APPENDIX 6A: General Mass Balance Model

Overview

The General Mass Balance (GMB) model is an attempt to combine the differential mass balance (DMB) model, the Tracer Mass Balance Regression (TMBR) model, the Tracer Mass Balance (TMB) model and the Chemical Mass Balance (CMB) model , all in a unified framework. Several other hybrid models can also be shown to be special cases of the GMB model. The model is designed to help calculate the fraction of an ambient aerosol species contributed by an emissions source. The particular special case of the GMB model to be employed is determined by the type of information that is presumed available.

The development of the GMB model as described in this appendix is purely mathematical and deterministic. 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 be addressed separately in appendices dealing with each of the special cases of the GMB model mentioned above.

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 airmass from source j to the receptor, corresponding to sampling period k.

r_{ijk} = a factor that accounts for conversion and 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.

We will also follow the convention that 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 SO4 then i^* will stand for SO2.

Model Equations

It follows from the definitions that for primary aerosol species we have

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

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} and r^*_{ijk} are functions of depositon and conversion rates as well as transport times. Under reasonable assumptions, the quantities r_{ijk} and r^*_{ijk} can be shown to have simple functional forms. These functional forms for r_{ijk} and r^*_{ijk} result from solving certain basic differential equations governing material deposition and conversion as described below.

Suppose $X(t)$ denotes the mass, at time t after emission, of a species i in a unit volume of aerosol mixture. We then 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 for the species under consideration. We have assumed here that the conversion and deposition rates 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 we actually have

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

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. Temporarily ignoring the dispersion factor, we assume that 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 trasnport 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, we get

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

From this relation we see 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

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

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, we have

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

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.

Recall the convention that i^* refers to the parent species corresponding to the species i.

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 , we have the mass balance equation

$$C_{ik} = C_{i1k} + C_{i2k} + \ldots + C_{ink} \tag{13}$$

for each sampling period $k = 1, 2, \ldots, s$. From this basic equation we now derive various special cases.

(a) DMB model.

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^* as usual. For example, if we were interested in SO4 , then i would stand for SO4 and i^* would stand for SO2. We are then interested in the quantity C_{ijk} for each of the sampling periods. We have, from equation (2) that

$$C_{ijk} = c_{ijk} r_{ijk} d_{jk} + c_{i^*jk} r^*_{ijk} d_{jk} \tag{14}$$

If i represents a primary species, then r^*_{ijk} is zero for all k. If i represents a secondary aerosol species, then the quantity c_{ijk} is assumed to be zero for all k, i.e., the secondary aerosol species of interest is not present at the source and it is only formed by conversion of its parent aerosol during transport from the source to the receptor. Therefore, the above equation simplifies to

$$C_{ijk} = c_{ijk} r_{ijk} d_{jk} \tag{15}$$

when i is a primary species and

$$C_{ijk} = c_{i^*jk} r^*_{ijk} d_{jk} \tag{16}$$

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. It is assumed to be a a primary aerosol species denoted by the subscript i_0. We then have

$$C_{i_0jk} = c_{i_0jk} r_{i_0jk} d_{jk}. \tag{17}$$

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

$$\frac{C_{ijk}}{C_{i_0jk}} = \frac{c_{ijk}}{c_{i_0jk}} \frac{r_{ijk}}{r_{i_0jk}} \tag{18}$$

when species i is a primary aerosol and

$$\frac{C_{ijk}}{C_{i_0jk}} = \frac{c_{i*jk}}{c_{i_0jk}} \frac{r^*_{ijk}}{r_{i_0jk}} \tag{19}$$

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

$$C_{ijk} = \frac{c_{ijk}}{c_{i_0jk}} \frac{r_{ijk}}{r_{i_0jk}} C_{i_0jk} \tag{20}$$

for primary aerosols i and

$$C_{ijk} = \frac{c_{i*jk}}{c_{i_0jk}} \frac{r^*_{ijk}}{r_{i_0jk}} C_{i_0jk} \tag{21}$$

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 be mathematically calculated. However, not all the required quantities are known with certainty. The measured quantities will have uncertainties associated with them. This calls for statistical methods of estimation and propagation of uncertainties. A detailed discussion of these issues can be found in a separate appendix dealing specifically with the application of the DMB model for the WHITEX.

(b) CMB Model.

Suppose our list of aerosol components includes only the trace elements. In this case we can assume that $K_c(i,j,k)$ 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. We then obtain,

$$\begin{aligned} \frac{C_{ijk}}{C_{i'jk}} &= \frac{c_{ijk} r_{ijk} d_{jk}}{c_{i'jk} r_{i'jk} d_{jk}} \\ &= \frac{c_{ijk}}{c_{i'jk}} \end{aligned} \tag{22}$$

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

We 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

$$a_{ijk} = \frac{C_{ijk}}{S_{jk}} \tag{23}$$

is then the fraction of the aerosol mixture at the receptor attributable to source j that is species i, during sampling period k. Thus the numbers a_{ijk} for $i = 1, 2, \ldots, m$, represent the source signature for source j for sampling period k. If the a_{ijk} for all the sources affecting the receptor site are known then we have the following set of simultaneous equations

$$C_{ik} = \sum_{i=1}^{n} a_{ijk} S_{jk} \tag{24}$$

in n unknowns $S_{1k}, S_{2k}, \ldots, S_{nk}$, for each of the sampling periods $k = 1, 2, \ldots, s$. If the rank of the above system of linear equations is n for each sampling period k, then these equations can be solved to determine the quantities S_{jk} which in turn will lead to the apportionment of each of the m aerosol components being considered. In practice however, neither the quantities a_{ijk} nor the ambient concentrations C_{ik} will be known without error. This necessitates the use of appropriate statistical methods for estimation of the fractional contributions of the various sources to the ambient concentrations of the aerosol components under consideration as well as to assign appropriate uncertainties to the estimates. Details of the actual application of this approach to the analysis of WHITEX data can be found in a separate appendix dealing with CMB modeling.

(c) TMBR Model.

In this section we show that under appropriate assumptions we can reduce the general mass balance model to a simpler linear model. We 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. Furthermore, the tracers are all primary aerosol components.

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

$$\begin{aligned} C_{1jk} &= r^*_{1jk} d_{jk} c_{2jk} \\ &= \frac{r^*_{1jk}}{r_{i_u jk}} \frac{c_{2jk}}{c_{i_u jk}} C_{i_u jk} \\ &= \beta_{i_u jk} C_{i_u jk} \end{aligned} \tag{25}$$

Therefore,

$$\sum_{j=n_{u-1}+1}^{n_u} C_{1jk} = \sum_{j=n_{u-1}+1}^{n_u} \beta_{i_u jk} C_{i_u jk} = \beta_{i_u k} C_{i_u k} \tag{26}$$

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

$$\beta_{i_u k} = \frac{\sum_{j=n_{u-1}+1}^{n_u} \beta_{i_u jk} C_{i_u jk}}{C_{i_u k}} \tag{27}$$

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

$$\beta_{0k} = \sum_{j=n_h+1}^{n} C_{1jk}. \tag{28}$$

The general mass balance equation then reduces to the equation

$$C_{1k} = \beta_{0k} + \sum_{u=1}^{h} \beta_{i_u k} C_{i_u k} \tag{29}$$

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

If the quantities $\beta_{i_u k}$ are all independent of k for each u, in which case we drop the subscript k and just write β_{i_u}, the above set of equations reduce to the set of equations

$$C_{1k} = \beta_0 + \sum_{u=1}^{h} \beta_{i_u} C_{i_u k} \tag{30}$$

The quantities $C_{i_u k}$ are ambient concentrations of the tracer species i_1, i_2, ..., 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 β_0, β_{i_1}, β_{i_2}, ..., β_{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

$$C_{1k} = \gamma_0 + \sum_{u=1}^{h} \gamma_{i_u} C_{i_u k} \phi_{jk} \tag{31}$$

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 (29).

In practice, however, none of the species concentrations at the receptor are known without error. This makes it necessary to use statistical techniques to estimate the beta or the gamma coefficients and to propagate the uncertainties in the measured values to obtain the uncertainties in the final apportionment fractions. Details of the TMBR modeling application to the WHITEX data can be found in a separate appendix.

(d) 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

$$C_{1k} = \beta_{0k} + \beta_{i_1 k} C_{i_1 k} \tag{32}$$

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

$$C_{1k} = \beta_0 + \beta_{i_1} C_{i_1 k} \tag{33}$$

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

$$C_{1k} = \gamma_0 + \gamma_{i_1} C_{i_1 k} \phi_{1k} \tag{34}$$

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.

In practice none of the aerosol component concentrations at the receptor are known without error. This necessitates the use of statistical techniques for estimating the beta coefficients and to assign appropriate uncertainties to the final results. The details of the statistical procedures and the use of TMB modeling for the WHITEX data are discussed in a separate appendix dealing with TMBR and TMB modeling.

Remarks

Model Assumptions, Potential deviations from assumptions and possible consequences of such deviations and uncertainty calculations are all dependent upon the particular special case of the GMB model that is being considered in an application. Consequently, these issues are discussed in detail separately in appendices dealing with the special case models.