Other Studies (Percent of TNMHC)	3-11
Table 3-X	VOC Species Apportionment (ppbC)	3-12
Table 4-I	Abbreviations of Site Names	4-6
Table 4-II	Species Included in the PM-10 Analysis for Each Site	4-6
Table 4-III	Non-uniform Deposit Correction Factors for PM-10	4-7
Table 4-IV	Number of Samples in the Fall PM-10 Data	4-7
Table 4-V	Species Included in the PM-2.5 Data Sets	4-7
Table 4-VI	Species Included in the Final Summer PM-10 Data Sets for 2D-SAFER Modeling	4-8
Table 4-VII	Source Composition Estimates for Summer PM-10.	4-8
Table 4-VIII	Source Contribution Estimates for Summer PM-10	4-8
Table "3-1"	Annual Average PM10 Source Contribution Estimates	4-9
Table "3-2"	Maximum 24-hr Average PM10 Source Contribution Estimates	4-10
Table 5-I	Sites Used in the SITEOF Analysis	5-3
Table 5-II	Sites where Upper Air Sounding Data Estimated Mixing Heights	5-1

List of Figures

Figure 3-1 SAFER plot of the data with species normalized to a mean of one. The xy coordinates are the normalized second and third principal components. The lines are the locus of points where a species =0, some key lines are labeled. The points are the observed data. The points representing the three vehicle-related sources are shown. All possible non-negative source compositions lie inside the convex hull formed by the lines. Non-negative source contributions are guaranteed if all the observed data points lie inside the triangle formed by the sources. The gasoline and gasoline vapor sources must lie on the line propene = 0, since it is known that there is no propene in these sources. The gasoline vapor source was chosen to minimize the concentration of heavy VOCs, and the roadway source has the smallest amount of propene possible while meeting the non-negative source contribution requirement. 3-13

Figure 3-2 Observed Total Non-Methane Hydro-Carbons versus that predicted by SAFER as the sum of the three vehicle-related species. 3-14

Figure 4-1 Plot showing different behavior of summer and fall elemental composition data. 4-11

Figure 4-2 2-dimensional SAFER plot for Anaheim summer PM-10 data. The region below the left hand solid line represents source compositions with one or more soil element less than zero. The area below the right hand line represents source compositions with elemental carbon less than zero. The left hand region between the solid and dotted lines contains all feasible roadway source vectors and the right hand region contains all feasible soil composition vectors. 4-12

Figure 4-3 Roadway source versus reduced total mass (PM-10 total mass minus sulfate, nitrate, and ammonium). 4-13

Figure 4-4 Unscaled soil source contributions versus reduced total mass minus the roadway source contribution estimates. 4-13

1. Introduction

Innovative multivariate receptor modeling techniques were applied to several types of data collected during the Southern California Air Quality Study (SCAQS). The purpose of the modeling was to deduce as much information about the sources as possible from the data alone and with the aid of some basic physical constraints. The data examined were the Volatile Organic Compound (VOC) gas phase data, the airborne particle composition data, and some routine monitoring data. The models applied were the Source Apportionment by Factors with Explicit Restrictions (SAFER) multivariate receptor model and the Source Identification Through Empirical Orthogonal Functions (SITEOF) hybrid source-receptor model. The following section gives a brief introduction to these models.

2. AN OVERVIEW OF RECEPTOR MODELING

By using air quality information at a certain site mathematical receptor models try to determine the possible sources of these air pollutants. It is hoped that the information gained can be useful in developing air quality management tools. Receptor models go about their work in a fashion fundamentally different from dispersion models. Commonly a dispersion model would be used to study the effects of a plume. A dispersion model works by taking the emissions at a certain point and trying to determine the concentrations of various pollutants downwind of that source. In order to do this, the meteorology of the area must be well understood. In contrast, a receptor model works by taking air quality information at a certain point and trying to determine the sources of the various air pollutants. A receptor model (along with dispersion models) works on the principle of mass conservation of elements between sources and ambient concentrations. By using receptor models, one can avoid reconstructing the transportation and dispersion patterns of the various species. In other words, a receptor model makes no assumptions as to wind speed or direction though this information can be useful.

Receptor models have developed rapidly in the last several years. The factors which have led to the rapid development of receptor models include:

1) the development of source profiles. These source profiles may need modification when used but they provide some necessary guidelines to receptor model applications;
2) the huge database of ambient monitoring data now available allows for the development of models based on statistical techniques; and
3) the advancement in computation techniques and equipment in the last several years that allows the computations to be done on desktop machines in a reasonable time.

2.1 RECEPTOR MODEL FUNDAMENTALS

The fundamental mathematical expression of a receptor model can be stated as

$$\mathbf{C} = \mathbf{AS} + \mathbf{E}, \tag{1.1}$$

where C is an n by m matrix of ambient concentration data; A is an n by p matrix of source compositions, whose columns correspond to the source profiles of the p sources; S is a p by m source contribution matrix, whose rows are the contribution factors from the p sources; and E is an n by m matrix composed of monitoring and analysis errors. Equation (1.1) is simply the mass balance relationship. In general, C is known, having been obtained from air quality monitoring, but A, S, and E remain unknown. A and S need to be solved by means of source apportionment.

Theoretically, there is no unique analytical solution for Equation (1.1). For solving A and S in practice, there are many approaches which have been developed and formulated. A set of complete

solution procedures, fundamental theories, and application methods form a mathematical model for solving equation (1.1), these models are called receptor models. In order to solve equation (1.1), several models have been investigated and developed. Among these models, there are two major groups which have been fully studied and widely applied: 1) CMB (chemical mass balance) models; and 2) multivariate models. This separation is based on the mathematical techniques for solving equation (1.1).

2.2 CMB Modeling Methods

The primary modeling approach of a CMB model is to apportion source contributions, $\boldsymbol{S}$, on the basis of known source profiles, $\boldsymbol{A}$, or an assumed $\boldsymbol{A}$. Mathematically, this procedure may be described in the following way. Because $\boldsymbol{E}$ in equation (1.1) is an n by m matrix composed of monitoring and analysis errors, the expectation value of $\boldsymbol{E}$ is equal to zero. Mathematically this may be expressed by the following equation:

$$E(\boldsymbol{E}) = \boldsymbol{0}, \tag{1.2}$$

where $\boldsymbol{0}$ is a zero 1×m raw vector. If $\boldsymbol{A}$ in Equation (1.1) is given, the unbiased estimation of the source contributions, $\boldsymbol{S}$, can be analytically given by the following equation:

$$\boldsymbol{S} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{C}. \tag{1.3}$$

where the superscript t stands for the matrix transpose. This is the fundamental formula for a CMB model. A CMB model is composed of the algorithms developed from equation (1.3) and their corresponding computation programs.

The CMB model features a mathematically simple and very direct procedure, which results in this model having been widely applied to a variety of source apportionment problems. On the other hand, the CMB models have several limitations for source identifications. First, without further assuming the atmospheric dispersion factors, the solution from a CMB model will be a set of source contributions with no emission rates. The source contributions being equal to the product of the dispersion factors and emission rates.

Second, although there is a huge library of source profiles which is helpful in defining $\boldsymbol{A}$ in equation (1.3), modifications are usually needed for a specific problem. The determination of the needed modifications can be a tedious, subjective process. In current CMB models, there is no constraint to mathematically guarantee that equation (1.3) can give a non-negative solution of $\boldsymbol{S}$. In order to avoid negative values in $\boldsymbol{S}$, a traditional method largely relies on the selection of $\boldsymbol{A}$, which undoubtedly increases the difficulty in determing $\boldsymbol{A}$.

Third, if there is a solution to Equation (1.3), $(A^tA)^{-1}$ must exist. However, a singularity may lead to no solution or a to very unstable solution to equation (1.3). Practically, this condition requires that the source- profile vectors are linearly independent. However, when selecting $\boldsymbol{A}$, there is no way to directly identify the collinearity in $\boldsymbol{A}$. On the other hand, collinearities often exist among the source profile vectors in actual analyses, which limits the accuracy and resolution of solutions from CMB models. A collinearity problem occurs when any two sources have similar source profiles.

2.3 Multivariate Analysis Methods

In comparison with the CMB modeling methods, multivariate approaches attempt to directly extract the information on both source contributions and profiles with the help of the singular value decomposition (SVD) of $\boldsymbol{C}$. Therefore, the utility of multivariate analysis is that the estimation results are not dependent on possibly erroneous source profile libraries. The principle of multivariate models is that a measured concentration is a linear sum of the products of source compositions and source contributions plus random errors created during monitoring and laboratory analyses. Mathematically, it may be expressed by equation (1.1), too. Instead of directly solving equation (1.1), the matrix $\boldsymbol{C}$ is transformed by means of SVD into a product of three special matrices:

$$\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{V}^t, \tag{1.4}$$

where U is an n by n matrix whose columns are the eigenvectors of CC^t; V is an m by n matrix whose columns are the eigenvectors of CC; and D is an n by n diagonal matrix composed of the corresponding singular values. Singular values are the square roots of eigenvalues. Eigenvectors of a matrix are special because these vectors are mapped by the matrix into themselves, i.e., the eigenvectors do not change in direction after being multiplied by the matrix. Only the length of the eigenvector is changed, and this amount is given by the eigenvalue. If the ambient pollutants are released from p sources, $p < m$, in the D matrix, the first p singular values will be much greater than the rest. Taking these characteristics of D into consideration, the three matrices in equation (1.4) are partitioned for the p sources:

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_p & \mathbf{U}_r \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} \mathbf{D}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_r \end{bmatrix}, \qquad \mathbf{V}^t = \begin{bmatrix} \mathbf{V}_p^t \\ \mathbf{V}_r^t \end{bmatrix}, \tag{1.5}$$

where U_p is an n by p matrix which includes the first p eigenvectors of U; D_p is a p by p diagonal matrix composed of the first p singular values of D; and V_p^t is a p by m matrix including the first p rows of V^t. Accordingly, U_r and V_r^t contain the rest of the eigenvectors of their corresponding matrices. D_r is an m-p by m-p diagonal matrix formed by the rest of the singular values. The three matrices in Equation (1.4) are substituted by equation (1.5), then equation (1.4) becomes

$$\mathbf{C} = \begin{bmatrix} \mathbf{U}_p & \mathbf{U}_r \end{bmatrix} \begin{bmatrix} \mathbf{D}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_r \end{bmatrix} \begin{bmatrix} \mathbf{V}_p^t \\ \mathbf{V}_r^t \end{bmatrix}.$$

This equation can be further simplified by expanding these matrices and abandoning the zero terms:

$$\mathbf{C} = \mathbf{U}_p\mathbf{D}_p\mathbf{V}_p^t + \mathbf{U}_r\mathbf{D}_r\mathbf{V}_r^t. \tag{1.6}$$

Equation (1.6) is the primary equation which multivariate analysis models attempt to solve by various means. If C represents a distribution of the concentration data in the multi-dimensional space established by elements or species in a sample, the physical meaning of SVD decomposition is a transformation of the space bases. The new space is a linear and orthogonal one which is established by the eigenvectors of C. These eigenvectors are practically the linear combinations of the original elements and species. Therefore, a source profile will be the linear combinations of these eigenvectors. The importance of a SVD transformation is to ascertain the eigenvectors because each of them is a part of the source profiles. The methods of determining these eigenvector combinations lead the multivariate analysis models to develop along various directions.

The mathematical approaches in multivariate analysis models may be grouped into six categories: 1) source apportionment by factors with explicit restrictions (SAFER) method (Henry and Kim, 1990; and Henry, 1991); 2) singular value decomposition of initial analysis data matrix or principle component analysis (Lawson and Hanson, 1976); 3) factor analysis based upon eigenvector analysis; 4) target transformation factor analysis (TTFA) to approach a transformation of eigenvectors until the difference between the transformed and the targeting vectors is minimized (Hopke, 1991); and 5) multiple linear regression which is actually a linear least-square fitting method (Kleinmann *et al.*, 1976).

The multivariate modeling approach seems to be mathematically more robust than the CMB approach. These multivariate modeling solutions do not depend on any previous assumption to obtain calculation results. Therefore, they are mathematically more meaningful and more objective. Although Equation (1.6) consists of a set of linear algebra equations, there are still more unknown variables than mathematical relationships or equations. Consequently, Equation (1.6) defines a mathematical problem which does not come to closure. The solution to Equation (1.6) will be actually some feasible regions of source profiles. A feasible region is defined as the space that includes all of the variables satisfying all of

the related equations and constraints in the problem. Under some constraints, these feasible regions may be limited to quite narrow areas.

Finally, the calculation accuracy of some multivariate models relies excessively on the quality of input data and the size of this data body. In other words, these methods do not have enough tolerance to random errors. This could be a serious difficulty in simulations dealing with practical problems.

2.4 *Hybrid Mathematical Modeling-SITEOF*

The hybrid dispersion-receptor model employed in this study, Source Identification Through Empirical Orthogonal Functions (SITEOF) was established on the basis of the gradient dispersion equation as the basic mass balance equation. Generally, a dispersion model usually solves concentration distributions according to emission rates coupled with some atmospheric transportation characteristics. Instead, the primary objective of this hybrid model is to estimate emission rates on the basis of ambient concentrations at several given sites and some atmospheric transportation characteristics. Conceptually, the two modeling approaches are the reverse of each other. However, the mathematical methods used in SITEOF are not the reverse mathematical procedure of a dispersion model. SITEOF utilizes multivariate analysis to extract atmospheric dispersion and transportation information from ambient concentrations. An SVD and some empirical orthogonal functions (EOFs) are used to perform this statistical computation. The empirical orthogonal functions are a group of linear orthogonal polynomials which can be utilized to form a basis for the expansion of analytical functions. In SITEOF, EOFs are determined by the SVD of the concentration data matrix, and the resulting EOFs are composed of the eigenvectors and singular values obtained from this SVD. The function of these EOFs is to substitute into the integral functions formed during an integration of the dispersion equation. The mathematical background for SITEOF follows.

For transportation in the atmosphere, the mass conservation equation is actually a dispersion equation. Let $c = c(t, x,y,z)$ be the concentration of the species of interest at point (x,y,z) and time t. For dispersion in the atmosphere the following equation can be obtained

$$\frac{\partial c}{\partial t} = -u\frac{\partial c}{\partial x} - v\frac{\partial c}{\partial y} - w\frac{\partial c}{\partial z} + K_x\frac{\partial^2 c}{\partial x^2} + K_y\frac{\partial^2 c}{\partial y^2} + K_z\frac{\partial^2 c}{\partial z^2} + Q + S + R. \quad (1.7)$$

This is a general expression of the dispersion equation often used for air pollution research. The left hand side of Equation (1.7) is the concentration change with time. The first three terms on the right side represent the concentration changes as a result of advection, or the so-called advection terms. The next three terms are the diffusion terms caused by atmospheric turbulence. Q, S and R are the source term, the sink term and the transformation term, respectively.

SITEOF uses a different form of the dispersion equation. In particular the diffusion terms in equation (1.7) are used for describing a stochastic diffusion process for a concentration distribution. However, from the receptor modeling point of view, the concentrations from a receptor site are a set of stochastic measurements or a random field, which have already included the effect of the stochastic diffusion. In this situation, the diffusion terms in equation (1.7) become unnecessary. Moreover, comparing the advection and diffusion terms, it can be found that the advection is generally one order greater than the diffusion. Because the concentrations are always associated with some random errors from sampling, and analysis, the inaccuracy created from the evaluation of the advection is generally similar in size to the value of the diffusion terms. As a result of this situation, the diffusion terms in equation (1.7) can be eliminated without decreasing the accuracy of the calculation.

For air pollutant concentration data, the concentration distributions are usually given as two dimensional fields, $c = c(t, x, y)$. Equation (1.7), then, is changed into equation (1.8)

$$\frac{\partial c}{\partial t} = -u\frac{\partial c}{\partial x} - v\frac{\partial c}{\partial y} + Q + S + R. \quad (1.8)$$

Rearranging equation (1.8) gives us the basic equation used in SITEOF

$$Q + S + R = \frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y}. \qquad (1.9).$$

Some further modifications are also done to better carry out the solution to the equation 1.9. These modifications are described elsewhere.

2.5 Source Apportionment by Factors with Explicit Restrictions-SAFER

SAFER (Henry and Kim, 1990) begins with a singular value decomposition (SVD) of the unscaled, uncentered, raw data matrix. Non-negativity constraints (positive source compositions and contributions) and other physical constraints on the predicted source compositions are represented as a linear programming in the eigenvector space determined from the SVD. The solution to this linear programming defines the boundary of the feasible region, i.e., the region within which predicted sources have only positive compositions and contributions and obey any additional physical constraints which have been imposed. Much of the the power of the SAFER (or any) multivariate model comes from restricting the range of possible source compositions by reducing the dimensionality of the data set. This comes about because the various species are commonly very highly correlated with each other. In this way, the candidate source contributions are found to lie in an n-dimensional subspace of the data.

3. Volatile Organic Gas Data

3.1 Introduction

A three dimensional version of the Source Apportionment by Factors with Explicit Restrictions (SAFER) model was applied to Volatile Organic Compound (VOC) data collected during the fall intensive sampling campaign of the SCAQS. The purpose of the analysis was to estimate the VOC composition of the three vehicle-related sources that dominate VOC concentrations. These are roadway emissions (vehicle exhaust plus running evaporative losses), whole gasoline, and gasoline vapor. Harley *et al.* (1992) and Watson *et al.* (1993) have shown the dominance of these sources in the SCAQS data and Henry *et al.* (1994) have demonstrated the same for VOC data from Atlanta. Source contributions for the three vehicle-related sources are also estimated by SAFER.

3.2 Data

A descriptive analysis of the VOC data for the summer and fall SCAQS intensives is given by Lurmann and Main (1992). For the current study, SAFER modeling was restricted to the data from the fall intensive for two reasons. First, there is less missing data because the VOC concentrations are much higher than in the summer, and second, the data from different sites could be combined into one data set containing 124 observations. Multivariate analysis of the summer VOC data showed significant differences between the sites that prohibited combining data from several sites together, which is necessary in order to obtain a data set that has sufficient observations for a reliable multivariate analysis.

The VOC species included in the analysis are given in Table 3-I. These species were chosen because they had few or no missing data. In addition, the Table indicates those species that were selected for application of the SAFER model because singular value decomposition showed this subset of species had three strong factors.

All the VOC concentrations in this report are in ppbC. Multiply ppbC by 0.4908 to convert to $\mu gC/m^3$. Because carbon makes up the vast majority of the mass of the VOCs, this is very close to $\mu g/m^3$.

3.3 Methods

3.3.1 Source Compositions and Contributions from SAFER

Source compositions were determined by a three-source version of the SAFER model (Henry and Kim, 1990) applied to the sixteen identified species in Table 3-I. Although developed entirely independently, this method is similar to the three factor self-modeling curve resolution model as described in the chemometrics literature by Sun *et al.* (1987). The conditions of the SAFER model ensure that the source compositions and contributions are non-negative; however, SAFER relies on additional physical constraints to find a unique solution among the infinite number of feasible solutions. The additional physical constraints applied in this case were:

- Propene is used a tracer for roadway emissions because the preferred tracer acetylene was missing almost half the time. Propene is more reactive than acetylene, but since the analysis is restricted to the fall when photochemisrty is reduced, this is not a problem.
- The gasoline and gasoline vapor sources do not contain propene.
- The roadway source was chosen to have the minimum possible amount of propene that is consistent with the data and the non-negativity constraint on the source contributions..
- The gasoline source was chosen to have the maximum possible amount low carbon number VOCs and the minimum amount of high carbon number VOCs.

- The gasoline vapor source was chosen to have the minimum possible amount of high carbon number VOCs.

The source contributions determined by SAFER have the correct relative VOC composition but each may be multiplied by an arbitrary factor yet to be determined. These factors are determined so as to best fit the observed Total Non-Methane Hydrocarbons (TNMHC), as explained next.

The relative source composition matrix determined by SAFER is normalized so that the sum of across each species is the average value of the species in the data. Next, the source contributions for this source composition matrix are determined by a linear least squares fit to the data. Because of this normalization, the three source contributions so obtained must have a mean value of one. TNMHC is regressed against these source contributions; the coefficients are then the average source contribution of each source to TNMHC. The final source composition matrix is calculated by dividing each column of the relative source composition matrix by the corresponding average source contribution. The final source contributions are determined by multiplying the normalized source compositions by the corresponding average source composition determined by regression on TNMHC. The resulting source compositions and source contributions are optimal in explaining the sixteen species used in the analysis and TNMHC.

3.3.2 Source Compositions from Regression

The source compositions of the remaining fifteen species in Table 3-I were determined by regressing the concentrations against the source contributions, the regression coefficients are the source compositions, assuming that there are no other significant sources to that species or other problems with the data. If the species have high R^2 values and the regression coefficients are reasonable, this is a good model validaton.

3.4 Results And Discussion

3.4.1 SAFER

Figure 3-1 shows the SAFER plot of the 16 vehicle-related species in model. Because some VOC species concentrations are much larger than others, the VOCs have been normalized to a mean of one so that a few species do not dominate the analysis. The lines are the locus of points where a species = 0, some key lines are labeled. All possible non-negative source compositions lie inside the convex hull formed by the lines. The points representing roadway, gasoline, and gasoline vapor are shown. Non-negative source contributions are guaranteed if all the observed data points lie inside the triangle formed by the sources. The gasoline and gasoline vapor sources must lie on the line propene = 0, since it is known that there is no propene in these sources. The gasoline vapor source was chosen to minimize the concentration of heavy VOCs, and the roadway source has the smallest amount of propene possible while meeting the non-negative source contribution requirement.

A set of relative source contributions normalized to mean one was calculated using the relative source composition matrix corresponding to the points in Figure 3-1 The regression of TNMHC against the normalized source contributions is given in Table 3-II. The R^2 values are over 99% indicating that TNMHC is very well explained by the three vehicle-related sources. A plot of the observed TNMHC versus the sum of the three vehicle-related sources is shown in Figure 3-2. The background TNMHC is given by the intercept is 90 ppbC, the average roadway source contribution is 891 ppbC, gasoline is 288 ppbC, and gasoline vapor is 227 ppbC. These average source contributions are then used to calculate the unnomalized source contributions and source compositions, as explained in the methods section above.

3.4.2 Source Contributions

Table 3-III gives the average source contribution estimates by site for the morning (0500 to 0900 starting time) and afternoon (1200 to 1600 starting time). The same figures expressed as a percentage of average TNMHC are given in Table 3-IV. Not too much should be read into differences between sites since the averages in the Tables represent widely different numbers of samples from perhaps different

days, as shown in Table 3-V. There is a significant difference between morning and afternoon period, however. In the morning, roadway emissions are about 70 percent of TNMHC, while in the afternoon this drops to only 44 percent. The difference is made up by roughly equal increases in gasoline and gasoline vapor. Also, the unexplained TNMHC goes from 4 percent in the morning to 10 percent in the afternoon, on average. The increased gasoline and gasoline vapor contribution in the afternoon is consistent with emissions from parked vehicles and the higher temperatures in the afternoon.

3.4.3 Source Compositions

Table 3-VI gives the source compositions for the 16 vehicle-related species in the SAFER model. The table also gives the source compositions for the remaining species determined by regression on the source contribution estimates. The adjusted R^2 for the regression is given for each species not included in SAFER; a low R^2, say less than 0.9, indicates that the species may have other sources, or that it has other data problems. For example, ethane and propane have R^2 values below 0.8 and are found in geogenic and commercial natural gas. Other species with low R^2 values are noctane and nnonane; these are present at low levels in the vehicle-related sources and may have other significant sources. The low R^2 for ethene is more likely caused by data problems, since no other large sources are known. The fact that so many species that were not used in the SAFER model are well fit by it is a validation of the model and its predictions.

The roadway VOC composition derived in this study is in reasonably good agreement with those reported elsewhere. Table 3-VII compares the roadway profile of this study with three others, the profile used for Chemical Mass Balance (CMB) by the Desert Research Institute (DRI), Watson *et al.* (1993); measurements in the Caldecott Tunnel as reported in Watson *et al.* (1993); and measurements taken in a freeway underpass in downtown Atlanta during the summer of 1990 (Conner *et al.*, 1994).

The comparison of the gasoline profile with other reports is not quite as good as for the roadway profile, but it is still very good, as can be seen in Table 3-VIII. Benzene and nbutane in the current profile are quite a bit larger than in other reported profiles. The gasoline vapor profile also shows more benzene and toluene, but less nbutane than other reports, as seen in Table 3-IX. Since most of the other major and minor VOC species compare quite well for gasoline and gasoline vapor, the inconsistencies in some species cannot be explained as all species being too low or too high.

Additional insight into the SAFER model results can be had by comparing the predicted and observed concentrations for the individual VOC species; these are seen in Table 3-X. Most species, including benzene and toluene, are well modeled, with less than a 10 percent difference between the average predicted and observed concentrations. The only species that is significantly over-predicted is nbutane, by 11 percent. Noctane and nnonane are significantly underpredicted, like propane and ethane in Table 3-VI, these species had low R^2 and may have other sources. The case of nbutane is less easy to understand. Perhaps the multivariate analysis was unable to determine a reasonable source composition for nbutane because the composition was varying during the period. This is possible because refiners increase the nbutane in gasoline in the late fall and winter to improve cold starting performance.

3.5 Conclusions

The primary conclusions of this study are:

- VOC data from the various sites from the SCAQS fall intensive can be consolidated into a single data set, i.e., all sites seem to be affected by the same sources. This is not true for summer VOC data.
- TNMHC and most of the individual VOC species are well explained by three vehicle-related sources: roadway emissions (direct exhaust and running evaporative), gasoline, and gasoline vapor.

- The three source SAFER model was validated by its success in explaining VOC species that were not used to develop the model.
- On average, 60 percent of the TNMHC was apportioned to roadway emissions, 19 percent to whole gasoline, 15 percent to gasoline vapor, and 6 percent were unexplained or due to background.
- There was a significant difference between the morning and afternoon source apportionment. In the morning, 71 percent of the TNMHC was apportioned to roadway emission, 14 percent to whole gasoline, 11 percent to gasoline vapor, and 4 percent were unexplained. In the afternoon, 44 percent of the TNMHC was apportioned to roadway emission, 27 percent to whole gasoline, 20 percent to gasoline vapor, and 10 percent wre unexplained.
- The source composition for roadway emissions as determined by the SAFER model is similar to the composition of roadway emissions seen in other studies. However, the SAFER profile shows some differences that are specific to the SCAQS data set.
- The composition of gasoline and gasoline vapor, as estimated by SAFER, is in general agreement with source samples. The SAFER estimate of the nbutane fraction of these sources is probably inaccurate since nbutane is over-predicted. This may be the result of variations in nbutane composition of gasoline made by refiners to adjust the cold start characteristics of the gasoline.
- Ethane and propane were not well modeled by the three-source model. They probably have significant non-vehicle related sources.

Tables

Table 3-1 VOC Species Included in the Multivariate Analysis

Name or Abbreviation	Chemical Name	Included in SAFER
Ethane	Ethane	
Ethene	Ethene	
Ethyne	Ethyne	
Propene	Propene	x
nButane	n-butane	
iPentane	2-methylbutane	x
nPentane	n-pentane	x
CyPentane	cyclopentane	
2MPentane	2-methylpentane	x
3MPentane	3-methylpentane	
nHexane	n-hexane	x
MCyPentane	methyl-cyclopentane	x
24DMPentane	2,4-dimethylpentane	
Benzene	benzene	
CyHexane	cyclohexane	
2MHexane	2-methylhexane	x
23DMPentane	2,3-dimethylpentane	x
3MHexane	3-methylhexane	x
224TMPent	2,2,4-trimetylpentane	x
nHeptane	nheptane	x
MCyHexane	methylcyclohexane	
234TMPent	2,3,4-trimethylpentane	x
Toluene	methylbenzene	
2MHeptane	2-methylheptane	x
3MHeptane	3-methylheptane	
nOctane	noctane	
EBenzene	ethylbenzene	x
MPXylene	1,3-dimethylbenzene +1,4-dimethylbenzene	x
OXylene	1,2-dimethylbenzene	x
nNonane	nnonane	
124TMBenz	1,2,4-trimethylbenzene	

Table 3-II Regression of TNMHC on Source Contributions Normalized to Mean 1.

Regression Statistics	
Multiple R	0.9968
R Square	0.9936
Adjusted R Square	0.9935
Standard Error	86.3068
Observations	124

	Coefficients	*Standard Error*	*t Stat*
Intercept	89.9	13.77	6.53
Roadway	890.7	9.17	97.06
Gasoline	288.3	10.01	28.79
Gasoline Vapor	227.0	9.02	25.16

Table 3-III Source Contributions by Site and Time (ppbC)

SITE		*Roadway*	*Gasoline*	*Gas Vapor*	*Unexplained*	*TNMHC*
ANAHEIM	Morning	1671	199	248	137	2255
	Afternoon	512	233	243	66	1054
	All	874	222	244	88	1429
BURBANK	Morning	2147	444	325	58	2975
	Afternoon	642	438	219	107	1405
	All	1112	440	252	92	1896
CENTRAL LA	Morning	1145	287	109	145	1687
	Afternoon	492	325	190	103	1110
	All	838	305	147	125	1415
HAWTHORNE	Morning	2583	280	128	-64	2927
	Afternoon	230	153	133	136	652
	All	701	178	132	96	1107
LONG BEACH	Morning	1638	330	401	65	2433
	Afternoon	375	224	230	114	943
	All	1064	281	323	87	1756
RIVERSIDE	Morning	1595	140	83	-20	1797
	Afternoon	199	129	83	35	445
	All	414	131	83	26	653
Grand Total	Morning	1551	303	249	92	2195
	Afternoon	420	258	188	94	959
	All	883	276	213	93	1466

Table 3-IV Source Contributions as a Percentage of TNMHC

SITE		Roadway	Gasoline	Gas Vaopr	Unexplained
ANAHEIM	Morning	74	9	11	6
	Afternoon	49	22	23	6
	All	61	16	17	6
BURBANK	Morning	72	15	11	2
	Afternoon	46	31	16	8
	All	59	23	13	5
CENTRAL LA	Morning	68	17	6	9
	Afternoon	44	29	17	9
	All	59	22	10	9
HAWTHORN	Morning	88	10	4	-2
	Afternoon	35	23	20	21
	All	63	16	12	9
LONG BEACH	Morning	67	14	16	3
	Afternoon	40	24	24	12
	All	61	16	18	5
RIVERSIDE	Morning	89	8	5	-1
	Afternoon	45	29	19	8
	All	63	20	13	4
Grand Average	Morning	71	14	11	4
	Afternoon	44	27	20	10
	All	60	19	15	6

Table 3-V Number of Observations by Site and Time

	SITE	No. Obs.
Morning	ANAHEIM	5
	BURBANK	5
	CEN. LA	18
	HAWTHORNE	3
	LONG BEACH	18
	RIVERSIDE	2
Afternoon	ANAHEIM	11
	BURBANK	11
	CEN. LA	17
	HAWTHORNE	8
	LONG BEACH	15
	RIVERSIDE	11
Grand Total		124

Table 3-VI Source Compositions Determined by Multivariate Receptor Modeling

(Percent of TNMHC)

	Roadway	Gasoline	Gas Vapor	Adj. R^2 [a]
Ethane	4.13	0.00	0.00	0.79
Ethene	4.79	0.00	0.00	0.87
Ethyne	4.01	0.00	0.00	0.93
Propene	2.16	0.00	0.00	
Propane	3.39	2.18	10.59	0.75
nButane	5.97	10.45	20.91	0.93
iPentane	6.52	7.08	15.34	
nPentane	2.97	2.31	8.12	
CyPentane	0.38	0.36	0.70	0.99
2MPentane	2.48	2.84	3.19	
3MPentane	1.74	2.06	2.00	0.92
nHexane	1.62	1.83	2.16	
MCyPentane	1.60	1.63	1.68	
24DMPentane	0.56	0.71	0.42	0.94
Benzene	3.30	3.04	2.52	0.94
CyHexane	0.38	0.46	0.44	0.86
2MHexane	1.03	1.35	0.36	
23DMPentane	0.82	1.15	0.45	
3MHexane	1.11	1.62	0.49	
224TMPent	1.70	2.18	0.41	
nHeptane	1.00	1.65	0.24	
MCyHexane	0.86	1.76	0.33	0.86
234TMPent	0.58	0.70	0.00	
Toluene	8.53	9.70	2.98	0.98
2MHeptane	0.39	0.66	0.09	
3MHeptane	0.46	0.59	0.02	0.92
nOctane	0.39	0.52	0.11	0.74
EBenzene	1.31	1.75	0.26	
MPXylene	5.56	5.11	0.12	
OXylene	1.94	1.99	0.23	
nNonane	0.18	0.40	0.21	0.76
124TMBenz	2.21	1.82	0.00	0.97
Sum	74.07	67.92	73.97	

[a] R^2 values are only given for species not included in the SAFER model, i.e., those determined by multiple regression.

Table 3-VII Comparison of Roadway Composition to Other Studies. (Percent of TNMHC)

	This Study	LA Winter 1987[a]	Caldecott Tunnel[a]	Atlanta Summer 1990[b]
Ethene	4.79	9.90	6.43	4.34
Ethyne	4.01	2.78	2.12	3.80
Propene	2.16	3.13	2.75	1.96
nButane	5.97	6.44	2.84	4.11
iPentane	6.52	4.93	7.79	8.64
nPentane	2.97	2.21	2.73	2.66
CyPentane	0.38	0.21	0.44	0.25
2MPentane	2.48	1.91	3.17	2.43
3MPentane	1.74	1.37	1.85	1.42
nHexane	1.62	0.92	1.72	1.09
MCyPentane	1.60	0.92	1.77	0.78
24DMPentane	0.56	0.82	0.00	0.70
Benzene	3.30	3.91	5.42	2.73
CyHexane	0.38	0.22	0.29	0.17
2MHexane	1.03	0.76	1.20	0.87
23DMPentane	0.82	0.76	0.46	0.90
3MHexane	1.11	1.38	1.25	0.89
224TMPent	1.70	2.91	1.59	2.51
nHeptane	1.00	0.78	0.93	0.54
MCyHexane	0.86	0.64	0.51	0.30
234TMPent	0.58	0.17	0.00	0.95
Toluene	8.53	7.10	8.27	6.59
2MHeptane	0.39	0.33	0.47	0.32
3MHeptane	0.46	0.55	0.00	0.36
nOctane	0.39	0.44	0.32	0.29
EBenzene	1.31	0.79	1.50	1.28
MPXylene	5.56	3.32	5.95	4.35
OXylene	1.94	2.22	2.46	1.66
nNonane	0.18	0.31	0.00	0.21
124TMBenz	2.21	3.33	2.27	2.16

[a] Watson *et al.* (1993), [b]Conner *et al.*(1993)

Table 3-VIII Comparison of Gasoline Composition with Other Studies (Percent of TNMHC)

	This Study	LA Winter 1987[a]	LA Summer 1987[b]	Atlanta Summer 1990[c]
Ethane	0.00	0.00	0.00	0.01
Ethyne	0.00	0.00	0.00	0.00
Propene	0.00	0.00		0.04
nButane	**10.45**	5.60	3.30	3.23
iPentane	7.08	6.19	6.90	7.37
nPentane	2.31	2.56		2.76
CyPentane	0.36	0.49	0.50	0.27
2MPentane	2.84	2.32	3.30	2.88
3MPentane	2.06	1.77	2.10	1.79
nHexane	1.83	1.69	2.00	1.50
MCyPentane	1.63	2.19	2.50	1.10
24DMPentane	0.71	0.98		0.75
Benzene	3.04	1.88	1.90	1.53
CyHexane	0.46	0.50	0.60	0.25
2MHexane	1.35	1.68		1.28
23DMPentane	1.15	1.34		1.02
3MHexane	1.62	1.51		1.27
224TMPent	2.18	0.00		2.82
nHeptane	1.65	1.58		0.85
MCyHexane	1.76	0.86	1.00	0.57
234TMPent	0.70	0.00		1.33
Toluene	9.70	9.64	10.20	8.11
2MHeptane	0.66	0.00		0.50
3MHeptane	0.59	0.00		0.61
nOctane	0.52	0.90	1.00	0.45
EBenzene	1.75	2.02	1.90	1.80
MPXylene	5.11	8.63	8.30	6.30
OXylene	1.99	3.42	3.10	2.60
nNonane	0.40	0.58		0.22
124TMBenz	1.82	3.87	3.20	4.18

[a] Watson *et al.* (1993), [b]Harley *et al.* (1992),
[c] Conner *et al.*(1993)

Table 3-IX Comparison of Gasoline Vapor Composition to Other Studies (Percent of TNMHC)

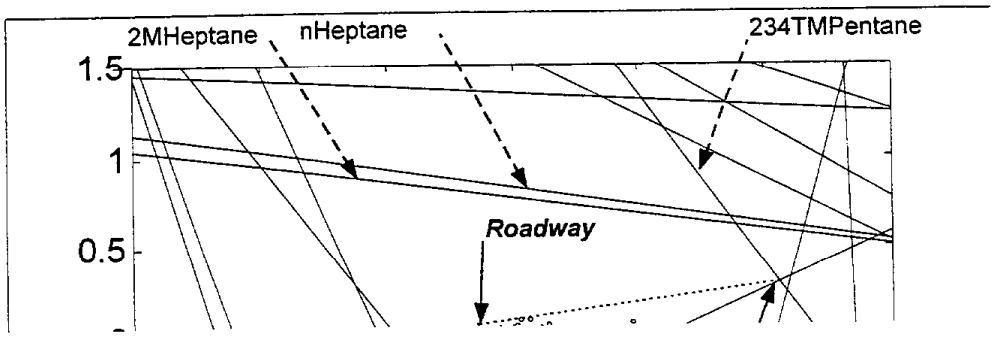
	This Study	LA Winter 1987[a]	LA Summer 1987[b]	Atlanta Summer 1990[c]
Ethane	0.00	0.00		0.01
Ethyne	0.00	0.00		0.00
Propene	0.00	0.00		0.13
nButane	20.91	30.48	30.00	21.80
iPentane	15.34	14.28	22.30	27.90
nPentane	8.12	4.40		7.40
CyPentane	0.70	0.53	0.50	0.43
2MPentane	3.19	2.03	2.80	3.53
3MPentane	2.00	1.49	1.60	1.93
nHexane	2.16	1.19	1.10	1.20
MCyPentane	1.68	1.49	1.10	0.81
24DMPentane	0.42	0.56		0.52
Benzene	2.52	1.20	0.70	0.87
CyHexane	0.44	0.27	0.30	0.12
2MHexane	0.36	0.87		0.46
23DMPentane	0.45	0.69		0.46
3MHexane	0.49	0.62		0.44
224TMPent	0.41	0.00		0.99
nHeptane	0.24	0.53	0.20	0.21
MCyHexane	0.33	0.26	0.10	0.12
234TMPent	0.00	0.00		0.23
Toluene	2.98	1.54	0.70	1.26
2MHeptane	0.09	0.00		0.06
3MHeptane	0.02	0.00		0.01
nOctane	0.11	0.13	0.00	0.04
EBenzene	0.26	0.15	0.00	0.10
MPXylene	0.12	0.51	0.10	0.32
OXylene	0.23	0.23	0.00	0.12
nNonane	0.21	0.01		0.01
124TMBenz	0.00	0.03	0.30	0.11

[a] Watson *et al.* (1993), [b]Harley *et al.* (1992),
[c] Conner *et al.*(1993)

Table 3-X VOC Species Apportionment (ppbC)

	Roadway	Gasoline	Gas Vapor	Sum	Observed	Predicted / Observed
Propene	19.21	0.00	0.00	19.21	19.59	0.98
nButane	53.21	30.12	47.45	130.78	118.05	**1.11**
iPentane	58.03	20.42	34.83	113.28	113.28	1.00
nPentane	26.44	6.65	18.42	51.51	51.44	1.00
CyPentane	3.34	1.04	1.59	5.98	6.05	0.99
2MPentane	22.12	8.18	7.24	37.54	37.53	1.00
3Mpentane	15.53	5.93	4.53	25.99	28.27	0.92
nHexane	14.40	5.29	4.90	24.59	24.77	0.99
MCyPentane	14.27	4.71	3.82	22.79	22.87	1.00
24DMPentane	5.01	2.03	0.96	8.00	7.60	1.05
Benzene	29.42	8.76	5.72	43.90	43.18	1.02
CyHexane	3.37	1.34	1.00	5.71	6.07	0.94
2MHexane	9.18	3.91	0.82	13.90	13.83	1.01
23DMPentane	7.30	3.33	1.02	11.65	11.60	1.00
3MHexane	9.93	4.68	1.11	15.72	15.82	0.99
224TMPent	15.10	6.29	0.93	22.32	22.35	1.00
nHeptane	8.90	4.76	0.54	14.20	14.23	1.00
MCyHexane	7.70	5.08	0.75	13.53	15.03	0.90
234TMPent	5.19	2.01	0.00	7.21	7.08	1.02
Toluene	75.99	27.98	6.76	110.72	115.50	0.96
2MHeptane	3.48	1.91	0.20	5.60	5.68	0.99
3MHeptane	4.12	1.70	0.05	5.88	6.06	0.97
nOctane	3.47	1.49	0.26	5.22	7.08	**0.74**
EBenzene	11.64	5.05	0.59	17.28	17.60	0.98
MPXylene	49.50	14.73	0.28	64.50	64.25	1.00
OXylene	17.25	5.74	0.52	23.52	23.70	0.99
nNonane	1.58	1.14	0.47	3.20	3.82	**0.84**
124TMBenz	19.66	5.25	0.00	24.92	24.21	1.03

Figures



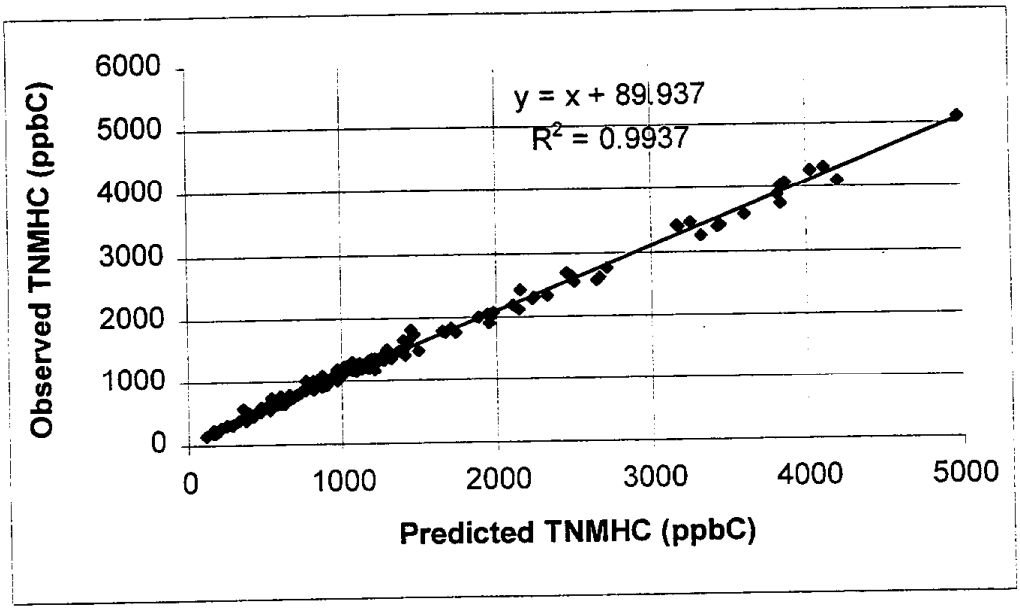


Figure 3-2 Observed Total Non-Methane Hydro-Carbons versus that predicted by SAFER as the sum of the three vehicle-related species.

4. Particle Composition Receptor Modeling

This section covers the multivariate receptor modeling of PM-10 and PM-2.5 composition data for the summer and fall SCAQS intensive periods. Limitations of the data make impossible receptor modeling of the summer PM-2.5 data and all the fall data. Also, significant differences between the summer and winter data make it impossible to consolidate these two data sets. Thus, this section is restricted to analysis of the PM-10 data from the summer intensive.

Multivariate analysis shows that two sources explain, at most sites, practically all of the non-secondary component of PM-10. These two sources can be identified as roadway and soil. The roadway source consists of all the PM-10 coming directly from roadways, a combination of resuspended road dust and direct emissions. Elemental carbon is assumed to be a tracer for roadway emissions. The soil source is simply all soil dust not directly associated with roadways. The composition of these two source types and the contributions of each to total PM-10 are estimated by a graphical version of the Source Apportionment by Factor with Explicit Restrictions (SAFER) model.

4.1 Data Screening

4.1.1 PM-10 Summer Data

We obtained 9 data files which included the original summer PM-10 data as LOTUS worksheets. They included the analysis results of the samples originally collected from 9 sampling sites: Anaheim; Azusa; Burbank; Claremont; Downtown Los Angeles; Hawthorn; Long Beach; Riverside; and San Nicholas Island. The abbreviations used for these site names are given in Table 4-I. Each worksheet is composed of 3 types of data: the original data body; the detection limits; and the 3-sigma error bars. These three data sets have been individually saved as three major data matrices of 55 rows by 33 columns. The rows of the original data matrix represent the 55 observed samples, and the columns represent the 33 target elements. The data are the analysis results of elemental concentrations. Therefore ,theoretically, 55×33 elemental concentrations have been provided for each sampling site. The matrices of the detection limits and the 3-sigma values include the data corresponding to the original detected data

In some of the nine files, there are several samples which have some problems caused by errors in their analyses or sampling processes. Also, it has been found that several samples include some significant outlier data points. The data from these two types of samples may contain some wrong statistical information, and may mislead the whole analysis to an incorrect results. For limiting any possible error sources to this analysis, these sample rows have been deleted from the data matrices.

At all sites, the original worksheet includes a large amount of BDL (Below Detection Limit) points. An element becomes unavailable to this analysis when it contains too many BDLs. Therefore, the next data preparation step is to screen out such elements. Based upon the above rule of data elimination, an element column was deleted from the analysis data body if there exist more than 3 BDL among the 55 data points. In order to provide some unbiased statistical information to the analysis, for every element, there should be a sufficient number of data above the detection limit in the original worksheet. For this analysis, to obtain a correct result, each involved element should contain more than 95% of originally detected data points. However, the actual situation is that, among the 33 target elements, most of their elemental concentrations are below the limit of detection. The remaining elements at each site are listed in Table 4-II. The few BDL's that remained in the data were replaced by the half values of the corresponding detection limits.

The PM-10 Teflon filters collected by the SCAQS samplers were non-uniform in deposit caused by an error in sampler design (Matsumura and Cahill, 1990). Some conversion factors to the individual elemental concentrations have been found by Matsumura and Cahill (1990) based upon their statistical analyses of several selected filters by the PIXE proton milliprobe. Therefore, the original data are

converted in the manner of dividing the elemental concentration columns by the corresponding conversion factors. The conversion factors used in this step are shown in Table 4-III.

4.1.2 PM-10 Fall Data

Six data files including the original fall PM-10 data have been obtained. The files are the analysis results of the samples originally collected from 6 sampling sites: 1) Anaheim; 2) Burbank; 3) Downtown of Los Angeles; 4) Hawthorn; 5) Long Beach; and 6) Riverside. Each file provides 2 types of data: the original data body and the detection limits. These 2 parts of data have been individually saved one after another on a day by day base. Thirty samples are included in the files, and their 31 elements have been analyzed. The data in these files are the analysis results of elemental concentrations. After rearranging these data, each file is now a 30×31 matrix, whose rows and columns are corresponding to 30 samples and 31 target elements, respectively. Similar to the preparation procedures for the summer data, these fall data have also been carefully examined. However, because the number of available samples in the 6 files is 30 or less (refer to Table 4-IV), these files don't have enough data to support multivariate analysis.

4.1.3 Summer and Fall Combined PM-10 Data

The six corresponding summer and fall data matrices are combined together, respectively, to form six lumped data matrices. Doing this work is based upon an assumption that a better analysis result could be obtained if the data body has more data points. However, another condition to make this assumption true is that the two parts of the data should contain similar statistical information. To examine whether there exists such a condition for the two parts, several scatter plots between some elements have been prepared. But these plots are showing the same fact, to a different extent, that the two parts of data are statistically different. Figure 4-1 is one of these plots showing this conclusion. Therefore, no further analyses were conducted for the consolidated data.

4.1.4 PM-2.5 Summer and Fall Data

The original PM-2.5 summer and fall data have the same format as the corresponding PM-10 data. However, these data have more BDLs, and therefore, more element columns were deleted from their original data body than PM-10 data. The PM-2.5 data were screened in the same way as the PM-10 data. The remaining elements and their numbers of sample rows in the summer data are listed in Table 4-V.

For the PM-2.5 summer data, very few elements are available after screening out the BDLs. Given the limited number of species with most observations above minimum detectable limits, the summer PM-2.5 data was judged unsuitable for multivariate analysis. After screening, the fall PM-2.5 data also do not included enough samples for multivariate analysis. After applying the data screening process to these data, between 22 and 24 observations were left at each site, much too few for a reliable application of multivariate receptor model.

4.1.5 Data Screening Conclusions

The end result of the data screening was that only the PM-10 data from the summer intensive was amenable to multivariate analysis. The fall PM-10 and the summer and fall PM-2.5 had too few observations for a reliable multivariate analysis. An attempt to overcome this problem by consolidating the fall and summer data was not possible because analysis showed that the sources for the two datasets were different.

Due to the limited number of the elements available in the summer PM-10 data, it was found that only two possible particulate sources, roadway and soil dust, can be identified from such data. Therefore, there was no reason to keep the data columns whose elements are normally not or very weakly related to any of these two sources, such as Na, Mg, P, Cl, K, Zn, and Ba, in the data matrices. In order to eliminate any possible interferences from these species, their columns are deleted from the data matrices.

For the identification of the roadway sources, the data of the elemental carbon concentrations may provide some valuable information to this analysis because the elemental carbon is a tracer element for

roadway sources, which can help in separating the roadway sources from soil dust. Fortunately, the elemental concentrations of carbon have been detected for the SCAQS samples and saved in another data file. The elemental carbon data were extracted from the file, and combined with other elements to form a new data matrix. The structure of the final data matrices for this analysis are summarized in Table 4-VI.

4.2 Methods and Results

A 2D (2-Dimensional) version of the SAFER model is employed to identify sources and to estimate the composition and contribution of these sources to PM-10 during the summer intensive. The 2D model is mathematically identical to the original version of the self-modeling curve resolution technique (Lawton and Sylvestre, 1971). Since all the non-negativity and additional constraints in a 2D model can be seen graphically, there is no need to use the linear programming methods of the full SAFER model.

4.2.1 Principle Component Analysis

Principle component analysis was applied to the data matrices listed in Table 4-VI. The results of this analysis provide the information about how many and how clearly separated the potential sources can be identified. The purpose of conducting a principle component analysis of these data is to examine whether each data matrix carries enough statistical information about the two particulate sources.

The results for Claremont do not show the necessary two clear factors from the eigenvalues of its data matrix. Reviewing the data of Claremont, it has been found that this data includes only 38 sample rows, which apparently is not enough for a reliable multivariate analysis. The Riverside data gives three clearly separated factors from their resulting eigenvalues. The third potential source is strongly associated with the first two, and a biased analysis will result from the 2-D model if the existence of the third factor is ignored. Therefore, this data set is not two-dimensional, and the 2-D model is certainly not applicable. After the previous data screening, there are only 4 elements in the San Nicholas Island data. Based upon these four elemental concentrations, the two particulate sources cannot be well apportioned because only the four elements cannot provide enough varieties to recognize the difference between the two analyzing particulate sources. Therefore, the data from Claremont, Riverside, and San Nicholas Island are excluded from this analysis. The principle component analysis resulting from the rest of the six data matrices have shown two significant factors according to their eigenvalues. The six sampling sites left for analysis are: Anaheim, Azusa, Burbank, Central LA, Hawthorne; and Long Beach.

4.2.2 Determination of the Feasible Region

The first two orthogonal eigenvectors from the singular value decomposition (SVD) of the data now can be utilized to compose a new analysis space. In order to ensure this source apportionment results to be physically meaningful, two additional physical constrains must be applied to limit the solution region:

- the source composition estimates must be positive; and,
- estimates of contribution of the two sources must be positive.

These two constraints can be expressed as two sets of lines in the eigenvector plane (Lawton and Sylvestre, 1971). This is shown in Figure 4-2 for the Anaheim data set, which shows the plot of the first two principal component scores for each data point. All source compositions with all positive values must lie above the solid lines in the Figure. All source compositions that give positive source contributions lie below the dotted lines in the Figure. These two constraints are satisfied in two regions of the Figure lying between the solid lines and the dotted lines. The left hand region is the contains all possible roadway source compositions, and the right hand region contains the possible soil compositions.

The determination the roadway and soil source compositions comes down to picking a point in each region. This is done in two parts, first the direction of the point is determined; this gives the relative source composition, i.e., the composition up to an arbitrary scale factor. The scale factor is determined by

the additional physical constraint that the model must explain the total mass. A graphical method for finding the directions and scale factors is given next.

4.2.3 Determination of the Source Compositions

Some additional physical or mathematical constraints are needed to determine the direction of the roadway and soil source vectors. Physically, the soil should have no or very little elemental carbon. The constraint that elemental carbon is zero is satisfied by all points on the right hand solid line. Thus, the soil source vector must lie along this direction. The roadway source vector is chosen to lie along the left hand dotted line. This line gives a roadway source that has a maximum amount of the soil elements, Si, Al, Fe, Ca, Ti, and Fe. Thus, this choice maximizes the amount of PM-10 explained by roadway. Also, this choice makes the two source vectors almost perpendicular, and, therefore, almost statistically independent.

Both the optimal directions of soil-dust and roadway have been decided above. However, scale factors need to be determined in order to obtain the final composition of the two source profiles. The procedure of determining the scale factors can be separated into two steps that ensure that the two sources explain a maximum of the observed total mass. Since sulfate, nitrate, and ammonium are significant parts of the total mass but are not explained by the two sources of interest, the reduced total mass is defined as the observed total (PM-10) mass minus the concentrations of these three species. Figure 4-3 is a plot of the unscaled roadway source contributions versus the observed reduced total mass. The slope of the lower edge of this plot, ignoring any outliers, is the required scale factor. The lower edge is used to account for the contributions of other sources to reduced total mass.

The scale factor for the soil source is determined by subtracting the roadway contribution from the reduced total mass and plotting this against the unscaled soil source contributions. The result is found in Figure 4-4. As with the roadway source, the slope of the lower edge of this plot (ignoring outliers) defines the scale factor for the soil source.

4.3 Results of Source Apportionment

The source composition estimates of the roadway and soil dust sources of PM-10 as estimated for five sites is given in Table 4-VII. In general, the source compositions are remarkably consistent between sites. Since the results for each site are estimated completely independently, this consistency is a measure of the validity of the modeling approach.

The source contributions, expressed as a percentage of the average, total PM-10 at each site is given in Table 4-VIII. Overall, the sum of roadway and soil explains about 50 percent of the PM-10, secondary inorganic species are about 35 percent and the remaining 15 percent is probably mostly secondary organic particulate matter.

4.3.1 Other Modeling Results

Limitations of the SAFER model can be best judged by comparing Tables 3-1 and 3-2 from South Coast Air Quality Management Report's PM10 Modeling and Visibility for the South Coast Air Basin (1994). These provide more detailed analysis and more specific sources such as motor vehicles instead of roadway and marine sources. These types of analysis require less intense data gathering efforts and seem to provide information more useful to control strategy planning.

4.4 Conclusions

Multivariate receptor modeling of the PM-10 particulate composition data showed:

- The PM-10 composition data for five sites from the summer intensive, Anaheim, Azusa, Burbank, Hawthorne, and Long Beach, is amenable to multivariate receptor modeling.

- Two source types, roadway and soil, account for most of the non-secondary PM-10. Roadway is defined as all PM-10 emissions from roadways, including direct tailpipe emissions and reentrained road dust. Soil is defined as all crustal material not directly associated with roadway emissions.
- As determined by multivariate receptor modeling, the composition of roadway and soil sources varies little between the five sites.
- The average PM-10 source apportionment allocates 30 percent of the total PM-10 to roadway, 18 percent to soil, 35 percent to inorganic secondary species, and the remaining 17 percent is, by process of elimination, an upper limit on the amount of secondary organic particulate matter.

Tables

Table 4-I Abbreviations of Site Names

Sampling Site	Abbreviation
ANAHEIM	ANAH
AZUSA	AZUS
BURBANK	BURK
CLAREMONT	CLAR
DOWNTOWN LA	DOWN
HAWTHORNE	HAWT
LONG BEACH	LONG
RIVERSIDE	RIVR
SAN NICKOLAS IS.	SNIC

Table 4-II Species Included in the PM-10 Analysis for Each Site

	Element	ANAH	AZUS	BURK	CLAR	DOWN	HAWT	LONG	RIVR	SNIC
1	Na	x	x	x	x	x	x	x	x	x
2	Mg	x	x	x	x	x	x	x	x	
3	Al	x	x	x	x	x	x	x	x	x
4	Si	x	x	x	x	x	x	x	x	x
5	P	x	x	x	x				x	
6	Cl	x	x	x	x	x	x	x	x	x
7	K	x	x	x	x	x	x	x	x	x
8	Ca	x	x	x	x	x	x	x	x	
9	Ti	x	x	x	x	x	x	x	x	
10	Mn		x		x				x	
11	Fe	x	x	x	x	x	x	x	x	x
12	Zn				x					
13	Ba	x	x	x	x	x		x	x	

Table 4-III Non-uniform Deposit Correction Factors for PM-10

	Element	Conversion Factor	No.	Element	Conversion Factor
1	Na	1.13	8	Ca	2.09
2	Mg	1.13	9	Ti	2.09
3	Al	1.13	10	Mn	1.00
4	Si	1.13	11	Fe	2.09
5	P	1.00	12	Zn	2.09
6	Cl	1.00	13	Ba	1.00
7	K	2.09			

Table 4-IV Number of Samples in the Fall PM-10 Data

Sampling Site	No. of Rows
ANAH	27
BURK	30
DOWN	28
HAWT	30
LONG	24
RIVR	29

Table 4-V Species Included in the PM-2.5 Data Sets

No.	Element	ANAH	AZUS	BURK	CLAR	DOWM	HAWT	LONG	RIVR	SNIC
1	Na	x	x	x		x	x	x		
2	Mg	x	x	x	x	x	x	x	x	
3	Al		x						x	
4	Si		x	x	x	x		x	x	
5	P		x	x	x	x			x	
6	K	x	x	x	x	x	x	x	x	
7	Ca		x	x	x	x			x	
8	Ti		x	x	x				x	
9	Fe	x	x	x	x	x			x	
10	Zn		x	x	x					
	Count	46	44	41	42	44	43	50	47	N/A

Table 4-VI Species Included in the Final Summer PM-10 Data Sets for 2D-SAFER Modeling

Element	ANAH	AZUS	BURK	CLAR	DOWN	HAWT	LONG	RIVR	SNIC
El. C	x	x	x	x	x	x	x	x	x
Al	x	x	x	x	x	x	x	x	x
Si	x	x	x	x	x	x	x	x	x
Ca	x	x	x	x	x	x	x	x	
Ti	x	x	x	x	x	x	x	x	
Fe	x	x	x	x	x	x	x	x	x
Mn		x		x				x	
No. of Obs.	50	43	47	38	41	44	41	45	44

Table 4-VII Source Composition Estimates for Summer PM-10.

	Anaheim		Azusa		Burbank	
Element	Roadway Wt%	Soil Wt%	Roadway Wt%	Soil Wt%	Roadway Wt%	Soil Wt%
El. C	13.61	0	7.72	0	13.98	0
Al	3.22	5.12	3.22	5.34	2.53	3.62
Si	7.58	15.23	7.58	14.29	6.34	10
Ca	1.98	4.05	1.72	2.91	1.96	2.89
Ti	0.27	0.42	0.22	0.39	0.23	0.51
Fe	2.28	4.57	2.38	4.04	2.29	3.57

	Hawthorne		Long Beach		Average	
Element	Roadway Wt%	Soil Wt%	Roadway Wt%	Soil Wt%	Roadway Wt%	Soil Wt%
El. C	10.9	0	12.73	0	11.79	0
Al	3.11	3.78	2.07	5.74	2.83	4.72
Si	6.37	11.67	4.34	15	6.44	13.24
Ca	1.27	3.25	1.59	3.21	1.70	3.26
Ti	0.18	0.37	0.14	0.61	0.21	0.46
Fe	1.65	3.43	1.58	4.34	2.04	3.99

Table 4-VIII Source Contribution Estimates for Summer PM-10

Source	Anaheim Wt%	Azusa Wt%	Burbank Wt%	Hawthorne Wt%	Long Beach Wt%	Average Wt%
Roadway	23	49	34	20	25	30
Soil	16	22	15	18	20	18
Sulfate	16	10	14	23	20	17
Nitrate	14	12	14	11	11	12
Ammonium	6	5	7	8	6	6
Other	25	2	16	20	18	16

TABLE 3-1

Annual Average PM10 Source Contribution Estimates ($\mu g/m^3$)

	DLAP[1]	BURK	LGBH	LENX	ANAH
Ammonium Sulfate	6.7 (11.3)	6.4 (11.5)	7.3 (14.6)	7.2 (15.3)	6.3 (12.0)[2]
Ammonium Nitrate	11.2 (18.8)	10.2 (18.4)	8.3 (16.7)	7.8 (16.7)	9.9 (18.8)
Secondary Carbon	7.8 (13.1)	8.0 (14.5)	5.5 (11.2)	5.5 (11.7)	5.4 (10.2)
Motor Vehicle	6.9 (11.5)	6.2 (11.3)	4.8 (9.7)	4.6 (9.8)	3.5 (6.7)
Geological	23.0 (38.6)	21.2 (38.4)	19.2 (38.6)	16.4 (34.9)	23.5 (44.6)
Residual Oil	1.3 (2.2)	1.1 (2.0)	1.5 (3.1)	1.6 (3.4)	1.0 (2.0)
Marine	2.7 (4.5)	2.2 (3.9)	3.0 (6.1)	3.8 (8.1)	3.0 (5.7)
Total Mass Predicted	59.5 +- 2.5	55.2 +- 2.2	49.6 +2.0	46.9+-2.3	52.7+-2.5[3]
Total Mass Observed	60.2+-3.6	56.6+-3.6	50.2+-2.9	47.0+-3.0	52.1+-3.1

	RUBI	UPLA	TAFT	SANI	DLAF[4]
Ammonium Sulfate	6.1 (6.9)	5.9 (10.3)	4.8 (14.9)	2.0 (12.1)	6.6 (18.9)
Ammonium Nitrate	22.7 (25.4)	14.4 (24.9)	3.9 (12.2)	2.0 (11.8)	6.3 (18.0)
Secondary Carbon	6.2 (7.0)	5.2 (9.1)	4.9 (15.2)	1.7 (9.9)	9.4 (26.9)
Motor Vehicle	6.8 (7.7)	3.6 (6.2)	2.1 (6.4)	0.5 (3.1)	8.6 (24.8)
Geological	37.6 (42.1)	26.1 (45.3)	13.3 (41.5)	3.9 (23.1)	2.8 (8.0)
Residual Oil	0.7 (0.8)	0.8 (1.4)	2.0 (6.3)	0.5 (2.7)	0.9 (2.6)
Marine	2.0 (2.2)	1.6 (2.9)	1.1 (3.5)	6.2 (37.1)	0.3 (0.9)
Limestone	7.0 (7.9)				
Total Mass Predicted	89.2+-4.0	57.7+-3.1	32.1+-1.5	16.8+-0.9	34.8+-1.3
Total Mass Observed	85.1+-6.0	58.0+-4.7	31.9+-2.5	21.1+-1.9	35.4+-2.1

[1]PM10 at downtown Los Angeles
[2]Values in parentheses indicates % of total mass predicted.
[3]Mean $\pm$ standard error
[4]PM2.5 at downtown Los Angeles

TABLE 3-2

Maximum 24-hr Average PM10 Source Contribution Estimates ($\mu g/m^3$)

	DLAP[1]	BURK	LGBH	LENX	ANAH
Ammonium Sulfate	3.8 (2.1)	2.1 (1.2)	28.9 (28.6)	7.9 (5.4)	3.4 (2.9)[2]
Ammonium Nitrate	67.8 (37.6)	69.4 (39.1)	19.2 (19.1)	63.3 (43.2)	49.1 (41.0)
Secondary Carbon	15.2 (8.4)	20.0 (11.3)	15.6 (15.5)	16.6 (11.3)	9.8 (8.1)
Motor Vehicle	21.1 (11.7)	30.8 (17.4)	7.4 (7.4)	19.6 (13.4)	14.0 (11.7)
Geological	68.6 (38.1)	52.7 (29.7)	27.0 (26.8)	32.8 (22.4)	40.7 (33.9)
Residual Oil	2.0 (1.1)	1.4 (0.8)	2.6 (2.6)	3.6 (2.5)	1.2 (1.0)
Marine	1.6 (0.9)	1.2 (0.7)	0.0 (0.0)	2.8 (1.9)	1.8 (1.5)
Total Mass Predicted	180.3+-9.2	177.5+-9.4	100.8+-4.6	146.6+-7.9	119.9+-6.3[3]
Total Mass Observed	186.9+-3.3	187.3+-3.2	131.0+-3.0	145.8 +-3.1	129.7 +- 3.0
Date	Dec. 4, '86	Dec. 4, '86	Mar. 27, '86	Dec. 4, '86	Dec. 4, '86

	RUBI	UPLA	TAFT	SANI	DLAF[4]
Ammonium Sulfate	26.0 (10.0)	20.0 (10.9)	9.7 (14.7)	1.2 (3.9)	3.9 (2.9)
Ammonium Nitrate	124.3 (47.7)	86.5 (47.3)	13.9 (21.1)	2.8 (9.3)	69.3 (52.4)
Secondary Carbon	18.8 (7.2)	15.5 (8.5)	14.6 (22.1)	0.0 (0.0)	19.1 (14.4)
Motor Vehicle	18.2 (7.0)	1.5 (6.3)	10.2 (15.4)	1.4 (4.6)	32.5 (24.6)
Geological	61.8 (23.7)	46.6 (25.5)	15.9 (24.0)	22.4 (74.7)	5.8 (4.4)
Residual Oil	1.6 (0.6)	2.5 (1.4)	1.8 (2.7)	1.1 (3.8)	1.0 (0.8)
Marine	0.7 (0.3)	0.2 (0.1)	0.0 (0.0)	1.1 (3.8)	0.7 (0.5)
Limestone	9.2 (3.5)				
Total Mass Predicted	260.6+-10.8	182.8+-10.4	66.0+-3.1	30.0+-1.1	132.3+-10.7
Total Mass Observed	298.7+-3.6	208.7+-3.2	75.1+-2.7	84.8+-3.3	138.8+-3.8
Date	Oct. 29, '86	Oct. 29, '86	Oct. 29, '86	Jan. 26, '86	Dec. 4, '86

[1] PM10 at downtown Los Angeles
[2] Values in parentheses indicate % of total mass predicted
[3] Mean $\pm$ standard error
[4] PM2.5 at downtown Los Angeles

Figures

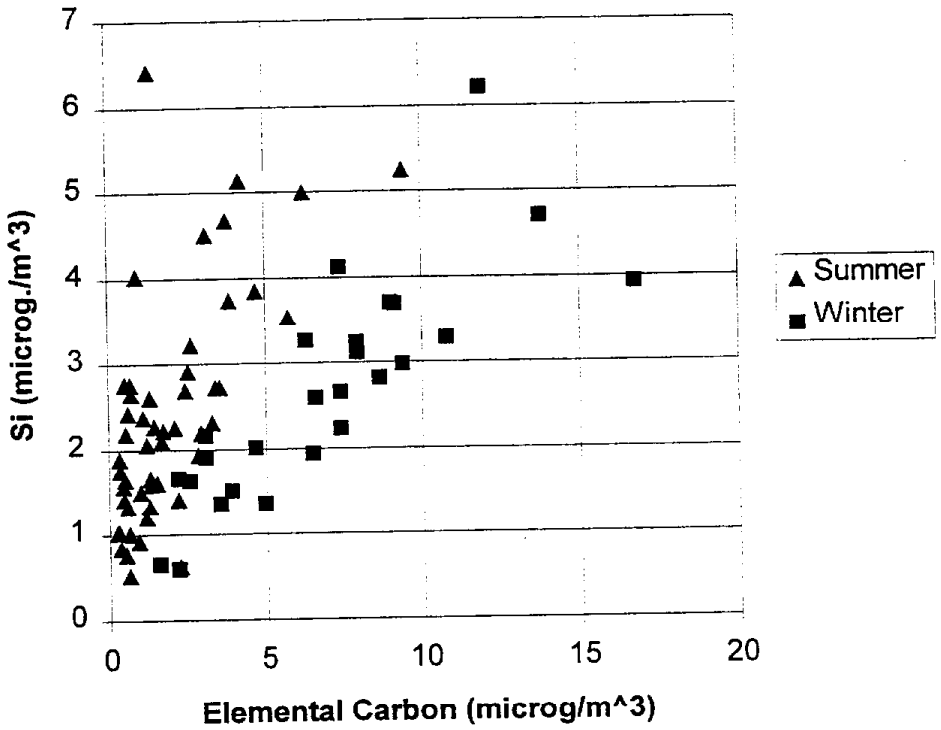


Figure 4-1 Plot showing different behavior of summer and fall elemental composition data.

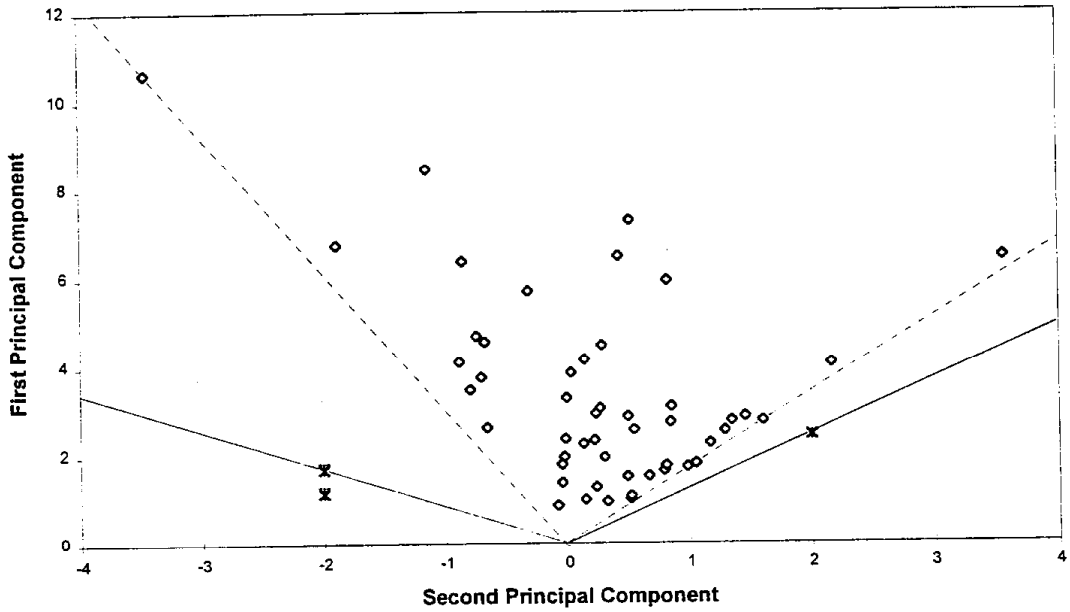


Figure 4-2 2-dimensional SAFER plot for Anaheim summer PM-10 data. The region below the left hand solid line represents source compositions with one or more soil element less than zero. The area below the right hand line represents source compositions with elemental carbon less than zero. The left hand region between the solid and dotted lines contains all feasible roadway source vectors and the right hand region contains all feasible soil composition vectors.

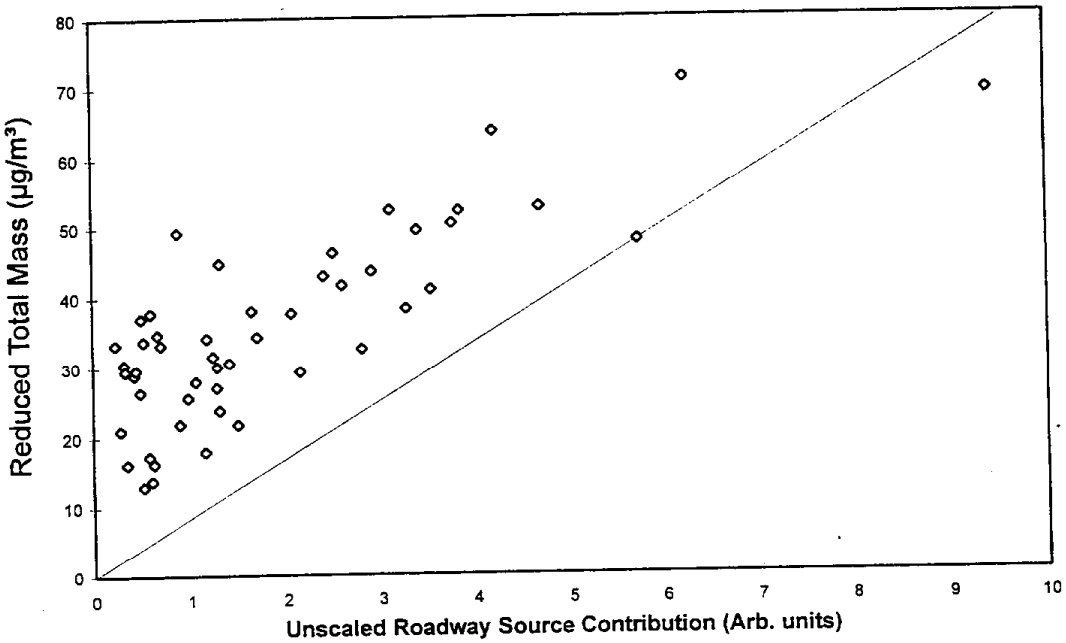


Figure 4-3 Roadway source versus reduced total mass (PM-10 total mass minus sulfate, nitrate, and ammonium).

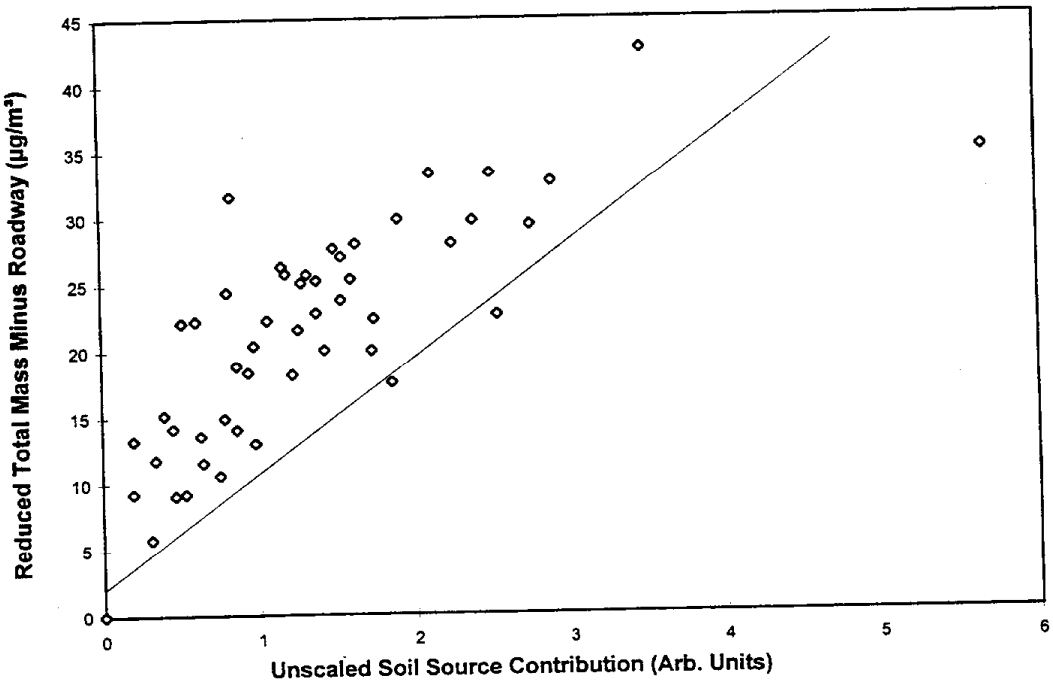


Figure 4-4 Unscaled soil source contributions versus reduced total mass minus the roadway source contribution estimates.

5. Hybrid Source – Receptor Modeling

5.1 INTRODUCTION

Hybrid source – receptor models use observed concentrations (like receptor models) and meteorological observations (like source – oriented models). The existence of the wind fields developed for the intensive study days of the SCAQS lead to the attempt to apply the Source Identification Through Empirical Orthogonal Functions (SITEOF) hybrid source – receptor model. Like all receptor models this model is at heart a mass balance model that seeks, with the aid of winds and mixing heights, to determine the source emission rate of a pollutant, less any sinks that pollutant may have. The mathematical details are described in the first section of this report.

The attempt was made to apply SITEOF to carbon monoxide and nitrogen oxide concentrations during the fall intensive days. Concentrations of carbon monoxide are too low during the summer to apply the model. Previous experience implied that at least one month of hourly data was needed to apply SITEOF successfully. In this case, only six of the fall intensive days had wind data that could be used. The decision was made to try to apply the model anyway. The final results, however, of applying the model to CO and NOX data was disappointing. No physical sense could be made of the results, the numbers were either much too big or much too small.

The following section details the considerable work that was put into obtaining correct wind fields and mixing heights.

5.2 DATA

5.2.1 Air Quality Data

Carbon monoxide and nitrogen oxide data from the fall intensive sampling periods of the SCAQS was used because the concentrations of carbon monoxide and the nitrogen oxides were higher in the fall. The requirements of SITEOF, described below, allowed us to choose 18 sites which would be useful in the analysis. The sites chosen along with their UTM coordinates are given in Table 5-I.

5.2.2 Meteorological Data

The wind fields used for this analysis were prepared by Systems Applications International (1991) for the California Air Resources Board . The wind fields used were hourly, gridded wind fields for each of the six fall intensive sampling days. The wind fields were available at six altitudes 10, 100, 300, 600, 1000, and 1500 meters above ground level. In general, the three lower levels were used for reasons related to the mixing height.

Mixing height data were supplied by the Air Resources Board. This data was not found to be sufficient and mixing heights were developed as outlined below. The mixing heights were developed using a method developed for this study. Temperature profile data provided by the California Air Resources Board was used to determine the mixing heights. The sites where data were made available and their UTM coordinates are given in Table 5-II.

5.3 METHODS

5.3.1 Mixing height fields

Mixing heights were determined by plotting the temperature profile data for the sites given in Table 5-II. After the temperature profile was plotted, the mixing height was determined by inspecting the resulting plot and looking for temperature inversions. Additionally, when the temperatures were near the adiabatic lapse rate, if there was no appreciable driving force to increase the mixing height, the mixing

height was set to be near where the temperature profile was hours before. It was found that by comparing the plots at different times of the day against each other, the mixing height could be determined from the plot. During the early morning hours and the late evening, about 2100 hours to 0700 hours the mixing height was found to be around 50 meters or less. The mixing height was set to 50 meters as a minimum mixing height.

5.3.2 Emission Rates and Locations Estimated from SITEOF

Emission rates were determined by using the SITEOF model applied to carbon monoxide and nitrogen oxides. SITEOF as described by Henry et. al. (1991) uses air quality data along with meteorological data to determine the average balance of sources and sinks of a species within a well defined area. Briefly some requirements of SITEOF are:

1) Time resolved air quality data given as an average over some time period (in this case hourly air quality data was supplied at various sites)
2) Time resolved wind fields (in this case hourly gridded wind fields were supplied)

with these two types of data defined one can find a rate of change in concentration/time period in a defined area. To determine absolute amounts over an area-region, the mixing height is needed. Thus another requirement is:

3) Gridded mixing heights or the ability to determine them.

5.4 RESULTS AND DISCUSSION

For CO the SITEOF estimates the area source emission rates since, on an urban scale, the sinks of CO are very small. For NOX, the SITEOF results would be affected by the substantial sinks of NOX as it is chemically converted in the atmosphere. However, for both NOX and CO, the SITEOF results did not make physical sense. The estimated results were too small, too large, or even negative, depending on the geographical region. The failure of SITEOF in this case could be caused by a number of factors, the most likely being the small number of days of data. SITEOF relies on the cancellation of errors that comes with calculations averaged over a long time period. In this case, six days may be too few for this to happen. Also, the possibility exists that, despite the best efforts of all involved that the wind fields were inaccurate. SITEOF is especially sensitive to poor wind input if there are only a few days of data, as was the case here.

Table 5-I Sites Used in the SITEOF Analysis

Location Symbol	UTM Coordinates(km)		Location
	Easting	Northing	
ANAH	415.0	3742.0	Anaheim
AZUS	414.9	3777.4	Azusa
BURK	379.5	3783.0	Burbank
CELA	386.9	3770.1	Central Los Angeles
COST	413.8	3724.2	Costa Mesa
HAWT	373.4	3754.3	Hawthorn
LA AV	398.0	3737.0	Los Angeles Rover
LAHB	412.1	3754.2	La Habra
LBCC	394.7	3743.7	Long Beach City College
LGBH	390.0	3743.0	Long Beach
LYNN	388.2	3754.8	Lynwood
PASA	396.1	3777.3	Pasadena
PICO	402.3	3764.1	Pico Rivera
POMO	430.7	3769.6	Pomona
RESE	358.7	3785.1	Reseda
TORO	432.0	3725.2	El Toro
WHIT	405.3	3754.0	Whittier
WSLA	365.7	3768.6	West Los Angeles

Table 5-II Sites where Upper Air Sounding Data Estimated Mixing Heights

Location Symbol	UTM Coordinates(km)		Altitude (m)	Location	Type	
	Easting	Northing				
BUR	374.8	3785.0	236	Burbank/Glendale/Pasadena Airport	SoCAB	
EMUA	402.1	3770.0	76	El Monte-9528 Telstar	SoCAB	
FUUA	412.3	3747.6	30	Fullerton-SCE Maintenance Yard	SoCAB	
LBCC	394.7	3743.7	17	Long Beach City College	SoCAB	
LMUA	271.7	3760.4	45	Loyola Marymount University-Engng Bldg	SoCAB	
ONT	577.6	3768.4	290	Ontario International Airport	SoCAB	

6. Conclusions

The primary conclusions for VOCs are:

- VOC data from the various sites from the SCAQS fall intensive can be consolidated into a single data set, i.e., all sites seem to be affected by the same sources. This is not true for summer VOC data.
- TNMHC and most of the individual VOC species are well explained by three vehicle-related sources: roadway emissions (direct exhaust and running evaporative), gasoline, and gasoline vapor.
- The three source SAFER model was validated by its success in explaining VOC species that were not use to develop the model.
- On average, 60 percent of the TNMHC was apportioned to roadway emission, 19 percent from whole gasoline, 15 percent from gasoline vapor, and 6 percent unexplained or background..
- There was a significant difference between the morning and afternoon source apportionment. In the morning, 71 percent of the TNMHC was apportioned to roadway emission, 14 percent to whole gasoline, 11 percent to gasoline vapor, and 4 percent were unexplained. In the afternoon, 44 percent of the TNMHC was apportioned to roadway emission, 27 percent to whole gasoline, 20 percent to gasoline vapor, and 10 percent were unexplained.
- The source composition for roadway emissions as determined by the SAFER model is similar to the composition of roadway emissions seen in other studies. However, the SAFER profile shows some differences that are specific to the SCAQS data set.
- The composition of gasoline and gasoline vapor as estimated by SAFER is in general agreement with source samples. The SAFER estimate of the nbutane fraction of these sources is probably inaccurate since nbutane is over predicted. This may be the result of variations in nbutane composition of gasoline made by refiners to adjust the cold start characteristics of the gasoline.
- Ethane and propane were not well modeled by the three source model. They probably have significant non-vehicle related sources.

Multivariate receptor modeling of the PM-10 particulate composition data showed:

- The PM-10 composition data for five sites from the summer intensive, Anaheim, Azusa, Burbank, Hawthorne, and Long Beach, is amenable to multivariate receptor modeling.
- Two source types, roadway and soil, account for most of the non-secondary PM-10. Roadway is defined as all PM-10 emissions from roadways, including direct tailpipe emissions and reentrained road dust. Soil is defined as all crustal material not directly associated with roadway emissions.
- As determined by multivariate receptor modeling, the composition of roadway and soil sources varies little between the five sites.
- The average PM-10 source apportionment allocates 30 percent of the total PM-10 to roadway, 18 percent to soil, 35 percent to inorganic secondary species, and the remaining 17 percent is, by process of elimination, an upper limit on the amount of secondary organic particulate matter.

The primary conclusions from the attempted SITEOF hybrid source – receptor modeling are:

- Six days is not enough time for SITEOF to overcome limitations in real world data. To use SITEOF more time would have to be devoted to sampling- around 30 days.

- No conclusions as to the carbon monoxide or nitrogen oxide emissions can be made from this study.
- A new method of determining mixing heights was developed.

7. References

Conner, T. L., W. A. Lonneman, and R. L. Seila 1994. Transportation-related Volatile Hydrocarbon Source Profiles Measured in Atlanta. Paper in Preparation.

Harley R. A., M. P. Hannigan, and Glen R. Cass 1992. Respeciation of Organic Gas Emissions and the Detection of Excess Unburned Gasoline in the Atmosphere. Environmental Science & Technology, 26, 2395-2408.

Henry R. C. and B. M. Kim 1990. 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.

Henry, R. C.;C. W. Lewis and John F. Collins 1994. Vehicle-Related Hydrocarbon Source Composition from Ambient Data: The GRACE/SAFER Method. Environ. Science & Technology 28:823-832.

Henry, R.C., Y. J. Wang, and K. A. Gebhart 1991. The Relationship Between Empirical Orthogonal Functions and Sources of Air Pollution, Atmospheric Environment, 25A, 503-509.

Hopke, Phillip K. 1991. An Introduction to Receptor Modeling, Chemometrics and Intelligent Laboratory Systems, Vol. 10, pp. 21-43.

Kleinmann, M.T., T.J. Kneip, and M. Eisenbud 1976. Seasonal Patterns of Airborne Particulate Concentrations in New York City. Atmospheric Environment, Vol. 10, pp. 9-11.

Lawson, C.L., and Hanson, R.J. 1976. Solving Least Square Problems, Prentice – Hall, Englewood Cliffs, New Jersey.

Lawton W. H. And E. A. Sylvestre 1971. Self Modeling Curve Resolution, Technometrics, 13, 617-633.

Lurmann F. W. And H. H. Main 1992. Analysis of the Ambient VOC DATA Collected in the Southern California Air Quality Study. Final Report Contract No. A832-130, Sonoma Technology, Inc., Santa Rosa, CA

Matsumura, R.T., and T.A. Cahill 1990 Spatial Inhomogenities in SCAQS Filters, Draft Final Report on ARB Contract No. A832-128, California Air Resources Board.

Sun, Y-P, D. F. Sears Jr., and J. Saltiel 1987. Three-component self-modeling technique applied to luminescence spectra. Analytical Chemistry, 59, 2515-19.

Systems Applications International, 1991. Diagnostic Analysis of Wind Observations Collected during the Southern California Air Quality Study.

Technical Report V-C, PM10 Modeling and Visibility for the South Coast Air Basin, September 1994, South Coast Air Quality Management District.

Watson J. G., J. C. Chow, Z-Q Lu, and A. W. Gertler 1993. Particulate and Gaseous Organic Receptor Modeling for the Southern California Air Quality Study. Final Report Contract N0. A832-132, Desert Research Institute, Reno, NV.