Chemometrics and Intelligent Laboratory Systems 129 (2013) 15–20

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Source apportionment of time- and size-resolved ambient particulate matter

Na Li [a], Philip K. Hopke [b,*], Pramod Kumar [b], Steven S. Cliff [c], Yongjing Zhao [c], Carmeliza Navasca [d]

[a] Department of Mathematics, Clarkson University, Box 5815, Potsdam, NY 13699, USA
[b] Center for Air Resource Engineering and Science, Clarkson University, Box 5708, Potsdam, NY 13699, USA
[c] Air Quality Research Center, University of California, Davis, CA 95616, USA
[d] Department of Mathematics, University of Alabama at Birmingham, 1300 University Boulevard, Birmingham, AL 35294, USA

ARTICLE INFO

Article history:
Received 31 December 2012
Received in revised form 16 March 2013
Accepted 18 April 2013
Available online 28 April 2013

Keywords:
Multiway receptor modeling
Size fractionated particles
Rotating drum impactor
SXRF
Airport emission sources

ABSTRACT

Size- and time-resolved particulate matter samples collected using eight-stage Davis Rotating-drum Universal-size-cut Monitoring (DRUM) impactors at the Washington-Dulles International Airport were analyzed for the elemental composition using synchrotron X-ray fluorescence. A physically realistic three-way factor analysis model consisting of the outer products of matrices (profiles) times a vector of mass contributions was applied to these data. The problem was solved using a weighted alternating least squares method. Five major emission sources: soil, road salt, aircraft landings, transported secondary sulfate, and local sulfate/construction were identified. The study shows that time- and size-resolved RDI data can assist in the identification of the airport emission sources and atmospheric processes leading to the observed ambient concentrations.

1. Introduction

Airport emissions are studied with regard to the local air quality in nearby area of an airport for years (see [1–3]). In order to reduce the exposure of pollutants which emitted from airport operations, different airport emission sources need to be controlled. It is also necessary to quantify the various airport sources (ground vehicles, landings, etc.) in order to develop a reliable emissions inventory.

Receptor modeling is the application of data analysis methods to elicit information on the sources of air pollutants. It employs methods of solving the mixture resolution problem using chemical composition data for airborne particulate matter samples to identify the pollution source types and estimate the contribution of each source type to the mass of each sample. The fundamental principle of receptor modeling is that mass conservation can be assumed and a mass balance analysis can be used to identify and apportion sources of airborne particulate matter (PM) in the atmosphere.

Time integrated measurements and instruments [4] have been applied to the study of airport-related PM. However, time resolved measurements are really needed to detect PM mass associated with airport operations [3]. Very little data are currently available to address the characteristics of particles emitted from airport operations and the potential impact on exposure and health in adjacent communities. PM compositional data from airport emissions can provide useful information for source apportionment.

Another significant factor in the study of airport emissions is particle size since there are different sizes of particle in the atmosphere. They originate from different sources, and they are continually distributed by atmospheric transport processes. According to an earlier study [5], source compositions are dependent on the particle size. Therefore, analyses of particle size distribution data have also been performed to identify air pollution sources. Such data that contain both size and compositional information require advanced data analysis tools. The bilinear multivariate receptor models that are used extensively for source apportionment of airborne particles are not applicable to size-resolved data since they do not ensure appropriate continuity of the values as a function of particle size [6,7].

Therefore, in this study, size- and time-resolved PM samples were collected using several eight-stage rotating DRUM impactor samplers at Washington-Dulles International Airport. These data were analyzed by using a DRUM receptor model [12], which can take the size-composition variation into account to properly resolve the ambient data for the apportionment of potential airport emission sources. A weighted alternating least squares method is introduced to solve this model and five emission sources are identified successfully.

2. Data description

The original size- and time-resolved aerosol samples were collected using eight-stage rotating DRUM impactor samplers at Washington-Dulles International Airport. Three measurement campaigns were conducted during 3 different seasons (i) April 17–28, 2009; (ii) January 16–24, 2010; and (iii) July 9–23, 2010. During April, 2009, samples were collected by deploying one Rotating Drum Impactors (RDI) at the Base

---

* Corresponding author.
E-mail address: phopke@clarkson.edu (P.K. Hopke).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.04.010

Station. In the winter and summer seasonal campaigns, two RDIs were deployed; one at the Fire Station and the other at the Stone House (New Base Station) sites [8].

Particulate matter samples were analyzed by synchrotron X-ray Fluorescence (s-XRF) [9] using a broad-spectrum X-ray beam generated on beamline 10.3.1 at the Advanced Light Source Lawrence Berkeley National Laboratory. The s-XRF analysis provides quantitative elemental data for 27 elements (Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Se, Br, Rb, Sr, Y, Zr, Mo, and Pb) in 8 size modes (0.1–0.26 μm, 0.26–0.34 μm, 0.34–0.56 μm, 0.56–0.75 μm, 0.75–1.15 μm, 1.15–2.5 μm, 2.5–5 μm, and 5–10 μm) with 3-hour time resolution for the samples collected in this campaign. A total of 357 samples were collected over three seasons: spring, summer, and winter. In addition, mass concentrations were measured using soft beta attenuation. The summary statistics for the measured variables are presented in Table 1.

The data were considered as a function of size, time, and chemical composition (i.e. elemental species), which can be organized a third-order tensor $\boldsymbol{\chi}_{\text{orig}} \in \mathbb{R}^{I \times J \times K}$. If i denotes the chemical species, j to express particle size, and k to be the time sample, then a datum point, x_{ijk}, can be expressed as the concentration value of the ith chemical species of the jth particle size of the kth time sample.

There are two problems that need to be addressed. First, the synchrotron XRF does not provide carbon and nitrate values, so the measured mass minus the reconstructed mass, termed the “unmeasured mass”, is introduced in the analysis. This approach has been used previously in the analysis of data from Denver [10] that provided reasonable estimates of carbonaceous sources. Another issue is the influence of high-noise variables (chemical species). For some variables, the data may consist almost entirely of noise which would increase the errors in computed factors. The question of accepting or rejecting individual chemical constituents has been studied by Paatero and Hopke [11]. The signal-to-noise ratio (S/N) and below detection level (BDL) were introduced to determine the noisy variables (containing much more noise than signal). For uncensored data, a variable is defined to be bad if S/N < 0.2. For censored data, a sufficiently large number of BDL values (>80%) may also indicate a noisy variable. Therefore, four chemical species (P, Ga, Y, Zr) were eliminated on the basis of S/N and large number of BDL values, so that the value of I index of the tensor data we are using in the analysis is 24. It includes 23 chemical elements and the unmeasured mass. Consequently, the dimension of the tensor $\boldsymbol{\chi}$ is 24 × 8 × 357.

Table 1
The summary statistics for the original data set.

Element	Mean (ng/m^3)	Standard deviation (ng/m^3)	Median (ng/m^3)	S/N[a]	Number of BDL values[b]
Mg	61.10	270.69	32.14	2.6503	53
Al	37.02	85.52	12.82	1.7189	6
Si	54.44	113.84	10.00	3.2904	5
P	7.65	29.77	4.75	0.7456	6
S	113.26	721.72	31.21	5.0237	2
Cl	5.46	22.14	0	0.9685	30
K	7.47	11.49	3.29	0.9954	9
Ca	30.64	61.97	3.30	1.0085	0
Ti	2.99	5.37	0.90	0.9978	0
V	0.11	0.22	0.05	0.9960	247
Cr	0.05	0.10	0.02	1.0000	402
Mn	0.46	0.87	0.17	0.9998	85
Fe	23.48	41.48	5.84	1.0004	0
Co	0.07	0.10	0.04	0.9981	497
Ni	0.12	0.33	0.06	1.0000	84
Cu	0.99	1.59	0.36	0.9996	0
Zn	1.69	1.80	1.14	0.9979	0
Ga	0.03	0.04	0.02	0.9994	572
As	0.12	0.26	0.03	0.9998	237
Se	0.22	0.35	0.14	0.9995	299
Br	2.00	0.83	1.81	0.9987	0
Rb	0.21	0.22	0.17	0.9999	482
Sr	0.48	0.24	0.43	0.9999	448
Y	0.49	0.43	0.38	0.9999	603
Zr	0.99	0.66	0.78	0.9997	537
Mo	2.32	1.10	2.07	0.9995	489
Pb	1.06	3.10	0.43	0.9995	331

[a] Signal to noise ratio as defined by Paatero and Hopke [11].
[b] Number of values below the method detection limit.

3. DRUM model description

In order to take full advantage of the size dependent composition behavior that exists in source emissions, a model had been previously developed for DRUM data [12]. In this model, the profile for a given source is a matrix of dimensions defined by the number of measured variables and the number of measured size fractions. For each source (factor), there is a vector of mass contributions in terms of time, so the outer product of the source profile matrix times the mass contribution vector produces a tensor whose dimensions are defined by the number of measured chemical species, the number of size, and the number of time samples. Then, tensor $\boldsymbol{\chi}$ can be factored into a summation of R outer products of the source profile matrix and the vector of mass contribution, where R denotes the number of independent sources (factors).

Therefore, the main equation of the model is as below:

$$\boldsymbol{\chi} = \sum_{r=1}^{R} \mathbf{A}^{(r)} \circ \mathbf{b}^{(r)} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\chi}$ is the third-order tensor of observed data, $\mathbf{A}^{(r)}$ is the rth source profile array and $\mathbf{b}^{(r)}$ is the corresponding rth contribution vector. The tensor $\boldsymbol{\varepsilon}$ having the same size as $\boldsymbol{\chi}$ contains the residuals. Fig. 1 schematically shows the model used in this study.

In its component form, the model equation becomes:

$$x_{ijk} = \sum_{r=1}^{R} A_{ij}^{(r)} b_k^{(r)} + e_{ijk} \tag{2}$$

where $A_{ij}^{(r)}$ is the ith species mass fraction of the jth particle size range from the rth source, $b_k^{(r)}$ is the rth source mass contribution during the time units for the kth sample, and e_{ijk} is the residual associated with the ith species concentration measured in the kth sample of the jth size range, and R is the total number of independent sources.

In this receptor model, the problem is to find matrices $\mathbf{A}^{(r)}$ and vectors $\mathbf{b}^{(r)}$, for $r = 1, \ldots, R$, to minimize the objective function:

$$Q = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(x_{ijk} - \sum_{r=1}^{R} A_{ij}^{(r)} b_k^{(r)} \right)^2}{u_{ijk}^2} \tag{3}$$

where u_{ijk} is the uncertainty value associated with data value x_{ijk}.

Tensor block term decomposition in rank $(L;L;1)$, BTD-$(L;L;1)$, is the another way to solve the receptor model in Eq. (1).

Definition (BTD-(L;L;1)). Given a third order tensor $\boldsymbol{\chi} \in \mathbb{R}^{I \times J \times K}$, a rank-$(L;L;1)$ block term decomposition of $\boldsymbol{\chi}$ is described by:

$$\boldsymbol{\chi} = \sum_{r=1}^{R} \left(\mathbf{A}_r \cdot \mathbf{B}_r^T \right) \circ \mathrm{c}_r \tag{4}$$

in which the matrices $\mathbf{A}_r \in \mathbb{R}^{I \times L}$, $\mathbf{B}_r \in \mathbb{R}^{I \times L}$, and vector $\mathrm{c}_r \in \mathbb{R}^{K}$.

Fig. 2 shows the decomposition structure of BTD-$(L;L;1)$. Therefore, comparing Eqs. (4) and (1), the product of matrices $\left(\mathbf{A}_r \cdot \mathbf{B}_r^{\mathrm{T}}\right)$ actually is the matrix $\mathbf{A}^{(r)}$. Thus, for a given tensor $\boldsymbol{\chi}$, the BTD-$(L;L;1)$ solution set $\{\mathbf{A}_r, \mathbf{B}_r, \mathbf{c}_r, r = 1, 2, \ldots, \mathrm{R}\}$, provides the solution of Eq. (1) as

$$\mathbf{A}^{(r)} = \mathbf{A}_r \cdot \mathbf{B}_r^T, \mathbf{b}^{(r)} = \mathrm{c}_r.$$