



## Application of receptor modeling methods

Philip K. Hopke<sup>1</sup>, David D. Cohen<sup>2</sup>

<sup>1</sup> Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY, USA

<sup>2</sup> Australian Nuclear Science and Technology Organisation, Locked Bag 2001, Kirrawee DC, NSW, 2232, Australia

### ABSTRACT

The use of atmospheric compositional data for the identification and apportionment of sources has been ongoing for more than 40 years. Beginning in the 1960s, it was recognized that data analysis techniques could be applied to data and resolve combination of constituents that represent sources. In the late 1970s, these data analysis tools came to be called Receptor Models. This paper traces the early history of receptor models through those early papers and provides a historical introduction to the paper in this special issue showing the state of the art in the field and the application of these modern tools to a variety of atmospheric data.

### Keywords:

Receptor models  
Chemical mass balance  
Factor analysis  
Unmix  
Positive matrix factorization

### Article History:

Received: 15 October 2010  
Revised: 11 January 2011  
Accepted: 16 January 2011

### Corresponding Author:

Philip K. Hopke  
Tel: +1-315-268-3861  
Fax: +1-315-268-4410  
E-mail: hopkepk@clarkson.edu

© Author(s) 2011. This work is distributed under the Creative Commons Attribution 3.0 License.

doi: 10.5094/APR.2011.016

### 1. Introduction

The management of ambient air quality is a difficult but important problem. In general, it involves the identification of the sources of materials emitted into the air, the quantitative estimation of the emission rates of the pollutants, the understanding of the transport of the substances from the sources to downwind locations, and the knowledge of the physical and chemical transformation processes that can occur during their transport. All of those elements can then be put together into a mathematical model that can be used to estimate the changes in observable airborne concentrations that might be expected to occur if various actions are taken. Such actions could include the initiation of new sources as new industries are built and begin to function, and the imposition of emission controls of existing facilities in order to reduce the pollutant concentrations. However, the atmosphere is a very complex system, and it is necessary to greatly simplify the descriptions of reality in order to produce a mathematical model capable of being calculated on even the largest and fastest computers. Thus, although significant improvements have been made over the past thirty years in the mathematical modeling of dispersion of pollutants in the atmosphere, there are still many instances when the models are insufficient to permit the full development of effective and efficient air quality management strategies. Thus, it is necessary to have alternative methods available to assist in the identification of sources and the apportionment of the observed pollutant concentrations to those sources. Such methods are called receptor-oriented or receptor models since they are focused on the behavior of the ambient environment at the point of impact as opposed to the source-oriented dispersion models that focus on

the transport, dilution, and transformations that begin at the source and then follow the pollutants to the sampling or receptor site.

All of the currently used receptor models are based on the assumption of mass conservation and the use of a mass balance analysis. Similarly, the assumption of mass conservation and the use of a mass balance analysis can be used to identify and apportion sources of airborne particulate matter in the atmosphere. This methodology has generally been referred to within the air pollution research community as *receptor modeling* (Hopke, 1985; Hopke, 1991; Hopke, 2010). For example, suppose a sample of air is passed through a filter and the collected particles are analyzed for the elemental concentrations in mass of element per unit volume of air sampled. Let us assume that measured total airborne particulate lead concentration ( $\text{ng}/\text{m}^3$ ), for example, can be considered to be the sum of contributions from independent source types such as motor vehicles, incinerators, smelters, etc:

$$Pb_T = Pb_{\text{auto}} + Pb_{\text{incin.}} + Pb_{\text{smelter}} + \dots \quad (1)$$

However, a motor vehicle burning leaded gasoline emits particles containing materials other than lead. Therefore, the atmospheric concentration of lead from automobiles in  $\text{ng}/\text{m}^3$ ,  $Pb_{\text{auto}}$ , can be considered to be the product of two cofactors; the gravimetric concentration ( $\text{ng}/\text{mg}$ ) of lead in automotive particulate emissions,  $f_{\text{pb,auto}}$ , and the mass concentration ( $\text{mg}/\text{m}^3$ ) of automotive particles in the atmosphere,  $g_{\text{auto}}$ :

$$Pb_{auto} = g_{auto} \cdot f_{Pb,auto} \quad (2)$$

The normal approach to obtaining a data set for receptor modeling is to determine a large number of chemical constituents such as elemental concentrations in a number of samples. The mass balance equation can thus be extended to account for all  $m$  elements in the  $n$  samples as contributions from  $p$  independent sources:

$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (3)$$

where  $x_{ij}$  is the  $i^{\text{th}}$  elemental concentration measured in the  $j^{\text{th}}$  sample,  $a_{ik}$  is the gravimetric concentration of the  $i^{\text{th}}$  element in material from the  $k^{\text{th}}$  source, and  $f_{kj}$  is the airborne mass concentration of material from the  $k^{\text{th}}$  source contributing to the  $j^{\text{th}}$  sample.

The solution to Equation (3) depends on what information is available *a priori*. At least one vector of ambient concentration data,  $x_j$ , must have been measured. If the sources in an area are known and an opportunity to sample them, then the number of sources,  $p$ , is known as are the source composition profile vectors,  $a_k$ . Thus, the problem becomes a multiple regression problem with only the vector of source contributions,  $f_j$ , to be estimated. This multiple regression problem is termed a chemical mass balance (CMB) analysis. However, if the number and nature of the source profiles are not known, then a factor analysis approach is needed in which the number of sources, the source composition vectors, and the source contribution vectors can be estimated only from the ambient data. However, many more than one sample are needed for such an analysis. The use of factor analysis and related techniques has been generally termed multivariate receptor models.

## 2. Chemical Mass Balance

Receptor models date back to the late 1960's when both mass balance analysis and multivariate statistical methods were first applied to air quality data. The concept of an atmospheric mass balance model was suggested independently by Miller et al. (1972) and by Winchester and Nifong (1971). In these initial models, specific elements were associated with particular source types to develop a mass balance for airborne particles. Subsequently, more chemical species than sources were used in a least-squares fit to provide estimates of the mass contributions of the sources (Friedlander, 1973).

There were a number of these early applications of the mass balance analysis including Gent, Belgium (Heindrycx and Dams, 1974), Heidelberg, Germany (Bogen 1973), and Chicago, Illinois (Gatz, 1975). Several major research efforts have subsequently resulted in substantially better source data. The source emission studies led to much improved resolution of the particle sources in Washington, D.C. (Kowalczyk et al., 1978; Kowalczyk et al., 1982). In the first of these studies, Kowalczyk et al. (1978) introduced weighted least-squares regression to fit six sources with eight elements for ten ambient samples. Subsequently, Kowalczyk et al. (1982) examined 130 samples using 7 sources with 28 elements included in the fit. They obtained an excellent fit of the ambient concentration data and a quite good understanding of the major sources of airborne particles in the Washington, D.C. area.

Mayrhoon and Crabtree (1976) presented the use of an iterative least-squares approach to apportion six sources of airborne hydrocarbon compounds in the Los Angeles basin. The sources were automotive exhaust, volatilization of gasoline and release of gasoline vapor, commercial natural gas, geological natural gas, and liquefied petroleum gas. They performed the least-squares fit to the hydrocarbon compound concentrations

using gas chromatography to determine the concentrations of eight compounds. Their ordinary least-squares source reconciliation algorithm recognized that not all sources may contribute to every sample, and, if negative contributions were obtained, a different configuration of sources was employed with certain qualifying assumptions. Each possible configuration with positive coefficients was considered and the one with the lowest standard error was chosen as the optimum solution. On the average, automotive exhaust was the source of almost 50% of observed hydrocarbons. Gasoline and its vapor contributed up to 30% by weight and the balance resulted from commercial and geological natural gas. Thus, automobiles and other highway related sources were responsible for the majority of these hydrocarbons. A similar study utilizing this mass balance approach for resolving hydrocarbon sources has been made by Nelson et al. (1983) in Sydney, Australia. Thus, it is possible to identify the impact of emission sources on gaseous as well as particulate pollutants. Similar work is reported in this special issue for Mumbai, India (Pandit et al., 2011). CMB analysis has also been used to determine the sources of benzene in rural New York State (Li et al., 2011).

In 1979, Watson (1979) and Dunker (1979) independently suggested a mathematical formalism called effective variance weighting that included the uncertainty in the measurement of the source composition profiles as well as the uncertainties in the ambient concentrations. As part of this analysis, a method was also developed to permit the calculation of the uncertainties in the mass contributions. Effective-variance least squares has been incorporated into the standard personal computer software developed by the U.S. EPA for receptor modeling. The most extensive use of effective-variance fitting has been made by Watson and colleagues in their work on data from Portland, OR (Watson et al., 1984). Since that study, a number of other applications of this approach have been made in a wide variety of locations and extensive libraries of compositional profiles of emission sources have been developed to be used in the mass balance models. These models are described in detail by Watson et al. (1991). In this issue, Pandit et al. (2011) have used the CMB approach to apportion non-methane hydrocarbons (NMHCs) concentrations in Mumbai, India.

## 3. Factor Analysis

Actually, the first type of receptor modeling analysis reported in the literature was factor analysis using eigenvector methods that had been developed in the social sciences for interpreting large data sets. Blifford and Meeker (1967) used a principal component analysis with several types of axis rotations to examine particle composition data collected by the National Air Sampling Network (NASN) during 1957–61 in 30 U.S. cities. They were generally not able to extract much interpretable information from their data. Since there are a very wide variety of particle sources among these 30 cities and only 13 elements were measured, it is not surprising that they were not able to provide much specificity to their factors. Prinz and Stratmann (1968) examined both the aromatic hydrocarbon content of the air in 12 West German cities and data on the air quality of Detroit using factor analysis methods. In both cases, they found solutions that yielded readily interpretable results.

However, there was not further use of factor analysis until it was reintroduced in the mid-1970's by Hopke et al. (1976) and Gaarenstroom et al. (1977) in their analyses of particle composition data from Boston, MA and Tucson, AZ, respectively. In the Boston data for 90 samples at a variety of sites, six common factors were identified that were interpreted as soil, sea salt, oil-fired power plants, motor vehicles, refuse incineration and an unknown manganese-selenium source. In the study of Tucson, whole filter data were analyzed separately at each site. They find factors that are identified as soil, automotive, several secondary aerosol materials such as  $(\text{NH}_4)_2\text{SO}_4$  and several unknown factors. They

also discovered a factor that represented the variation of elemental composition in their aliquots of their neutron activation standard containing Na, Ca, K, Fe, Zn, and Mg. This finding illustrates one of the important uses of factor analysis; screening the data for noisy variables or analytical artifacts. It can provide useful insight into several possible problems that may exist in a data set including incorrect single values and some types of systematic errors.

With the use of atomic and nuclear methods to analyze environmental samples for a multitude of elements, very large data sets have been generated. Because of the ease in obtaining these results with computerized systems, the elemental data acquired are not always as thoroughly checked as they should be, leading to some, if not many, bad data points. It is advantageous to have an efficient and effective method to identify problems with a data set before it is used for further studies. Eigenvector analysis is described in detail by Henry (1991).

A problem that exists with these forms of factor analysis is that they do not permit quantitative source apportionment of particle mass or of specific elemental concentrations. In an effort to find alternative methods that would provide information on source contributions when only the ambient particulate analytical results are available, Henry and coworkers (Henry and Kim, 1989; Kim and Henry, 1999; Kim and Henry, 2000; Henry, 2003) have developed alternative methods based on eigenvector methods called Unmix. In these analyses, resolution similar to that obtained from a CMB analysis can be obtained although there remain problems (Paatero and Tapper, 1993). An alternative least-squares method has been developed by Paatero (1997) called Positive Matrix factorization (PMF). The PMF approach uses an explicit least-squares fit to the data to obtain the source profiles and mass contributions. Both PMF and Unmix have been recognized as applicable to air quality management efforts and are distributed by the United States Environmental Protection Agency ([www.epa.gov/ttn/scram/receptorindex.htm](http://www.epa.gov/ttn/scram/receptorindex.htm)). There are a number of examples of the application of PMF in articles in this Special Issue.

#### 4. Incorporation of Meteorological Information

It is often helpful to be able to combine the results of a source apportionment obtained from a CMB or factor analysis approach with local wind direction data to assist in the identification of specific sources or source types. There are several approaches for making such analyses including Conditional Probability Function (CPF) and Non-Parametric Regression (NPR). These methods are described and applied to highly time resolved data by Wang et al. (2011).

None of the source apportionment methods (CMB, Unmix, PMF) can provide a definite indication of the sources of secondary particles such as sulfate, nitrate, or secondary organic materials. The usual results of a CMB analysis are to list "sulfate" as a source or possibly describe it as "regional sulfate". Similar results are typically obtained through factor analysis. In order to really develop effective control strategies, it will be necessary to attribute the secondary particle mass to the original gaseous precursor sources. In order to make such an apportionment, additional information must be included in the analysis. This information is generally in the form of spatial/temporal information or in terms of meteorology as defined by air parcel back trajectories. A number of methods are available to analyze an ensemble of back trajectories along with the related air quality data (Hopke, 2003). Back trajectories have been applied in several of the studies presented in this Special Issue (Begum et al., 2011; Cohen et al., 2011).

The collection of studies presented in this issue provides useful illustrations of the application of a variety of data analysis

tools to multiple types of air quality data. They demonstrate, by example, the current trends and directions in air pollution characterization, source fingerprinting and source apportionment. The current analytical tools applied in these articles can reduce such data to information that may be useful in developing and applying air quality management strategies to improve air quality in an effective and efficient manner.

#### References

- Begum, B.A., Biswas, S.K., Pandit, G.G., Saradhi, I.V., Waheed, S., Siddique, N., Seneviratne, M.C.S., Cohen, D.D., Markwitz, A., Hopke, P.K., 2011. Long range transport of soil dust and smoke pollution in the south Asian region. *Atmospheric Pollution Research* 2, 151-157.
- Blifford, I.H., Meeker, G.O., 1967. A factor analysis model of large scale pollution. *Atmospheric Environment* 1, 147-157.
- Bogen, J., 1973. Trace elements in atmospheric aerosol in the Heidelberg area, measured by instrumental neutron activation analysis. *Atmospheric Environment* 7, 1117-1125.
- Cohen, D.D., Steker, E., Garton, D., Crawford, J., 2011. Fine particle characterisation, source apportionment and long range dust transport into the Sydney basin: A long term study between 1998 and 2009. *Atmospheric Pollution Research* 2, 182-189.
- Dunker, A.M., 1979. A Method for Analyzing Data on the Elemental Composition of Aerosols. General Motors Research Laboratories Report MR-3074 ENV-67, Warren, MI.
- Friedlander, S.K., 1973. Chemical element balances and identification of air pollution sources. *Environmental Science and Technology* 7, 235-240.
- Gaarenstroom, P.D., Perone, S.P., Moyers, J.L., 1977. Application of pattern recognition and factor analysis for characterization of atmospheric particulate composition in southwest desert atmosphere. *Environmental Science and Technology* 11, 795-800.
- Gatz, D.F., 1975. Relative contributions of different sources of urban aerosols: Application of a new estimation method to multiple sites in Chicago. *Atmospheric Environment* 9, 1-18.
- Heindryckx, R., Dams, R., 1974. Continental, marine and anthropogenic contributions to the inorganic composition of the aerosol of an industrial zone. *Journal of Radioanalytical and Nuclear Chemistry* 19, 339-349.
- Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. *Chemometrics and Intelligent Laboratory Systems* 65, 179-189.
- Henry, R.C., 1991. Multivariate receptor models. In *Receptor Modeling for Air Quality Management*, Hopke PK (ed). Elsevier: Amsterdam, 117-147.
- Henry, R.C., Kim, B.M., 1989. Extension of self-modeling curve resolution to mixtures of more than three components: Part 1. Finding the basic feasible region. *Chemometrics and Intelligent Laboratory Systems* 8, 205-216.
- Hopke, P.K., 2010. The application of receptor modeling to air quality data, *Pollution Atmospherique* Special Issue, September 2010, 91-109.
- Hopke, P.K., 2003. Recent developments in receptor modeling. *Journal of Chemometrics* 17, 255-265.
- Hopke, P.K., 1991. *Chapter 1 an Introduction to Receptor Modeling*, (Ed.) Philip, K.H., Data Handling in Science and Technology, Elsevier, pp. 1-10.
- Hopke, P.K., 1985. *Receptor Modeling in Environmental Chemistry*, A Wiley-Interscience publication, New York.
- Hopke, P.K., Gladney, E.S., Gordon, G.E., Zoller, W.H., 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.
- Kim, B.M., Henry, R.C., 2000. Extension of self-modeling curve resolution to mixtures of more than three components: Part 3. Atmospheric aerosol data simulation studies. *Chemometrics and Intelligent Laboratory Systems* 52, 145-154.

- Kim, B.M., Henry, R.C., 1999. Extension of self-modeling curve resolution to mixtures of more than three components: Part 2. Finding the complete solution. *Chemometrics and Intelligent Laboratory Systems* 49, 67-77.
- Kowalczyk, G.S., Gordon, G.E., Rheingrover, S.W., 1982. Identification of atmospheric particulate sources in Washington, D.C. using chemical element balances. *Environmental Science and Technology* 16, 79-90.
- Kowalczyk, G.S., Choquette, C.E., Gordon, G.E., 1978. Chemical element balances and identification of air pollution sources in Washington, D.C. *Atmospheric Environment* 12, 1143-1153.
- Li, R., Kalenge, S., Hopke, P.K., Lebouf, R., Rossner, A., Benedict, A., 2011. Source apportionment of benzene downwind of a major point source. *Atmospheric Pollution Research* 2, 138-143.
- Mayrsohn, H., Crabtree, J.H., Kuramoto, M., Sothorn, R.D., Mano, S.H., 1977. Source reconciliation of atmospheric hydrocarbons 1974. *Atmospheric Environment (1967)* 11, 189-192.
- Miller, M.S., Friedlander, S.K., Hidy, G.M., 1972. A chemical element balance for the Pasadena aerosol. *Journal of Colloid and Interface Science* 39, 165-176.
- Nelson, P.F., Quigley, S.M., Smith, M.Y., 1983. Sources of atmospheric hydrocarbons in Sydney a quantitative determination using a source reconciliation technique. *Atmospheric Environment* 17, 439-449.
- Paatero, P., 1997. Least squares formulation of robust, non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems* 37, 23-35.
- Paatero, P., Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems* 18, 183-194.
- Pandit, G.G., Sabu, S.K., Puranik, V.D., 2011. Distribution and source apportionment of atmospheric non-methane hydrocarbons in Mumbai, India. *Atmospheric Pollution Research* 2, 231-236.
- Prinz, B., Stratmann, H., 1968. The possible use of factor analysis in investigating air quality. *Staub, Reinhaltung der Luft* 28, 33-39.
- Wang, G., Hopke, P.K., Turner, J.R., 2011. Using highly time resolved fine particulate compositions to find particle sources in St. Louis, MO. *Atmospheric Pollution Research* 2, 219-230.
- Watson, J.G., Chow, J.C., Pace, T.G., 1991. *Chapter 4 Chemical Mass Balance*, (Ed.) Philip, K.H., Data Handling in Science and Technology, Elsevier, pp. 83-116.
- Watson, J.G., Cooper, J.A., Huntzicker, J.J., 1984. The effective variance weighting for least-squares calculations applied to the mass balance receptor model. *Atmospheric Environment* 18, 1347-1355.
- Watson, J.G., 1979. *Chemical Element Balance Receptor Model Methodology for Assessing the Source of Fine and Total Suspended Particulate Matter in Portland, Oregon*, Ph.D. Thesis, Oregon Graduate Center, Beaverton, OR.
- Winchester, J.W., Nifong, G.D., 1971. Water pollution in Lake Michigan by trace elements from pollution aerosol fallout. *Water, Air, and Soil Pollution* 1, 50-64.