Atmospheric Environment Vol. 26A, No. 9, pp. 1701–1711, 1992.
Printed in Great Britain.

0004-6981/92 \$5.00+0.00

THE APPLICATION OF THREE-MODE FACTOR ANALYSIS (TMFA) TO RECEPTOR MODELING OF SCENES PARTICLE DATA

YOUSHENG ZENG* and PHILIP K. HOPKE†
Department of Chemistry, Clarkson University, Potsdam, NY 13699-5810, U.S.A.

(*First received* 16 *December* 1990 *and in final form* 5 *November* 1991)

Abstract—In the previous applications of eigenvector mathematical methods such as factor analysis, principal components analysis, and empirical orthogonal function analysis, the analysis has been made on a two-dimensional set of data. These data sets could be the chemical composition of a series of particle samples taken at a single location over time or the concentration of a single species measured over multiple locations at multiple times. However, there have not been methods previously available to examine a data set of chemical compositions measured at multiple sites over a series of sampling time intervals. Three-mode factor analysis permits the reduction of a three-dimensional data set into three two-dimensional matrices and a three-dimensional core matrix that presents how the system variance is partitioned among the three modes (chemical species, location and time). The technique will be illustrated with data from the SCENES program that is measuring particle compositions at a number of sites in the southwestern United States.

Key word index: Source apportionment, three-mode matrix, empirical orthogonal function.

INTRODUCTION

Dispersion modeling and receptor modeling are two types of techniques used in air quality management. Dispersion models have been used in this field for a long time. For dispersion modeling, it is necessary to obtain a survey of emissions over the studied area, collection of meteorological data, field tests to determine the dispersion parameters, mechanistic studies on the chemical transformation and scavenging between phases, as well as the studies of the models themselves. There are a number of difficulties associated with these steps in dispersion modeling (Budiansky, 1980; Gordon, 1980, 1988). Budiansky (1980) has discussed the problems of dispersion models, for example, the error in emission inventories, the uncertainty in the horizontal and vertical dispersion parameters, low availability of complete information and high cost for complex models, etc. In addition, dispersion models can only handle the sources that are known to contribute to the system. It cannot include the sources that are unknown or where lack of emission data makes a source unmodelable.

In the last two decades, receptor modeling has played a larger and larger role in air quality management. Receptor models, in contrast to dispersion models, utilize the data collected at receptor sites to elucidate information on properties and influence of emission sources. In typical receptor models, ambient samples are analysed, yielding their chemical compositions. These data can then be analysed so that the sources can be identified and their contributions to the collected sample mass can be estimated. This procedure is also called source apportionment. The estimates of the source contributions for an airshed are useful information in developing air pollution control strategies.

A variety of receptor models have been developed. There have been several major review articles (Gordon, 1980, 1988; Cooper and Watson, 1980), a number of symposium books (Macias and Hopke, 1981; Dattner and Hopke, 1983; Pace, 1986; Watson, 1989), and two comprehensive books (Hopke, 1985, 1991) reviewing and describing the principles and applications of these models.

Air pollution systems are typically characterized by taking samples over time at a number of sampling sites. These samples are then characterized by a variety of analytical methods so that the data set can be represented by a data block as illustrated in Fig. 1. We can use the terms "three-way data table", "three-way matrix" or "three-mode matrix" to refer to this kind of data set. Such a data set provides information on the simultaneous temporal and spatial variations in the system. However, conventional (two-mode) components analysis methods are really designed to analyse a data rectangle obtained by taking a slice of the block in one of the three orthogonal directions. For factor and principal component analysis, this slice is typically the measured concentrations at a single location over time. The empirical orthogonal function (EOF) approach analyses a single species concentration over sites and time (Peterson, 1970). One can also perform a spatial analysis of elemental concentrations over

* Present address: National Environmental Technologies Inc., One Pine Brook Plaza, 9101 Southern Pine Blvd, Suite 280, Charlotte, NC 28217, U.S.A.
† To whom correspondence should be addressed.

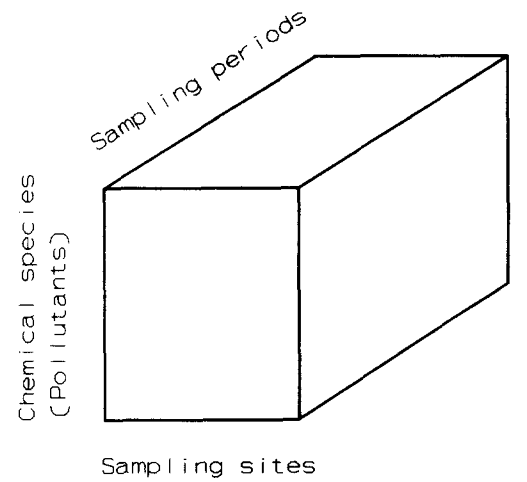

Fig. 1. Typical structure of the data gathered for air quality management.

sites for a single time interval. However, each of these methods is only looking at a single plane in the data block.

There have been several applications of three-way principal components analysis to meteorological data. These data included barometric pressure, surface temperature and precipitation (Kutzbach, 1967) and sea surface temperature, sea level and wind field data (Barnett, 1981). The technique used in these studies is presented in detail by Preisendorfer (1988). However, these methods are designed to decompose the data into three two-way matrices that really do not provide a direct method for examining the partitioning of the system variance among the three different dimensions of the space. Thus, an alternative method that can provide the three-dimensional partitioning of the system variance may be helpful in identifying the processes that control the observed airborne particle compositions at a series of sites over time.

This other three-way factor analysis method is called three-mode factor analysis (TMFA) and was originally developed by Tucker (1963, 1964, 1966) and Kroonenberg (1983) to interpret educational psychological data. Since then, many applications have appeared in the social science literature (Kroonenberg, 1983). In these applications, there was no underlying physical model for the studied systems. Therefore, the interpretation of the results has only been presented in a statistical sense. Because of a lack of an underlying physical model, application of TMFA in natural sciences has been very limited. Only a few applications are found in the literature (Hohn, 1979; Hohn and Friberg, 1979; De Ligny *et al.*, 1984; Spanjer *et al.*, 1985). The model in these applications "is not a physical, but a mathematical-statistical model" (De Ligny *et al.*, 1984).

In very recent years, analysis of multi-way data became more popular in chemistry. Geladi (1989) provided a systematic overview of this area. Several three-way analysis methods have been used in analytical chemistry and TMFA is one of them. As Geladi (1989) commented on these methods, "many publications in the scientific literature only go as far as the example problem". This problem can be called GLOGA (General Lack Of Good Applications) problem. GLOGA and the failure to understand multimode mathematics hinder the common use and acceptance of multi-way methods of data analysis. TFMA is still not a clear model that natural scientists can meaningfully relate to their physical systems. However, it is an attractive method because it opens a new dimension to extract potentially more information from available data. Since this method works directly on the full three-way data matrix as its input, and thus can examine variations among all three modes simultaneously, it appears to be a technique that may be particularly useful in interpreting airborne particle compositional data sets, and thus, appears to warrant further investigation.

In order to explore the application of TMFA to three-way air quality monitoring data and examine its interpretation in terms of the physical and chemical processes operating within an airshed system, methodological studies on simulated data with a known, well-defined structure have been performed (Zeng and Hopke, 1990). The studies were then extended to real data and the results are presented in this report.

THEORY OF THREE-MODE FACTOR ANALYSIS

The basic concept of TFMA is an extension of two-mode (conventional) factor analysis model to three-way data. Each mode corresponds to a class of variables. The term "mode" is used to mean a "set of indices by which data might be classified" (Tucker, 1966). Through TMFA, the data block is decomposed into three two-mode matrices (called factor weight matrices) and one three-mode matrix (called the core matrix). Figure 2 schematically presents this decomposition process. Similar to conventional factor analysis, most of the variations of measured variables are compressed into a few factors according to the covariance among the variables. The results can determine: (1) how many underlying causal factors are controlling the system, (2) what relationships exist between the factors and the variables, and (3) how much of the system variance is accounted for by the factors.

The TMFA model is presented in terms of elements of the factor weight matrices, **A**, **B** and **C**, and the core matrix, $\mathbf{G}_{npq}$. The nomenclature for the complex mathematics of this method is described in the Appendix:

$$\mathbf{x}_{ijk}=\sum_{n=1}^{N}\sum_{p=1}^{P}\sum_{q=1}^{Q}\mathbf{a}_{in}\mathbf{b}_{jp}\mathbf{c}_{kq}\mathbf{g}_{npq}. \tag{1}$$

This model is often referred to as the Tucker-3 model. It can also be expressed in matrix form as a Kronecker product (Pease, 1965):

$$_i\mathbf{X}_{(jk)}={}_i\mathbf{A}_n\mathbf{G}_{(pq)}({}_p\mathbf{B}_j\odot{}_q\mathbf{C}_k). \tag{2}$$

The notational device that Tucker (1966) employed has been used here. The subscript (e.g. i) is used in several related, but distinct roles: (1) as a general identification of the mode, (2) as a subscript identifying the mode to which an element belongs, and (3) as a variable identification symbol for the elements in the mode. The pre-subscript letter denotes the row mode while the post-subscript letter is the designation

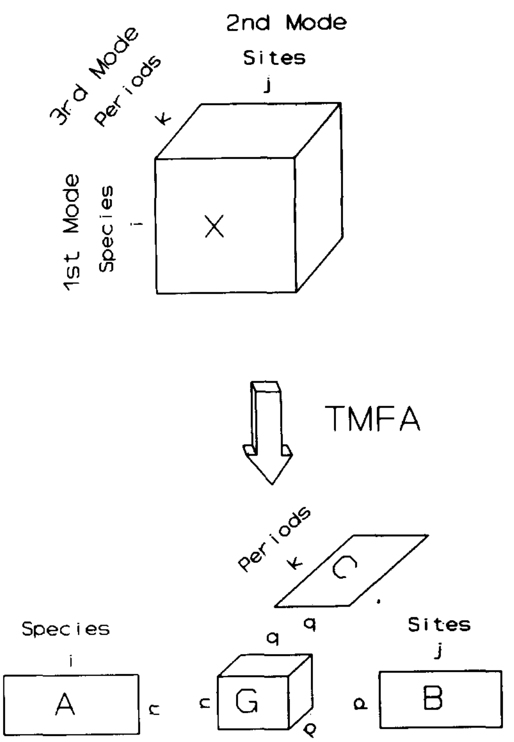

Fig. 2. Outline of three-mode factor analysis model.

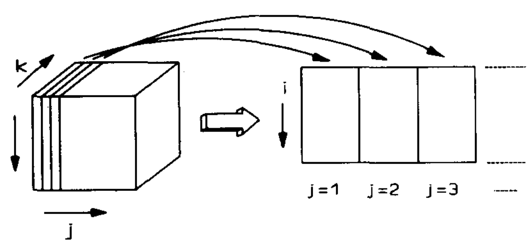

Fig. 3. Unfolding a three-way matrix to a two-way matrix.

for the column mode. Reversal of the subscripts indicates transposition, for example, ${}_i\mathbf{A}_n$ is the transpose of ${}_n\mathbf{A}_i$. The matrix ${}_i\mathbf{X}_{(jk)}$ is the three-mode data matrix rearranged as a two-mode matrix by sequentially placing all of the vertical planes that are parallel to the side planes of the three-mode matrix, side by side along the row direction (Fig. 3). The designation (jk) is called combination mode. The order (jk) may be read as "j-outer loop, k-inner loop".

The Kronecker product of ${}_p\mathbf{B}_j$ and ${}_q\mathbf{C}_k$, denoted as ${}_p\mathbf{B}_j \odot {}_q\mathbf{C}_k$, yields ${}_{(pq)}\mathbf{H}_{(jk)}$. Then ${}_{(pq)}\mathbf{H}_{(jk)}$ can be represented as below as a supermatrix containing submatrices proportional to the matrix ${}_q\mathbf{C}_k$:

$$ {}_{(pq)}\mathbf{H}_{(jk)} = {}_p\mathbf{B}_j \odot {}_q\mathbf{C}_k = \begin{vmatrix} (b_{11\,q}\mathbf{C}_k) & (b_{12\,q}\mathbf{C}_k) & \cdots \\ (b_{21\,q}\mathbf{C}_k) & (b_{22\,q}\mathbf{C}_k) & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{vmatrix}. \quad (3)$$

Using these definitions and notation, the three-mode problem is reduced to a two-mode problem as in Fig. 3. Tucker (1966) showed that the matrices ${}_i\mathbf{A}_n$, ${}_j\mathbf{B}_p$ and ${}_k\mathbf{C}_q$ in Equation (2) could be obtained from the matrices ${}_i\mathbf{M}_i$, ${}_j\mathbf{P}_j$ and ${}_k\mathbf{Q}_k$. These latter matrices are analogous to covariance matrices in conventional factor analysis.

$$\begin{aligned} {}_i\mathbf{M}_i &= {}_i\mathbf{X}_{(jk)}\mathbf{X}_i \\ {}_j\mathbf{P}_j &= {}_j\mathbf{X}_{(ik)}\mathbf{X}_j \\ {}_k\mathbf{Q}_k &= {}_k\mathbf{X}_{(ij)}\mathbf{X}_k. \end{aligned} \quad (4)$$

After the eigenvalues and eigenvectors of each of the covariance matrices are computed, the factor weight matrices ${}_i\mathbf{A}_n$, ${}_j\mathbf{B}_p$ and ${}_k\mathbf{C}_q$ can be obtained in this way: N significant eigenvectors from ${}_i\mathbf{M}_i$ constitute ${}_i\mathbf{A}_n$; P significant eigenvectors from ${}_j\mathbf{P}_j$ constitute ${}_j\mathbf{B}_p$; and Q significant eigenvectors from ${}_k\mathbf{Q}_k$ constitute ${}_k\mathbf{C}_q$.

The core matrix ${}_n\mathbf{C}_{(pq)}$ is given by

$$ {}_n\mathbf{G}_{(pq)} = {}_n\mathbf{A}_i^{+}\mathbf{X}_{(jk)}({}_j\mathbf{B}_p^{+} \odot {}_k\mathbf{C}_q^{+}), \quad (5)$$

where

$$\begin{aligned} {}_n\mathbf{A}_i^{+} &= ({}_n\mathbf{A}_i\mathbf{A}_n)^{-1}{}_n\mathbf{A}_i \\ {}_p\mathbf{B}_j^{+} &= ({}_p\mathbf{B}_j\mathbf{B}_p)^{-1}{}_p\mathbf{B}_j \\ {}_q\mathbf{C}_k^{+} &= ({}_q\mathbf{C}_k\mathbf{C}_q)^{-1}{}_q\mathbf{C}_k. \end{aligned} \quad (6)$$

Since matrices **A**, **B** and **C** are column-wise orthonormal (a property of the eigenvectors):

$$ {}_n\mathbf{A}_i\mathbf{A}_n = {}_n\mathbf{I}_n, \quad {}_p\mathbf{B}_j\mathbf{B}_p = {}_p\mathbf{I}_p, \quad {}_q\mathbf{C}_k\mathbf{C}_q = {}_q\mathbf{I}_q. \quad (7)$$

Then

$$ {}_n\mathbf{A}_i^{+} = {}_n\mathbf{A}_i, \quad {}_p\mathbf{B}_j^{+} = {}_p\mathbf{B}_j, \quad {}_q\mathbf{C}_k^{+} = {}_q\mathbf{C}_k. \quad (8)$$

Equation (5) becomes

$$ {}_n\mathbf{G}_{(pq)} = {}_n\mathbf{A}_i\mathbf{X}_{(jk)}({}_j\mathbf{B}_p \odot {}_k\mathbf{C}_q). \quad (9)$$

In principle, all of the system information is contained in the three factor weight matrices (**A**, **B** and **C**) and the core matrix ($\mathbf{G}_{npq}$) in terms of the factors that give rise to the system variation. As in conventional factor analysis, each weight matrix (e.g. ${}_i\mathbf{A}_n$) will show the relationship between the variables (i) and factors (n) in that mode. The squares of the elements in the core matrix are an indication of how much the system variance is accounted for by the corresponding three-way combination of factors (n, p and q) from each mode. This relationship can be seen from the following equation:

$$ \sum_i \sum_j \sum_k x_{ijk}^2 = \sum_n \sum_p \sum_q g_{npq}^2. \quad (10)$$

In conventional factor analysis, a rotation is usually needed before interpreting the results. Tucker (1966) also presented a transformation scheme. Let the matrices ${}_n\mathbf{T}_{n*}$, ${}_p\mathbf{T}_{p*}$ and ${}_q\mathbf{T}_{q*}$ be square, non-singular matrices, and let

$$\begin{aligned} {}_i\mathbf{A}_n\mathbf{T}_{n*} &= {}_i\mathbf{A}_{n*} \\ {}_j\mathbf{B}_p\mathbf{T}_{p*} &= {}_j\mathbf{B}_{p*} \\ {}_k\mathbf{C}_q\mathbf{T}_{q*} &= {}_k\mathbf{C}_{q*}. \end{aligned} \quad (11)$$

The n^*, p^* and q^* stand for transformed modes. For core matrix:

$$ {}_{n*}\mathbf{G}_{(p*q*)} = ({}_n\mathbf{T}_{n*})^{-1}{}_n\mathbf{G}_{(pq)}[({}_{p*}\mathbf{T}_p)^{-1} \odot ({}_{q*}\mathbf{T}_q)^{-1}]. \quad (12)$$

In transformed form, Equation (2) becomes

$$ {}_i\mathbf{X}_{(jk)} = {}_i\mathbf{A}_{n*}\mathbf{G}_{(p*q*)}({}_{p*}\mathbf{B}_j \odot {}_{q*}\mathbf{C}_k). \quad (13)$$

In terms of algorithms, there are several approaches other than the method described above to perform TMFA. Kroonenberg (1983) presented these approaches, particularly, an alternating least squares algorithm (TUCKALS3, see Appendix A in Zeng and Hopke, 1990). The alternating least-squares algorithm was used in these studies.

METHODOLOGICAL STUDIES WITH SIMULATED DATA SETS

In applying TMFA to the receptor modeling, it is desirable to relate not only statistically, but physically, the TMFA model to the actual physical system. In order to examine these relationships and develop the TMFA method as a receptor model, it is extremely useful to employ a simulated data set with a known underlying structure. These studies have been reported by Zeng and Hopke (1990). Studies without rotation have shown some applicability of TMFA in receptor modeling. It could identify sources, classify sites as pollution zones, group periods into primary meteorological regimes, and indicate the contributions of the combinations of factors from three modes. However, some aspects of the results are difficult to interpret.

These results suggest the need for a rotation of the abstract factor matrices to achieve a kind of "simple structure" as in conventional factor analysis (Hopke *et al.*, 1976). The rotated results should be more interpretable. Axis rotation of the initially derived factor solution is a very important step to make the results more interpretable. Varimax rotation without row normalization has been selected as the best rotation method. Good agreement between the TMFA model and the simulated physical system was achieved by this rotation.

The application of TMFA as a receptor model was tentatively established with the analysis of these simulated data sets. It is a qualitative model like conventional two-mode factor analysis, but works with three-way data. In general, the three modes of a data set are associated with measured species, sampling sites and periods. With TMFA, the data set can be decomposed into three two-way factor weight matrices and one three-way core matrix. This decomposition is based on simultaneous variations of causal factors in the system. The factors in the mode corresponding to species can be recognized as pollution sources. Thus, the presence of several elements known to be emitted by a particular source type would permit the identification of that source type as a contributor to the variation observed in the elemental compositions of the collected particle samples. The factors in site mode denote zones in which similar sources or processes dominate the observed elemental variations. The factors in time period mode appear to be related to the meteorological causes of variation. In this way, the whole complicated pollution system can be summarized by a few of "sources", "zones" and "regimes", and their relationships are shown in the core matrix.

APPLICATION OF THREE-MODE ANALYSIS TO SCENES DATA

In this section, TMFA is applied to a real data set from SCENES (McDade and Tombach, 1986) database. SCENES is a large sampling and analysis program that has yielded a large atmospheric environmental database. Its sampling network covers the region where California, Nevada, Arizona and Utah intersect one another (Fig. 4). The data residing in the database include the chemical compositions of two size fractions of particles, visibility measurements, meteorological measurements and gaseous air pollutant measurements. The chemical composition data of particles are suitable to TMFA application. Since the total particulate matter samples ($2.5\ \mu m \leqslant$ aerodynamic dia. $\leqslant 15\ \mu m$) have more chance to be influenced by a number of local sources while the fine particulate samples ($<2.5\ \mu m$ aerodynamic dia.) should have sources distributed over the whole region, the fine particulate matter data are used in this study.

The samples were collected on Teflon and quartz filters and analysed using established methods: X-ray fluorescence (XRF) for trace elements, thermal extraction for carbonaceous materials (organic and elemental carbon), and ion chromatography for ions (sulfate, nitrate and ammonium). The ionic species data are not available at most of the sites. The concentration of some elements are always near the filter blank values. These data are not used in this study. Eighteen species are included in the analysis. They are Al, Si, P, S, Cl, K, Ca, Ti, V, Mn, Fe, Cu, Zn, Br, Cd, Pb, ROC (residual organic carbon) and EC (elemental carbon).

There are 11 observation sites (Table 1). Aerosol samples were collected at seven of them. The site MC (see Fig. 4) has very limited data so that six sites (BY, GC, HI, HP, MV and SM) are included. The data used are for 24-h samples collected every day. These data are available from March 1986 to February 1987. In order to get a complete three-way data set, the sampling periods (dates) were selected such that all six sites have valid data within common time periods. Using this criterion, 46 periods were selected. Therefore, the three-way data matrix is 18 (species) $\times$ 6 (sites) $\times$ 46 (periods). Screening the data found that the Si concentration at site SM on 2 April 1986 is unusually high (2 orders of magnitude higher than the most of the samples). It is considered as an outlier. This sampling period was removed. The final data matrix is then 18 (species) $\times$ 6 (sites) $\times$ 45 (periods).

The data matrix was analysed with each element centered to its mean value and scaled by elemental variance. Thus, each standardized elemental concentration has a zero mean and unit variance (R-mode), i.e. equivalent to standard two-mode factor analysis. Other scaling methods had been examined in our studies of simulated data and it was found that this approach to scaling provided a model that could be readily interpreted in terms of the system under study (Zeng, 1989). The fitting and residual data (Table 2) do not show the number of factors in each mode as clearly as in the simulation case. Since three factors give a better fit [$SS(Fit_3) = 84\%$] than two factors

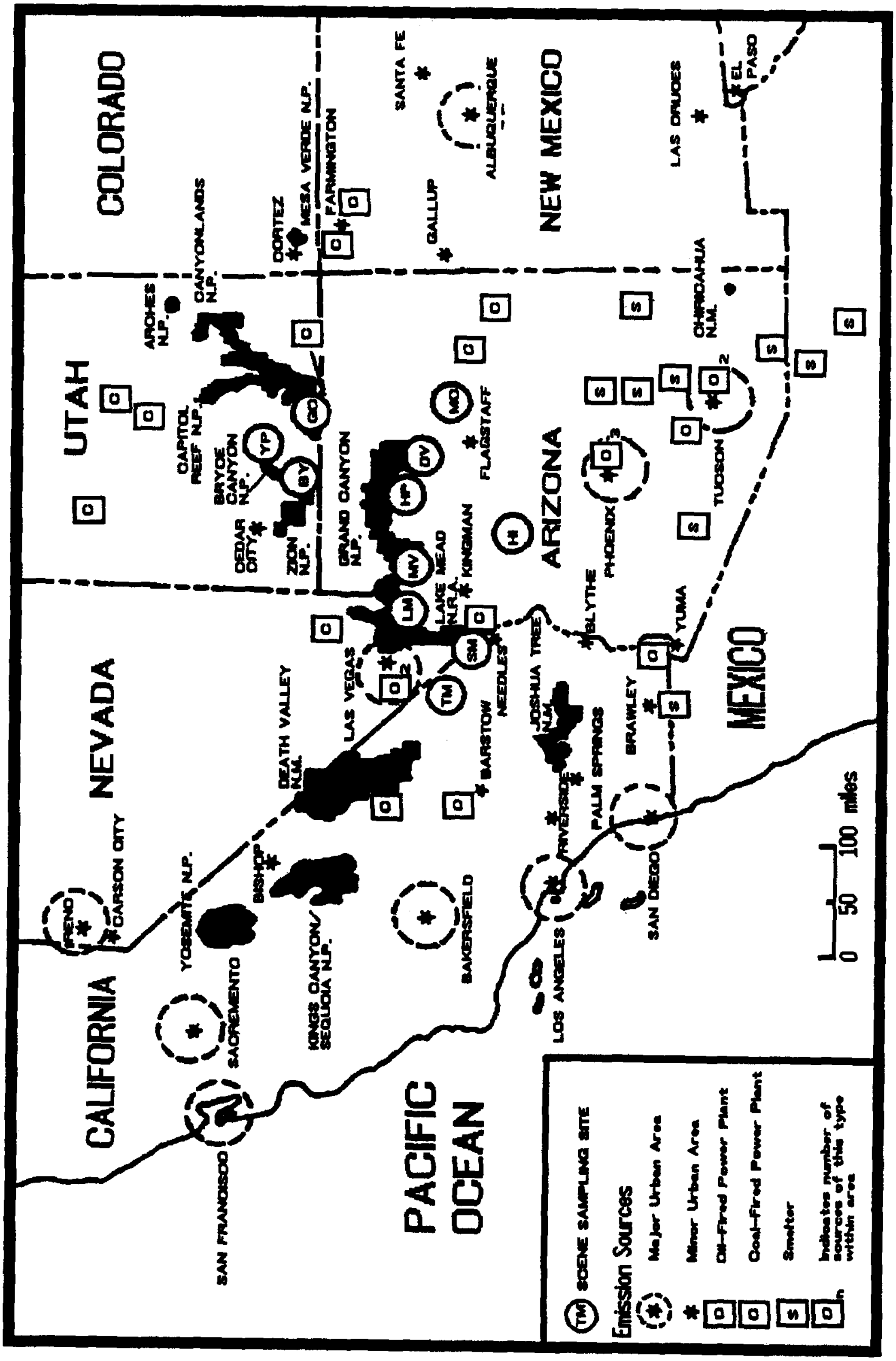

Fig. 4. Air quality and visibility observation sites in the SCENES area.

Table 1. SCENES observatories and types of aerosol samples collected

Site designation	Map description	Aerosol*
HI	Hillside, AZ	TP, FP
MC	Meteor Crater, AZ	TP, FP
TM	Turquoise Mountain, CA	—
SM	Spirit Mountain, NV	TP, FP; PM-10 (HV)
MV	Mead View, NV	TP, FP; PM-10 (HV)
LM	Lake Mead, NV	—
HP	Hopi Point, AZ	TP, FP; CP, FP (SFU)
DV	Desert View, AZ	—
GC	Glen Canyon, AZ	TP, FP; CP, FP
BY	Bryce Canyon, UT	TP, FP; CP, FP (SFU)
YP	Yovimpa Point, UT	—

*TP = Total (< 15 μm AD) particulate matter, FP = fine (< 2.5 μm AD) particulate matter, CF = coarse (> 2.5 μm AD) particulate matter, SFU = stacked filter units.

Table 2. Sum of squares of fitting (SS) in three modes of the SCENES data set

Test for mode 1		Test for mode 2		Test for mode 3	
N	SS(Fit_1)	P	SS(Fit_2)	Q	SS(Fit_3)
1	0.44	1	0.56	1	0.39
2	0.60	2	0.72	2	0.48
3	0.68	3	0.84	3	0.59
4	0.74	4	0.90	4	0.65
5	0.79	5	0.96	5	0.70
6	0.84	6	1.00	6	0.74
7	0.87			7	0.77

and reflect the main feature of the sites (see discussion below), three factors are retained in mode 2. By examining the explained variance (rotated core matrix) with different numbers of factors, four and five factors are retained for mode 1 and mode 3, respectively. The SS(Fit) is 63% and SS(Residual) = 37%. Thus, unexplained system variance in this analysis is 37% of total system variance (also see discussion and Fig. 5 later in this section). The unexplained portion consists of possible unique factors (Harman, 1976; Hopke, 1985) and noise.

Table 3 presents the Varimax rotated factor weights for mode 1 of the SCENES fine particulate data set. The factor 1 has large weights of crustal elements (Al, Si, Ca, Ti, Mn and Fe). It is identified as soil. The weights look smaller than factor loadings in usual two-mode factor analysis, but they indicate the existence of strong relationships. The interpretation of the TMFA results is primarily based on (1) the factor weight (e.g. a_{in}) being the fraction of the factor (n) on the variable (i); (2) the square of an element in the core matrix, g_{npq}^2, being a measure of the system variance accounted for by the three-way combination of factors n, p and q. The interpretation of three-mode factor weights is similar to, but not exactly the same as that of two-mode factor loadings. In two-mode factor analysis, factor loadings are derived from the eigenvectors by

$$a_{in} = v_{in} \cdot \sqrt{\lambda_n}, \qquad (14)$$

Table 3. Varimax rotated factor weights of mode 1 of the SCENES data set

i	Species	Factor 1 (source 1)	Factor 2 (source 2)	Factor 3 (source 3)	Factor 4 (source 4)
1	Al	0.390	0.013	0.004	0.018
2	Si	0.400	−0.028	−0.033	0.009
3	P	0.034	0.470	0.088	0.040
4	S	−0.044	0.519	0.114	0.077
5	Cl	−0.063	−0.080	−0.713	0.042
6	K	0.099	0.046	−0.570	0.043
7	Ca	0.364	−0.059	−0.012	0.058
8	Ti	0.398	−0.021	0.002	0.012
9	V	0.249	0.125	0.013	−0.025
10	Mn	0.372	0.023	0.130	−0.050
11	Fe	0.401	−0.030	−0.018	0.014
12	Cu	−0.006	−0.026	0.041	−0.883
13	Zn	−0.050	0.258	0.052	0.175
14	Br	0.124	0.242	−0.180	−0.026
15	Cd	0.012	−0.042	0.084	−0.184
16	Pb	−0.030	0.411	0.048	0.007
17	ROC	−0.014	0.351	−0.220	−0.099
18	EC	0.026	0.244	−0.185	−0.357

where v_{in} = the ith element of the nth eigenvector;
λ_n = the nth eigenvalue.

It can be shown (Harman, 1976) that this two-way factor loading a_{in} (or the one that has been subjected to an orthogonal rotation) is the correlation coefficient between variable i and factor n. In TMFA, however, the factor weight is simply the eigenvector itself, i.e.

$$a_{in} = v_{in}. \quad (15)$$

Therefore, the weight is not the correlation coefficient although it still shows the relationship between the factor n and the variable i (Kroonenberg, 1983).

Another point needs to be noted. Since the weights in TMFA are simply the eigenvectors, their sum of the squared values for each factor is equal to 1. Therefore, the large loadings commonly observed in the two-mode case (e.g. values of 0.8 or 0.9) should not be expected if three or more variables are associated with the factor. The weights need to be examined in a relative sense rather than as absolute correlation coefficients with possible maximum values of 1.

Factor 2 is associated with S, P, Pb, ROC, EC, Zn and Br. This factor is attributed to be a general urban area source, mainly autoemission along with the SO_2 converted to sulfate by photochemical oxidants in the transport process. The weight of P is likely due to the peak P X-ray being interfered with by the large S peak in XRF analysis. Factor 3 is assigned to vegetative combustion aerosol since it contains Cl, K, ROC and EC. It is not clear why Br would also be associated with this factor. Factor 4 is a Cu source with a medium weight for EC. It could be the copper smelters in southern Arizona or the contamination from the motor brushes of the sampler (Gordon, 1980) if care has not been taken to vent the pump exhaust well away from the samplers. Since the samplers used in the SCENES program were protected from contamination of the motor brushes, this factor is considered to be the copper smelters.

The results are compared with that of two-mode factor analyses individually performed for each site (each lateral slice of the three-way matrix). The two-mode analysis results are summarized in Table 4. All four sources discussed above are also identified by the two-mode analyses. In two-mode analysis, more small sources show up and some of them seem to be from breaking down of the sources considered as the urban source in TMFA. In general, the results of two-mode and three-mode agree well, but TMFA covers more common variations in the whole region and two-mode analyses reveal more details of the local variations.

Table 5 gives factor weights of mode 2 of the three-way data set. It is quite clear that site HI forms a zone by itself. This site is about 100 miles south of the Grand Canyon National Park and about 100 miles northwest of the Phoenix urban area (see Fig. 4). Another site SM near the edge of the park and reservation area in the southern California side, represents another zone that is polluted differently. The other four sites (BY, HP, GC and MV) belong to a relatively cleaner zone. The sites MV and GC also have small weights on factor 1, which is a pollution zone represented by site SM. From Fig. 4, it can be seen that MV is close to SM and there is a coal-fired power plant near GC. These weights reflect the configuration of the sources and the sites. If one more factor is taken in this mode, sites MV and GC are separated from BY and HP (Table 5). However, it does not seem necessary to separate them in this six-variable mode. The sites are then grouped into three zones.

The factor weights of mode 3 are presented in Table 6 along with the corresponding sampling periods. The

Table 4. Summary of two-mode factor analysis for the SCENES data set

Site	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
BY	Si, Al, Fe, Ti, Ca, K, (soil)	S, P, Br, Pb, ROC, EC	Zn, Cu	Cd, Pb	Cl, K, EC, ROC (marine)	V, EC, Mn		
GC	Si, Al, Fe, Ti, Ca, K (soil)	EC, ROC, Zn, Br	Cu	Cl, K (marine)	Pb	S, P	Cd	V
HI	Si, Al, Fe, Ti, Ca (soil)	Cu, Pb, Zn, S, P, ROC, EC	Br, ROC, EC, K	Cd	V	Cl, K, EC (marine)	S, P	
HP	Si, Al, Fe, Ti, Ca, K (soil)	S, P, Pb	Cd	Zn	Cl, K (marine)	ROC, EC, Br	V	
MV	Si, Al, Fe, Ti, Ca, V, K (soil)	EC, ROC, Zn	Cl, K (marine)	Cd	Cu	S, P		
SM	Si, Al, Fe, Ti, Ca, K, V, Mn (soil)	S, P, Pb, Br	Cd	Cu	Zn	Cl, K, Br (marine)	EC, ROC, Br	

Table 5. Varimax rotated factor weights of mode 2 in the SCENES data set

		If three factors are retained			If four factors are retained			
j	Site	Factor 1 (zone 1)	Factor 2 (zone 2)	Factor 3 (zone 3)	Factor 1	Factor 2	Factor 3	Factor 4
1	BY	− 0.162	0.111	0.714	− 0.071	− 0.093	0.756	0.030
2	GC	0.230	− 0.022	0.348	− 0.053	− 0.019	− 0.010	0.756
3	HI	− 0.019	− 0.982	− 0.021	− 0.010	0.988	− 0.008	0.001
4	HP	− 0.003	− 0.133	0.511	0.089	0.118	0.654	− 0.043
5	MV	0.305	− 0.069	0.320	0.074	0.033	0.020	0.653
6	SM	0.910	0.027	− 0.067	0.989	− 0.010	− 0.006	− 0.003

most significant weights are in boldface. Since there are 45 periods, the weights are relatively small. Weights larger than 0.3 may be considered to be significant. The interpretation of this mode is difficult because of lack of related information. Each factor represents a particular regime. The regimes could be characterized by meteorological conditions (wind direction and speed, temperature, inversion, etc), or emission events. These conditions or events can be related to the regime (the factor) by the sampling periods. For

Table 6. Varimax rotated factor weights of mode 3 in the SCENES data set

k	Sampling date	Factor 1 (regime 1)	Factor 2 (regime 2)	Factor 3 (regime 3)	Factor 4 (regime 4)	Factor 5 (regime 5)
1	14/04/86	0.029	0.280	0.009	0.110	0.091
2	17/04/86	0.109	0.233	− 0.030	− 0.046	0.015
3	20/04/86	0.124	0.144	0.088	− 0.014	− 0.086
4	23/04/86	0.031	− 0.025	− 0.002	0.078	0.067
5	05/05/86	0.276	0.239	0.007	− 0.050	− 0.022
6	08/05/86	0.041	0.274	0.057	− 0.040	0.130
7	14/05/86	0.146	− 0.157	0.007	0.108	0.065
8	17/05/86	0.010	− 0.065	0.078	− 0.022	− 0.008
9	20/05/86	0.033	− 0.233	0.012	**0.325**	0.134
10	23/05/86	**0.395**	− 0.007	− 0.008	0.118	0.013
11	29/05/86	**0.306**	− 0.015	− 0.075	**0.319**	0.293
12	07/06/86	− 0.055	0.014	0.019	**0.810**	− 0.031
13	04/07/86	0.037	− **0.311**	0.038	0.011	− 0.126
14	07/07/86	− 0.003	0.184	0.081	− 0.007	0.012
15	10/07/86	0.041	− 0.048	0.077	0.076	− 0.050
16	13/07/86	**0.591**	0.069	− 0.008	− 0.111	− 0.045
17	19/07/86	− 0.010	− 0.073	0.086	0.006	− 0.042
18	18/08/86	− 0.044	− 0.155	0.077	− 0.002	0.001
19	21/08/86	**0.450**	− 0.219	0.159	− 0.129	0.030
20	30/08/86	0.057	− 0.228	0.015	− 0.109	0.109
21	02/09/86	0.001	0.032	0.019	− 0.001	0.162
22	05/09/86	0.021	0.050	0.075	0.011	0.201
23	08/09/86	0.004	− 0.220	− 0.032	− 0.114	0.165
24	11/09/86	− 0.003	0.077	0.032	0.027	0.059
25	14/09/86	0.021	0.004	0.119	0.035	0.043
26	17/09/86	− 0.013	0.089	0.051	0.022	0.079
27	20/09/86	0.016	− 0.050	0.065	0.005	0.031
28	23/09/86	− 0.032	0.234	0.110	− 0.014	0.062
29	26/09/86	− 0.027	− 0.040	0.121	0.002	0.020
30	08/10/86	− 0.035	− 0.071	0.020	− 0.097	0.279
31	29/10/86	0.008	− 0.204	0.016	− 0.055	0.362
32	10/11/86	− 0.052	0.153	0.118	− 0.012	0.121
33	04/11/86	− 0.067	− 0.121	**0.370**	− 0.002	− 0.045
34	07/11/86	0.049	0.006	**0.653**	0.024	− 0.025
35	10/11/86	− 0.014	0.052	**0.457**	0.020	− 0.034
36	13/11/86	− 0.029	0.139	− 0.111	− 0.027	**0.500**
37	16/11/86	− 0.003	− 0.008	0.014	− 0.060	0.305
38	22/11/86	− 0.048	0.269	0.055	0.024	0.119
39	13/12/86	− 0.055	0.092	0.134	− 0.006	0.079
40	16/12/86	− 0.096	− 0.149	0.094	− 0.013	0.110
41	19/12/86	− 0.085	− 0.135	0.028	− 0.053	0.258
42	22/12/86	− 0.072	− 0.025	0.072	− 0.044	0.164
43	25/12/86	− 0.078	0.024	0.107	− 0.014	0.069
44	28/12/86	− 0.079	0.000	0.088	− 0.012	0.074
45	31/12/86	− 0.065	0.102	0.116	− 0.007	0.039

example, factor 1 represents a regime whose typical cases occurred on 13/7/86, 21/8/86, 23/5/86, 29/5/86, etc. A survey of major emission events, meteorological data or air parcel trajectories may lead to the characterization of this regime. It is anticipated that these regimes could be characterized once the related data become available. Presently, they will be treated as virtual regimes to discuss the rest of the results.

The core matrix is listed in Table 7. The four largest entries in the core matrix are g_{111}, g_{324}, g_{232} and g_{131}. They indicate that the most important events are: (1) pollution by soil (source 1) at zone 1 in the periods of regime 1, (2) pollution by marine (source 3) at zone 2 in regime 4, (3) pollution by urban (source 2) at zone 3 in regime 2, and (4) pollution by soil at zone 3 in regime 1.

The results may be examined zone by zone. In zone 1 (SM site), the largest contributor is the soil source (source 1). The urban (source 2) and Cu (source 4) sources also play significant roles. For zone 2 (HI site), the marine source (source 3) becomes important, particularly in regime 4. Another significant source in this zone is the urban source (in regimes 5 and 2). The soil and urban sources show influence on zone 3 (BY, HP, GC and MV). If the core is examined regime by regime, it can be seen that, in general, the regime 1 represents favorable conditions for the soil source to dominate and regime 2 for dominance by the urban source.

An alternative presentation can be helpful in the interpretation of the core matrix. The square of each entry of the core represents the system variance explained by the corresponding combination. The ratio of the square to the total system variance [SS(Total)] can give the percentage of the variance explained by the combination. Summation of these percentages over two index series (two modes) will give the explained variance by factors in the third mode. The results are shown in Fig. 5.

In terms of source contributions, soil is a major source (Fig. 5). It contributes 30.6% of the system variance. It should be noted that the variance is not equivalent to mass contribution. A source contributing a large amount of system variance does not necessarily contribute a large amount of mass to the sample because the original data matrix is based on species (system variance is based on species) and the composition of the source will affect its contribution to the system variance. The second largest contributor is the

Table 7. Varimax rotated core matrix of SCENES data set

	$p^* = 1$ (zone 1)	$p^* = 2$ (zone 2)	$p^* = 3$ (zone 3)
$q^* = 1$ (regime 1)			
$m^* = 1$ (source 1)	26.63	− 6.80	15.65
$m^* = 2$ (source 2)	8.50	− 2.44	0.19
$m^* = 3$ (source 3)	− 5.40	4.57	− 0.98
$m^* = 4$ (source 4)	− 1.03	− 0.22	1.48
$q^* = 2$ (regime 2)			
$m^* = 1$ (source 1)	− 5.54	0.12	− 7.15
$m^* = 2$ (source 2)	− 12.60	9.40	− 16.38
$m^* = 3$ (source 3)	2.34	− 0.99	1.60
$m^* = 4$ (source 4)	3.02	− 0.35	0.71
$q^* = 3$ (regime 3)			
$m^* = 1$ (source 1)	− 3.06	6.94	− 7.71
$m^* = 2$ (source 2)	− 3.33	5.24	− 7.02
$m^* = 3$ (source 3)	1.26	− 1.90	3.46
$m^* = 4$ (source 4)	− 12.35	− 2.18	2.82
$q^* = 4$ (regime 4)			
$m^* = 1$ (source 1)	9.16	− 0.58	11.83
$m^* = 2$ (source 2)	3.65	− 1.09	0.17
$m^* = 3$ (source 3)	− 3.25	18.11	− 7.61
$m^* = 4$ (source 4)	− 0.82	0.01	0.41
$q^* = 5$ (regime 5)			
$m^* = 1$ (source 1)	− 0.72	5.02	− 6.22
$m^* = 2$ (source 2)	5.08	− 13.06	− 2.48
$m^* = 3$ (source 3)	0.39	1.60	3.23
$m^* = 4$ (source 4)	− 1.68	1.51	1.67

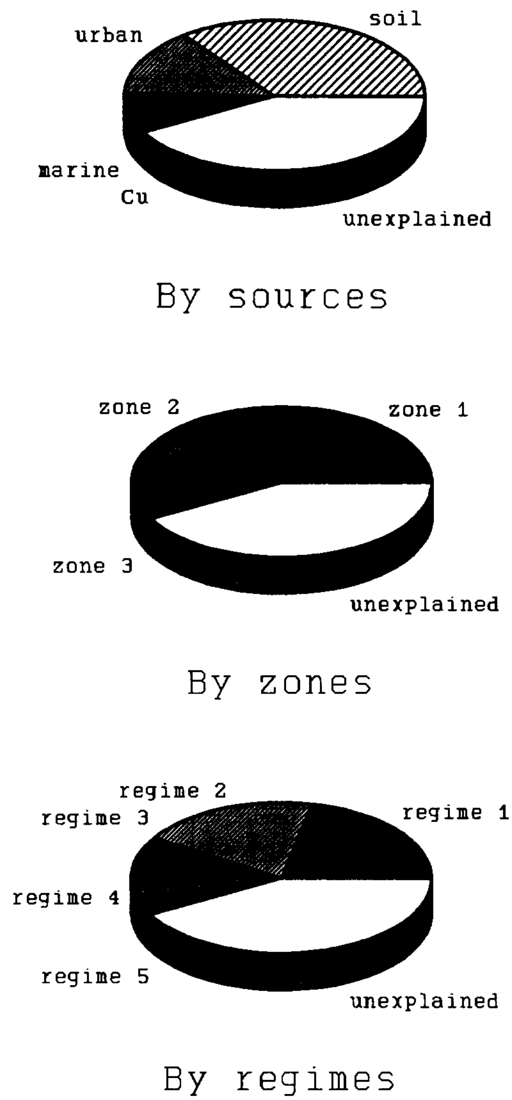

Fig. 5. The system variance explained by the factors in the SCENES data set.

urban source, and the third is the marine source. The Cu source is not important to system variance because it only involves a single element.

The three zones could be compared. Figure 5 indicates that zone 1 receives the largest particle weight (the particles are predominantly soil particles). The other two zones have lower weights. Among the five regimes, regime 1 is the most important one (23.3% variance). Regimes 2 and 4 include about 13% of variance each. Regimes 3 and 5 only contribute 8 and 6%, respectively.

CONCLUSIONS

The TMFA method has been developed for receptor modeling purpose, and its application to a real data set shows good initial results. Like two-mode factor analysis, TMFA is a qualitative method. It can be used to find those factors that cause variations of the observed values in a three-way data set. A large data set can be explained by a few of these principal factors. The method provides not only the information of possible particle sources, but also indications of the interactions between the sources and spatial and temporal domains. This latter function is the main feature of three-mode analysis that two-mode factor analysis does not possess.

In the interpretation of TMFA, three factor weight matrices are used to identify respectively the nature of the factors, and the entries of the core matrix show the explained system variance. It should be noted that the weights are not scaled by the corresponding square root of the eigenvalue, so that their values are usually lower than those commonly observed in two-mode analyses, particularly when more variables are associated with a factor. A Varimax rotation is a necessary step to obtain interpretable results. Another concept that should not be misunderstood is that the explained variance in general is not equivalent to source contribution to the ambient aerosol mass.

In the application of TMFA to the SCENES data set, four sources were identified. They are soil, urban dust, vegetative combustion aerosol and a Cu source. The first three sources are major sources. The region of interest is divided into three zones according to the manner in which the various sources contributed to the observed aerosol compositions. Zone 1 (SM), which is closer to southern California, is highly polluted compared with the other two zones. Five factors were extracted from the third mode. However, these factors were not characterized because of lack of related information (meteorology or emission events). Thus, the interpretation was not complete.

Acknowledgements—This work was supported in part by the National Science Foundation under grant ATM 89-96203 and the Ministry of the Environment of the Province of Ontario under grant 433G. The views and ideas expressed in this report are those of the authors and do not necessary reflect the views and policies of the Ministry of the Environment nor does the mention of trade names or commercial products constitute an endorsement or recommendation for use.

REFERENCES

Barnett T. P. (1981) Statistical relations between ocean/atmosphere fluctuations in the tropical Pacific. *J. phys. Oceanogr.* **11,** 1043–1058.

Budiansky S. (1980) Dispersion modeling. *Envir. Sci. Technol.* **14,** 370–374.

Cooper J. A. and Watson J. G. (1980) Receptor-oriented methods of air particulate source apportionment. *J. Air Pollut. Control Ass.* **30,** 1116–1125.

Dattner S. L. and Hopke P. K. (1983) *Receptor Models Applied to Contemporary Pollution Problems,* Specialty Conf. SP-48, Air Pollution Control Association, Pittsburgh, PA.

De Ligny C. L., Spanjer M. C., Van Houwelingen J. C. and Weesie H. M. (1984) Three-mode factor analysis of data on retention in normal-phase high-performance liquid chromatography. *J. Chromat.* **301,** 311–324.

Geladi P. (1989) Analysis of multi-way (multi-mode) data. *Chemometrics Intell. Lab. Syst.* **7,** 11–30.

Gordon G. E. (1980) Receptor models. *Envir. Sci. Technol.* **14,** 792–800.

Gordon G. E. (1988) Receptor models. *Envir. Sci. Technol.* **22,** 1132–1142.

Harman H. H. (1976) *Modern Factor Analysis,* 3rd edn. University of Chicago Press, Chicago.

Hohn M. E. (1979) Principal components analysis of three-way table. *J. Int. Ass. Math. Geol.* **11,** 611–626.

Hohn M. E. and Friberg L. M. (1979) A generalized principal components model in petrology. *Lithos* **12,** 317–324.

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

Hopke P. K. (ed) (1991) *Receptor Modeling For Air Quality Management.* Elsevier Science, Amsterdam.

Hopke P. K., Gladney E. S., Gordon G. E., Zoller W. H. and Jones A. G. (1976) The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. *Atmospheric Environment* **10,** 1015–1025.

Kroonenberg P. M. (1983) *Three-Mode Principal Component Analysis.* DSWO Press, Leiden, the Netherlands.

Kutzbach J. E. (1967) Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *J. appl. Met.* **6,** 791–802.

Macias E. S. and Hopke P. K. (1981) *Atmospheric Aerosol: Source/Air Quality Relationships,* Symp. Series No. 167, American Chemical Society, Washington, DC.

McDade C. E. and Tombach I. H. (1986) Goals and initial findings from SCENES. In *Visibility Protection* (edited by Bhardwaja P. S.), pp. 76–86. Publication TR-10, Air Pollution Control Association, Pittsburgh, PA.

Pace T. G. (1986) *Receptor Methods for Source Apportionment.* Publication TR-5, Air Pollution Control Association, Pittsburgh, PA.

Pease M. C. III (1965) *Methods of Matrix Algebra.* Academic Press, New York.

Peterson J. T. (1970) Distribution of sulfur dioxide over metropolitan St. Louis as described by empirical eigenvectors and its relation to meteorological parameters. *Atmospheric Environment* **4,** 501–518.

Preisendorfer R. W. (1988) *Principal Component Analysis in Meteorology and Oceanography.* Elsevier Science, Amsterdam.

Spanjer M. C., De Ligny C. L., Van Houwelingen J. C. and Weesie J. M. (1985) Simultaneous description of the influence of solvent, reaction type, and substituent on equilibrium constants by means of three-mode factor analysis. *J. Chem. Soc. Perkin Trans.* **2,** 1401–1411.

Tucker L. R. (1963) Implications of factor analysis of three-way matrices for measurement of change. In *Problems in Measuring Change* (edited by Harris C. W.), pp. 122–137. University of Wisconsin Press, Madison, WI.

Tucker L. R. (1964) The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology* (edited by Gulliksen H. and Frederisken N.), pp. 109–127. Holt, Rinehart and Winston, New York.

Tucker L. R. (1966) Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**, 279–311.

Watson J. G. (1989) *Receptor Models in Air Resources Management*, Specialty Conf., TP-5, Air and Waste Management Association, Pittsburgh, PA.

Zeng Y. (1989) Methodological studies and applications of several new receptor models for air quality management. Ph.D. thesis, University of Illinois at Urbana-Champaign, Urbana, IL.

Zeng Y. and Hopke P. K. (1990) Methodological study applying three-mode factor analysis to three-way chemical data sets. *Chemometrics Intell. Lab. Syst.* **7**, 237–250.

APPENDIX

NOMENCLATURE AND LIST OF SYMBOLS

Nomenclature

The term "mode" or "way" denotes a set of indices by which data might be classified. A data set with two sets of indices is called two-mode or two-way data, and it can be presented as a regular matrix (i.e. two-mode matrix). Boldface capital letters with subscripts represent generalized matrices including regular two-mode matrices and any multimode (multi-way) matrices. The number of the subscripts reflect the number of modes of the matrices. The subscripts are generally omitted if the matrices are two-mode. For example, both $\mathbf{A}_{ij}$ and $\mathbf{A}$ can be used for same two-mode matrix; $\mathbf{X}_{ijk}$ represents a three-mode matrix. A subscript is used in several related, but distinct roles: (1) as a general identification of the mode, (2) as a subscript identifying the mode to which an element belongs, and (3) as a variable identification symbol for the elements in the mode. As a element subscript, i may have a value from 1 to I (upper case of the same letter as the subscript stands for the upper limit). In some circumstances, a multi-mode matrix can be unfolded to a two-dimensional matrix. The subscripts can be split into a pre-subscript and a post-subscript to reflect this unfolded arrangement. The pre-subscript associates with the row variables and the post-subscript reflects the column variables. For instance, $\mathbf{X}_{ijk}$ is a three-mode matrix, ${}_i\mathbf{X}_{(jk)}$ is the three-mode matrix unfolded as a two-dimensional I by $(J \times K)$ matrix. The transpose of ${}_i\mathbf{X}_{(jk)}$ then can be represented by ${}_{(jk)}\mathbf{X}_i$. This notation can also be used for two-mode matrix so that both $\mathbf{A}^T$ and ${}_j\mathbf{A}_i$ can be used for the transpose of ${}_i\mathbf{A}_j$. $\mathbf{A}^{-1}$ stands for the inverse of $\mathbf{A}$. Boldface lower-case letters represent column vectors, e.g. $\mathbf{x}$, and their transpose for row vectors, $\mathbf{x}^T$. For a matrix $\mathbf{X}$, $\mathbf{x}_j$ is its jth column vector. Similarly, $\mathbf{X}_i$ or $\mathbf{X}_j$ or $\mathbf{X}_k$ could be the ith or jth or kth slice (two-mode matrix) taken from the three-mode matrix $\mathbf{X}_{ijk}$ along the first or the second or the third mode of the matrix. The $\mathbf{x}_i$, $\mathbf{x}_{ij}$ and $\mathbf{x}_{ijk}$ denote the elements of the vector $\mathbf{x}$, the two-mode matrix $\mathbf{X}_{ij}$ and the three-mode matrix $\mathbf{X}_{ijk}$, respectively. The signs $\otimes$ $\searrow\backslash\swarrow$ $\odot$ represent an outer product and a Kronecker product, respectively. The sign {variable} means a collection of the variables (e.g. $\{\mathbf{x}_j\}$ is a collection of vectors $\mathbf{x}_j$). The term "average" or "mean" is used for arithmetic mean.

Symbols

$\mathbf{A}$	factor loading matrix in two-mode analysis; factor loading matrix of mode 1 in three-mode analysis
$\bar{\mathbf{A}}$	rotated factor loading matrix in two-mode analysis
$\mathbf{B}$	factor loading matrix of mode 2 in three-mode analysis
$\mathbf{C}$	factor loading matrix of mode 3 in three-mode analysis
$\mathbf{D}$	observed data matrix standardized to a mean of zero and a variance of one
$\mathbf{F}$	factor score matrix in two-mode analysis
$\bar{\mathbf{F}}$	rotated factor score matrix
$\mathbf{G}_{npq}$	core matrix in three-mode analysis
i	index for mode 1; index for species
I	number of variables in mode 1; number of species
$\mathbf{I}$	identity matrix
j	index for mode 2; index for sampling site
J	number of variables in mode 2; number of sampling sites
k	index for mode 3; index for sampling period
K	number of variables in mode 3; number of sampling periods
mass_j	total mass of sample j
n	index for factors; index for sources
n^*	index for rotated factors in mode 1
N	number of sources; number of factors
p	index for factors in mode 2
p^*	index for rotated factors in mode 2
q	index for factors in mode 3
q^*	index for rotated factors in mode 3
$\mathbf{R}$	correlation matrix
s_n	scaling coefficient for source (factor) n
$\mathbf{T}$	transformation matrix
TMFA	three-mode factor analysis
$\mathbf{U}$	eigenvector matrix (its columns are eigenvectors)
x, $\mathbf{x}$, $\mathbf{X}$, $\mathbf{X}_{ijk}$	observed ambient data in scalar, vector, two-way matrix or three-way matrix form
x_i	ambient concentration of species i
x_{ij}	concentration of species i in the ambient sample collected at site j
x_{ijk}	concentration of species i in the ambient sample collected at site j in period k
$\mathbf{y}$, $\mathbf{Y}$	source profile or source profile matrix
y_{in}	source profile, i.e. concentration of species i in the particles coming from source n
z, $\mathbf{z}$, $\mathbf{Z}$, $\mathbf{Z}_{njk}$	mass contribution of sources to samples in scalar, vector, two-way matrix or three-way matrix form
z_{njk}	mass contribution of source n to the sample collected at site j in period k
ε, $\boldsymbol{\varepsilon}$, $\mathbf{E}$, $\mathbf{E}_{ijk}$	model error terms for corresponding dimensional models
Λ	diagonal matrix with diagonal elements being eigenvalues