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

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.

© 2013 Elsevier B.V. All rights reserved.

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).

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 $\chi_{orig} \in \mathbb{R}^{I \times J \times K}$. 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 χ is $24 \times 8 \times 357$.

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 χ 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:

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

where χ 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 χ 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 $r = 1, ..., R$, 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 $\chi \in \mathbb{R}^{I \times J \times K}$, a rank-$(L;L;1)$ block term decomposition of χ is described by:

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

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

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

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

Table 1
The summary statistics for the original data set.

Element	Mean (ng/m³)	Standard deviation (ng/m³)	Median (ng/m³)	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].
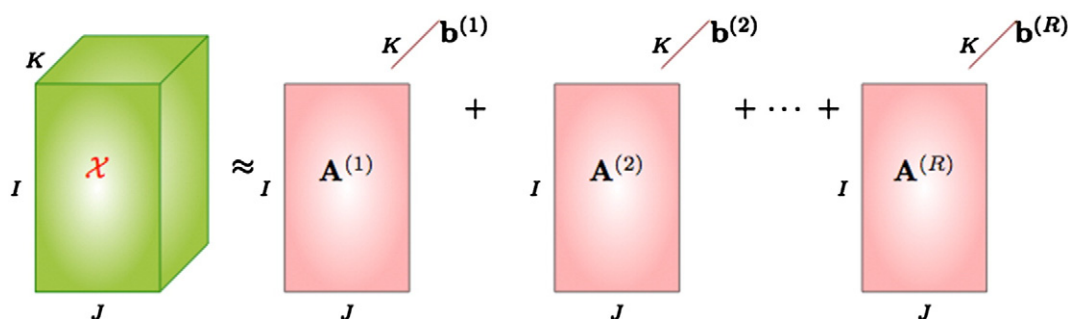[b] Number of values below the method detection limit.

Fig. 1. Receptor model for the observed tensor. $\mathbf{A}^{(r)}$ is the source profile matrix and $\mathbf{b}^{(r)}$ is the mass contribution.

To decrease the rotational freedom in the solution, non-negativity constraints were applied to the factors. The uncertainty estimation in Eq. (3) also provides a useful tool to decrease the weight of missing and below detection limit data in the solution. The procedures of Polissar et al. [13] will be used to assign measured data and the associated uncertainties as the input data.

For convenience, the uncertainties u_{ijk} are organized into a third-order tensor U. The objective function in Eq. (3) can then be written as:

$$Q = \left\| \frac{\left(\chi - \sum_{r=1}^{R} \mathbf{A}^{(r)} \circ \mathbf{b}^{(r)} \right)}{U_F^2} \right\|^2 \tag{5}$$

where the division between the two tensors is element-wise division.

The problem was solved based on the objective function (Eq. (5)). A weighted alternating least-squares (WALS) algorithm similar to that described by Wentzell et al. [14] is used. Variants like the BTD-ALS [15,16] can be used to produce the decomposition in Eq. (1). However, these three-way tensor block model solvers incur higher operational costs and memory for this specific model (Eq. (1)). The WALS framework is based on ALS with only two subproblems that are appropriately chosen matricizations of Eq. (1). In addition, the WALS formulation offers more flexibility than BTD-ALS with respect to constraint integration and column-wise reformulation to further reduce operational costs. A detailed discussion on WALS is presented in the supplemental information of this paper.

To reduce rotations, uncertainty estimates and non-negativity constraints were used. The standard MATLAB least squares function, 'lsqnonneg', was used to impose the non-negativity constraints. However, according to Paatero et al. [17], those are generally insufficient to wholly eliminate the rotational problem. Several methods are proposed by Paatero et al. [17] to control the rotations. One way is constraining individual factor elements, either scores and/or loadings, toward zero values based on some external information about acceptable or desirable shapes of the factors. Therefore, constraints based on the priori information were imposed. Based on an initial analysis [8], one factor

is dominated by large particles with high concentrations of chlorine. This factor should only be contributing during the January sampling campaign and can be associated with the use of salt and sand on snow and ice. Therefore, the mass contribution vector $\mathbf{b}^{(r)}$ was constrained to zeros for the summer and spring samples.

For each chemical element, the mass fraction values across the eight particle size ranges should be relatively smooth, which means that there should not be very low or zero concentrations in an intermediate particle size while both its adjacent particle sizes (smaller one and larger one) have high concentrations. Thus, another constraint was imposed on the size mode to make sure that the change across the size mode is smooth.

The resulting apportionments are only good to a scale constant so the results were normalized by regressing the apportioned masses for each source for each sampling period to the total measured mass as per Hopke et al. [18].

4. Numerical results and interpretation

We apply the weighted alternating least squares to the sample tensor $\chi \in \mathbb{R}^{24 \times 8 \times 357}$ to calculate the source profile matrices $\mathbf{A}^{(r)}$ and the corresponding contribution vectors $\mathbf{b}^{(r)}$. Two criteria were used to assess the number of factors. The fits to the data are examined by reviewing the distributions of scaled residuals. These distributions should be symmetric and the values should generally range from -3 to $+3$. In addition, the profiles have to be physically realistic. The interpretability of the factors includes appropriate behavior across the particle size dimension since there should be a degree of smoothness in that direction. The pattern of elements and their appearance in physically meaningful size ranges were used to assess the appropriateness of the various solutions. In our experiment, five factors were ultimately chosen to adequately reproduce the data and provide interpretable factors.

Fig. 3 shows the profiles for each factor (source) for all three sampling campaigns. They are shown as grouped bar plots so that the size variation of the chemical species in each source profile can be observed. The time series of the source contributions are shown in Fig. 4. The average mass contributions of each source for each season to particulate
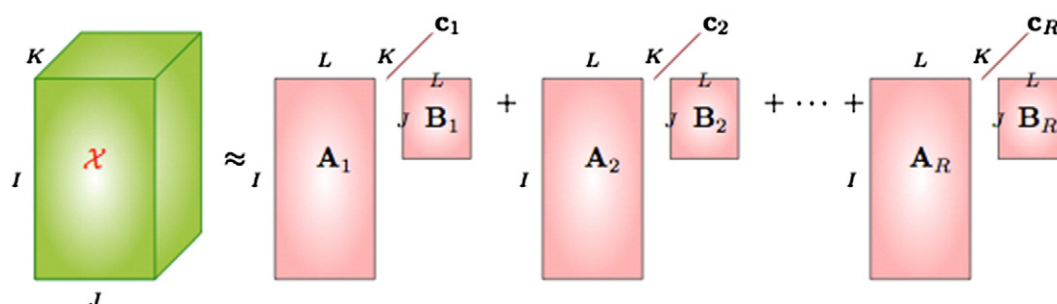

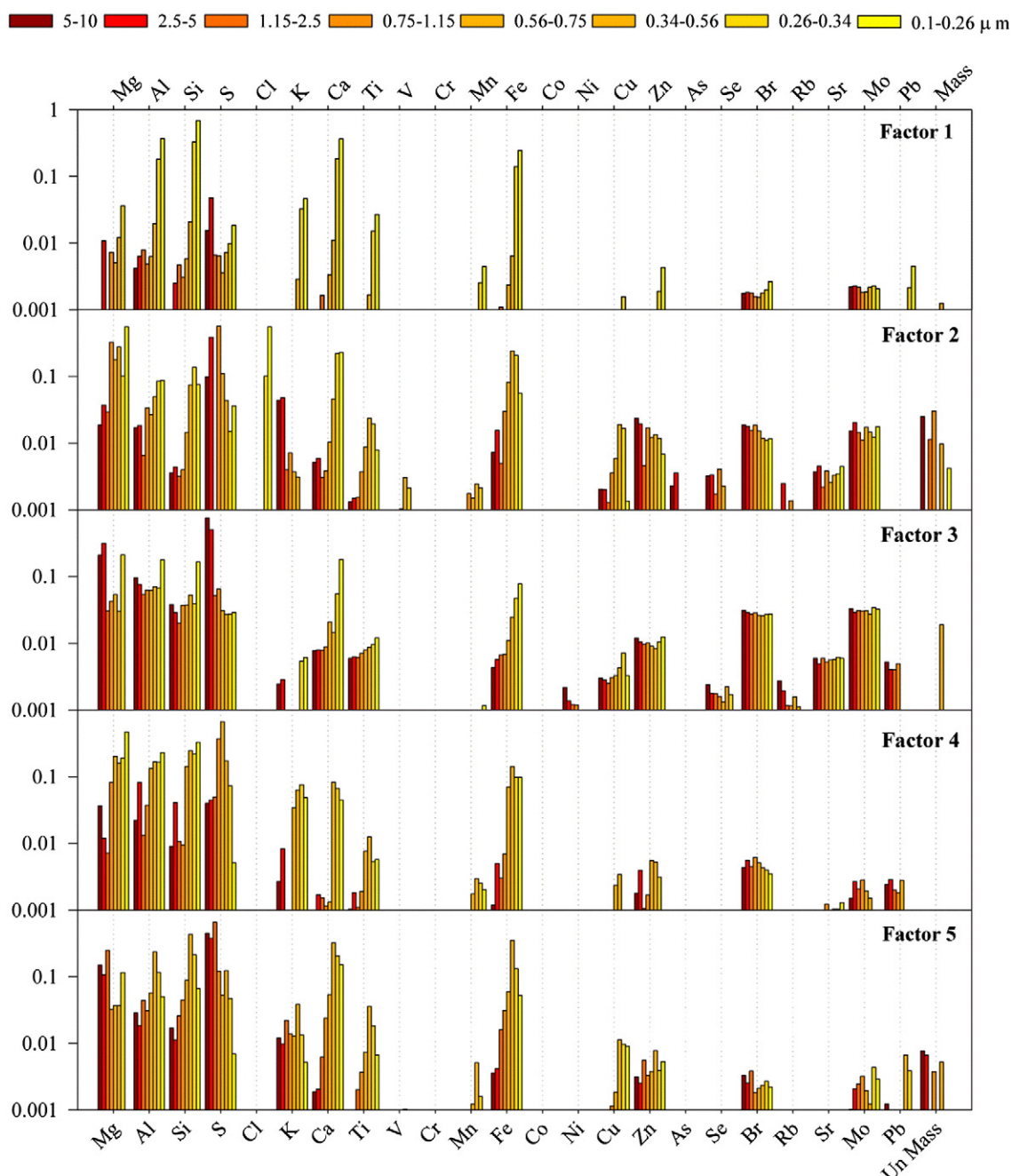**Fig. 2.** The BTD-$(L;L;1)$ structure for a third-order tensor $\chi \in \mathbb{R}^{I \times J \times K}$.

Fig. 3. Source profiles for the resolved factors.

matter less than 10 μm in aerodynamic diameter (PM₁₀) are presented in Table 2.

The first factor shows high concentrations of crustal elements (Al, Si, Fe, Ca, Ti) peaking in the two largest size ranges. This factor can therefore be attributed to "soil". There can be some wind-driven aerosolization of surface soils, but more of the soil is probably re-suspended by various forms of traffic including the cars bringing passengers to the airport, ground activities at the airport, along with the taxiing, take-off, and landings of the aircraft.

The second factor is dominated by large particles with high concentrations of chlorine, calcium, and magnesium with some iron and other crustal species. This factor only had contributions only during the January sampling campaign and can be associated with the use of salt and sand on snow and ice. This factor was only seen during the January campaign and we constrained the other values to be zero in the final model.

Factor 3 shows a very different pattern with small particles of sulfur, zinc, bromine, zirconium and molybdenum. This factor is assigned to particles that are emitted during landings. The sulfur and zinc come from tire wear. These elements are key constituents in tires. Often a visible puff of smoke is observed at touchdown. There is considerable frictional heat produced at this instant and particles are generated across the particle size range. Both zirconium and molybdenum are used in high temperature greases as might be used to lubricate bearings that would undergo significant heat stress. The energy deposited in the bearings can be expected to liberate particles from the lubricants.

Factors 4 and 5 have the highest values of S, but in different size fractions. S in factor 4 peaks in the middle size ranges. Such sizes are indicative of cloud processed sulfur and a similar factor was observed by Peré-Trepat et al. [12]. There is some intermixing of the sulfate with coarse particle soil. This sulfate is transported to the site given the
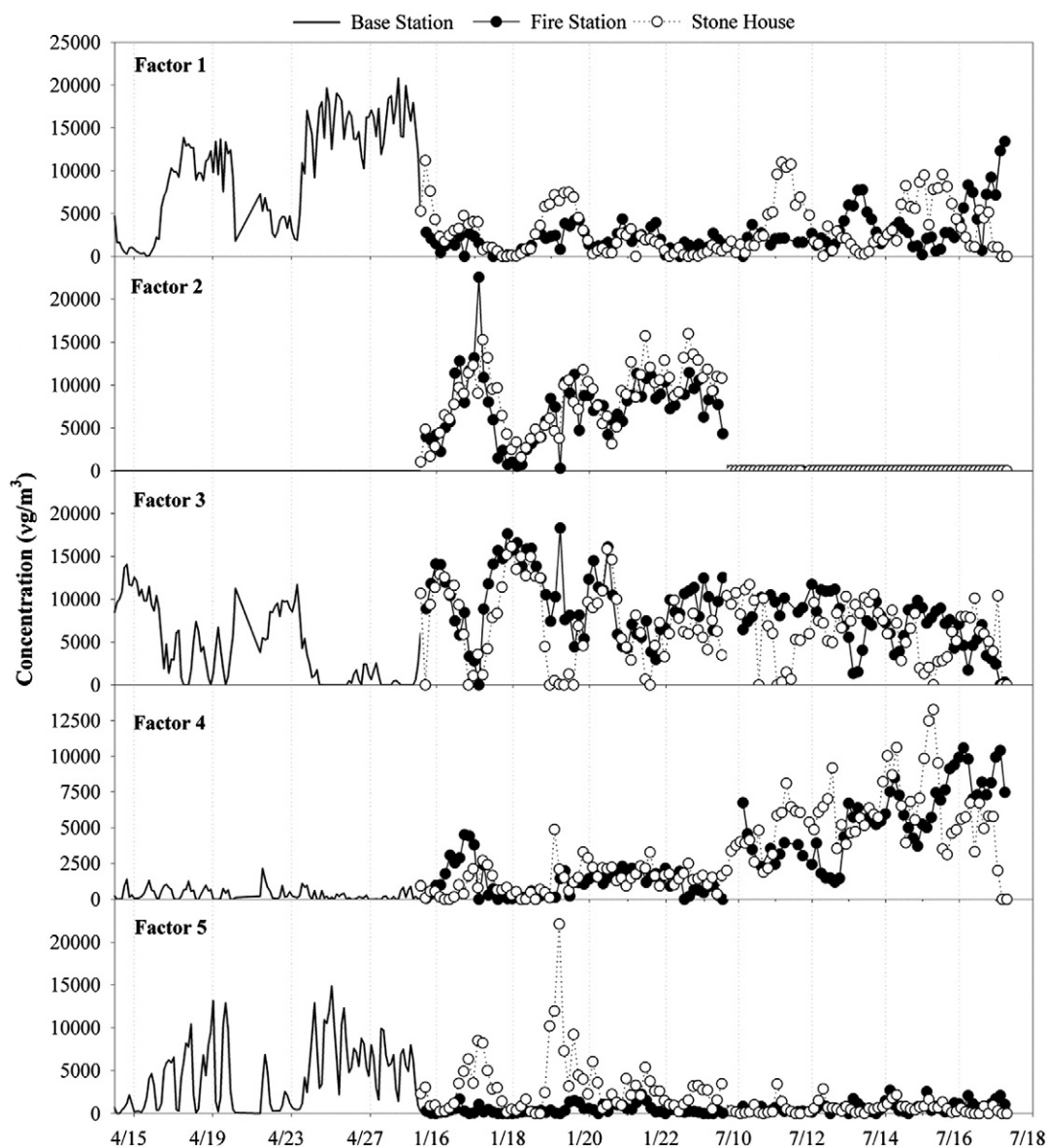
Fig. 4. The time series of source contributions.

Table 2
Apportionment of PM_{10} for each site during each sampling campaign.

		PM_{10} (ng/m³)	Base Station April	Fire Station January	Stone House January	Fire Station July	Stone House July
Soil	Mean		10,000	1708	2422	3622	3758
	Std dev		6044	1145	2488	2929	3173
	Median		10,995	1495	1664	2733	2412
Salt	Mean		0	7240	8338	0	0
	Std dev		0	3883	3782	0	0
	Median		0	7703	9040	0	0
Landings	Mean		4277	9870	7225	6920	6238
	Std dev		4254	4147	4717	2995	3398
	Median		2619	10,277	7084	7362	6662
Homogeneous sulfate	Mean		350	1210	1322	5695	5565
	Std dev		403	1072	958	2543	2611
	Median		190	1063	1392	5628	5603
Local sulfate	Mean		4498	431	3175	759	606
	Std dev		3814	498	3545	619	681
	Median		4153	259	2379	623	481

uniformity of contributions at multiple sites. This factor was an important contributor of particle mass during the summer sampling campaign.

Factor 5 has high sulfur concentrations in the smallest size bins and shows contributions from crustal species in intermediate sized particles. Small size sulfate is usually attributed to homogeneous sulfate formation. Given that off-road diesel fuel has a significantly higher sulfur content than on-road fuel, there may be some contribution from local diesel vehicles such as aircraft tugs and other ground vehicles. It is not clear what the source of the soil particles might be in sizes around 1.0 to 2.5 μm. This factor was primarily observed in the April 2009 sampling period with a peak at the Stone House site during the winter. It is likely that there is some admixture of sand or construction material. Diesel construction equipment could also provide small particle sulfate.

5. Conclusions

From the analysis of size- and time-resolved particle sample compositional data, five emission sources were identified using a weighted

alternating least squares method: soil, deicing road salt, aircraft landings, transported secondary sulfate, and local sulfate/construction. The largest source associated with the airport operations was aircraft landing that had not been previously considered as a significant source of particles.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.chemolab.2013.04.010.

References

[1] H. Peace, J. Maughan, B. Owen, D. Raper, Identifying the contribution of different airport related sources to local urban air quality, Environmental Modelling and Software 21 (2006) 532–538.

[2] B.S. Cohen, A.L. Bronzaft, M. Heikkinen, J. Goodman, A. Nadas, Airport-related air pollution and noise, Journal of Occupational and Environmental Hygiene 5 (2008) 119–129.

[3] D. Westerdahl, S.A. Fruin, P.L. Fine, C. Sioutas, The Los Angeles International Airport as a source of ultrafine particles and other pollutants to nearby communities, Atmospheric Environment 42 (2008) 3143–3155.

[4] S. Barbosa, E. Ayers, M. Belknap, Air monitoring study at Los Angeles International Airport, in: S.C.A.Q.M. District (Ed.), South Coast Air Quality Management District, 1999, p. 39.

[5] J.A. Dodd, J.M. Ondov, G. Tuncel, T.G. Dzubay, R.K. Stevens, Multimodal size spectra of submicrometer particles bearing various elements in rural air, Environmental Science and Technology 25 (1991) 890–903.

[6] P.K. Hopke, Recent developments in receptor modeling, Journal of Chemometrics 7 (2003) 255–265.

[7] P.K. Hopke, The application of receptor modeling to air quality data, Pollution Atmosphérique, Special Issue (September, 2010) 91–109.

[8] B. Kim, J. Rachami, D. Robinson, B. Robinette, K. Nakada, S. Arunachalam, N. Davis, B.-H. Baek, U. Shankar, K. Talgo, D. Yang, A.F. Hanna, R.L. Wayson, G. Noel, S.S. Cliff, Y. Zhao, P.K. Hopke, P. Kumar, Guidance for Quantifying the Contribution of Airport Emissions to Local Air Quality, Appendix A, ACRP Report 71, Transportation Research Board, Washington, DC, 2012.

[9] A. Knochel, Basic principles of XRF with synchrotron radiation, 2nd International Workshop on XRF and PIXE Applications in Life Science, Capri, Italy, World Scientific Publishing Co, Singapore, 29–30 June, 1989.

[10] W. Zhao, P.K. Hopke, E.W. Gelfand, N. Rabinovitch, Use of an expanded receptor model for personal exposure analysis in schoolchildren with asthma, Atmospheric Environment 41 (2007) 4084–4096.

[11] P. Paatero, P.K. Hopke, Discarding or downweighting high-noise variables in factor analytic models, Analytica Chimica Acta 490 (2003) 277–289.

[12] E. Peré-Trepat, P.K. Hopke, P. Paatero, Source apportionment of time and size resolved ambient particulate matter measured with a rotating DRUM impactor, Atmospheric Environment 41 (2007) 5921–5933.

[13] A.V. Polissar, P.K. Hopke, P. Paatero, W.C. Malm, J.F. Sisler, Atmospheric aerosol over Alaska 2. Elemental composition and sources, Journal of Geophysical Research 103 (1998) 19045–19057.

[14] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.A. Allen, M.W. Washburne, Multivariate curve resolution of time course microarray data, BMC Bioinformatics 7 (2006) 343.

[15] C. Navasca, S. Kindermann, L. De Lathauwer, Swamp reducing technique for tensor decomposition, In the 16th Proceedings of the European Signal Processing Conference, Lausanne, August 2008.

[16] L. De Lathauwer, D. Nion, Decompositions of a higher-order tensor in block terms — Part III: alternating least squares algorithms, SIAM J. Matrix Anal. Appl. 30 (2008) 1067–1083.

[17] P. Paatero, P.K. Hopke, X.H. Song, Z. Ramadan, Understanding and controlling rotations in factor analytic models, Chemometrics and Intelligent Laboratory Systems 60 (2002) 253–264.

[18] P.K. Hopke, R.E. Lamb, D.F.S. Natusch, Multielemental characterization of urban roadway dust, Environmental Science and Technology 14 (1980) 164–172.