SIZE-RESOLVED SOURCE APPORTIONMENT OF AMBIENT PARTICLES BY POSITIVE MATRIX FACTORIZATION

J. S. Han[1], K. J. Moon[1], S. J. Lee[1], Y. J. Kim[2], S. Y. Ryu[2], S. S. Cliff[3] and S. M. Yi[4]

[1]Department of Air Quality Research, National Institute of Environmental Research, Environmental Research Complex, Kyeongseo-dong, Seo-gu, Incheon, 404-170, Republic of Korea, first-author: nierhan@me.go.kr
[2]ADvanced Environmental Monitoring Research Center (ADEMRC) at Gwangju Institute of Science and Technology (GIST), Oryong-dong, Buk-gu, Gwangju, 500-712, Republic of Korea
[3]The DELTA Group (DAS), University of California, Davis, CA, USA.
[4]Department of Environmental Health, Seoul National University, Yongun-dong, Jongno-gu, Seoul, 110-799, Republic of Korea

ABSTRACT

The size- and time-resolved aerosol samples were collected using an eight-stage DRUM sampler from 29 March to 29 May in 2002 at Gosan, Jeju Island, Korea, which is one of the representative background sites in East Asia. The size-resolved 3-hr average concentrations of 19 elements including S, Si, Al, Fe, Ca, Cl, Cu, Zn, Ti, K, Mn, Pb, Ni, V, Se, As, Rb, Cr, Br were then analyzed using the positive matrix factorization (PMF) technique, in order to identify the possible sources and estimate their contribution to particulate matter mass. Fifteen sources were then resolved in eight size ranges (0.07~12 μm), including Chinese aerosol, soil dust, sea salt, biomass burning, coal combustion, oil heating furnace, residual oil fired boiler, municipal incineration, nonferrous metal source, ferrous metal source, gasoline vehicle, diesel vehicle, copper smelter, and volcano emission.

Key Words : Drum sampler, Size distribution, Trace element, PMF, Source apportionment

1. INTRODUCTION

Generally, it is necessary to identify the aerosol sources and estimate their influence on ambient PM concentration, in order to formulate effective control strategies for ambient particulate matter. For that reason, source apportionment has been derived from various statistical methods, such as chemical mass balance (CMB), factor analysis, and multiple linear regression methods. Among them, receptor modeling using aerosol chemical composition data is a reliable method, which can provide information on aerosol sources. However, the commonly used multivariate receptor models such as principal component analysis (PCA) has several drawbacks. The factors of PCA are not always physically realistic, because negative values may appear among factor loadings and factor scores. In addition, PCA can't handle

missing and below detection limit data often observed in the environmental measurements.

Positive Matrix Factorization (PMF) method (Paatero and Tapper, 1994) complements the weak points in the previous receptor modeling. Adjustment of the corresponding error estimates also allows it to handle missing and below detection limit data. Moreover, non-negative constraints are implemented in order to obtain more physically explainable factors (Paatero, 1993).

Composition of ambient aerosol varies with size even in the fine size range (<2.5 μm) depending on the type of anthropogenic source. Therefore, more size-segregated composition data are needed in the fine size range in order to improve the accuracy of anthropogenic source apportionment. From this point of view, PMF analysis has been performed in this study on the Gosan aerosol data collected by a DRUM sampler which has merit to collect the fine particles in five stages below 2.5 μm. Therefore, it is expected that size-segregated aerosol data collected by a DRUM sampler with high time resolution will improve the efficiency of PMF analysis.

2. SAMPLING AND ANALYSIS

Ambient aerosol collection using an eight-stage Davis Rotating Unit for Monitoring (DRUM) sampling system was made at the western tip of Gosan, Jeju Island, Korea (33° 17' N, 126° 10' E, 70m asl), which is a representative background site in East Asia, from 29 March to 29 May 2001. During the measurement period, two Asian dust (AD) outbreaks were observed on 8-10 April and 17 April. In this study, aerosol data pertaining to those AD periods were not subject to PMF and ACWT analysis in order to focus on the estimation of anthropogenic aerosol sources.

The DRUM sampler collects size-resolved aerosol samples on Apiezon^{TM} coated Mylar^{TM} strips in eight stages, having the equivalent aerodynamic cut-off diameters 0.07, 0.26, 0.34, 0.56, 0.75, 1.15, 2.5, 5.0, and 12 μm (Cahill et al. 1985). The DRUM sampler was operated continuously during the 61-day sampling period. The collected aerosol samples were then analyzed for inorganics (19 elements between aluminum and lead) using synchrotron X-ray fluorescence (S-XRF) at the Lawrence Berkeley National Laboratory Advanced Light Source (Perry et al., 2004). A detailed description of the sampling and analysis methods is provided by Cahill et al. (1993).

3. DATA ANALYSIS BY PMF

Positive Matrix Factorization (PMF) method was developed by Paatero (Paatero and Tapper, 1993; Paatero, 1994) to provide flexible modeling approach that effectively uses the information in the data. In PMF, all data matrix X of dimension n rows and m columns, where n and m are the number of samples and species, respectively, can be factorized into two matrix, namely G (n$\times$p) and F (p$\times$m), and the residual part E, where p represents the number of factors extracted.

$$X = GF + E \tag{1}$$

G is source contribution matrix with p sources, and F is a source profile matrix. PMF provides a solution that minimizes an object function, Q based upon uncertainty for each observation (Paatero, 1994), which is defined as

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{e_{ij}}{s_{ij}} \right)^2 , \quad (2)$$

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

where s_{ij} is the uncertainty in the measured data x_{ij}. PMF uses a least squares approach to solve the factor analysis problem with integrating non-negativity constraints into the optimization process, meaning that sources cannot have negative species concentration ($f_{kj} \geq 0$) and sample cannot have a negative source contribution ($g_{ki} \geq 0$). The solution of Eq. (2) is obtained using an iterative minimization algorithm PMF2 (two way PMF) (Paatero, 1994). PMF2 uses the error of measurement in the data to provide optimum data point scaling and permits better treatment of missing and below-detection-limit values. For measurement values below detection limit, x_{ij} was replaced by half of the detection limit and an error corresponding to a relative uncertainty of 100% was assigned to the original error estimate.

A robust mode of PMF2 has been selected for handling outlier data, in order to degrade the disproportional affect of excessively large data points, especially for analyzing environmental data. This can be achieved by introducing a filter function h_{ij} in Eq. (2), the least-squares minimization of Q (Paatero, 1994),

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{e_{ij}}{h_{ij} s_{ij}} \right)^2 \quad (4)$$

, where

$$h_{ij} = \begin{cases} 1 & if \left| e_{ij} / s_{ij} \right| \leq \alpha, \\ \left| e_{ij} / s_{ij} \right| / \alpha & otherwise. \end{cases} \quad (5)$$

The parameter α is called the outlier threshold distance, and the value $\alpha=4$ was chosen in the present study as done by Lee et al. (1999) and Hien et al. (2004). The other important parameter of PMF2 was F_{peak}, which can be used to control rotations and yield more physically realistic solutions (Paatero et al., 2002). In this study, the rotation was controlled by F_{peak} until an appropriate distribution of the edges is achieved and G space plotting for PMF modeling reveals the independence of the contributions reducing the rotational ambiguity.

4. RESULTS AND DISCUSSIONS

4. 1. Determination Of The Number Of Factors And Rotational Freedom

The information from the scaled residual matrix (**R**) in PMF is used to reduce the ambiguity, due to manual judgment on the number of factors. For each specific number of factors, two parameters are obtained from **R**: *IM*, the maximum individual

column mean, and *IS*, the maximum individual column standard deviation (Lee et al., 1999), where

$$IM = \max_{j=1...m}\left(\frac{1}{n}\sum_{i=1}^{n}r_{ij}\right),$$ (6)

$$IS = \max_{j=1...m}\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_{ij}-\bar{r}_j)^2}\right).$$ (7)

When the number of factors increases to a critical value, *IM* and *IS* will sharply drop. Figure 1a and 1b show the variation of *IM* and *IS* with different number of factors in eight size ranges. Finally, four to eight factors were found to generate the most reasonable results. After the source profiles were identified without transformation, PMF was run with different F_{peak} values in order to find out a range within which the objective function *Q* value in Eq. (4) remains relatively constant (Song et al., 2001). The largest element in **Rotmat**, a rotational matrix in PMF, is also used to reveal if factors have excessive rotational freedom (Lee et al., 1999). Figure 2.c and 2.d show the variation of *Q* value and largest element in **Rotmat** with the different F_{peak} values. As a result, F_{peak} values of –0.4 ~ –0.9 provided the most physically reasonable source profiles in the eight size ranges.
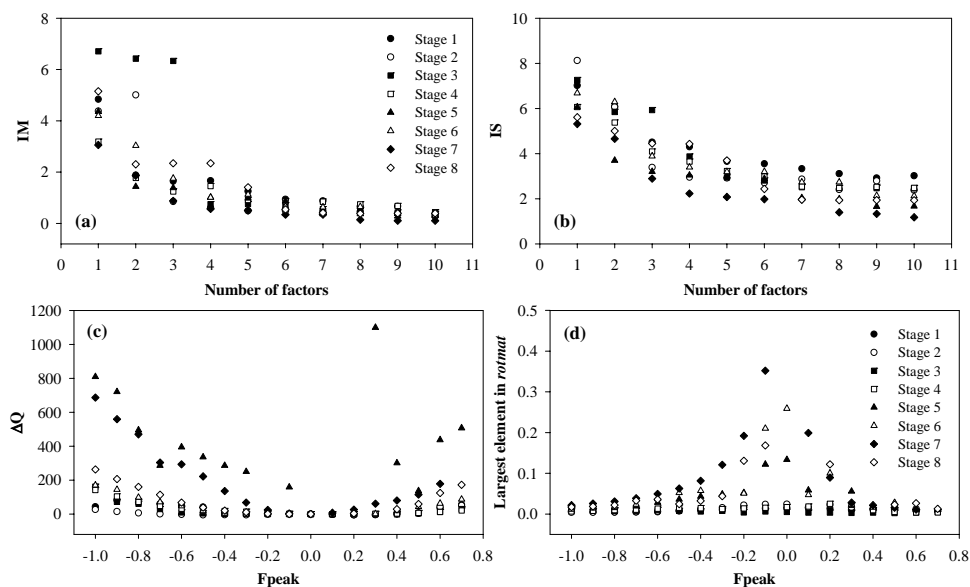

Figure 1. Determination of the number of factors and rotational freedom in eight size ranges by (a) maximum individual column mean (IM), (b) standard deviation (IS) of standardized residuals, (c) Q, and (d) largest element in the rotational matrix.

4. 2. Source Profiles And Temporal Variations

Overall, fifteen distinct primary sources were resolved for the ambient aerosols collected at the Gosan site in spring of 2001 excluding AD periods. Secondary sources were not resolved in this study, because the chemical composition data of secondary particles including sulfate, nitrate, ammonium, and organic carbon matters was not used in PMF analysis. The parameters and results of PMF analysis are summarized in Table 1. The resolved sources include not only natural sources such

as soil dust, sea salt, Chinese aerosol, and volcano emission, but also eleven anthropogenic sources such as biomass burning, municipal incineration, coal combustion, oil heating furnace, residual oil fired boiler, gasoline vehicle, diesel vehicle, nonferrous and ferrous metal sources, and copper smelter. The influence of most anthropogenic sources was observed in the fine size range (0.07~1.15 μm) while the coarse particles (1.15~12 μm) mainly originated from natural sources. Especially, several anthropogenic sources such as gasoline vehicle, diesel vehicle, nonferrous metal source, and residual oil combustion were only resolved in the ultra-fine size range (0.07~0.75 μm).

Table 1. The sources identified in each of eight size ranges.

	Source	Major components	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7	Stage 8
1	Chinese soil	Si, Al, Fe, Ca, K, S, Pb	+	+	+	+	+	+	+	+
2	Soil dust	Si, Al, Fe, K, Ti, Ca	+	+	+					
3	Sea salt	Cl, S, K, Br	+	+	+					
4	Biomass burning	S, K, Cl, Si, Al				+	+	+	+	+
5	Municipal incineration	Cl, Fe, S, Al, Ca, Zn, Br, Pb	+				+		+	
6	Coal combustion	S, Si, K, Zn, Ca, Fe, As, Se				+	+	+	+	+
7	Oil heating furnace	S, Si, K, Ca, Fe, V, Pb				+	+	+		
8	Oil fired boiler	S, V, Si, Ni						+	+	+
9	Gasoline vehicle	S, Si, Ca Fe, Zn, Cl, K					+	+	+	+
10	Diesel vehicle	Si, S, Al, K								+
11	Ferrous metal source-C	Fe, Zn, Cu, Pb	+							
12	Nonferrous metal source	Cu, S, Zn, Fe, Cr, Pb							+	+
13	Ferrous metal source-F	Fe, Si, Al, K, Zn, Mn		+	+	+	+			
14	Copper smelter	S, Si, K, Fe, Zn, Cu, As, Pb						+	+	
15	Volcano emission	Si, Al, S, K, Ca, Fe, Ti					+			
	Number of samples		411	411	411	411	411	335	381	411
	Number of factor		5	4	5	5	8	6	8	7
	Q (Chi square)		37015	31996	30992	37007	12773	23331	5067	12175
	Fpeak		-0.4	-0.5	-0.9	-0.7	-0.7	-0.6	-0.7	-0.7

The profiles of fifteen sources were determined by taking an average of the similar profiles in different size ranges. The average mass profiles, obtained from PMF analysis, are shown in Figure 2 with the known profiles from the previous works. The resolved source profiles had very similar chemical composition to the corresponding existing source profiles: Chinese aerosol (He et al., 2001), soil dust (EPA profile 41340), marine aerosol (Watson, 1979), volcano emission (Davis et al., 1981), oil heating furnace (Mamoro et al, 1979a), oil fired boiler (EPA 13505), coal combustion (Mamuro et al., 1979a), field burning (EPA profile 42320), municipal incineration (EPA profile 17106), gasoline vehicle (Watson et al., 1994), diesel vehicle (Watson et al., 1994), nonferrous metal source (Mamuro et al.,1979b), ferrous metal related sources (Watson, 1979; Mamuro et al., 1979b), and copper smelter (Small et al., 1981). At this time, the known source profile and the resolved one didn't have exactly the same composition, because the source composition is slightly different from each individual emission source and the known profiles from previous works are simply averaged source composition regardless of aerosol size range. Figure 3 represents the corresponding temporal variations of these possible sources. The mass concentration of each source was calculated from the sum of scaled intensity values in the resolved size ranges. Overall, apparent differences in

temporal variations of these anthropogenic sources confirm the independence of the estimated source contributions.
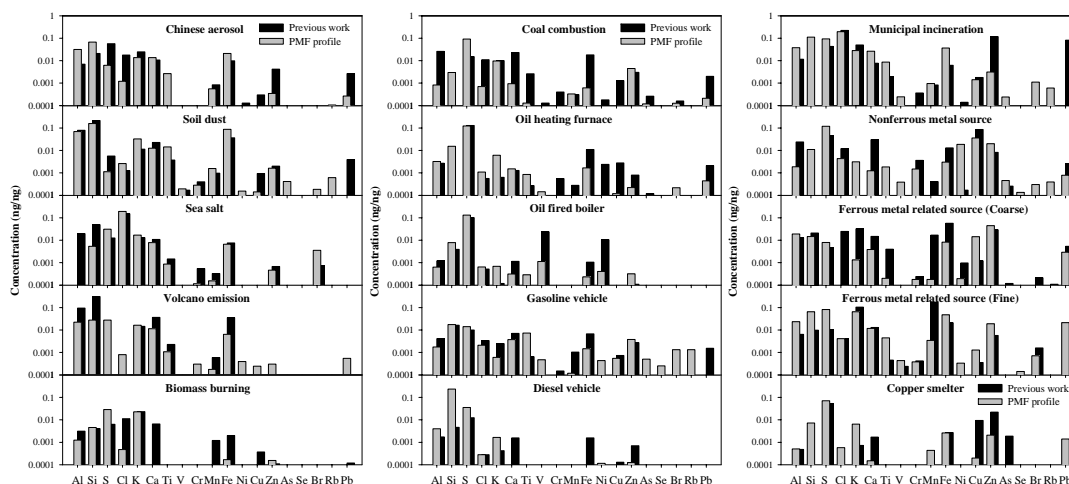

Figure 2. Average source profiles resolved from the size-segregated aerosol samples in this study and the corresponding known profiles.
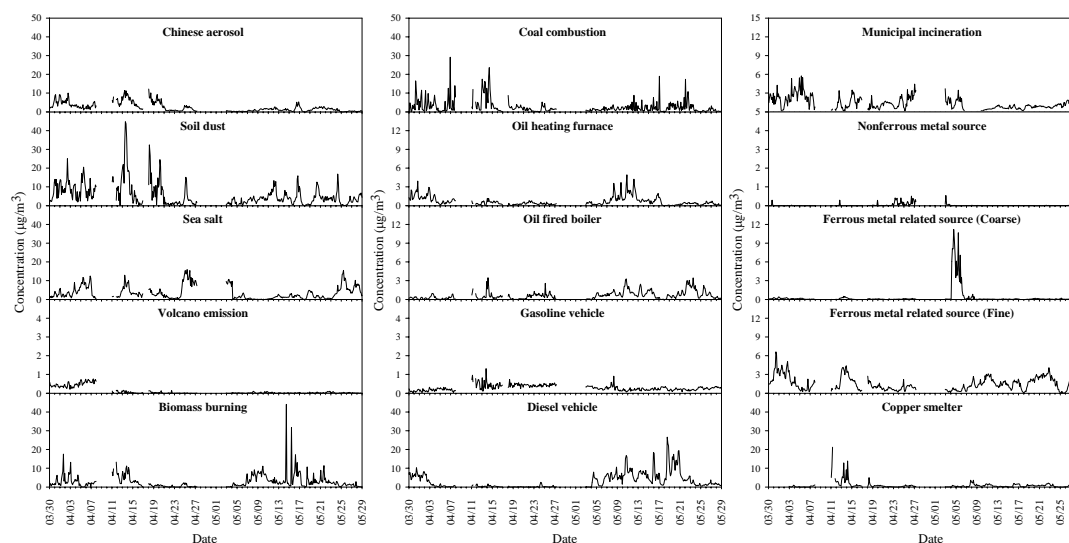

Figure 3. Temporal variations of total mass concentration by each of the resolved sources during the non-AD periods.

Chinese aerosol source usually has higher concentration of S and higher ratio of Ca to Al than general soil sources influenced by various industrial activities (He et al., 2001). This difference helped to identify these factors into different sources in the PMF analysis. These sources commonly contain the characteristic elements Si, Al, Fe, Ca, K and mainly contributed during the first half of the measurement period, especially before and after the AD outbreaks. However, they had different patterns in the temporal variation and size distribution; Chinese aerosol source has lower values and relatively small variations during the entire measurement period while soil dust reveals many sharp peaks over the entire measurement period. The sea salt factor

1394

characterized by high Cl and Br was also resolved in the same coarse size range as the soil dust source.

The factors characterized by high S and V in the fine size range (0.07~2.5 μm) could be represented by oil combustion source (Watson, 1997). However, these factors were separated into two different sources such as residual oil fired boiler and industrial oil heating furnace sources in order to account for the differences in their chemical composition. In the ultra-fine size range (0.07~0.56 μm), the factor containing larger amount of V and Ni represents the influence of residual oil combustion source (Swietlicki and Krejci, 1996). The factor characterized by K, Ca, and Pb in the relatively large size range (0.56~2.5 μm) represents the industrial oil combustion source (Kang, 2002). The resulting different temporal variations of the two sources support the separate treatment of these sources.

The coal combustion factor shows the presence of S and As in the fine size range (0.07~1.15 μm). Biomass burning source was characterized by K, S and Cl (Song et al., 2001) in the same size range. The municipal incineration source was described by Cl, S, Fe, Br, and Zn (Kang, 2002) in both coarse (5~12 μm) and fine size ranges (0.26~0.76 μm).

The motor vehicle sources were classified further into gasoline and diesel vehicle sources. The gasoline vehicle source, represented by S, Si, Ca, Fe, Cl, was resolved in the fine size range (0.07~0.75 μm). On the other hand, the diesel vehicle source characterized by Si, S, Al, K was only resolved in the ultra-fine size range (0.07~0.26 μm). The factor associated with S, Cu, Zn, and Pb which represents the industrial emission source mostly from the non-ferrous smelter (Lee et al., 1999) was resolved in the ultra-fine size range (0.07~0.34 μm). The ferrous metal source, mainly loaded with Fe, Mn, and Zn, was resolved in both coarse (5~12 μm) and fine size range (0.56~2.5 μm). In two size ranges, the ferrous metal related source revealed significantly different source composition and temporal variation, indicating that the detailed emission source and the source region are different in each size range. The other metallurgical source, copper smelter, was also resolved in the fine size range (0.56~0.75 μm). Finally the volcano emission source, containing large amount of Al, Si, K, Ca, and Fe, was resolved in the fine size range (0.56~0.75 μm).

The effectiveness of PMF analysis has been evaluated by the comparison of the predicted primary PM mass with the measured one as shown in Figure 4. The measured primary PM mass was obtained from the sum of element mass concentrations measured in eight size ranges, while the predicted one was calculated from the sum of scaled source contribution values for each sample. The correlation coefficient between the measured and predicted primary PM masses was 0.82, indicating that the resolved factors effectively accounted for the most of the variations in mass concentration of particulate elements.
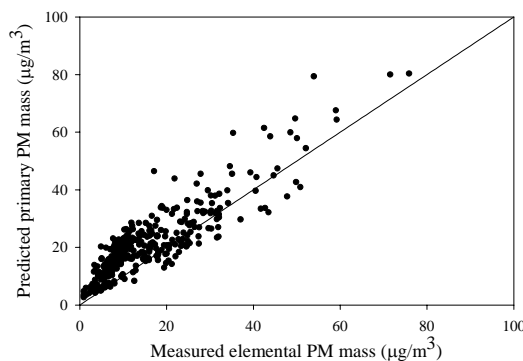
Figure 4. Comparison of the predicted primary PM mass from PMF analysis with the measured one.

4. 3. Source Contributions

The average contribution of each source to the measured total PM mass during the non-AD periods are shown in Figure 5. On average, the estimated fifteen sources from PMF analysis contributed to about 53% of total PM_{12} mass. At this time, soil dust contributed 11%, the largest portion of PM_{12} mass, and natural sources including soil dust, sea salt, and Chinese aerosol occupied about 23%. On the other hand, anthropogenic sources contributed to rather larger portion than natural sources, accounting for 30% of total PM_{12} mass. Especially, the contributions of diesel vehicle, biomass burning, coal combustion, ferrous metal source, and municipal incineration were large in the order, accounting for 6.7%, 6.1%, 5.4%, 3.0%, 2.7%, respectively.

Considering the fact that water soluble ion and carbon components are not included in PMF analysis, the rest fractions of total PM_{12} mass (47%), which can't be described by estimated sources, could be the portion of secondary aerosol components such as ammonium sulfate, ammonium nitrate, and secondary organic carbon matters
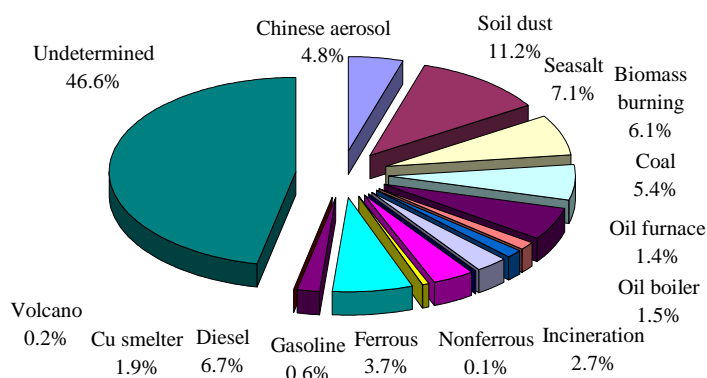

Figure 5. Average source contributions to the total particle mass concentration.

Temporal variations of the total PM_{12} mass and the estimated source contributions are shown in Figure 6, where the contribution by each source is accumulated. It shows that many large peaks in total PM mass are described by the contributions of the resolved sources from PMF analysis. However, the undetermined fraction of

PM$_{12}$ mass reveals rather different temporal variation from those estimated by sources, implying that this fraction originated not from primary emission sources but from secondary particulate matter. In addition, the contributions of various sources reveal different temporal variations, according to the aerosol size range. Source contributions in three different size ranges are assessed separately in Figure 8. Figure 8.b shows that the contribution of natural sources including soil dust, sea salt, and Chinese aerosol was dominant, making up for 79% in the coarse size range (2.5~12 μm). On the other hand, the contribution of anthropogenic sources, such as coal combustion, biomass burning, and diesel vehicle, sharply increased in the fine and ultra-fine size range (0.07~2.5 μm) as shown in Figure 8.c and 8.d. Especially, the contribution of diesel vehicle (52%) was mainly observed in the ultra-fine size range (0.07~0.56 μm) while the coal combustion source (33%) contributes mostly in the fine size range (0.56~2.5 μm). The effect of anthropogenic sources increased up to 98% of the total contribution in the ultra-fine size range (0.07~0.56 μm).
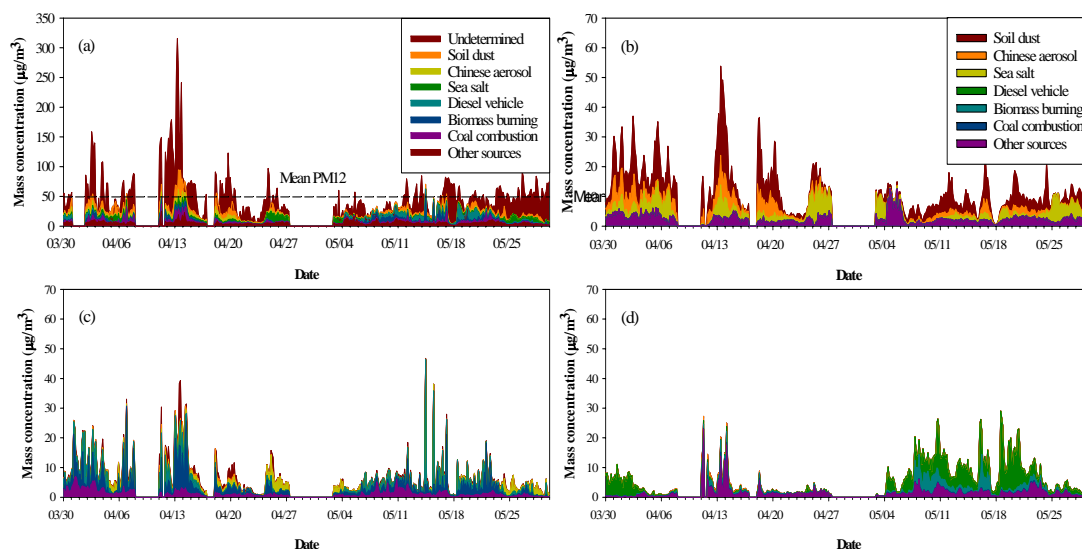

Figure 6. Temporal variation of PM mass contributed by different source types (a) in total size ranges, (b) in the coarse size range (2.5~12 μm), (c) in the fine size range (0.56~2.5 μm), and (d) in the ultra-fine size range (0.07~0.56 μm).

REFERENCES
Cahill T. A., Goodart, C., Nelson, J. W., Eldred, R. A., Nasstrom, J. S., Feeny, P. J., 1985. Design and evaluation of the DRUM impactor. Proceedings of the International Symposium on Particulate and Multi-Phase Processes (vol. 2). Ariman, T. and Nejat, T. (Eds.), Taylor and Francis, Philadelphia, Pa., pp. 319-325.

Cahill, T. A., and Wakabayashi, P.: Compositional analysis of size-segregated aerosol samples, in Measurement Challenges in Atmospheric Chemistry, edited by L. Newman, Adv. Chem. Ser., 232, 211-228, 1993.

Davis, B. L., Johnson, L. R., Griffen, D. T., Phillips, W. R., Stevens, R. K., and Maughan, D.: Quantitative Analysis of Mt. St. Helens Ash by X-Ray Diffraction and X-Ray Fluorescence Spectrometry, J. Appl. Meteorol., 20, 922-933, 1981.

He, K., Yang, F., Ma, Y., Zhang, Q., Yao, X., Chan, C. K., Cadle, S., Chan, T., Mulawa, P.: The characteristics of PM2.5 in Beijing, China, Atmos. Environ., 35, 4959-4970, 2001.

Hien, P. D., Bac, V. T., Thinh, N. T. H.: PMF receptor modeling of fine and coarse PM10 in air masses governing monsoon conditions in Hanoi, northern Vietnam, Atmos. Environ., 38, 189-201, 2004.

Kang, C. M.: Characteristics of the Fine Particles and Source Apportionments using the CMB model in Seoul Area. Doctoral dissertation, Konkuk University, Seoul, Korea, 2002.

Lee, E., Chan, C. K., Paatero, P.: Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, Atmos. Environ., 33, 3201-3212, 1999.

Mamuro, T. A., Mizohata, T. K.: Elemental Compositions of Suspended Particles Released from Various Boilers, *Annual Report of the Radiation Center of Osaka Prefecture* 20: 9-17, 1979a.

Mamuro, T. A., Mizohata, T. K.: Elemental Compositions of Suspended Particles Released from Iron and Steel Works, *Annual Report of the Radiation Center of Osaka Prefecture* 20: 19-28, 1979b.

Paatero, P. and Tapper, U.: Analysis of different modes of factor analysis as least squares fit problem, Chemometrics and Intelligent Laboratory Systems, 18, 183-194, 1993.

Paatero, P. and Tapper, U.: Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 11-126, 1994.

Paatero, P., Hopke, P. K., Song, X. H., Ramadan, Z.: Understanding and controlling rotations in factor analytic models, Chemometrics and Intelligent Laboratory Systems, 60, 253-264, 2002.

Perry, K. D., Cliff, S. S., and Jimenez-Cruz, M. P.: Evidence for hygroscopic mineral dust particles from the Intercontinental Transport and Chemical Transformation Experiment, J. Geophys. Res., 109, D23S28, 2004.

Small, M., Germani, M. S., Zoller, W. H., and Moyers, J. L.: Fractionation of Elements During Copper Smelting, Environ. Sci. Technol., 15, 299-304, 1981.

Song, X. H., Pollissar, A. V., Hopke, P. K.: Sources of fine particle composition in the northeastern US, Atmos. Environ., 35, 5277-5286, 2001.

US EPA (U.S. Environmental Protection Agency): Protocol for applying and validating the CMB model, EPA-450/4-87-010, 1987.

Watson, J. G.: Chemical Element Balance Recetor Model Methodology for Assessing the Source of Fine and Total Suspended Particulate Matter in Portland, Oregon, Ph. D. Thesis, Oregon Graduate Center, Beaverton, 1979.

Watson, J. G., Chow, J. C., Lu, Z., Fujita, E. M., Lowenthal, D. H., Lawson, D. R., Ashbaugh, L. L.: Chemical Mass Balance source apportionment of PM$_{10}$ during the

Southern California Air Quality Study, Aerosol Science and Technology, 21:1-36, 1994.