Addendum 6: Presented at the Air and Waste Management Association International Specialty Conference 'Visibility and Fine Particles', October 1989, Estes Park, CO.

SURVEY OF A VARIETY OF RECEPTOR MODELING TECHNIQUES

William C. Malm
National Park Service, Air Quality Division
Cooperative Institute for Research in the Atmosphere
Colorado State University
Ft. Collins, CO 80523

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

John Watson
EEEC, Desert Research Institute
Reno, NV 89506

Douglas A. Latimer
Latimer & Associates
P.O. Box 4127
Boulder, CO 80306-4127

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 concentration 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. Assumptions 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 aerosol 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

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

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

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

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

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

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

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

and

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

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

$$C_{ik} = \sum_{j=1}^{n} C_{ijk} \quad (13)$$

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

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

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

Let $S_{jk} = \sum_{i=1}^{m} C_{ijk}$. 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 Equations (13) and (15) it follows that the set of Equations in (16) below also holds.

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

If the a_{ijk} for all the sources affecting the receptor sites are known, then 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; 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 this similar source compositions, which is known as "collinearity." He uses the "singular value decomposition" 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 colinearities. 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 be 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, 16, 17, 18, 19, 20, 21, 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 obviously 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 measurement uncertainties), there are no results available from verification or evaluation studies. Every least squares solution to the CMB equations requires this assumption, as demonstrated by the derivation of Watson.[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; and 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 contribution estimates (S_j) of each source type; 2) the standard errors of these source contribution estimates. 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 ex-

pected 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 (1987).[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 concentrations 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 $j = 2$ thru $j = n_2$ have an associated tracer species i_2, sources $j = n_2 + 1$ thru $j = n_3$ have an associated tracer species i_3 etc., and sources $j = n_{h-1} + 1$ thru $j = n_h$ have an associated tracer i_h. Sources $j = n_h + 1$ thru $j = n$ 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

$$C_{1k} = \beta_0 + \sum_{u=1}^{h} \beta_{i_u} C_{i_u k} \tag{22}$$

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 it is more reasonable to assume the quantities $\beta_{i_u k}/\phi_{jk}$ are constant for all sampling periods rather than the quantities $\beta_{i_u k}$. 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: (i) A distinguished source labeled $j = 1$ with associated tracer species $i = i_1$, and (ii) All other sources. In this case, the TMBR model reduces to

$$C_{1k} = \beta_{0k} + \beta_{i_1 k} C_{i_1 k} \tag{24}$$

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

$$C_{1k} = \beta_0 + \beta_{i_1} C_{i_1 k} \tag{25}$$

If the quantities C_{1k} and $C_{i_1 k}$ 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 $\beta_{i_1 k}/\phi_{1k}$ are constant for all sampling periods rather than the quantities $\beta_{i_1 k}$. In such cases we define $\gamma_{i_1} = \beta_{i_1 k}/\phi_{1k}$. For uniformity of notation we define γ_0 to be equal to β_0. This results in the system of linear equations

$$C_{1k} = \gamma_0 + \gamma_{i_1} C_{i_1 k} \phi_{1k} \tag{26}$$

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

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

$$\{exp(-K_d(i, j, k)t_{jk}) - exp(-[K_c(i^*, j, k) + K_d(i^*, j, k)]t_{jk})\} \tag{28}$$

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

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

so that the ratio r^*_{1jk}/r_{i_ujk} reduces to $K_c(i^*,j,k)t_{jk}$ after using the approximation

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

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

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

and a first order approximation 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 consider a case where $K_c(i^*,j,k)$ is proportional to RH_{uk} with proportionality constant B_{i^*j}. Then the ratio r^*_{1jk}/r_{i_ujk} is equal to $B_{i^*j}t_{jk}RH_{uk}$ which gives

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

Defining

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

and assuming that γ_{i_uk} are constant for all sampling periods rather than the quantities β_{i_uk} 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

$$C_{SO_4k} = \gamma_0 + \sum_{u=1}^{h} \gamma_{i_u} C_{i_uk} RH_k \tag{34}$$

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 $S_T(j = 1)$ 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 $\tilde{\phi}_{uk}$. We then assume that

$$\begin{aligned} \tilde{C}_{1k} &= C_{1k} + \epsilon_{C_{1k}} \\ \tilde{C}_{i_u k} &= C_{i_u k} + \epsilon_{C_{i_u k}} \\ \tilde{\phi}_{uk} &= \phi_{uk} + \epsilon_{\phi_{uk}} \end{aligned}$$

The quantity $\epsilon_{C_{1k}}$ is a random error with mean 0 and standard deviation $\sigma_{C_{1k}}$. The quantity $\epsilon_{C_{i_u k}}$ is a random error with mean 0 and standard deviation $\sigma_{C_{i_u k}}$. 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 $u = 1, \cdots, h$ 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 $\hat{C}^*_{1uk}$ 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_uk} 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.)

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

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

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

The quantities $\sigma_{A_{i_u k}}$ are the uncertainties in the measured values $A_{i_u k}$ and is assumed to be known. In the special case discussed in the previous section, these uncertainties are part of the WHITEX data base. The quantities $\hat{\gamma}_{i_u}$ may be obtained as outputs from the regression packages that are used. Errors in $A_{i_u k}$ 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}'s summed over all the reference aerosol species i_u and the intercept $\hat{\gamma}_0$.

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

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

$$\sigma_{C_{1k}} = \sqrt{\sigma^2_{\gamma_0} + \sum_{u=1}^{h} \sigma^2_{\hat{C}^{*}_{1uk}}} \tag{38}$$

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:

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

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

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

where

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

and

$$\overline{C} = \frac{1}{s}\sum_{k=1}^{s} C_{1k}. \tag{43}$$

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

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

and

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

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

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

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 model 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 of not a linearization scheme is appropriate can be examined through goodness of fit tests of the proposed model and possible 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 ratioing 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_1 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, 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_t. 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_0k} which is the ambient concentration of species i_0 at the receptor and can be measured. If furthermore 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

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

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

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

$$\{exp(-K_d(SO_4,1,k)t_{1k}) - exp(-[K_c(SO_2,1,k) + K_d(SO_2,1,k)]t_{1k})\} \quad (59)$$

and

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

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

$$K_1 = \frac{v_1}{H_m} \quad (61)$$

$$K_2 = \frac{v_2}{H_m} \quad (62)$$

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 Calculations

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. $C_{T,1,k}$ 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 unknown 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 $C_{SO_4,S_t,k}$ 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

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

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 values 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_t contributions to SO_4 and SO_2 at the receptor. By calculating the ratio of the total S_t 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_t during the experimental period.

Uncertainty Calculations

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

- Uncertainties in t_jk.
- 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 ranging from l_1 to u_1 cm/sec and v_2 for SO_4 ranging 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

$$v_1 = l_1 \text{ to } u_1 \text{ in increments of } \delta_1.$$

$$v_2 = l_2 \text{ to } u_2 \text{ in increments of } \delta_2.$$

$$K_c = l_c \text{ to } u_c \text{ in increments of } \delta_c.$$

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 judgement, as 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 $\mathcal{A}$. S_t contributions can be calculated for each of the parameter combinations in the set $\mathcal{A}$. This will result in a whole range of values for the daily 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 concentrations 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 $C_{SO_4,k}$, (3) The ambient SO_2 concentration $C_{SO_2,k}$, (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 particulates 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.G., J.C. Chow and N.F. Robinson, *Western States Acid Deposition Project Phase I: Volume 4–An Evaluation of Ambient Aerosol Chemistry in the Western United States.* 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, NY, 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, Massachusetts, 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 weighting 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 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, NC, 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, Pittsburg, 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, NY, 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, PA, 1989.

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