ADDENDUM

Source Apportionment of Organic and Light Absorbing Carbon Using Receptor Modeling Techniques

by

William C. Malm
Kristi A. Gebhart
National Park Service
Air Quality Division
CIRA/Colorado State University
Fort Collins, CO 80523

INTRODUCTION

An intensive field monitoring program was undertaken during the summer of 1990 to study the contribution of emission sources to ambient particulate concentrations and regional haze at a number of National Parks in the Pacific Northwest. The acronym for the Pacific Northwest Regional Visibility Experiment Using Natural Tracers study is PREVENT. One of the objectives of PREVENT was to apportion (or attribute) the summertime haze observed in federal Class I areas in Washington to the regional emissions from all sources in the Pacific Northwest and British Columbia. This paper will focus on the apportionment of organic and light absorbing carbon to various sources at Mount Rainier National Park.

Slash and prescribed burning are forest management tools used throughout the area. These burns are conducted by federal land management agencies, states, Indian tribes, and Canadian provinces. Wildfires can also occur throughout the region, depending on weather conditions. Because viewing of scenic vistas is integral to the enjoyment of the National Parks there is concern over the potential contribution of smoke associated with fire to visibility degradation.

Table 1 summarizes state emissions of total and fine particulate released from agricultural and forest burning in 1984 for Oregon, Washington and Idaho. Figure 1 shows the trends in prescribed burning in Western Washington. Since 1977 there has been about a three-fold decrease in the amount of material burned.

FIELD MONITORING PROGRAM

The PREVENT study focused on Mount Rainier and North Cascades National Parks. Two primary sites were located at Mount Rainier at low and high elevation sites. One low elevation primary site was located at North Cascades. At these three primary sites, a complete characterization of the ambient atmospheric extinction coefficient, aerosol and gaseous concentrations, meteorology, and visual air quality conditions was performed. In addition to the primary sites, elements in the fine particle mode were measured at 31 secondary sites throughout western Washington. The locations of the monitoring sites are shown in Figure 2.

The sampler[1,2] system used in PREVENT consisted of four independent filter modules controlled by a common microprocessor system. Each of the three fine modules (0.0-2.5 μm) has a dedicated inlet, cyclone, flow control device, and pump. The Module A Teflon filter was used to measure fine mass, H, Na to Pb, and the coefficient of optical absorption (b_{abs}). Module B contained an acidic vapor diffusion denuder followed by a nylon filter and was used to measure nitrate ions. Module C contained tandem prefired quartz filters and was used to estimate carbon aerosol concentrations. The second filter was used to monitor filter artifacts. A Teflon filter in Module D (0-10.0 μm) was used to determine PM_{10} mass concentrations. Only those measurements made by Module A (elements) and C (carbon) are needed for the data analyses carried out in this paper.

Details of the analytic procedures as well as a discussion of minimum quantifiable limits, minimum detectable limits, and precision and accuracy of measured variables can be found in papers by Eldred *et al.*[2] and Chow *et al.*[3]

Organic carbon can be estimated from measurements of H and S and will be denoted by OMH. If it is assumed that an average ambient organic particle contains constant fractions of carbon and hydrogen by weight, and that during exposure to a vacuum during the analysis procedures all nitrates and water volatilize,[2] then:

$$OMH = 13.75\ (H - 0.25S) \qquad (1)$$

Concentrations of carbon species are determined by heating the filter from Module C in an oven and measuring carbon as it is vaporized from the filter. Carbon concentrations are reported as a function of oven temperature. The four reported variables are low and high temperature organic carbon (oclt and ocht) and low and high temperature elemental carbon (eclt and echt). The oclt and ocht are assumed to be in the form of organic carbon, and eclt and echt are assumed to be light absorbing carbon (LAC):

$$OC = (oclt + ocht) \qquad (2)$$

$$LAC = (eclt + echt) \quad (3)$$

An alternate estimate of light absorbing carbon can be calculated by dividing b_{abs} by an assumed carbon particle absorption efficiency of 10 m^2/gm. The resulting "concentration" then has units of mass per unit volume. To avoid confusion with LAC as determined by eclt and echt, the name b_{abs} is retained. These b_{abs} "concentrations" and OMH (rather than OC) were used for analysis of spatial patterns since only the three receptor sites were equipped with Module C.

THE DATA SET

Figures 3a through 3d present the variables used in the Mount Rainier analysis. These figures show low and high temperature organic carbon (eclt and echt), and low and high temperature elemental carbon (eclt and echt) in units of $\mu g/m^3$. The remainder of the trace elements are in units of ng/m^3.

SOURCE ATTRIBUTION TECHNIQUES AND MODELS

Most visibility impairment is associated with secondary aerosols such as sulfates, nitrates, and secondary organics. Therefore, any visibility source apportionment scheme must address secondary as well as primary particles. Traditionally, the chemical mass balance (CMB) formalism has been used to apportion primary particles, and a number of review articles have addressed the associated assumptions.[4-6] CMB relies on known physical and chemical characteristics of aerosols, such as ratios of tracer species, natural or manmade, at the receptors and sources to attribute aerosols to single sources or source types. Important assumptions associated with CMB are: composition of source emissions are known and constant; aerosol components do not react with each other (implicitly this means that the ratios of various tracer species associated with a source remain constant over time); the number of sources is less than the number of tracer species; and compositions of sources are linearly independent of each other. If any of these assumptions are violated, the apportionment is degraded.[7-9]

CMB can also be used to estimate a source contribution to secondary aerosols if additional assumptions are imposed: all primary gas is converted to secondary aerosol of interest; the relative deposition of primary and secondary species is the same for all sources; the ratios of secondary aerosol to trace element species equilibrate at constant values; and the equilibrated ratios are in proportion to the relative emission rates of the primary gas[7]. The net effect of these assumptions is to apportion the secondary aerosol in direct proportion to the apportioned primary mass and the relative emission strength of each source. For instance, in a given sampling period, if 50% of the primary mass is associated with source A and source A emits 80% of all SO_2 emissions, then source A is estimated to contribute 40% of the ambient sulfate (0.5*0.8=0.4).

CMB modeling apportions aerosol species on a sampling-period by sampling-period basis. However, if the data set contains an adequate number of samples, regression techniques, along with less restrictive assumptions, can be used to estimate apportionment of secondary species.[10-12] In a regression approach, the secondary species is the dependent variable while the independent variables are tracers that are unique to a single source or group of sources. Assumptions associated with this approach are: the chemical species used as tracers are assumed to be uniquely emitted by non-overlapping groups of sources; the composition of source emissions are constant over the period of ambient sampling; deposition and conversion are constant from one sampling period to the next for each subgroup but don't have to be the same across all sources; and measurement errors are random, uncorrelated, and normally distributed.[10]

It is highly unlikely that the regression coefficients are constant for all sampling periods. This will inflate the uncertainty in the final apportionments, but the extent to which the inflation occurs will depend on the variability of the coefficients. Whitmore *et al.*[13] explored the effects of source emissions being collinear with each other, the effect of variability of deposition and conversion, and meteorological factors on apportionment of secondary aerosols to a single source. Under worst case assumptions, the relative error was about a factor of two, while the average relative error, for variation in the input variables chosen was less than one.

If emitted tracer species are not unique to a given source or source type, CMB can be used to partition the ambient concentrations into components attributable to the various source groups.

Hybrid models that combine elements of deterministic and receptor oriented models can be employed.[10,14-17] In this approach, dispersion is accounted for by rationing ambient trace material concentrations attributed to a source by known trace material release rates, while deposition and conversion are explicitly calculated. Assumptions associated with this approach are: the chemical species used as tracers are assumed to be uniquely emitted by non-overlapping groups of sources; the composition of source emissions are constant over the period of ambient sampling; and deposition and conversion rates are usually assumed to be first order and invariant in space and time.[7] However, since transport time is explicitly taken into account, deposition and conversion can vary from one sampling period to the next.

Rates for deposition and conversion probably are not first order nor invariant in space and time. For example, dry deposition will not occur until the plume has been mixed to the ground. Dry and wet deposition velocities are known to vary depending on atmospheric stability and on surface type. Conversion processes are dependent on other chemical species which vary in space and time. To date a systematic examination of the effect of assumptions associated with hybrid modeling has not been reported.

The objective of this source attribution is to determine the fraction of each light-scattering or absorbing species contributed by a given source or source category. In PREVENT a variety of attribution techniques and models are used in parallel to:

- supplement each other
- allow the strengths of one technique to offset the weaknesses and limitations of others, and
- reduce the overall uncertainties of the analysis.

For example, it has been recommended that source models and receptor models be used to expand the capabilities of each.[17] Elements of both receptor and deterministic source models have been combined by Stevens[16] and his colleagues at the Environmental Protection Agency (EPA) and by Malm[18] and his colleagues at the National Park Service (NPS).

EMPIRICAL STATISTICAL RELATIONSHIPS

The investigation of relationships between aerosol species is complicated by the inherent interdependence between variables due to underlying common physical processes. For instance, at high elevation sites, a whole array of species may rise and fall together because of the diurnal variation of mixing height. As the air, which is trapped in night-time inversions, mixes up to higher elevation monitoring sites, concentrations of all aerosol species may rise together. However, not all observables vary together. They are also influenced by spatial locations of sources and receptors and relative winds.

This section explores the relationship between the various aerosol species using multivariate techniques. In the first section, the relationship between variables is explored in a qualitative way using factor analysis techniques.[19,20] In the next section, relationships between variables are investigated using regression techniques.

Factor Analysis

The following factor analysis is exploratory in nature, seeking to find relationships among variables that may have underlying physical interpretations. Specifically, it is of interest to examine relationships between carbon and trace element species to determine whether there are any distinctive sources associated with aerosol concentrations at the receptor sites.

The underlying development of factor analysis methodologies will not be presented here. There are a number of excellent treatments of the subject.[19,20] For the application here, factors are extracted and followed by a Varimax orthogonal rotation.

The objective of factor analysis is to reduce the number of variables needed to explain most of the variance in the data set. Furthermore, by examining the variables with large weightings it is often possible to associate a component with a physical process or with specific source types. Such an examination also reveals underlying collinearity between variables.

Table 2 presents the results of factor extraction followed by a Varimax rotation. The table lists the factors in descending order according to the amount of variance explained. Only those factors corresponding to eigenvalues of the correlation matrix greater than one are retained.

Factor loadings greater than 0.60 are highlighted for reference. Factor 1 explains 35% of the variance and consists of trace elements normally associated with urban areas or transportation. Lead, manganese, zinc, and nickel all have factor loadings of greater than 0.8. High temperature organic carbon and low temperature elemental carbon are loaded highest in factor 1. This suggests an urban influence.

Factor 2, explaining 27% of the variance, seems to be a soil factor. Furthermore, relative humidity is loaded negatively. Relative humidity is a good indicator of time of day. As daytime temperatures rise the relative humidity decreases while during the evening hours the relative humidity nearly always approaches 100%. The negative loading on RH suggests that as RH increases concentrations of trace elements in factor 2 decreases. Therefore, transport of trace elements associated with factor 2 into the Tahoma Woods area may be occurring during daytime hours.

The third factor consists of two variables which are highly loaded; nitrate and sodium. Sodium is normally thought to be associated with sea salt, whereas, nitrates are not. However, there are a number of source profiles that are rich in sodium and nitrates. These include lime and cement kilns and oil combustion.

Other factors are less interesting. Factor 5 is primarily associated with low temperature organic carbon, and factor 7 with zirconium.

Regression Analysis

Linear regression analysis is often used to empirically relate a variable of interest, such as sulfates, to other aerosol species or to other atmospheric variables such as wind speed and direction. The relationships may have underlying physical interpretations such as when the wind blows from source A to receptor B, emissions from the source are transported to the receptor. Also, a source may directly emit sulfates or precursors to sulfates and a number of trace elements. The usual goal of multiple regression on observables is to find the best regression, i.e. the equation containing the fewest number of variables and explaining the greatest variation in the independent variable. A number of techniques, such as stepwise multiple linear regression, are available to meet this objective.[21] However, when applying regression analysis to environmental variables caution must be exercised because a number of variables may be highly correlated with each other.

An alternative to merely searching for the combination of variables that explain the maximum variance is to select unique aerosol species that are known to be associated with one or more sources, or to estimate the fraction of a species associated with a specific source and use the derived or unique variable in the regression analysis. In the following section, relationships among variables will be explored in an empirical way and where possible physical interpretations will be proposed.

Two strategies for selecting variables in the regression analysis were used. In the first approach one variable was selected to represent each of the factors presented in Table 2. Furthermore, if a variable was known to represent a potential source and wasn't loaded in one of the factors, it was included in the regression. In the second approach, composite variables were formed by adding together those variables in each of the factors and weighting them according to their factor loadings. All variables and only those with weighting greater than 0.70 were used. However, neither of these approaches yielded results different from just using one variable to represent each factor. Therefore, the results of only the single variable approach will be presented here.

The variables used in the regressions are extracted from Table 2. For Tahoma Woods, Pb was used to represent factor 1, Si for factor 2, Na for factor 3, K for factor 4, and Zr for factor 7. In addition, Br was included as a variable because it is known to be associated with automobile emissions; and Se because of its association with coal-fired power plants.

Stepwise regressions were carried out using the above species as independent variables and organic mass and LAC as dependent variables. Tables 3 and 4 present the results. The variable parameter estimate, standard error, and F are listed in the table as well as the order in which the variables were entered, the partial R^2, model R^2 and the average fraction of the aerosol species (dependent variable) associated with the independent variable.

Interpretation of these results should be semi-quantitative. Even though variables used in the regression models were selected on the basis of a factor analysis and should not be excessively collinear, each of the variables is associated with more than one source. The analysis is meant to give some insight into the source types associated with secondary species at the two receptor sites.

Br is the first variable to be added to the organic model. Pb is the second variable added. Br and Pb together strongly suggest transportation and burning as the sources of organic aerosol. On the other hand, LAC is more closely tied to Pb emissions suggesting a transportation source. Bromine accounts for the second largest fraction of LAC. Silicon is also linked to carbon emissions. As pointed out, significant amounts of silicon are emitted by fire-related activity and coal-fired boilers. However, it is not expected that significant amounts of organics are emitted by coal burning activity, but they may be re-entrained with blowing dust.

Another unanticipated relationship is the strong link between transportation activity and organic and light absorbing carbon levels. Organic aerosols are not only associated with agricultural or wood burning, but also to a large degree with transportation activities.

CHEMICAL MASS BALANCE (CMB)

CMB consists of a least-squares solution to a set of linear equations which expresses the ambient concentration of a given species at a given receptor as a linear sum of products of source profile species and source contributions. The source profiles are the fractional amount of each species in the emissions from each source type. Since different source categories can have quite different profiles, one can distinguish the contributions of individual species. The output consists of the amount contributed by each source type to each chemical species concentration. Traditionally, CMB has been applied solely to conservative species (e.g., those that do not form in the atmosphere).

The primary strategy taken in this section for purposes of estimating various source contributions to measured aerosol levels will be to use CMB modeling in combination with Tracer Mass Balance Regression (TMBR). CMB will be used to establish relative amounts of each tracer species associated with various sources and these fractional levels can then be use in a TMBR model to establish estimates of each source or source types contribution to the aerosol species of interest. This approach will be referred to as CMB regression (CMBR). Furthermore, results of the CMB analysis can be used to investigate how much each of the trace elements that showed a statistical link with visibility reducing aerosols is associated with various sources.

PREVENT data used in the CMB analysis consists of twelve-hour averages in the PM 2.5 μm particle fraction. The sources examined as to their potential contribution to primary particles were:

Coal-fired power plant	Kraft recovery furnace	Cement kiln
Lime kiln	Sulfite recovery boiler	Aluminum processing
Copper smelter	Residual oil combustion	Transportation
Oil-fired power plant	Marine aerosol	Wood-fired boiler
Agricultural burning	Broadcast burning	Municipal incinerator
12-soil dust profiles		

The twelve-soil dust profiles are quite similar, as are a number of other profiles such as burning and recovery boiler profiles.

CMB Source Contributions to Trace Elements

The results of the regression analysis are presented along with the measured and predicted concentrations of the various trace element species. Additionally, the concentration of trace element species associated with each source type is also presented. Species used in the analysis are Br, Cu, Fe, K, Mn, Na, Pb, V, Zn, Si, Al, Ca, Ti, Zr, Ni, As, and Se. Other elements are not used because they are below the detectable limit much of the time.

In most cases, five or less sources were adequate to develop a reasonable model as judged by the R^2 value. Addition of more sources resulted in increased standard errors or negative regression coefficients suggesting multicollinearity between variables. A variety of soil profiles were used with varying degrees of model fit. The Richland and Dog Mountain profiles in conjunction with the coal-fired power plant profile gave the best overall fit on elements such as Si, Al, Ca, and Fe. The lime kiln and kraft recovery profiles were almost interchangeable, while in many cases the interchange of some "burn profiles" didn't change the model fit significantly.

Since the analysis is being used more to derive a semi-quantitative understanding for the relative importance of various source types the CMB analysis was carried out with five generic sources: soil dust (Richland, Wash), coal-fired power plant, kraft recovery or lime profile, transportation, and burning (agricultural burning). A program was written to include all five sources in the analysis on the first "pass" and then remove, in a sequential manner, sources with negative coefficients. It is emphasized that on any given day the CMB model may be improved slightly by using different sources within each source group. However, the general trends do not change.

The first plot in Figures 4 and 5 are temporal plots of organics ($\mu g/m^3$), and light absorbing carbon ($\mu g/m^3$), respectively. The remaining temporal plots in each figure correspond to the relative contributions of burning, transportation, and soil. The variable plotted is K for burning, Pb for transportation and Si for soil. These trace elements are plotted to show relative contributions of each source type. Other trace elements could be used or the sum of the masses of all elements for a source could be used. However, the relative temporal trends would be the same since the relative ratios of one trace element to another remain constant for any given source.

Table 5 presents the fractional contribution of each source type to the trace elements used in the exploratory regression analysis, presented in the previous section. The element Si is split about 60-40% between soil and coal, while coal contributed about 90% of the Se and greater than 90% of the As. Over 85% of the K and 95% of the Cu is associated with burning. Burning is also responsible for about 20% of the observed Br. Transportation is responsible for the bulk of the Br at about 70% and Pb at greater than 90%. Either the kraft or lime profile accounts for most of the measured Na. At inland sites, such as Mount Rainier, Na may be a good tracer for pulp and paper mill activity.

Comparing Table 5 to Tables 3 and 4 gives some additional insight to the source of the various visibility reducing aerosols. The implication of the analysis as to the source of organic aerosols is interesting. Over 70% of the organics are associated with Br and Pb, and 70-90% of these two elements are linked to urban (transportation) emissions. In the empirical regression analysis, K didn't show up as a significant independent variable, and as such, burning is not linked to organics. Furthermore, most of the light absorbing carbon (approximately 70%) is associated with urban activity.

These relationships can be further explored using regression techniques with the aerosols of interest as the dependent variables and the relative source emissions as independent variables. These results are summarized in Tables 6 and 7.

The organic and light absorbing carbon models are highly significant with R^2s of 0.80 and 0.91, respectively. The results are surprising but consistent with the empirical regression model. The CMBR model

implies that transportation is by far the largest contributor to both organic and light absorbing carbon. Transportation activity contributes about 50% of measured organics and 63% of light absorbing carbon while burning contributes only about 11% and 13%, respectively. Surprisingly, the soil signature is associated with a significant fraction of organics and light absorbing carbon. Possibly, carbon material is re-entrained in the atmosphere along with soil during periods of elevated winds.

SPATIAL AND TEMPORAL TRENDS

Although not a quantitative technique, spatial and temporal trend analyses can be used to determine whether it is likely that a local source or a combination of regional influences is dominating a given impact. If concentrations are relatively homogeneous throughout a region and maxima occur at roughly the same time, one might expect distant source contributions. However, if concentrations are quite nonuniform and maxima occur at different times, it would be difficult to postulate that long-range transport was the key contributor rather than local sources.

One of the objectives of PREVENT was to determine the spatial and temporal patterns of aerosol concentration and chemical composition which occurred in western Washington during the summer of 1990 and to see if these patterns were consistent with the statistical and CMBR analyses. These patterns were examined systematically using empirical orthogonal function (EOF) analysis. EOF analysis reduced all observed spatial patterns for each measured parameter to a few typical patterns which explain most of the variance in the data. Also, the spatial and temporal patterns were examined using simple time plots of simultaneous measurements at the receptor sites, and by creating contour plots of mean concentrations for each species.

Empirical orthogonal function (EOF) analysis can be used to systematically examine the spatial and temporal patterns in measured concentrations (or other data) across several monitoring sites in a region. While these patterns can be observed simply by creating isopleth maps of the concentrations for each measurement period, EOF analysis has the advantage of simplifying these many maps into a few patterns (EOFs) which can be linearly recombined to reproduce the observations for all time periods.

A single data matrix (time by site) of pollutant concentrations is decomposed into two matrices, one independent of time (EOF by site) and the other independent of location (time by EOF). Multiplication of the resulting two matrices regenerates the original matrix.

The EOF by site matrix can be thought of as a series of isopleth maps (one for each EOF), of common concentration patterns in space, while the time by EOF matrix, plotted as a series of time plots, also one for each EOF, reveals the relative importance of each EOF for each sampling period. The advantage of the analysis is that while there are as many EOFs as there are sites, in general only a few are needed to explain most of the variance in the data.

The EOF analysis is useful because there are usually only a few similar physical conditions that occur during a field study in a given region and time period. For example, one hypothetical condition might be stagnation around the pollutant sources with low wind speeds and relatively little transport. In this case, the concentrations near sources would be much higher than the concentrations at receptors farther distant from any source. Another condition which occurs during the same study might be moderate westerly winds that transport pollutants eastward. In this case, receptors west of sources would have low or zero concentrations. Receptors east of sources would have concentrations which decreased as the distance from the source increased. These are simple examples. In the real world, interpretation of spatial concentration patterns is complicated by other factors such as complex terrain, changing emission rates, and variables affecting chemical transformation such as relative humidity, clouds, and temperature.

Although identical meteorological and chemical conditions will never occur repeatedly, in general, all spatial patterns in the pollutant concentrations can be expressed as linear combinations (with different weights for each spatial pattern for each monitoring period) of just a few conditions which occur repeatedly in a similar manner. Sources are usually located near areas where the spatial gradients in the concentration field (and EOF field) are largest.[22] The EOF analysis does not explicitly reveal the physical conditions which are associated with each spatial pattern in the concentrations, but often the physics can be inferred as described in the

examples above. The hypotheses can then be further investigated by examining additional information, such as meteorology and inter-species relationships.

Procedure

A matrix of centered concentrations of a single chemical species (mean for the site subtracted from each concentration) is decomposed into two matrices as follows:

$$Z = A \times P \quad (4)$$

where:

Z = Centered concentration matrix with the rows corresponding to times of observation and columns to locations.

A = Orthonormal matrix of dimensionless time weighing factors. Rows correspond to time periods and columns correspond to EOFs.

P = Orthogonal matrix of spatial EOFs. Rows are EOFs; columns correspond to sites. Values are weighted deviations from the mean concentration at each site.

P is the matrix of eigenvectors of Z^TZ multiplied by the matrix which contains the square roots of the eigenvalues of Z^TZ on the diagonal and zeros elsewhere. The eigenvalues have units of concentration squared so entries in the P matrix are units of concentration. A is found by postmultiplying the centered data, Z, by the inverse of P. More details of the decomposition can be found in texts on multivariate analysis.[23,24]

EOFs are orthogonal, statistically simple, and each explains a decreasing amount of the total variance, but they are not a unique decomposition of the data and frequently the patterns associated with each are more physically interpretable and better represent the patterns in the input data after rotation. The advantages and disadvantages of orthogonally and obliquely rotated and unrotated EOFs are discussed in detail by Richman.[25] For this study, the EOFs were rotated using the Varimax criteria which forces each site to load as strongly as possible onto one EOF while loading weakly onto all others.[26] This can be useful for determining the source areas and meteorological conditions which most influence each site. The number of factors to rotate was determined by the number of eigenvalues of the correlation matrix which were greater than one.

Missing Data

Data must exist for every site for every time period which is included in an EOF analysis. Data collection at most PREVENT sites began June 21, 1990, and ended September 3, 1990. However, nine sites had missing data on September 3, and data collection at two sites, Sauvie Island and Paradise did not begin until July 18 and 20, respectively. Therefore, to include all sites, the time period chosen for analysis was July 20 through September 2, 1990.

Two steps were taken to deal with missing data during this analysis period. First, missing concentrations preceded and followed by valid data at the same site were filled by linear interpolation across time. Approximately 1% of the data were filled in this way. Other missing data were replaced by a distance weighted (inverse of the squared distances) average of concentrations at sites with non-missing data for the same time period. Approximately 2% of the data were replaced by this method.

Results

Organic Matter

A contour plot of mean OMH is presented in Figure 6. The highest mean concentrations were at Puyallup (6.4 $\mu g/m^3$), North Bend (6.1 $\mu g/m^3$) and Sultan (5.8 $\mu g/m^3$), with values generally decreasing outward

from these sites. The mean values at the receptor sites were 4.9 $\mu g/m^3$ at Tahoma Woods, 2.9 $\mu g/m^3$ at Paradise, and 4.4 $\mu g/m^3$ at Marblemount. (See Figure 1 for location of the monitoring sites.)

Four EOFs, explaining a total of 82.3% of the variance were rotated for OMH. These are shown in Figures 7 through 10, and the corresponding time factors are shown in Figure 11. The first EOF, indicates an urban source for organic matter. Also of note for EOF 1 is the cyclic nature of the time factors. The cycle appears to be approximately weekly, especially for the first four weeks, with minimums on Saturday and maximums early in the week, on Sunday, Monday, or Tuesday. The pattern shifts forward a day or two in the last 2 1/2 weeks, when minimums are on Monday and Tuesday, and maximums around Wednesday and Thursday. The first day in Figure 11 is Friday, and the days are labeled every 7 days, i.e., July 20, July 27, etc.

The second EOF explaining 22.6% of the variance has the strongest gradient in a strip running north-south through the entire length of the state through and just west of the urban corridor. This pattern is also suggestive of urban sources. In this case, the associated meteorology is predominantly westerly winds.

The third and fourth EOFs for OMH explain 11.8 and 9.2% of the variance, respectively and show strong gradients around single monitoring sites, North Bend on EOF 3 and Puyallup on EOF 4.

The time factors show that EOF 3 is weighted highest on August 10 and 23. The two highest OMH concentrations at North Bend, 19.4 and 17.8 $\mu g/m^3$, respectively, were also on these days. August 10 also had the highest fine mass (22.9) and b_{abs} (3.6) concentrations measured for this site. On August 23 the fine mass concentration at North Bend was 14.3 $\mu g/m^3$ and b_{abs}, was 1.8 $\mu g/m^3$. Both these concentrations are about average for this site. It seems likely that there was a nearby fire on August 10. The cause of the relatively high OMH concentration on August 23 is less clear, but it may also have been due to a fire.

The EOF 4 pattern results from an unusually high OMH concentration of 30.5 $\mu g/m^3$ at Puyallup on a single day, August 27. The next highest OMH concentration at any site was 19.4 $\mu g/m^3$. The concentration, though high, appears to be real. The fine mass concentration at the same site on this day was 32 $\mu g/m^3$ (highest fine mass concentration anywhere during the study was 35.2) and the b_{abs} concentration was 8 $\mu g/m^3$ (highest anywhere during the study). The pattern shown in EOF 4 is probably the result of a nearby fire.

Absorption (b_{abs})

Because LAC was not measured at the receptor sites, spatial gradient analysis is carried out using b_{abs}. The contour plot of mean fine b_{abs} is shown in Figure 12. The highest mean concentration was 2.2 $\mu g/m^3$ measured at Puyallup. Three other sites, North Bend, Kent, and Sultan also had mean concentrations greater than 1.5 $\mu g/m^3$. Packwood, near the southeastern corner of the network had a mean concentration of 1.3 $\mu g/m^3$, which is higher than the other sites nearby. Mean b_{abs} concentrations at the receptor sites were 1.1, 0.8, and 0.9 $\mu g/m^3$, at Tahoma Woods, Paradise, and Marblemount, respectively.

Six b_{abs} EOFs, accounting for a total of 86.8% of the variance were retained for Varimax rotation. Contour plots of the first four are shown in Figures 13 through 16, and the corresponding time factors are shown in Figure 17. EOF 1 for b_{abs}, which accounts for 27.3% of the variance shows several areas of strong gradients, the strongest is around Puyallup, with others around Rochester and Sultan and in the southwestern corner of the network near Ohanapacosh and Packwood. This pattern suggests that b_{abs} concentrations may often be due to several localized sources throughout the network. Slash and control burns are one possibility. The second EOF, accounting for 21.4% of the variance shows a gradient centered around Kent. This pattern was weighted the strongest on August 11, when the b_{abs} concentration was the second highest measured at Kent.

EOFs 3 and 4, which account for 12.5 and 12.1% of the variance, respectively, are indicative of urban sources, though this interpretation is not as clear as were EOFs 1 and 2 for OMH. The strongest gradients on EOF 4, for example appear to radiate outward from the Seattle area, but are somewhat far north to implicate Tacoma. The apparent gradient around Packwood on EOF 3 may be an artifact of nine days of missing b_{abs} data at this site. A qualitative look at the time plots of b_{abs} at Packwood and other sites nearby, including Paradise, Ohanapacosh, and Glenoma show no obvious differences in the concentrations between these sites.

Potassium

Spatial patterns with potassium are explored because CMB analysis showed it to be almost uniquely associated with burning. All potassium concentrations except two were 611 ng/m^3 or less. The two highest concentrations, 3529 ng/m^3 on August 3 at Stampede Pass and 1371 ng/m^3 on August 17 at Sauvie Island were examined to try to determine whether or not they were valid data. The 1379 ng/m^3 value was determined to be probably valid because the potassium concentrations at 23 other sites were also higher than usual on the same day. This indicates that there was some physical reason for the high concentration.

It was more difficult to determine the validity of the highest concentration. EOF analyses were conducted with this concentration both included and deleted. When the value is included, it loads exclusively onto rotated EOF 1 and the spatial patterns associated with the remaining EOFs are not much different than if it is removed. Therefore, only the results with this concentration included are presented.

A contour plot of the mean potassium concentrations is shown in Figure 18. It is unusual compared to the other species examined so far, in that the highest mean value, 155 ng/m^3, is at Sauvie Island on the southern edge of the PREVENT network. Even if the high value measured on August 17 is removed, Sauvie Island still has the highest mean concentration (127 ng/m^3). The next highest concentration is 122 ng/m^3 at Stampede Pass. However, without the high value on August 3, the average at Stampede Pass is only 49 ng/m^3. The next highest mean concentrations are in the center of the network at Puyallup (98 ng/m^3), Skookumchuck (95 ng/m^3) and North Bend (90 ng/m^3). Sauvie Island is near Portland, Oregon.

Six EOFs, explaining 97.4% of the variance were rotated for potassium. The first four are shown in Figures 19 through 22 and the time factors are in Figure 23. As discussed above, the first EOF is associated nearly exclusively with the high concentration at Stampede Pass on August 3. EOFs 2, 3, and 5 (5 is not shown) all appear to be indicative of transport of potassium into the region from the southwest, and EOF 4 indicates transport from the southeast. EOF 2 also has a relatively strong gradient around Carbon River and Mud Mountain.

Summary of EOF Analysis

Urban areas are clearly associated with sources of organic matter. Although it appears that fires on occasion do produce significant quantities of organic carbon at individual monitoring sites over short time periods. The quantity b_{abs} also appears to be urban in nature in that it is most closely associated with monitoring sites near population centers. However, b_{abs} seems to be more localized in nature. Whereas OMH is uniformly elevated at many sites b_{abs} tend to be elevated over smaller spatial scales.

Pollutants emitted in the Seattle-Tacoma urban area appear to be transported inland as far as the Cascade Range where the high terrain then acts as a barrier to further transport. This is evident in the spatial patterns of both organics and b_{abs}.

Potassium which is primarily associated with burning activity is predominantly associated with transport from the south. Evidently emissions from fire-related activity in Oregon are transported into Washington.

CONCLUSIONS

Organics and light absorbing carbon were the single largest contributors to measured fine mass and are the second largest contributor to visibility reduction at Mount Rainier National Park. The inferred origin of organics and light absorbing carbon was surprising. Both the empirical and CMB regression models have high R^2 values. The empirical regression model attributed most organics and light absorbing carbon to either lead or bromine both of which were shown to be primarily associated with transportation activity. Very little carbon was associated with potassium which was mostly linked to burning.

The CMB regression model suggests that about 50% of the organics at Mount Rainier have an urban-transportation origin with fire-related activity accounting for only about 10%. On the other hand, almost 40% of organics are associated with the soil signature suggesting substantial re-entrainment of organic material along

with wind blown dust. Light absorbing carbon is also most closely associated with the transportation signature at a greater than 60% contribution. Again only a small fraction (13%) of LAC is linked to burning. The soil signature accounts for about 25% of LAC.

Very little of the carbon material is associated with burning. However, because over 90% of the potassium is linked to burning, it was of interest to examine spatial patterns of potassium to try to determine where the burning occurred. The first EOF is apparently associated with a fire within the State of Washington. However, all other EOFs imply transport of potassium into Washington from Oregon, and by inference the organics and light absorbing carbon linked to burning are also transported to Washington from Oregon.

DISCLAIMER

The assumptions, findings, conclusions, judgements, and views presented herein are those of the authors and should not be interpreted as necessarily representing official National Park Service policies.

REFERENCES

1. D. B. Joseph, J. Metsa, W.C. Malm, and M. Pitchford, "Plans for IMPROVE: A federal program to monitor visibility in class I areas," Transactions of the Air Pollution Control Association's Specialty Conference on Visibility Protection: Research and Policy Aspects. Jackson Hole, WY, 8-10 September 1986

2. R.A. Eldred, T.A. Cahill, L.K. Wilkinson, P.J. Feeney, and W.C. Malm, "Particulate characterization at remote sites across the U.S.: first year results of the NPS/IMPROVE network," in Proceedings of the 82nd Annual Air & Waste Management Association Meeting and Exhibition. 88-151.3, Anaheim, CA, 1988.

3. J.C. Chow, J.G. Watson, L.C. Pritchell, "The DRI thermal/optical reflectance carbon analysis system: description, evaluation, and applications in U.S. air quality studies." Atmos. Environ., in press, 1992.

4. J.G. Watson, "Overview of receptor model principles," JAPCA, 34:620 (1984).

5. G.E. Gordon, "Receptor models," Environ. Sci. Technol. 14:795 (1980).

6. J.A. Cooper, J.G. Watson, "Receptor-oriented methods of air particulate source apportionment," JAPCA, 30:1116 (1980).

7. L.A. Currie, R.W. Gerlach, C.W. Lewis, et al., "Interlaboratory comparison of source apportionment procedures: results for simulated data sets," Atmos. Environ., 18:1517 (1984).

8. T.G. Dzubay, R.K. Stevens, W.D. Balfour, et al., "Interlaboratory comparison of receptor model results for Houston aerosol," Atmos. Environ., 18:1555 (1984).

9. J.G. Watson and N.F. Robinson, "A method to determine accuracy and precision required of receptor model measurements," APCA Quality Assurance in Air Pollution Measurements, Pittsburgh, PA, 1984.

10. W.C. Malm, H. Iyer, J. Watson, et al., "Survey of a variety of receptor modeling techniques," in Transactions of the AWMA/EPA International Specialty Conference on Visibility and Fine Particles. Estes Park, CO, 1989.

11. W.C. Malm and H.K. Iyer, "Application of tracer mass balance regression to WHITEX data," in Transactions of the AWMA/EPA International Specialty Conference on Visibility and Fine Particles, Estes Park, CO, 1989.

12. P.K. Hopke, "Receptor modeling in environmental chemistry," in Chemical Analysis, John Wiley & Sons, New York, NY, 1985, pp. 76.

13. J.B. Whitmore, W.C. Malm, and H.K. Iyer, "Sensitivity analysis of tracer mass balance regression," in Proceedings of the 84th Annual Meeting of the Air & Waste Management Association, Vancouver, B.C., 1991.

14. T. Dzubay, R.K. Stevens, G.E. Gordon, et al., "A composite receptor method applied to Philadelphia aerosol," Environ. Sci. & Technol., 22:1 (1988).

15. C.W. Lewis and R.K. Stevens, "Hybrid receptor model for secondary sulfate from an SO_2 point source," Atmos. Environ. 19:917 (1985).

16. R.K. Stevens and C.W. Lewis, "Hybrid receptor modeling," in Extended Abstracts for the Fifth Joint Conference on Applications of Air Pollution Meteorology, APCA, American Meteorological Society, Boston, 1987.

17. J. Watson, editor, Transactions, Receptor Models in Air Resources Management. Air & Waste Management Association, Pittsburgh, PA, (1989).

18. W.C. Malm, K. Gebhart, D. Latimer, T. Cahill, R. Eldred, R. Pielke, R. Stocker, and J. Watson, "National Park Service Report on the Winter Haze Intensive Tracer Experiment," Draft Final Report, National Park Service, Fort Collins, Colorado, December 1989.

19. O.M. Essenwanger, Applied statistics in atmospheric science, part A. frequencies and curve fitting, Elsevier, New York, NY, 1976, pp. 252-288.

20. H.H. Harmon, Modern factor analysis. Third Edition, Revised, University of Chicago Press, Chicago, 1976.

21. D.A. Belsley, E. Kuh, R.E. Welsch. Regression diagnostics, Edited by John Wiley and Sons, New York, NY.

22. R.C. Henry, Y.-J. Wang and K.A. Gebhart, "The relationship between empirical orthogonal functions and sources of air pollution," Atmos. Environ., 25A(2):503-509 (1991).

23. C.L. Lawson and R.J. Hanson, Solving least squares problems, Princeton-Hall, Englewood Cliffs, NJ, 1974.

24. P.E. Green, Mathematical tools for applied multivariate analysis, Academic Press, New York, NY, 1976.

25. M.B. Richman, "Rotation of principal components," J. Climatology, 6:293-335 (1986).

26. H.F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," Psychometrika, 23(3):187-199 (1958).

27. R.W. Beck and Associates, "PANORAMAS Volume 1, Final Report," prepared for the PANORAMAS Steering Committee, States of Washington, Oregon, and Idaho, February 1986.

28. F. Van Haren, "Presentation to the Mt. Rainier Study Group," December 13, 1989.

Table 1. Agricultural and forest burning emission inventory summary for the pacific northwest states units are tons/year (source: Beck & Associates[27]).

	-------Oregon-------		-----Washington-----		-------Idaho-------	
Fire Type	TSP	Fine*	TSP	Fine	TSP	Fine
Wildfires	12,182	9,746	7,330	5,864	42,727	34,182
Slash Burning	101,936	81,549	87,459	69,967	27,273	21,818
Field Burning	20,366	16,293	926	741	5,105	4,084
Totals	134,484	107,588	95715	76,572	75,105	60,084

*Fine particle fraction less than 2.5μm; estimated here as 80% of TSP emissions.

Table 2. Results of factor analysis on the Tahoma Woods data set. Factor extraction was followed by a Varimax rotation. Variables with factor loadings greater than 0.6 are shaded for easy reference.

TAHOMA WOODS	FACTOR 1	FACTOR 2	FAC	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
$(NH_4)_2SO_4$	0.09	0.66	0.13	0.21	-0.09	0.25	-0.01
NH_4NO_3	0.27	0.50	0.77	0.06	-0.00	-0.09	-0.02
OCLT	0.11	0.08	0.05	-0.01	0.92	0.01	0.09
OCHT	0.72	0.50	-0.08	0.17	0.15	-0.02	-0.01
ECLT	0.85	0.41	-0.00	0.15	-0.04	0.10	0.08
ECHT	0.23	0.37	-0.13	-0.04	0.69	-0.05	-0.10
b_{abs}	0.69	0.56	-0.08	0.14	0.18	0.10	0.11
Br	0.60	0.55	0.16	0.17	0.15	0.15	0.07
Cu	0.52	0.12	0.04	-0.04	-0.11	-0.52	0.01
Fe	0.57	0.74	0.07	-0.01	0.23	-0.10	-0.01
K	0.19	0.10	-0.02	0.73	-0.09	-0.07	0.30
Mn	0.84	0.43	-0.05	-0.04	0.11	-0.00	-0.02
Na	-0.10	0.02	0.94	-0.01	-0.06	0.06	0.05
Pb	0.92	0.15	-0.01	0.08	0.15	0.00	0.03
V	0.75	-0.06	-0.00	0.30	0.15	0.07	-0.17
Zn	0.86	0.20	0.02	0.08	-0.01	-0.07	0.02
Si	0.40	0.85	0.01	0.05	0.22	-0.10	-0.01
Al	0.45	0.74	-0.14	-0.02	0.20	-0.15	0.05
Ca	0.40	0.72	0.40	0.09	0.18	-0.13	0.01
Ti	0.24	0.83	0.04	0.15	0.12	-0.19	-0.11
Rb	0.38	0.25	-0.02	0.59	-0.05	0.01	0.01
Sr	-0.05	0.03	0.31	0.64	0.18	-0.40	-0.31
Zr	-0.06	-0.09	0.03	0.10	0.04	-0.04	0.92
Cr	0.07	-0.08	0.05	-0.14	-0.04	0.63	-0.03
Ni	0.81	0.32	0.10	0.14	0.12	0.19	-0.05
As	0.34	-0.14	-0.16	0.43	0.00	0.47	-0.03
Se	0.60	0.31	0.25	0.12	0.03	-0.20	-0.17
RH	-0.09	-0.67	-0.29	0.11	0.01	0.24	0.11
Variance Explained	7.57	5.90	2.04	1.83	1.73	1.41	1.16

Table 3. Tahoma Woods: Results of a stepwise ordinary least square regression analysis with organic carbon as the dependent variable and Si, Se, Pb, and Br as independent variables. Variables are presented in the table is the order in which they were added to the regression model. The partial R^2 and percent contribution associated with each variable are also shown. For instance the amount of organic carbon associated with Br is derived by multiplying the Br regression coefficient by the average Br concentration. The percent organic carbon identified with Br is then calculated by dividing the "Bromine organic carbon" by the total predicted organic carbon.

Variable	Estimate	Std Error	Sum of Squares	F	Partial R^2	Model R^2	% Contribution
INTERCEPT	0.19	0.18	1.20	1.11	-----	----	--
Br	0.88	0.15	35.15	32.53	0.675	0.68	45
Pb	0.28	0.05	32.57	30.14	0.069	0.74	23
Si	0.01	0.00	35.21	32.58	0.050	0.79	21
Se	1.09	0.70	2.64	2.44	0.003	0.80	6

Table 4. Tahoma Woods: Results of a stepwise ordinary least square regression analysis with light absorbing carbon as the dependent variable and Si, Se, K, Pb, and Br independent variables. Variables are presented in the table is the order in which they were added to the regression model. The partial R^2 and percent contribution associated with each variable are also shown.

Variable	Estimate	Std Error	Sum of Squares	F	Partial R^2	Model R^2	% Contribution
INTERCEPT	0.07	0.03	0.14	5.04	-----	----	--
Pb	0.09	0.01	3.81	133.27	0.731	0.73	36
Si	0.0019	0.0003	1.05	36.54	0.107	0.84	20
Br	0.09	0.03	0.35	12.12	0.016	0.85	23
Se	0.23	0.11	0.12	4.23	0.004	0.86	6
K	0.0005	0.0003	0.09	3.12	0.003	0.86	4

Table 5. Fraction that each source type contributes to the various trace elements listed. Elements were measured at Tahoma Woods and Kraft profile was used.

	Soil	Coal	Burn	Kraft	Trans
Si	0.5900	0.3736	0.03045	0.005357	0.0005117
Pb	0.000	0.01369	0.07388	0.009496	0.9029
Se	0.000	0.8849	0.000	0.1151	0.000
Br	0.000	0.000	0.2254	0.1014	0.6731
Na	0.000	0.000	0.03780	0.9622	0.000
K	0.04450	0.02922	0.8560	0.07027	4.064E-05
Mn	0.5554	0.1506	0.09072	0.05465	0.1486
Cu	0.000	0.04154	0.9544	0.002091	0.001944
As	0.000	0.9454	0.000	0.05464	0.000

Table 6. The first part of this table are the results of an OLS regression using organics as the dependent variable and the relative source strengths of burning, transportation, and soil as independent variables. The second half of the table, part b, presents the estimated organic mass associated with each source type along with the fraction of organics attributed to each source type. Units of elemental organics are in $\mu g/m^3$.

a. Dependent variables = Organics, $R^2 = 0.80$

Variable	Estimate	Std Error	t-value	Prob t	Std Est	Cor with Dep Var
Constant	0.23	0.17	1.37	0.17	--	--
Burn	0.0075	0.0017	4.41	0.00	0.174	0.295
Trans	0.54	0.04	13.19	0.00	0.596	0.81
Soil	0.026	0.003	8.55	0.00	0.38	0.68

b.

Variable	Mean	Std Dev	Variance	Minimum	Maximum	Fraction
Burn	0.34	0.39	0.15	0.00	3.36	0.11
Trans	1.50	1.34	1.80	0.00	6.67	0.51
Soil	1.15	0.88	0.78	0.00	4.00	0.38

Table 7. The first part of this table are the results of an OLS regression using light absorbing carbon as the dependent variable and the relative source strengths of burning, transportation, and soil as independent variables. The second half of the table, part b, presents the estimated light absorbing carbon mass associated with each source type along with the fraction of light absorbing carbon attributed to each source type. Units of elemental light absorbing carbon are in $\mu g/m^3$.

a. Dependent variable = Carbon, $R^2 = 0.91$

Variable	Estimate	Std Error	t-value	Prob t	Std Est	Cor with Dep Var
Constant	-0.11	0.02	-5.38	0.00	--	--
Burn	0.0015	0.0002	7.64	0.00	0.20	0.35
Trans	0.12	0.005	24.68	0.00	0.75	0.91
Soil	0.0029	0.0004	7.99	0.00	0.24	0.62

b.

Variable	Mean	Std Dev	Variance	Minimum	Maximum	Fraction
Burn	0.07	0.08	0.006	0.00	0.69	0.13
Trans	0.34	0.30	0.089	0.00	1.48	0.63
Soil	0.13	0.10	0.009	0.00	0.44	0.24

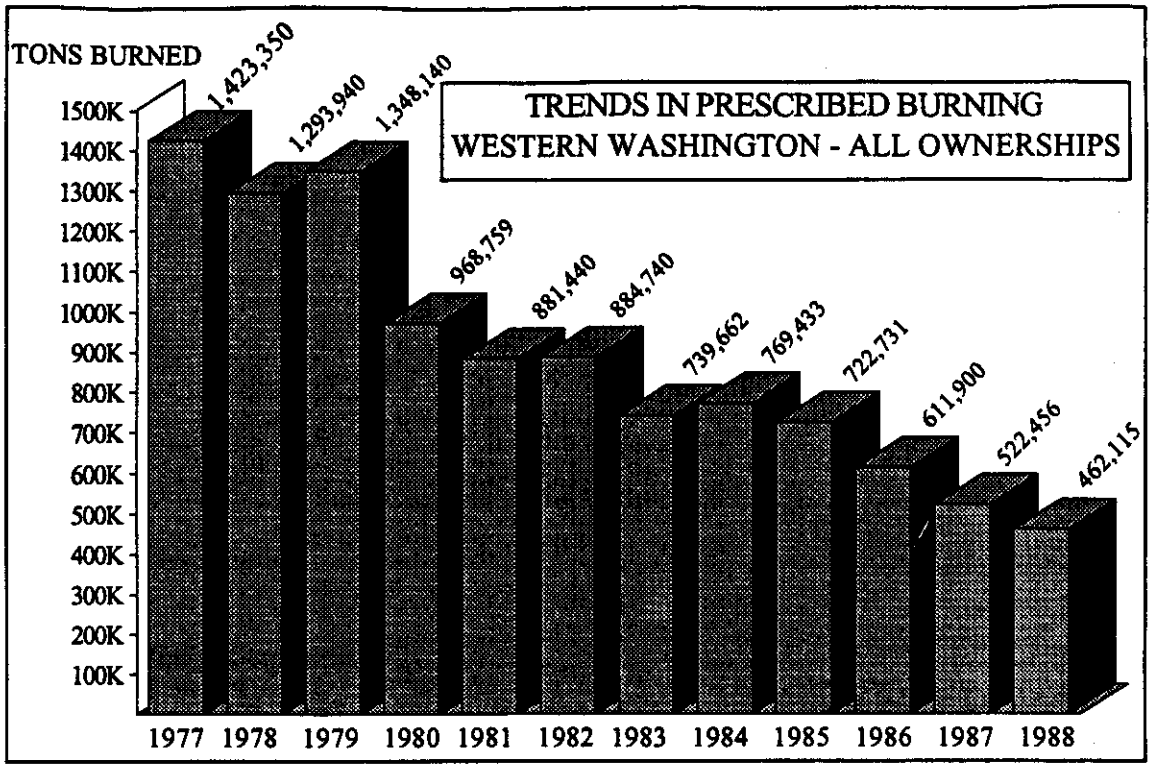

Figure 1. Trends in prescribed burning in western Washington, 1977-1988 (source: Van Haren[28]).

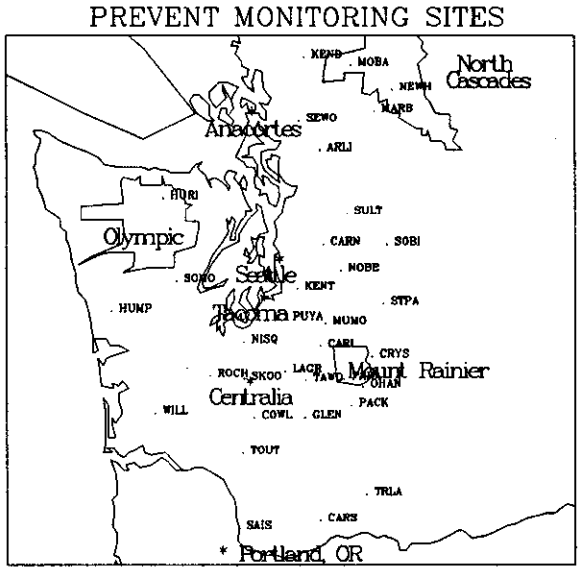

Figure 2. PREVENT particulate monitoring sites, National Parks in Washington State, and major source areas. Scale is the same as all contour maps in this paper.

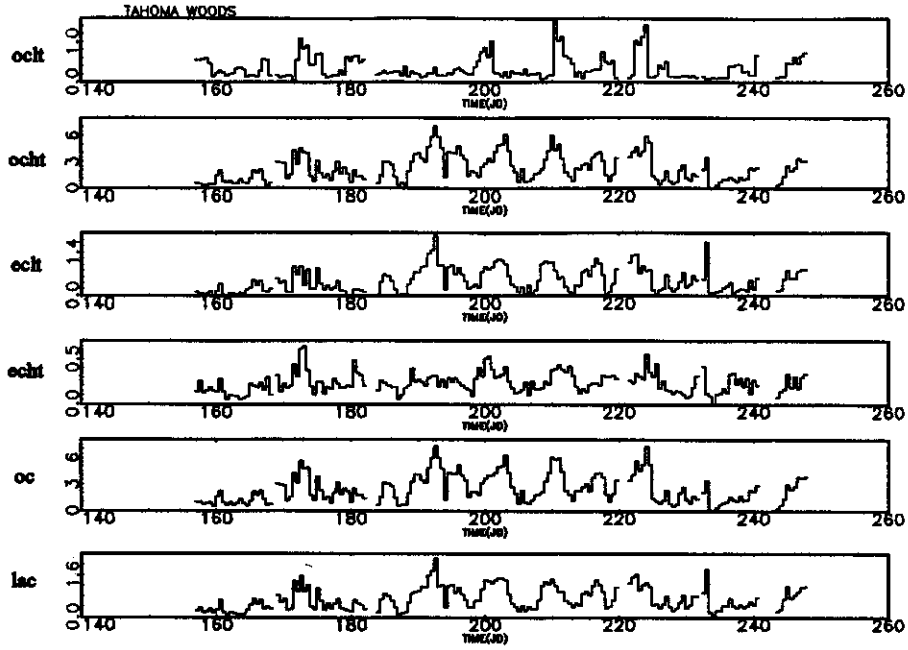

Figure 3a. Tahoma Woods: Temporal plots of oclt, ocht, eclt, echt, oc, and lac. Units are in μg m^{-3}.

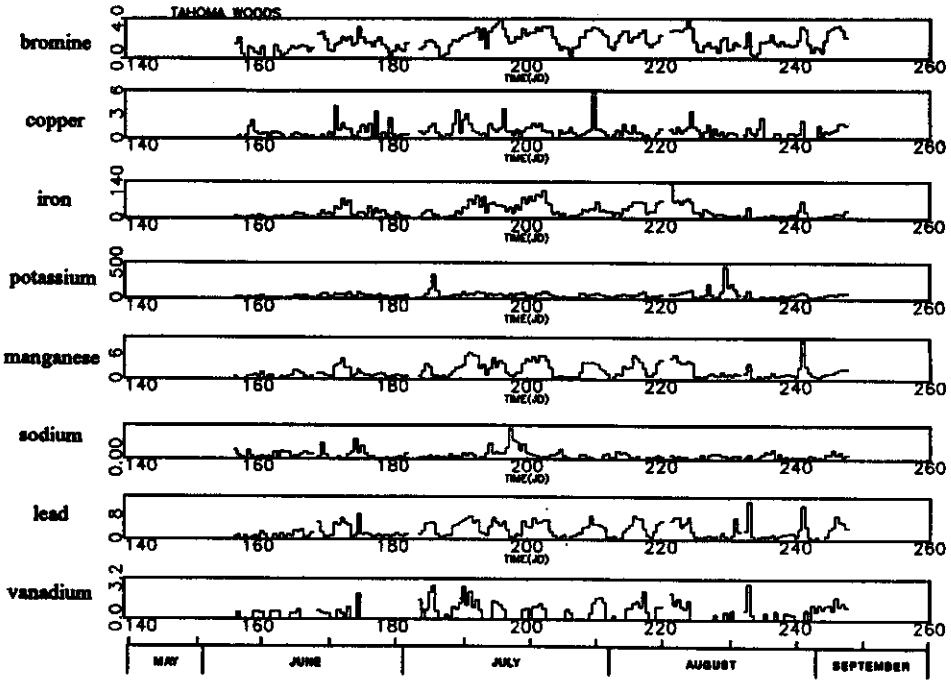

Figure 3b. Tahoma Woods: Temporal plot of bromine, copper, iron, potassium, manganese, sodium, lead, and vanadium. Fine mass, sulfur, ammonium nitrate, hydrogen, and optical absorption are all presented in concentration units of μg/m^3 while trace elements are in ng/m^3. Optical absorption "mass" is calculated assuming an absorption efficiency of 10 m^2/gm.

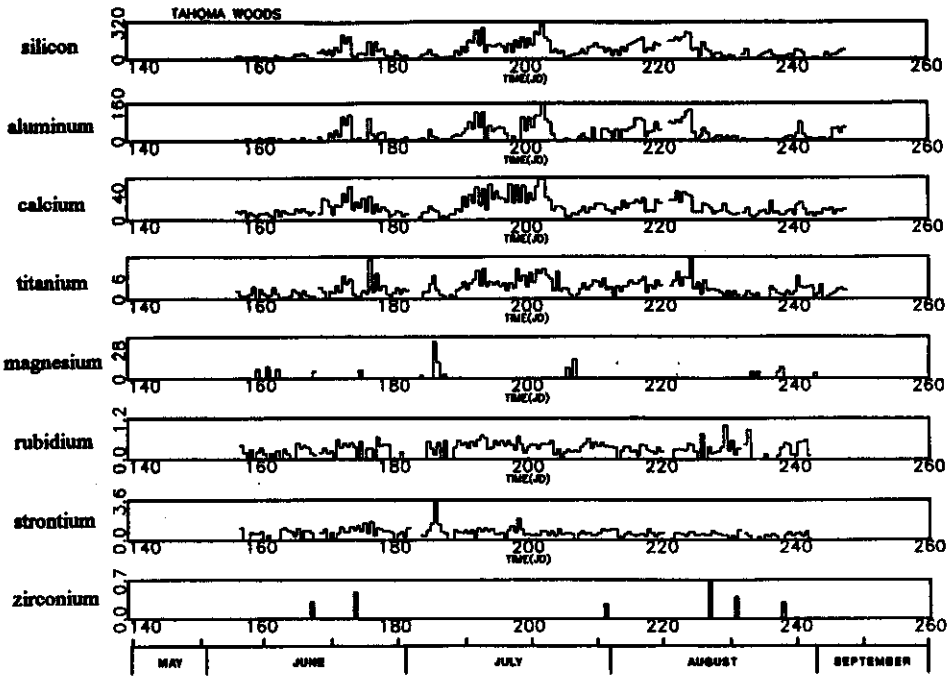

Figure 3c. Tahoma Woods: Temporal plot of silicon, aluminum, calcium, titanium, magnesium, rubidium, strontium, and zirconium. Fine mass, sulfur, ammonium nitrate, hydrogen, and optical absorption are all presented in concentration units of $\mu g/m^3$ while trace elements are in ng/m^3. Optical absorption "mass" is calculated assuming an absorption efficiency of 10 m^2/gm.

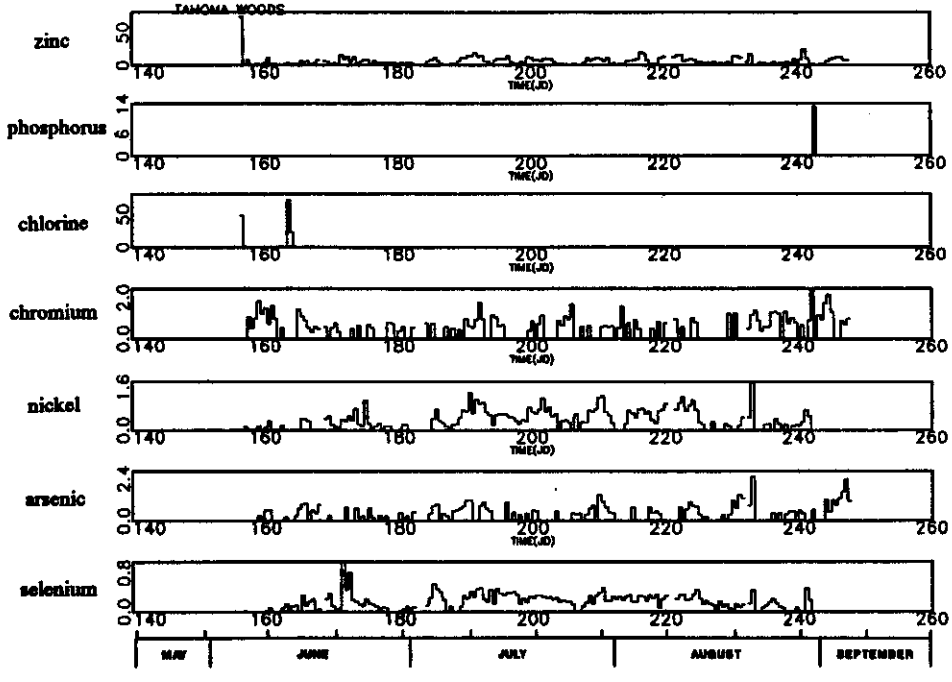

Figure 3d. Tahoma Woods: Temporal plot of zinc, phosphorus, chlorine, chromium, nickel, arsenic, and selenium. Fine mass, sulfur, ammonium nitrate, hydrogen, and optical absorption are all presented in concentration units of $\mu g/m^3$ while trace elements are in ng/m^3. Optical absorption "mass" is calculated assuming an absorption efficiency of 10 m^2/gm.

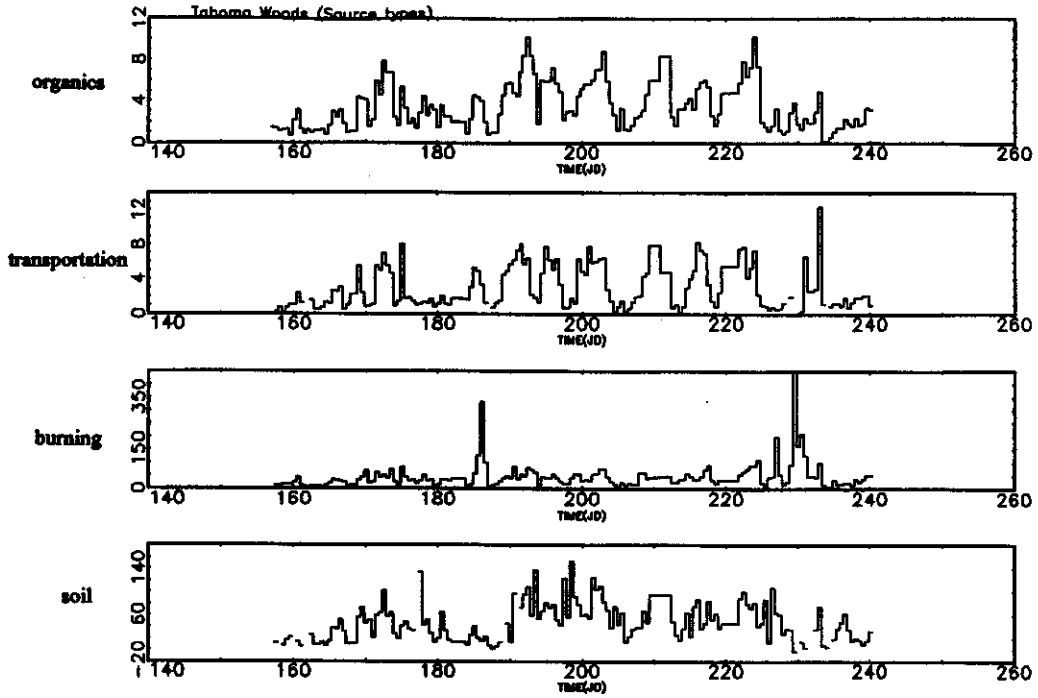

Figure 4. Temporal plot of organics and the relative source strength of burning, transportation, and soil at Mount Rainier. Units of organics are in $\mu g/m^3$.

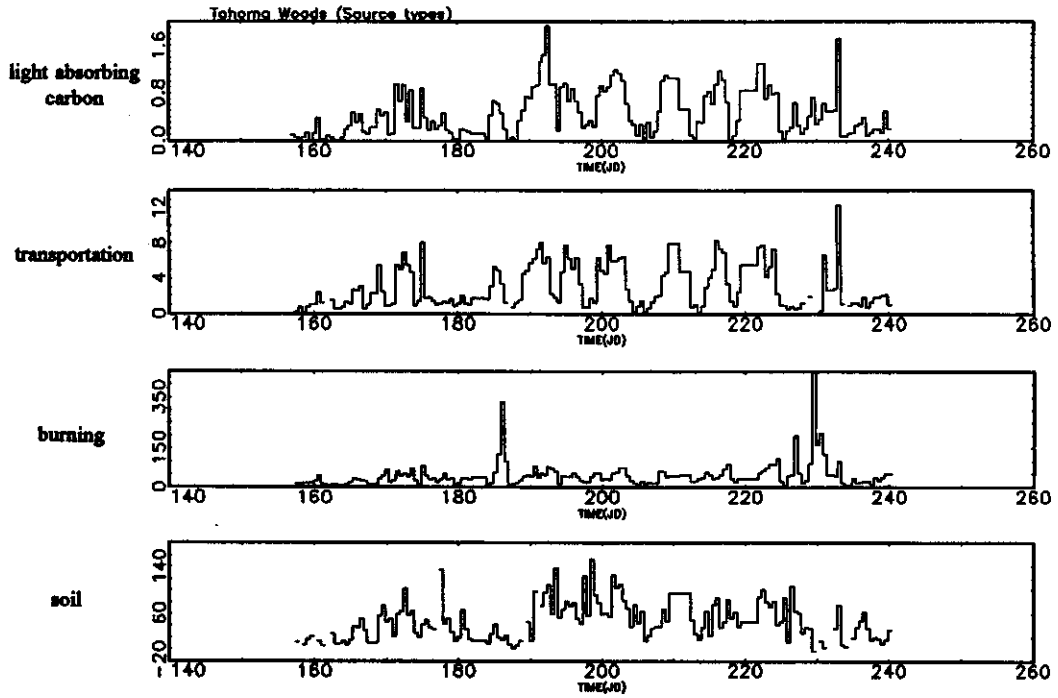

Figure 5. Temporal plot of light absorbing carbon, and the relative source strength of burning, transportation, and soil. Units of light absorbing carbon are in $\mu g/m^3$.

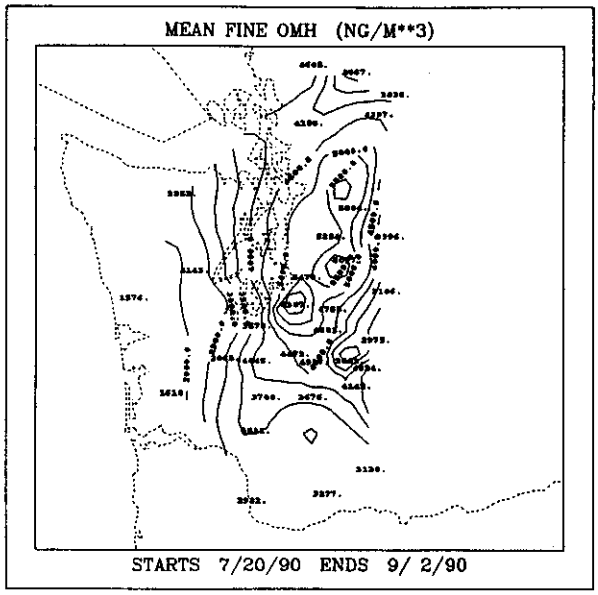

Figure 6. Contour plot of mean fine OMH. Contour interval is 500 ng m^{-3}.

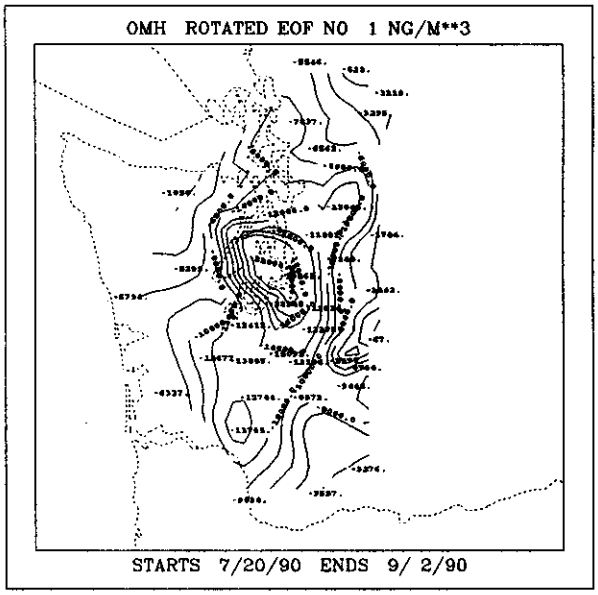

Figure 7. Rotated EOF 1 for OMH explaining 38.7% of the variance.

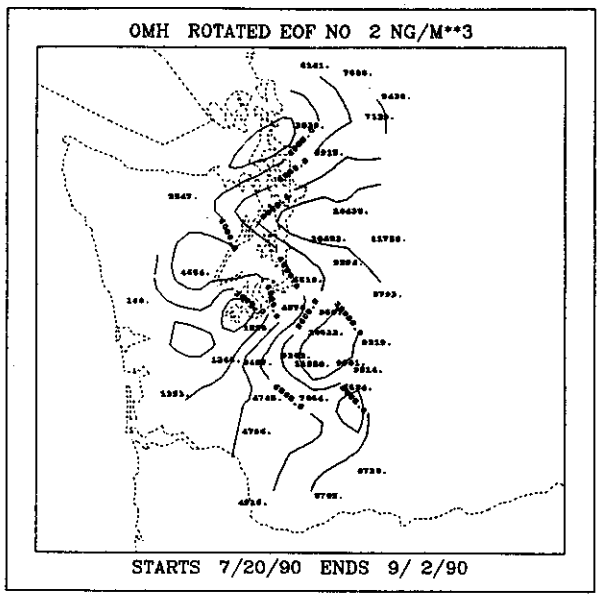

Figure 8. Rotated OMH EOF 2 explaining 22.6% of the variance. Contour interval is 2000 ng m^3.

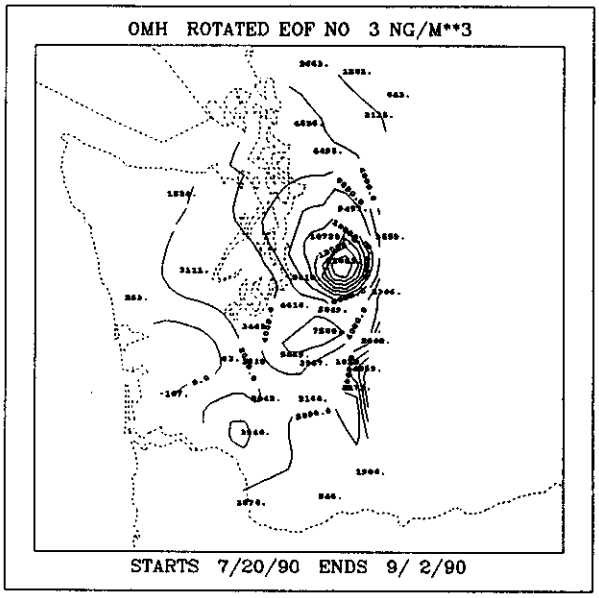

Figure 9. Rotated OMH EOF 3 explaining 11.8% of the variance. Contour interval is 2000 ng m^3.

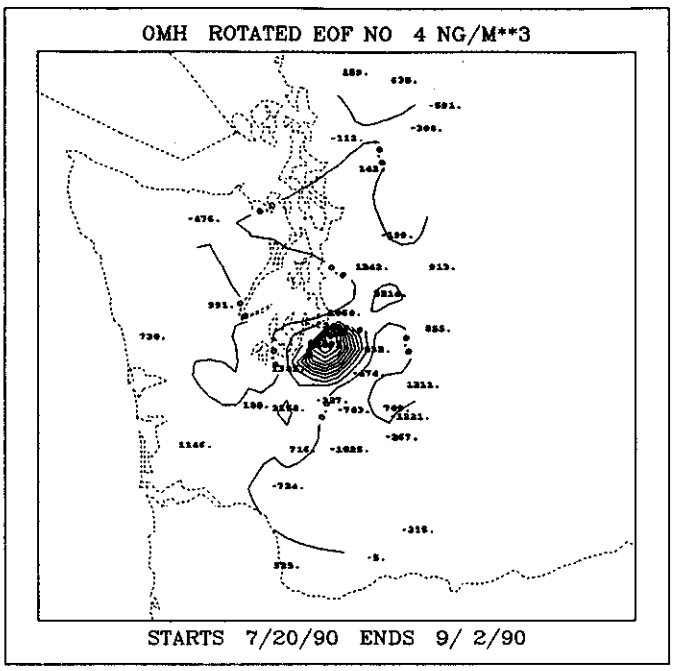

Figure 10. Rotated OMH EOF 4 explaining 9.2% of the variance. Contour interval is 2000 ng $^{-3}$.

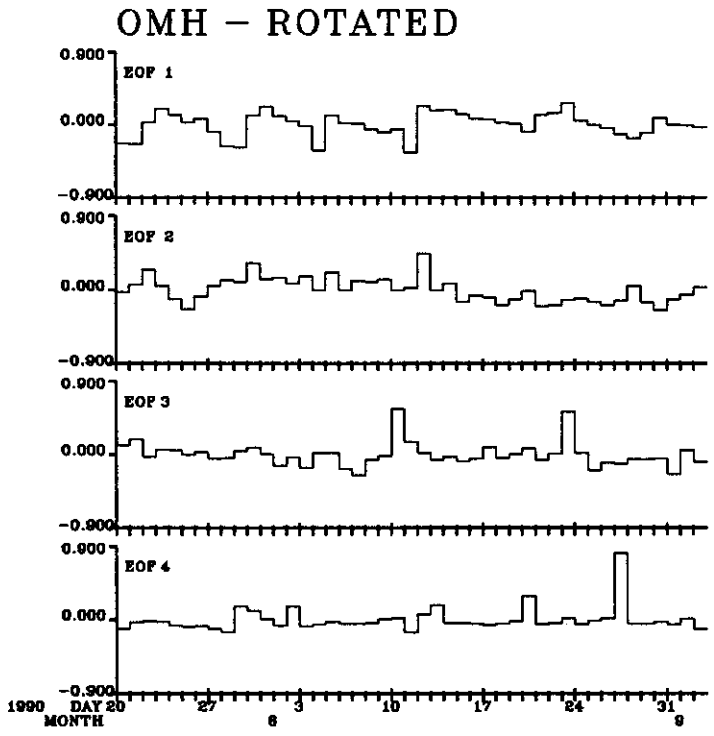

Figure 11. Time factors for first four rotated EOFs for OMH.

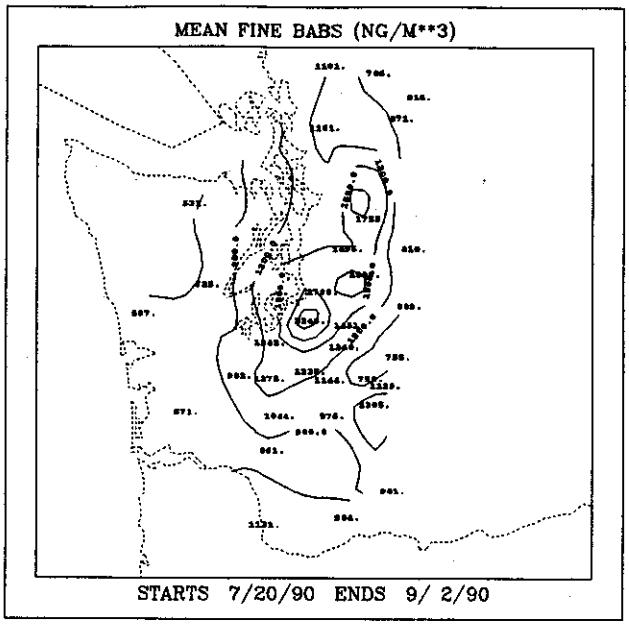

Figure 12. Contour plot of mean fine Babs. Contour interval is 300 ng m^{-3}.

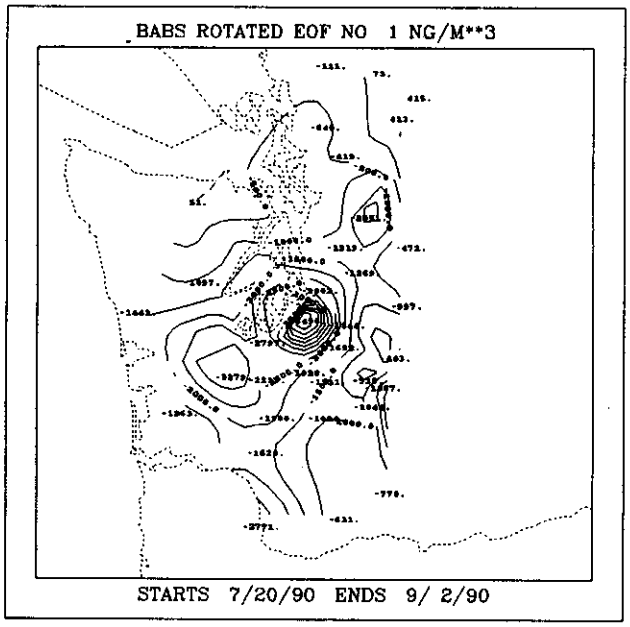

Figure 13. Rotated EOF 1 for Babs explaining 27.3% of the variance. Contour interval is 500 ng m^{-3}.

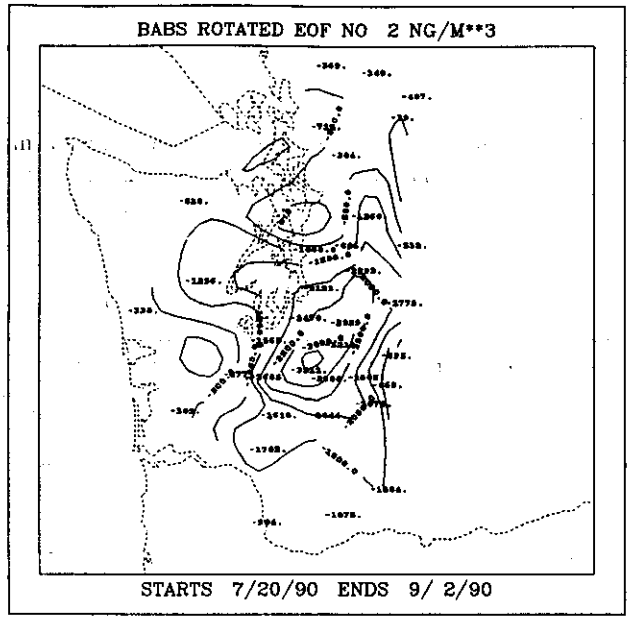

Figure 14. Rotated EOF 2 for Babs explaining 21.4% of the variance. Contour interval is 500 ng m^{-3}.

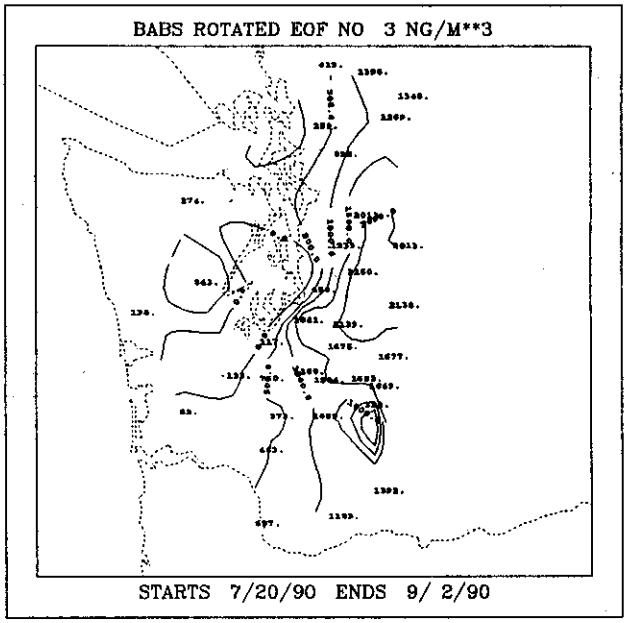

Figure 15. Rotated EOF 3 for Babs explaining 12.5% of the variance. Contour interval is 500 ng m^{-3}.

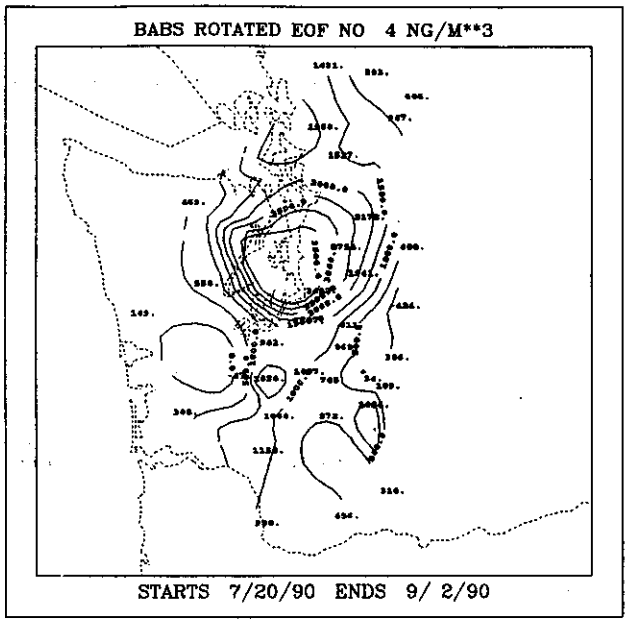

Figure 16. Rotated EOF 4 for Babs explaining 12.1% of the variance. Contour interval is 500 ng m^{-3}.

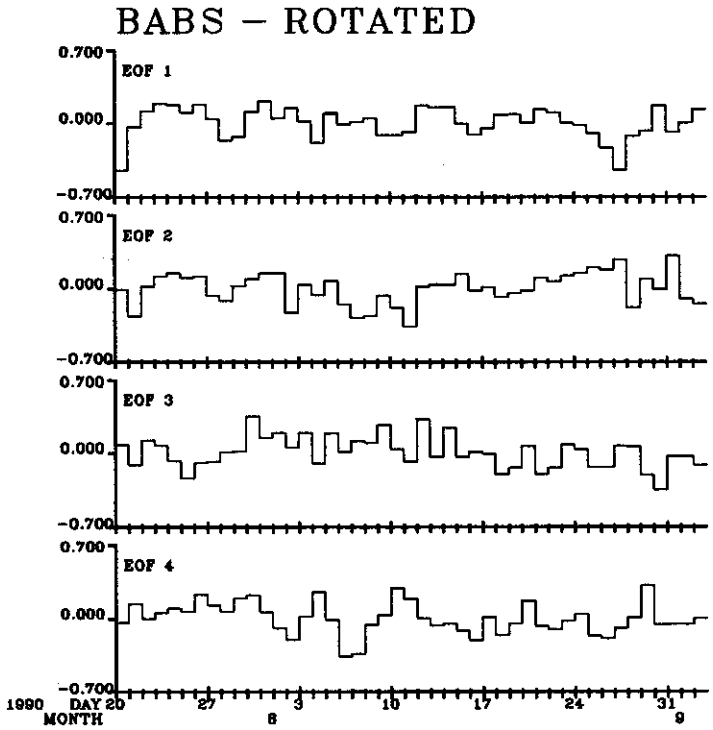

Figure 17. Time factors for the first 4 rotated EOFs for Babs.

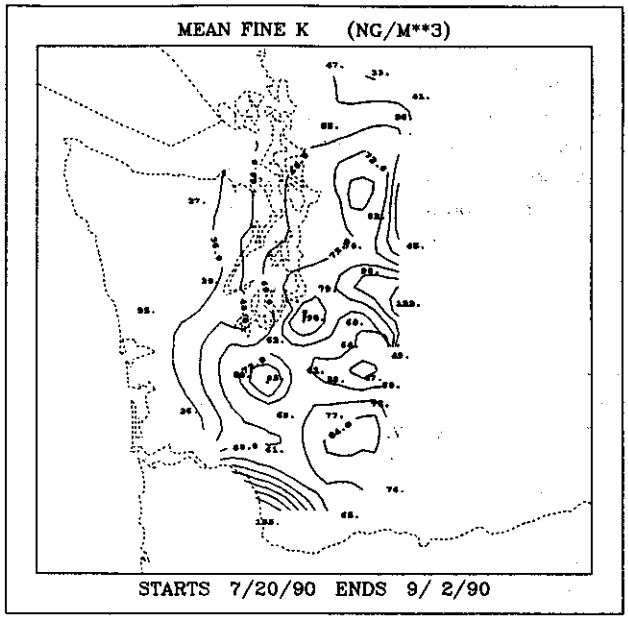

Figure 18. Mean potassium concentrations. Contour interval is 12 ng m^{-3}.

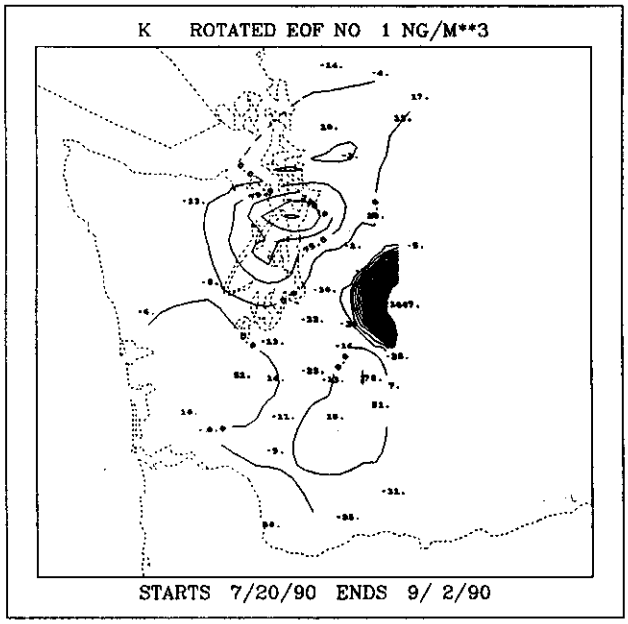

Figure 19. Rotated EOF 1 for potassium, explaining 68.3% of the variance. Contour interval is 75 ng m^{-3}.

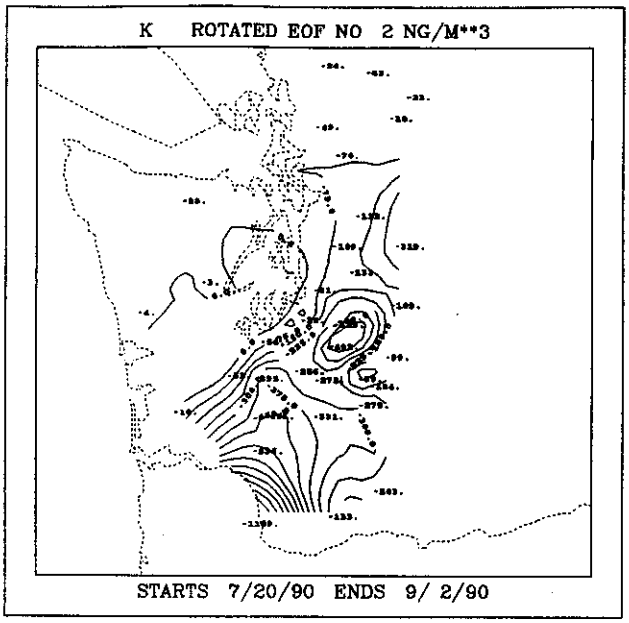

Figure 20. Rotated EOF 2 for potassium explaining 17.0% of the variance. Contour interval is 75 ng m^3.

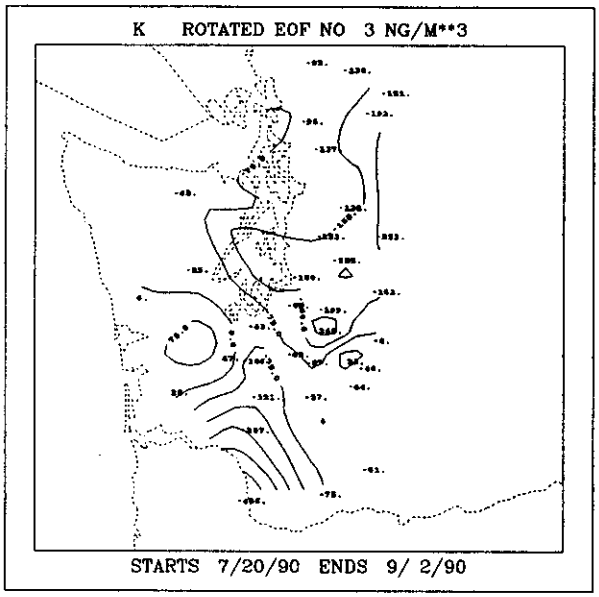

Figure 21. Rotated EOF 3 for potassium explaining 4.1% of the variance. Contour interval is 75 ng m^3.

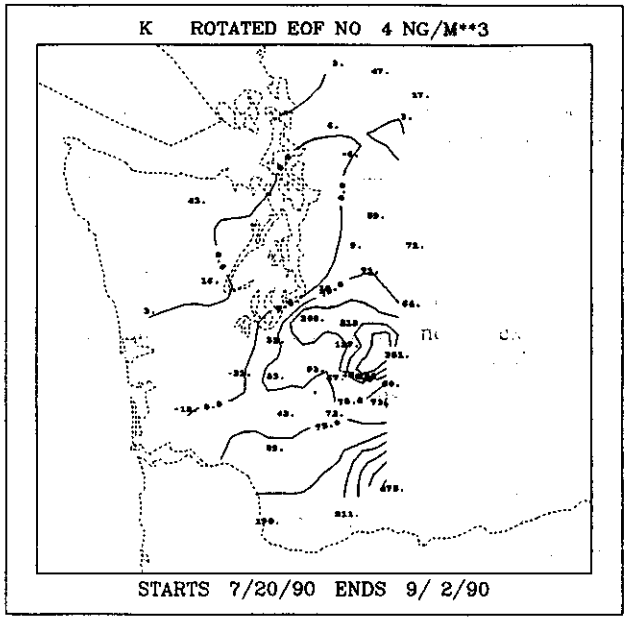

Figure 22. Rotated EOF 4 explaining 3.9% of the variance. Contour interval is 75 ng m^{-3}.

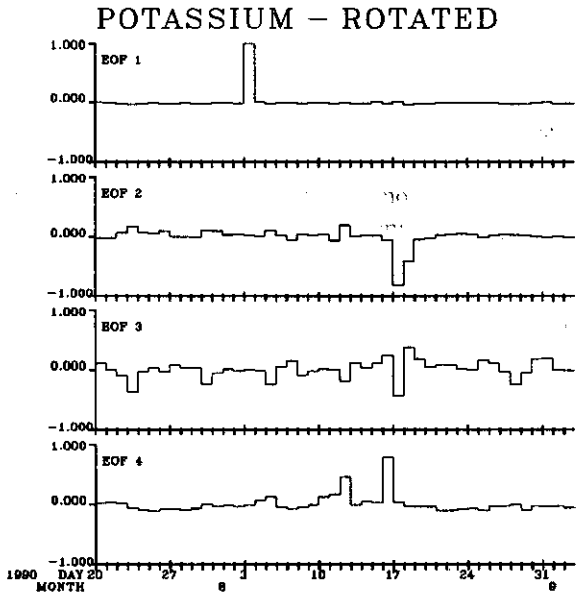

Figure 23. Time factors for the first 4 rotated EOFs for potassium.

APPENDIX 1

SURVEY OF A VARIETY OF RECEPTOR MODELING TECHNIQUES

William C. Malm
National Park Service, Air Quality Division
CIRA/Colorado State University
Foothills Campus
Ft. Collins, CO 80523

Hari K. Iyer
Department of Statistics
Colorado State University
Ft. Collins, CO 80523

John Watson
EEEC, Desert Research Institute
Reno, NV 89506

Douglas A. Latimer
Latimer & Associates
679 Quince Circle
Boulder, CO 80304

Abstract

The chemical mass balance (CMB) formalism has been used on a semi-routine basis to apportion emissions used to mass concentrations at specific receptor sites. Recently, two other techniques, differential mass balance (DMB) and tracer mass balance regression (TMBR) have been used to apportion secondary aerosols to sources and source types of a variety of receptor areas. CMB uses known source and receptor measured tracer profiles (gradients in tracer concentrations at one point in time) to apportion sources at one point in time. DMB uses gradients in trace elements across space, while TMBR uses changes in tracers across time to achieve apportionment of primary as well as secondary aerosol species. Assumption and limitations of each approach will be addressed and a unified formalism building on strengths of all three approaches will be presented.

SURVEY OF A VARIETY OF RECEPTOR MODELING TECHNIQUES

Introduction

Receptor modeling approaches rely on known physical and chemical characteristics of gases and particles at receptors and sources to attribute aerosols to a source or source type. Historically, the CMB formalism has been used to apportion primary particles. This formalism uses known relationships between emitted tracers and an assumption that various tracer profiles stay constant as material is transported from source to receptor. These tracer profiles are then used to apportion primary species for each time period that a measurement is made at a receptor site. Other common types of models include principal component analysis (PCA) and multiple linear regression (MLR). Explanations of these models are given by Watson,[1,2,3] Chow,[4] and Hopke.[5] All these models are special cases of a General Mass Balance (GMB) model which is deterministic in nature. A regressional model similar to MLR is derivable from the GMB equations and will be referred to as the TMBR model. The TMBR model incorporates changes in tracer material over time to apportion both primary and secondary aerosols. Finally, the DMB model, a special case of GMB and referred to here as a receptor oriented model, is really a hybrid model in that it relies on tracer material to establish atmospheric dispersion characteristics but deterministically accounts for deposition and oxidation. Stevens and Lewis,[6] Lewis and Stevens,[7] and Dzubay et al.,[8] have used models similar to the TMBR and GMB to create a hybrid model which they have used for source apportionment.

General Mass Balance Equations

Each special case of GMB has its own set of limiting assumptions and special requirements for solution. The assumptions that need to be satisfied for the mathematical model to be valid will be apparent during the process of derivation of the model equations. Nevertheless, the assumptions will be explicitly stated after the derivations of the model equations have been explained. The statistical aspects of the estimation of the fractional contribution by a given source and the calculation of the associated uncertainties will also be presented.

Notational Conventions

The following notation will be used throughout.

Total number of species under consideration = m.
Total number of sources under consideration = n.
Total number of sampling periods = s.

The subscript i will be used for indexing the species, j for sources and k for sampling periods.

c_{ijk} = concentration of aerosol species i at source j corresponding to sampling period k.

C_{ijk} = concentration of aerosol species i at the receptor, attributable to source j corresponding to sampling period k.

t_{jk} = travel time for the air mass from source j to the receptor, corresponding to sampling period k.

r_{ijk} = a factor that accounts for deposition of aerosols species i from source j, for sampling period k.

r^*_{ijk} = a factor that accounts for the formation of aerosol species i from a parent species i * emitted by source j as well as its deposition during transport for sampling period k.

d_{jk} = a factor that accounts for dispersion of aerosol mixture from source j during sampling period k, as the mixture travels from the source to the receptor.

Whenever a subscript i denotes a secondary aerosol species, then the subscript i^* will denote the corresponding parent aerosol species. For instance, if i denotes SO_4 then i^* will stand for SO_2.

Model Equations

It follows from the definitions that for primary aerosol species

$$C_{ijk}=c_{ijk}r_{ijk}d_{jk} \tag{1}$$

and for secondary aerosol components we have

$$C_{ijk}=c_{ijk}r_{ijk}d_{jk}+c_{i^*jk}r^*_{ijk}d_{jk}. \tag{2}$$

The quantities r_{ijk} are a function of deposition rates and transport time while r^*_{ijk} are functions of deposition and transport times as well as conversion rates. Simple functional forms for r_{ijk} and r^*_{ijk} can be derived if it is assumed that chemical conversion and deposition are governed by first order mechanisms, and conversion and deposition rates are constant in space over some finite increment in time.

Let $X(t)$ denote the mass, at time t after emission, of a species i in a unit volume of aerosol mixture. Assume, ignoring dispersion temporarily,

$$\frac{dX(t)}{dt}=-(K_c+K_1)X(t) \tag{3}$$

which, when solved, yields

$$X(t)=X(0)\exp(-(K_c+K_1)t) \tag{4}$$

where X(0) is the mass at time 0 in unit volume of aerosol mixture, i.e. the concentration of the species at the source. The quantities K_c and K_1 are the conversion and deposition rates, respectively, for the species under consideration. The conversion and deposition rates have been assumed to remain constant throughout the transport path in space and time. If *d(t)* denotes the dispersion factor corresponding to *t* time units after emission of the aerosol mixture, then

$$X(t)x=X(0)\exp(-(K_c+K_1)t)d(t). \tag{5}$$

The factor accounting for conversion and deposition thus has the form *exp* $(-(K_c + K_1)t)$ in this case.

Suppose Y(*t*) is the concentration of a secondary aerosol species at time *t* after the parent species *i* is emitted by the source. Let X(*t*) be the concentration of the parent species at time *t* after emission. If the dispersion factor is temporarily ignored the following pair of differential equations hold:

$$\frac{dY(t)}{dt}=K_c X(t)-K_2 Y(t) \tag{6}$$

and

$$\frac{dX(t)}{dt}=-(K_c+K_1)X(t) \tag{7}$$

where K_2 refers to the deposition rate of the secondary aerosol which is assumed to be nonconverting. Once again, it has been assumed that the deposition and conversion parameters are constant throughout the transport path in space and time. Solution of the pair of differential equations yields the relation

$$Y(t)=\frac{K_c}{K_1+K_c-K_2}\{\exp(-K_2t)-\exp(-(K_1+K_c)t)\}. \tag{8}$$

If now the dispersion factor *d(t)* is taken into account

$$Y(t)=\frac{K_c}{K_1+K_c-K_2}\{\exp(-K_2t)-\exp(-(K_1+K_c)t)\}d(t). \tag{9}$$

From this relation it becomes evident that the factor accounting for the formation of the secondary aerosol species from its parent species as well as its deposition during transport is of the form

$$\frac{K_c}{K_1+K_c-K_2}\{\exp(-K_2t)-\exp(-(K_1+K_c)t)\}. \tag{10}$$

Based on the above arguments, when the conversion and deposition rates of the various species remain constant throughout the duration of transport from the source to the receptor

$$r_{ijk}=\exp(-(K_c(i,j,k)+K_d(i,j,k))t_{jk}) \tag{11}$$

and

$$r^{*}_{ijk}=\frac{K_c(i^{*},j,k)}{K_c(i^{*},j,k)+K_d(i^{*},j,k)-K_d(i,j,k)} \times \{\exp(-K_d(i,j,k)t_{jk})-\exp(-[K_c(i^{*},j,k)+K_d(i^{*},j,k)]t_{jk})\} \tag{12}$$

where

$K_c(i,j,k)$ = conversion rate of species i from source j to its secondary form, during sampling period k.

$K_d(i,j,k)$ = deposition rate of species i from source j during sampling period k.

Let C_{ik} = concentration of aerosol component i at the receptor during sampling period k. Since the concentration of aerosol component i at the receptor is the sum of the concentrations attributable to various sources, the mass balance equation becomes

$$C_{ik}=\sum_{j=1}^{n} C_{ijk} \tag{13}$$

for each sampling period k = 1,2,...,s. From this basic equation various special cases can be derived.

CMB Model

The first special case of the GMB equations to be examined is the Chemical Mass Balance formalism.

Model Equations

Suppose our list of aerosol components includes only material that is nonreactive and maintains relative ratios between various species as material is transported from source to receptor. In this case K_c (i,j,k) are all zero and K_d $(i,j.k)$ are the same for all elements i. Their

common value is denoted by $K_d(j,k)$ indicating the nondependence on i. This implies that the quantities r_{ijk} do not depend on i. Then,

$$\frac{C_{ijk}}{C_{i'jk}} = \frac{c_{ijk} r_{ijk} d_{jk}}{c_{i'jk} r_{i'jk} d_j k} = \frac{c_{ijk}}{c_{i'jk}} \tag{14}$$

which implies that the signature for source j at the source equals the signature for source j as perceived at the receptor.

Let $S_{jk} = \Sigma_{i=1}^{m} C_{ijk}$. The quantity S_{jk} is the concentration of the aerosol mixture at the receptor during sampling period k that is attributable to source j. The fraction a_{ijk} defined by

$$a_{ijk} = \frac{C_{ijk}}{S_{jk}} \tag{15}$$

is then the fraction of species i in the aerosol mixture at the receptor attributable to source j during sampling period k. Assuming Equation (14) is valid, the numbers a_{ijk} for $i = 1,2,\ldots,m$ represent the source signature for source j for sampling period k. From Equation (13) and (15) it follows that the set of Equations in (16) below also holds.

$$C_{ik} = \sum_{j=1}^{n} a_{ijk} S_{jk} \tag{16}$$

If the a_{ijk} for all the sources affecting the receptor sites are known, then Equation (16) is a system of linear simultaneous equations in n unknowns S_{1k}, $S_{2k},\ldots,S_{nk}$, for each of the sampling periods $k = 1,2,\ldots,s$. These are in fact the chemical mass balance equations. The rank of the system of equations for each k must be equal to n in order to uniquely solve these equations. In particular, the numbers of equations must be greater than or equal to the number of chemical species (i).

Solutions to the CMB equations that have been used are: 1) a tracer solution; 2) a linear programming solution; 3) an ordinary weighted least squares solution with or without an intercept; 4) a ridge regression weighted least squares solution with or without an intercept; and 5) an effective variance least squares solution with or without an intercept. An estimate of the uncertainty associated with the source contributions is an integral part of several of these solution 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; and 3) they are capable of analytically estimating the uncertainty of the source contributions.

CMB software in current use[9] applies the effective variance solution developed and tested by Watson[11] 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) chemical species with higher precisions in both the source and receptor measurements are given greater influence than are species with lower precisions. The effective variance solution is derived [10] by minimizing the weighted sums of the squares of the differences between the measured and calculated values of C_{ik} and a_{ij}. The solution algorithm is an iterative procedure which calculates a new set of S_{jk} based on the S_{jk} estimated from the previous iteration.

Watson[12] found that individual sources with similar source profiles would yield unreliable values if included in the same chemical mass balance. Henry[13] proposed a quantitative method of identifying this interference between these similar source compositions, which is known as "collinearity." He uses the "singular value decomposition" to define an "estimable space into which resolvable sources should lie." The sources which do not fall into this estimable space are collinear, or too similar to be resolved from the sources which do lie within the estimable space.

Williamson and Dubose[14] claimed that the ridge regression reduces collinearities. Henry[13] tested the ridge regression solution with respect to the separation of urban and continental dust and found that the bias resulted in physically unrealistic negative values for several of the a_{ij}. The ridge regression solution has not been used in the CMB since these tests were published.

CMB Model Assumptions

The CMB model assumptions are:

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

Effects of Deviations from CMB Model Assumptions

Assumptions 1 through 6 for the CMB model are fairly restrictive and will never to totally complied within actual practice. Fortunately, the CMB model can tolerate reasonable 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.[3,12,13,15-22] 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 emission 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 the 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:

- 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.

- For random errors, the magnitude of the source contribution errors decreases as the number of components increases.

The most recent and systematic tests are those of Javitz[22] which apply to a simple four-source urban airshed and a complex ten-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 specifically 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 or 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[12] 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.

Recent research suggests that Assumption 3 should be restated to specify that source contributions above detection limits should be included in the CMB. At this time, however, it is not yet possible to determine the "detection limit" of a source contribution at a receptor since this is a complicated and unknown function of the other source contributions, the source composition uncertainties and the uncertainties of the receptor measurements. Additional model testing is needed to define this "detection limit."

The linear independence of source compositions required by Assumption 4 has become a subject of considerable interest since the publication of Henry's[13] 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[13] also proposed an optimal linear combination of source contributions that have been determined to be collinear.

Other "regression diagnostics" have been summarized by Belsley[23] and have been applied to the CMB by DeCesar.[19,20] Kim and Henry[24] show that most of these diagnostics are useless because they are based on the assumption of zero uncertainty in the source profiles. They 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 obvious collinear source compositions typically result in positive and negative values for the collinear source types as well as large standard errors on 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.[22]

Gordon[15] 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 percent, Watson[12] 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 measurment 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.[11] 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 for 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.

Model Input and Output Data

The chemical mass balance modeling procedure requires: 1) identification of the contributing sources types; 2) selection of chemical species to be included; 3) estimation of the fraction of each of the chemical species which is contained in each source (i.e., the source compositions); 4) estimation of the uncertainty in both ambient concentrations and source compositions; 5) solution of the chemical mass balance equations; and 6) validation and reconciliation. Each of these steps requires different types of data.

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[25] (identify most of the available source profiles) are examined to select the chemical species to be measured. Watson[12] demonstrates that the more species measured, the better the precision of the CMB apportionment.

The ambient concentrations of these species, C_i, and their fractional amount in each source-type emission, 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 contributions 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.

TMBR Model

The TMBR model is a multiple regression based model which may be used to apportion an aerosol species of interest measured at a receptor site to the various contributing sources. The actual regression analysis may be performed using the method of ordinary least squares.

However, since the independent variables in this model are ambient concentrations of various aerosol components which are measured with error, the method of Orthogonal Distance Regression (ODR) is expected to give better estimates of the source contributions. A detailed theoretical discussion of the method of ODR may be found in the book by Fuller.[26]

Model Equations

In this section it is shown that, under appropriate assumptions, the general mass balance model can be reduced to a simpler linear model. Let aerosol component $i = 1$ be a secondary aerosol with $i^* = 2$ denoting the corresponding parent species. It is of interest to determine the fractional contribution to the ambient concentration of this secondary aerosol component by a distinguished source which will be denoted by the subscript $j = 1$. We will also assume that aerosol species i_1 is a tracer for this distinguished source. Let sources $j = 2$ through $j = n_2$ have an associated tracer species i_2, sources $j = n_2 + 1$ through $j = n_3$ have an associated tracer species i_3 etc., and sources $j = n_{h-1} + 1$ through $j = n_h$ have an associated tracer i_h. Sources $j = n_h + 1$ through $j = n$ may be unknown sources or may be known sources with tracers that are not measured at the receptor. For the sake of uniformity of notation we let $n_1 = 1$. Thus the n sources have been partitioned into $h + 1$ groups, each of the first h groups of sources being associated with a unique tracer species or with a fraction of some reference species that has been calculated using CMB or some other appropriate model.

In general for $1 \leq u \leq h$ and $n_{u-1} + 1 \leq j \leq n_u$ we have

$$C_{1jk} = r^*_{1jk} d_{jk} c_{2jk} = \frac{r^*_{1jk}}{r_{i_u jk}} \frac{c_{2jk}}{c_{i_u jk}} C_{i_u jk} = \beta_{i_u jk} C_{i_u jk} . \tag{17}$$

Therefore,

$$\sum_{j=n_{u-1}+1}^{n_u} C_{1jk} = \sum_{j=n_{u-1}+1}^{n_u} \beta_{i_u jk} C_{i_u jk} = \beta_{i_u k} C_{i_u k} \tag{18}$$

where $\beta_{i_u k}$ is defined as

$$\beta_{i_u k} = \frac{\sum_{j=n_{u-1}+1}^{n_u} \beta_{i_u jk} C_{i_u jk}}{C_{i_u k}} . \tag{19}$$

For $n_h+1 \leq j \leq n$ let

$$\beta_{0k}=\sum_{j=n_h+1}^{n} C_{1jk} \,. \tag{20}$$

The general mass balance equation then reduces to the equation

$$C_{1k}=\beta_{0k}+\sum_{u=1}^{h} \beta_{i_u k} C_{i_u k} \tag{21}$$

for each sampling period $k = 1,2,\ldots,s$.

If the quantities $\beta_{i_u k}$ are all independent of k for each u, $\beta_{i_u k} = \beta_{i_u}$, and the above set of equations reduce to

$$C_{1k}=\beta_{0k}+\sum_{u=1}^{h} \beta_{i_u k} C_{i_u k} \tag{22}$$

The quantities $C_{i_u k}$ are ambient concentrations of the tracer species $i_1, i_2, \ldots, i_h$ and are assumed known. The quantities C_{1k} are the ambient concentrations of the aerosol species being apportioned and are also assumed known. We thus have a set of s linear equations in $h + 1$ unknowns $\beta_0, \beta_{i_1}, \beta_{i_2}, \ldots, \beta_{i_h}$. If the system of equations has rank $h + 1$, then these unknown beta coefficients may be obtained by solving the above system of linear equations. The apportionment of the species of interest to the various groups of sources is then carried out by calculating the individual terms of the equations above.

In certain instances it is known that the beta coefficients will differ significantly from one time period to another. In such cases it may be possible to determine, based on physical and chemical reasons, a function of the field measurements, the sampling period and the source, which we denote by ϕ_{jk}, such that is is more reasonable to assume the quantities $\beta_{i_u k}/\phi_{jk}$ are constant for all sampling periods rather than the quantites $\beta_{i_u k}$. In such cases we define $\gamma_{i_u} = \beta_{i_u k}/\phi_{jk}$. For uniformity of notation we define γ_0 to be equal to β_0. This results in the system of linear equations

$$C_{1k}=\gamma_0+\sum_{u=1}^{h} \gamma_{i_u} C_{i_u k} \phi_{jk} \,. \tag{23}$$

We may refer to this set of equations as the **TMBR model**. Again, if this set of equations has rank $h + 1$ then we may solve for the gamma coefficients and consequently calculate the individual terms of the equations. This will yield the apportionment we seek. Note that if we take $\phi_{jk} = 1$ then this set of equations reduces to the set of equations in (22).

Tracer Mass Balance (TMB) Model

This is a special case of the TMBR model and is obtained by partitioning the sources contributing a particular secondary aerosol species, (say species $i = 1$ with associated parent species designated as species $i^* = 2$), into two groups rather than $h + 1$ groups. That is, we take $h = 1$ in the TMBR model. The two groups are: (1) a distinguised source labeled $j = 1$ with associated tracer species $i = i_1$, and (2) all other sources. In this case, the TMBR model reduces to

$$C_{1k}=\beta_{0k}+\beta_{i_1k}C_{i_1k} \ . \tag{24}$$

As before, if we assume that the beta coefficients are independent of the sampling period, then the TMB model equations further reduce to

$$C_{1k}=\beta_{0}+\beta_{i_1}C_{i_1k} \ . \tag{25}$$

If the quantities C_{1k} and C_{i_1k} are known, and if the set of linear equations in (32) have rank 2 then we can solve for the unknown beta coefficients and consequently carry out the apportionment of species 1 by computing the individual terms of the above equations.

In certain instances it is known that the beta coefficients will differ significantly from one time period to another. In such cases it may be possible to determine, based on physical and chemical reasons, a function of the field measurements, the sampling period and the source, which we denote by ϕ_{1k}, such that it is more reasonable to assume the quantities β_{i_1k}/ϕ_{1k} are constant for all sampling periods rather than the quantities β_{i_1k}. In such cases we define $\gamma_{i_1} = \beta_{i_1k}/\phi_{1k}$. For uniformity of notation we define γ_0 to be equal to β_0. This results in the system of linear equations

$$C_{1k}=\gamma_{0}+\gamma_{i_1}C_{i_1k}\phi_{1k} \ . \tag{26}$$

We may refer to the above system of equations as the **TMB model.** Again, if this set of equations has rank 2, then we may solve for the gamma coefficients and consequently calculate the individual terms of the equations. This will yield the apportionment we seek.

A Special Case

The simplest versions of the TMBR model use $\phi_{uk} = 1$ for all time periods and source groups. However, if K_c or K_d are dependent on other variables such as solar radiation, concentration of key atmospheric chemicals and so forth, it may be possible to chose a form of ϕ_{uk} that will linearize the TMBR model.

In apportioning a secondary aerosol, the constant $\beta_{i_u jk}$ derived from the GMB model had the form

$$\beta_{i_u jk} = \frac{r^*_{1jk}}{r_{i_u jk}} \frac{c_{2jk}}{c_{i_u jk}} \tag{27}$$

with

$$r^*_{ijk} = \frac{K_c(i^*j,K)}{K_c(i^*j,k)+K_d(i^*j,k)-K_d(i_u j,k)} \times \{\exp(-K_d(i_u j,k)t_{jk}) - \exp(-[K_c(i^*j,k)+K_d(i^*j,k)]t_{jk})\} \tag{28}$$

and

$$r_{i_u jk} = \exp(-(K_c(i_u j,k)+K_d(i_u j,k))t_{jk}) \ . \tag{29}$$

If the species i_u does not convert and its deposition rate is the same as that of the secondary aerosol species i being apportioned, then

$$r_{i_u jk} = \exp(-K_d(i_u j,k)t_{jk}) \tag{30}$$

so that the ratio $r^*_{ijk}/r_{i_u jk}$ reduces to $K_c(i^*j,k)t_{jk}$ after using the approximation

$$\exp(x) \approx 1+x \ (\textit{when } x \textit{ is sufficiently small}). \tag{31}$$

The full infinite series expansion for *exp(x)* is given by

$$\exp(x) = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots$$

and a first order appromixation has been used in (31). It is possible to use higher order approximations of *exp(x)* in these derivations, but this is not pursued here.

An an example of the above approximation considers a case where $K_c(i^*j,k)$ is proportional to RH_{uk} with proportionality constant B_{i^*j}. Then the ratio $r^*_{ijk}/r_{i_u jk}$ is equal to $B_{i^*j}t_{jk}RH_{uk}$ which gives

$$\beta_{i_u jk} = B_{i*j} t_{jk}\ RH_{uk} \frac{c_{2jk}}{c_{i_u jk}}. \tag{32}$$

Defining

$$\gamma_{i_u k} = \beta_{i_u k} / RH_{uk} \tag{33}$$

and assuming that $\gamma_{i_u k}$ are constant for all sampling periods rather than the quantities $\beta_{i_u k}$ suggests the use of RH_{uk} as a linear factor in the TMBR model equation (23).

The use of RH as a linearization parameter does not necessarily imply that the RH dependence of K_c is grounded in some basic chemical process. Rather, in the case of SO_2 to SO_4 oxidation, RH may be thought of as a surrogate variable depicting the amount of time that SO_2 spends in contact with clouds where oxidation is accelerated. Therefore, assuming $RH_{uk} = RH_k$, the TMBR model for the SO_2 - SO_4 system becomes

$$C_{SO_4 k} = \gamma_0 + \sum_{u=1}^{h} \gamma_{i_u} C_{i_u k} RH_k \tag{34}$$

where:

$C_{SO_4 k}$	=	concentration of sulfate sulfur for time period k
$C_{i_u k}$	=	concentration of trace element i_u for time period k
γ_{i_u}	=	fractionation coefficient associated with trace element i_u
γ_0	=	background concentration of the species being apportioned, due to all sources not accounted for explicitly
h	=	the number of source groups or types, each source group having a unique tracer
RH_k	=	the relative humidity at the sampling site for time period k.

γ_0 and the γ_{i_u}'s can be estimated by various least square estimation techniques. Since both independent and dependent variables have error associated with them ODR is the method of choice.

Model Calculations and Uncertainties

We outline two approaches for the calculation of uncertainties associated with estimated contributions and fractional contributions of the species of interest (species 1, say) by the source of interest $S_7 (j = 1)$ and the associated uncertainties. The first approach is computationally intensive as it involves computer simulation while the second approach is computationally simpler but relies on several approximations being sufficiently accurate.

First Approach. In Equation (23), the quantities C_{1k}, $C_{i_u k}$ and ϕ_{uk} are all observed with error. We shall denote the true values by C_{1k}, $C_{i_u k}$ and ϕ_{uk} and the observed values by the quantities $\tilde{C}_{1k}$, $\tilde{C}_{i_u k}$ and ϕ_{uk}. We then assume that

$$\begin{aligned} C_{1k} &= C_{1k} + \epsilon_{C_{1k}} \\ C_{i_u k} &= C_{i_u k} + \epsilon_{C_{i_u k}} \\ \hat{\phi}_{uk} &= \phi_{uk} + \epsilon_{\phi_{uk}} . \end{aligned} \tag{35}$$

The quantity $\epsilon_{C_{1k}}$ is a random error with mean 0 and standard deviation $\sigma_{C_{1k}}$. The quantity $\epsilon_{C_{1k}}$ is a random error with mean 0 and standard deviation $\sigma_{C_{1k}}$. Likewise, the quantity $\epsilon\phi_{uk}$ is a random error with mean 0 and standard deviation $\sigma_{\phi uk}$. All random errors are assumed to be normal and mutually independent.

The unknown quantities γ_0, and $\gamma_{i_u k}$ for $u = 1,\ldots,h$ are estimated using the method of Orthogonal Distance Regression (ODR) with input data consisting of the measured values as well as the measurement uncertainties. Estimates of the true values C_{1k}, $C_{i_u k}$ and ϕ_{uk} are also obtained as output from ODR. The estimated values will be denoted with 'hats' over the corresponding true values. Then the estimated value of C^*_{1uk}, denoted by C^*_{1uk} is given by,

$$\hat{C}^*_{1uk} = \hat{\gamma}_{i_u} \hat{C}_{i_u k} \hat{\phi}_{uk} .$$

From this we obtain the estimated fractional contribution F_{uk} of species 1 by source group u for sampling period k as

$$F_{uk} = \frac{\hat{C}^*_{1uk}}{\hat{C}_{1k}} .$$

The estimated fractional contribution for the entire sampling period, by source group u, is denoted by F_u and is calculated as

$$F_u = \frac{\sum_{k=1}^{s} \hat{C}^*_{1uk}}{\sum_{k=1}^{s} \hat{C}_{1k}} .$$

To calculate the uncertainties to be associated with these estimates we may use the following procedure. We construct several (say, 100) synthetic data sets by perturbing the estimates of the true values C_{1k}, $C_{i_u k}$ and ϕ_{uk} using Gaussian random deviates with mean zero and standard deviations equal to the respective measurement uncertainties. Each such synthetic data set is subjected to an ODR analysis to obtain estimates of contributions and fractional contributions of the various source groups to the receptor as explained above. This procedure

results in a whole collection of estimates (say, 100) for the various quantities of interest. The root mean square error is then calculated for each quantity of interest using the collection of estimates obtained from perturbed synthetic data sets and using the initial estimates obtained from the actual data set as if they were the true values. This root mean square error associated with a given quantity of interest is used to quantify the uncertainty associated with that quantity. Recall that if θ represents the true value of a quantity and θ^*_q represents an estimate of θ obtained from the q^{th} synthetic data set, then the root mean square error is calculated by

$$\textit{Root Mean Square Error} = \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}(\theta_q^* - \theta)^2} \ .$$

Alternatively, we may quantify the uncertainty associated with a given estimate using confidence intervals but we do not discuss that approach here.

Second Approach. In this section we discuss an approximate method of calculating the uncertainties associated with the model outputs. The concentrations C^*_{1uk} of species 1 (secondary species of interest) associated with each trace element i_u for each time period may be calculated by multiplying the measured values of $A_{i_uk} = C_{i_uk}\phi_{uk}$ for each trace element by the respective estimated regression coefficients as follows. ($\hat{\gamma}_0$ would just be the estimated intercept representing the estimated contribution from all sources not explicitly accounted for by any of the reference species used in the TMBR model.)

$$\hat{C}^*_{1uk} = \hat{\gamma}_{i_u} \times A_{i_uk} \ . \tag{35}$$

The uncertainties for each of these concentrations C^*_{1uk} may be calculated by:

$$\sigma_{C^*_{1uk}} = \sqrt{A^2_{i_uk}\sigma^2_{\gamma i_u} + \gamma^2_{i_u}\sigma^2_{Ai_uk} + \sigma^2_{\gamma i_u}\sigma^2_{Ai_uk}} \tag{36}$$

The quantities σ_{Ai_uk} are the uncertainties in the measured values A_{i_uk} and are assumed to be known. In the special case discussed in the previous section, these uncertainties are part of the WHITEX data base. The quantities $\hat{\gamma}_{iu}$ may be obtained as outputs from the regression packages that are used. Errors in A_{i_uk} and the estimated regression coefficients have been assumed to be independent in the calculation of Equation (36).

The total calculated amount of species 1, C_{1k} for each time period is the sum of the C^*_{1uk} summed over all the reference aerosol species i_u and the intercept $\hat{\gamma}_0$.

$$C_{1k}=\hat{Y}_0+\sum_{u=1}^{h}\hat{C}^{*}_{1uk}\ . \tag{37}$$

The uncertainty associated with the total calculated concentration of species 1 for each time period is:

$$\sigma_{C_{1k}}=\sqrt{\sigma^2_{Y_0}+\sum_{u=1}^{h}\sigma_{C^{*}_{1uk}}} \tag{38}$$

assuming the covariance terms arising in the derivation are negligible.

The estimated fraction F_{uk} of species 1 from each source for any given time period is equal to the amount of species 1 associated with the trace element divided by the total calculated concentration of species 1:

$$F_{uk}=\frac{\hat{C}^{*}_{1uk}}{C_{1k}}\ . \tag{39}$$

The uncertainty for each of these fractions is:

$$\sigma_{F_uk}=\sqrt{\frac{\sigma^2_{C^{*}_{1uk}}}{C^2_{1k}}+\frac{\hat{C}^{*2}_{1uk}\sigma^2_{C_{1k}}}{C^4_{1k}}}\ . \tag{40}$$

The mean fraction $\bar{F}_u$ of species 1 attributed to each source type is estimated by the mean species 1 concentration $\bar{C}_u$ for that source type divided by the mean total calculated concentration of species 1, $\bar{C}$ as follows:

$$\bar{F}_u=\frac{\bar{C}_u}{\bar{C}} \tag{41}$$

where

$$\bar{C}_u=\frac{1}{S}\sum_{k=1}^{S}\hat{C}^{*}_{1uk} \tag{42}$$

and

The uncertainties for $\bar{C}_u$ and $\bar{C}$ are calculated by:

and

The uncertainties associated with the mean fractions $\bar{F}_u$ are calculated by

$$\bar{C}=\frac{1}{S}\sum_{k=1}^{S} C_{1k} \,. \tag{43}$$

$$\sigma_{\bar{C}_u}=\frac{1}{K}\sqrt{\sum_{k=1}^{S}\sigma^2_{C^*_{1uk}}} \tag{44}$$

$$\sigma_{\bar{C}}=\frac{1}{K}\sqrt{\sum_{k=1}^{S}\sigma^2_{C_{1k}}} \,. \tag{45}$$

$$\sigma_{\bar{F}_u}=\sqrt{\frac{\sigma^2_{\bar{C}_u}}{\bar{C}^2}+\frac{\bar{C}_k^2\sigma^2_{\bar{C}}}{\bar{C}^4}} \,. \tag{46}$$

The uncertainty formulas are all derived using propagation of error methods and assuming the covariances between various terms occurring in the derivation are negligible. These assumptions will not be true in practice and so the usefulness of the above approximations will depend upon how severely the assumptions used in the above derivations are violated.

Model Assumptions

The TMBR assumptions are:

- The chemical species used as tracers in the model are assumed to be uniquely emitted by non-overlapping groups of sources. In particular none of the species other than the tracer associated with the source of interest can be emitted by another source unless there is an independent method such as CMB to partition the ambient species concentrations into components attributable to the various groups of sources.

- The composition of source emissions are constant over the period of ambient sampling.

- Deposition and conversion are constant from one sampling period to the next for each subgroup *u*.

- Measurement errors are random, uncorrelated, and normally distributed.

For the special case where k_c was assumed to be proportional to RH the additional assumptions are:

- Exponential forms of deposition and conversions can be represented by first order approximations.

- The RH at the receptor site is indicative of the amount of time that air parcels spend in contact with clouds and therefore can be used as an indicator of oxidation rate.

Potential Deviations from Assumptions

It is highly unlikely that deposition and conversion are constant in space and time and in many cases one can expect source profiles to change over the course of the study. These assumptions are implicit to the assumption that background and fractionation coefficients are time independent. Whether or not a linearization scheme is appropriate can be examined through goodness of fit tests of the proposed model and possibly by direct experimental verification. The uniqueness of tracer species can be assessed by source testing and by releasing unique tracers from sources of interest.

Deviation from any of the assumptions will increase the calculated uncertainty in the final apportionments. The extent to which the inflation of uncertainty occurs will depend on how variable the regression coefficients are. Research into the effect of deviation from assumptions on apportionments is needed.

Model Inputs

The model requires the following quantities as inputs:

- The ambient concentrations of the aerosol species being apportioned.

- The ambient concentrations of the reference tracer species.

- Relative humidity at the receptor for each of the sampling periods, when $\phi_{uk} = RH_{uk}$ is used in the model rather than $\phi_{uk} = 1$.

- The uncertainties in the above quantities, when ODR is used to estimate the γ coefficients, rather than OLS.

Model Outputs

The model outputs include:

- Estimates of the actual amount of the contribution and the fractional contribution of the aerosol species of interest by the source or source type of interest to the receptor, along with the associated uncertainty estimates.

- Estimates of the average amount and the average fractional amount of the aerosol species of interest contributed by each source or source type of interest along with the associated uncertainty estimates.

Differential Mass Balance (DMB) Model

The DMB model is a receptor model combined with elements of a deterministic model. In this approach dispersion is accounted for by rationing ambient trace material concentrations attributed to a source by known trace material release rates while deposition and conversion are explicitly calculated. The name "Differential Mass Balance" refers to the use of difference in trace material concentration to account for dispersion.

Model Equations

Suppose a particular source is of interest and we wish to determine the fractional contribution of some aerosol species to the receptor by that source. We shall designate the aerosol species of interest by the subscript i, and the source of interest by j. If species i is a secondary species, then the corresponding parent species will be denoted by the subscript i^*. For example, if SO_4 is of interest, then i stands for SO_4 and i^* stands for SO_2. We are then interested in the quantity C_{ijk} for each of the sampling periods. We have, from Equation (2) that

$$C_{ijk}=c_{ijk}r_{ijk}d_{jk}+c_{i^*jk}r^*_{ijk}d_{jk} \tag{47}$$

If i represents a primary species, then r^*_{ijk} is zero for all k. If i represents a secondary aerosol species that is not emitted as a primary aerosol, then the quantity c_{ijk} is zero for all k. Therefore, the above equation simplifies to

$$C_{ijk}=c_{ijk}r_{ijk}d_{jk} \tag{48}$$

when i is a primary species and

$$C_{ijk}=c_{i^*jk}r^*_{ijk}d_{jk} \tag{49}$$

when i is a secondary species. A characteristic feature of DMB model applications is that the dispersion factor d_{jk} is determined based on field measurements. If a unique tracer is available for source j, then d_{jk} may be calculated based on this unique tracer. It can also be calculated based on a reference aerosol species that may not be a unique tracer for source j by first calculating the amount of this reference species contributed to the receptor by the source of interest. Chemical mass balance model may be applied for this purpose. Other approaches are also possible.

The following discussion assumes that a unique tracer is available for source j of interest. This source will be referred to as S_j. The tracer material may be a naturally emitted primary aerosol species or may be introduced artificially. The aerosol species is denoted by the subscript i_0. Therefore, Equation (48) becomes

$$C_{i_0jk}=c_{i_0jk}r_{i_0jk}d_{jk}\ . \tag{50}$$

Dividing the quantity C_{ijk} by the quantity C_{i_0jk} we get,

$$\frac{C_{ijk}}{C_{i_0jk}}=\frac{c_{ijk}}{c_{i_0jk}}\frac{r_{ijk}}{r_{i_0jk}} \tag{51}$$

when species i is a primary aerosol and

$$\frac{C_{ijk}}{C_{i_0jk}}=\frac{c_{i*jk}}{c_{i_0jk}}\frac{r^*_{ijk}}{r_{i_0jk}} \tag{52}$$

when species i is a secondary aerosol. It follows from this that

$$C_{ijk}=\frac{c_{ijk}}{c_{i_0jk}}\frac{r_{ijk}}{r_{i_0jk}}C_{i_0jk} \tag{53}$$

for primary aerosols i and

$$C_{ijk}=\frac{c_{i*jk}}{c_{i_0jk}}\frac{r^*_{ijk}}{r_{i_0jk}}C_{i_0jk} \tag{54}$$

for secondary aerosols.

Since aerosol component i_0 is a tracer for source j, the quantity C_{i_0jk} is the same as the quantity C_{i_0jk} which is the ambient concentration of species i_0 at the receptor and can be measured. Furthermore, if the quantities $K_d(i,j,k)$, $K_c(i,j,k)$ are known when species i is primary, or, $K_c(i^*,j,k)$, $K_d(i^*,j,k)$ and $K_d(i,j,k)$ are known when species i is secondary, and if in addition, $K_d(i_0,j,k)$, K_c (i_0,j,k), t_{jk} as well as the ratio c_{i*jk}/c_{i_0jk} are known, then the contribution of the source of interest to the concentrations of the species of interest at the receptor can, in principal, be calculated.

If T represents a unique nonconverting, nondepositing tracer for source $j = 1$, then for a species that is directly emitted by source $j = 1$, Equation (53) for primary aerosols reduces to

$$C_{i1k}=\frac{c_{1ik}}{c_{T,1,k}} r_{i1k} C_{T,k} \ . \tag{55}$$

If the ratio $c_{i1k}/c_{T,1,k}$ is known, the form of r_{i1k} is

$$r_{i1k}=\exp(-K_d(i,1,k)t_{1k}) \ .$$

For a species that is not directly emitted, but is a secondary species which is absent at the source, the equation for the DMB reduces to

$$C_{i1k}=\frac{c_{i*1k}}{c_{T,1,k}} r^*_{i1k} C_{T,k} \ . \tag{56}$$

The ratio $c_{i*1k}/c_{T,1,k}$ is assumed known and the form of r^*_{i1k} in this case is

$$r^*_{i1k}=\frac{K_c(i^*,1,k)}{K_c(i^*,1,k_+K_d(i^*,1,k)-K_d(i,1,k)} \times$$

$$\{\exp(-K_d(i,1,k)t_{1k}-\exp(-[K_c(i^*,1,k)+K_d(i^*,1,k)]t_{1k})\}$$

where

$K_c(i,1,k)$ = conversion rate of species i from source 1 to its secondary form, during sampling period k.

$K_d(i,1,k)$ = deposition rate of species i from source 1 during sampling period k.

Considering a specific example for SO_4 and SO_2 Equation (56) becomes

$$C_{SO_4,1,k}=\frac{c_{SO_2,1,k}}{c_{T,1,k}} r^*_{SO_4,1,k} C_{T,k} \tag{57}$$

and

$$C_{SO_2,1,k}=\frac{c_{SO_2,1,k}}{c_{T,1,k}} r_{SO_2,1,k} C_{T,k} \tag{58}$$

where

$$r^{*}_{SO_4,1,k}=\frac{K_c(SO_2,1,k)}{K_c(SO_2,1,k)+K_d(SO_2,1,k)-K_d(SO_4,1,k)}\times \tag{59}$$

$$\{\exp(-K_d(SO_4,1,k)t_{1k})-\exp(-[K_c(SO_2,1,k)+K_d(SO_2,1,k)]t_{1k})\}$$

and

$$r_{SO_2,1,k}=\exp(-(K_c(SO_2,1,k)+K_d(SO_2,1,k))t_{1k})\ . \tag{60}$$

From now on we shall use the notation $K_c = K_c(SO_2,1,k), K_1 = K_d(SO_2,1,k)$ and $K_2 = K_d(SO_4,1,k)$. Furthermore, these parameters may be related to deposition velocities v_1 for SO_2 and v_2 for SO_4, and SO_2 oxidation rate K_c by the equations

$$K_1=\frac{v_1}{H_m} \tag{61}$$

$$K_2=\frac{v_2}{H_m} \tag{62}$$

where H_m is the mixing height. K_1, K_2, or K_c may be functions time of day, surface conditions, meteorological conditions, relative humidity and a number of other variables.

Model Calculation

Again consider the specific example of calculating the fractional contribution of SO_4 that is associated with a specific source emission of SO_2. The contribution of SO_4 by the source of interest ($j = 1$) to the receptor is calculated using Equation (57). The value of $c_{SO_2,1,k}/c_{T,1,k}$ is estimated from field measurements. The plume ages t_{1k} can be estimated from plume streakline analysis. C_{T1k} are ambient concentrations of unique tracer, T, at the receptor and are assumed to be measured during the experiment. The values of K_c, K_1 and K_2 are known and may be estimated based on literature values of deposition velocity for SO_2 and particles and pseudo-first-order SO_2 oxidation rates. Alternatively, they may also be empirically derived from the measurements made during the experiment.

To judge if a particular combination of these parameters is consistent with the field measurements, the following procedure may be adopted. Using the chosen combination of values for these parameters we first calculate the SO_4 contributions $C_{SO_4,S_1,k}$ of source $j = 1$ for each sampling period. This procedure can be repeated for as many sources for which there are unique tracers or for as many sources for which the relative contribution of those sources to an ambient trace element concentration have been established. Relative contributions of sources to a specific tracer species could be established by CMB or deterministic approaches.

Therefore, the regression model

$$C_{SO_4,k}=\beta_0+\beta_1 C^*_{SO_4,1,k}+\sum_{u=2}^{h}\beta_u C^*_{SO_4,i_u,k}+error \qquad (63)$$

may be fitted and the adequacy of the fit judged by the resulting R^2 value and the closeness of the beta coefficients to one. $C^*_{SO_4,i,u,k}$ refers to the total contribution of SO_4 by source group u to the receptor. If the chosen parameter combination results in a high R^2 value and beta values are not significantly different from one, then the chosen parameter values v_1, v_2, K_c may be judged as being consistent with observed data. The best possible value of R^2 obtained, by varying the value of v_1, v_2 and K_c over their entire range of values suggested in the literature, may be denoted by R^2_{opt}. The values $v_1 = v_{1,opt}$, $v_2 = v_{2,opt}$, and $K_c = K_{c,opt}$ which result in the best R^2 may be used to calculate the daily S_i contributions to SO_4 and SO_2 at the receptor. By calculating the ratio of the total S_i contribution over the entire sampling period to total ambient concentrations over the same period we can calculate the fractional SO_4 and SO_2 contributions by S_i during the experimental period.

Uncertainty Calculations

Uncertainties in the final results are primarily due to three sources.

- Uncertainties in *tjk*.
- Uncertainties in the model parameters such as K_c, K_1, and K_2.
- Uncertainties in the measured values.
- Uncertainties in the extent to which the model assumptions are violated.

Uncertainties in the Model Parameters. The model parameters in question are K_c, K_1 and K_2 which are not known. Suppose a review of the literature suggests deposition velocities v_1 for SO_2 ranges from l_1 to u_1 cm/sec and v_2 for SO_4 ranges from l_2 to u_2 cm/sec. In addition suppose the sulfur dioxide oxidation rates varied from $K_c = l_c$ *to* $K_c = u_c$ percent per hour.

Clearly, not all combinations of values of v_1, v_2 and K_c are physically possible. To judge which combinations of these parameters are reasonable, the following procedure may be adopted. A grid of values for v_1, v_2 and K_c may be chosen by taking all possible combinations of these parameters resulting from

$v_1 = l_1$ to u_1 in increments of δ_1.
$v_2 = l_2$ to u_2 in increments of δ_2.
$K_c = l_c$ to u_c in increments of δ_c.

To decide whether a particular combination of values of v_1, v_2 and K_c are reasonable the regression model suggested by Equation 63 can be exercised and the adequacy of the fit may be judged by closeness of beta values to one and the resulting R^2. The best possible value of R^2 for β values close to one over the range of these parameters is denoted by R^2_{opt}. A value R^2_0 less than R^2_{opt} but close to it is chosen based on subjective judgements, is a criterion value for judging the reasonableness of various combinations of the parameter values. Parameter combinations resulting in an R^2 equal to R^2_0 or greater may be considered reasonable. The set of all such parameter combinations will be denoted by the symbol A. This will result in a whole range of values for the daily S_t contributions can be calculated for each of the parameter combinations in the set A. S_t contributions and the overall average S_t contributions. The mean and the standard deviation for this range of values (as well as the minimum and the maximum values) may be calculated to assess the uncertainty in the estimated S_t contributions due to imprecise knowledge of the model parameters. The measured values of concentration of species are assumed to be exact in these calculations.

Uncertainties in the Measured Values. To assess the effect of errors in measurements on the estimated S_t contributions to SO_4 and SO_2 at the receptor, the values of v_1, v_2 and K_c are fixed at their optimum values obtained as explained in the previous subsection. The measured values used in the calculations are (1) the ambient T concentration, $C_{T,k}$; (2) the ambient SO_4 concentration $C_{SO4,k}$; (3) the ambient SO_2 concentration $C_{SO2,k}$; (4) relative humidity RH_k at the receptor; and (5) transport time $t_{S_t,k}$ for the aerosol mixture from S_t to arrive at the receptor. Suppose each of these measurements have associated with them a standard deviation characterizing the uncertainty in the respective measurements. We generate a number of synthetic data sets (one hundred is sufficient for most purposes) on the computer by perturbing the measured values using random Gaussian deviates with zero means and standard deviations associated with each of the measured values. For each synthetic data set thus generated, the daily S_t contribution to SO_4 and SO_2 at the receptor as well as the average contributions over the entire sampling period are calculated. The range of values thus obtained for each of these quantities gives an indication of the uncertainty that would be due to imprecise measurements alone. The results are reported in the form of means and standard deviations of each of the quantities of interest calculated from the synthetic data sets. Throughout this exercise, the model parameters, viz, the conversion and deposition parameters, are to be kept constant at their optimum values.

Uncertainties in the Extent to which the Model Assumptions are Violated. Assessment of the uncertainties in reported results arising from model violations can be evaluated by conducting extensive sensitivity studies involving various perturbations in the model assumptions themselves.

Overall Uncertainties. Since the first two categories of uncertainties are expected to be "independent," the total uncertainty due to these two sources may be characterized by the effective total standard deviation

$$\sigma_{Total} = (\sigma_1^2 + \sigma_2^2)^{1/2}$$

where σ_1 and σ_2 are the standard deviations associated with the two categories of uncertainties, respectively.

Model Assumptions

The DMB model assumes that the rates for deposition and conversion processes in the atmosphere are first-order and invariant in space and time. In particular, these rates are assumed to be constant at every point in space and time along the transport path. Deposition is dominated by dry deposition; precipitation scavenging is small by comparison. It is also assumed that the ratio of the emission rates for the species of interest (or its parent species) and the tracer is known.

Potential Deviations from Assumptions

Rates for deposition processes may not be first-order and invariant in space and time. For example, no dry deposition will occur until the plume has been mixed to the ground. Dry deposition velocities are known to vary depending on atmospheric stability, and the type of surface (vegetation or rock, dry or wet). Wet deposition during periods of intense precipitation may deposit more material than dry processes do. The ratio of the emission rates for the species of interest (or its parent species) and the tracer may not be known precisely.

Model Inputs

The inputs to the model are, as a function of time, the relative emission rates of SO_2 and tracer, ambient concentrations of tracer, primary and secondary aerosols, and plume ages, deposition velocities for particulate and reactive gases (v_d) , mixing height (H_m), SO_2 pseudo-first-order oxidation rate.

Model Outputs

Ambient concentrations and fractions of total ambient concentrations of aerosols of interest associated with a given source.

Conclusion

A set of deterministic general mass balance (GMB) equations describing how primary and secondary aerosols and gases are transported and transformed as they pass through the atmosphere were formulated. From the GMB equations it is possible, with a variety of limiting assumptions, to derive the chemical mass balance, the differential mass balance equations, and the tracer mass balance regression model. Derivation of these receptor modeling approaches from a first principle model allows for an examination of model assumptions and deviations from

assumptions. With assumptions identified it possible to make a better determination of how to incorporate measurement uncertainty and how to estimate model uncertainty associated with an imperfect knowledge of model parameters.

References

1. J.G. Watson, Overview of receptor model principles. *JAPCA*, **34**, 620, 1984.

2. J.G. Watson, J.C. Chow, D.L. Freeman, R.T. Egami, P. Roberts and R. Countess, *Model and Data Base Description for California's Level I PM10 Assessment Package.* DRI Document 8066-002.1D1, Draft Report, Prepared for the California Air Resources Board, Sacramento, CA., 1987.

3. J.G. Watson, 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. Prepared for the Western States Acid Deposition Project by Systems Applications, Inc., San Rafael, Ca., SYSAPP-87/064, 1987.

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

5. P.K. Hopke, Receptor modeling in environmental chemistry. *Chemical Analysis*, 76, John Wiley & Sons, New York, N.Y., 1985.

6. R.K. Stevens, C.W. Lewis, Hybrid receptor modeling. *In: Extended Abstracts for the Fifth Joint Conference on Applications of Air Pollution Meteorology with APCA*, November 18-21, 1986, Chapel Hill, N.C. Published by the American Meteorological Society, Boston, MA., 1987.

7. C.W. Lewis, R.K. Stevens, Hybrid receptor model for secondary sulfate from an SO_2 point source. *Atmos. Environ.* **19**,6:917-924, 1985.

8. T. Dzubay, R.K. Stevens, G.E. Gordon, I. Olmez, A.E. Sheffield, W.J. Courtney, A composite receptor method applied to Philadelphia aerosol. *Environ. Sci. & Technol.*, **22**, 1, 1988.

9. J.G. Watson, *Transactions, Receptor Models in Air Resources Management, Air and Waste Management Assoc.*, Editor, Pittsburgh, PA., 1989.

10. H.I. Britt, and R.H. Luecke, 1973: The estimation of parameters in nonlinear, implicit models. *Technometrics*, **15**, 233, 1973.

11. J.G. Watson, J.A. Cooper and J.J. Huntzicker, The effective variance weighing for least squares calculations applied to the mass balance receptor model, *Atmos. Environ.*, **18**, 1347, 1984.

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

13. R.C. Henry, Stability analysis of receptor models that use least squares fitting. Receptor Models Applied to Contemporary Air Pollution Problems, Air Pollution Control Association, Pittsburgh, PA., 1982.

14. H.J. Williamson, and D.A. DuBose, 1983: Receptor model technical series, volume III: User's manual for chemical mass balance model. EPA-450/4-83-014, U.S. Environmental Protection Agency, Research Triangle Park, N.C., 1983.

15. G.E. Gordon, W.H. Zoller, G.S. Kowalczyk and S.H. Rheingrover, Composition of source components needed for aerosol receptor models. *Atmospheric Aerosol: Source/Air Quality Relationships*. Edited by E.S. Macias and P.K. Hopke, American Chemical Society Symposium Series #167, Washington, D.C., 1981.

16. L.A. Currie, 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. *Atmos. Environ.* 18, 1517, 1984.

17. T.G. Dzubay, 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. *Atmos. Environ.*, 18, 1555, 1984.

18. J.G. Watson, and N.F. Robinson, A method to determine accuracy and precision required of receptor model measurements. Quality Assurance in Air Pollution Measurements, Air Pollution Control Association, Pittsburgh, PA., 1984.

19. R.T. DeCesar, S.A. Edgerton, M.A.K. Kahlil and R.A. Rasmussen, Sensitivity analysis of mass balance receptor modeling: methyl chloride as an indicator of wood smoke. *Chemosphere*, **14**, 1495, 1985.

20. R.T. DeCesar, S.A. Edgerton, M.A. Khalil and R.A. Rasmussen, A tool for designing receptor model studies to apportion source impacts with specified precisions. Receptor Methods for Source Apportionment: Real World Issues and Applications, Air Pollution Control Association, Pittsburgh, PA., 1986.

21. H.S. Javitz, J.G. Watson, J.P. Guertin and P.K. Mueller, Results of a receptor modeling feasibility study, *JAPCA*, **38**, 661, 1988.

22. H.S. Javitz, J.G. Watson, and N. Robinson, Performance of the chemical mass balance model with simulated local-scale aerosols, *Atmos. Environ.*, **22**, 2309, 1988.

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

24. B. Kim, and R.C. Henry, Analysis of multicollinearity indicators and influential species for chemical mass balance receptor model, *Transactions, Receptor Models in Air Resources Management*, J.G. Watson, ed., Air and Waste Management Assoc., Pittsburgh, PA., 1989.

25. J.C. Chow, and J.G. Watson, Summary of particulate data bases for receptor modeling in the United States, *Transactions, Receptor Models in Air Resources Management*, J.G. Watson, ed., Air and Waste Management Assoc., Pittsburgh, P.A., 1989.

26. W.A. Fuller, *Measurement Error Models*, John Wiley and Sons, New York, N.Y., 1987.