Mining airborne particulate size distribution data by positive matrix factorization

Liming Zhou,[1] Eugene Kim, and Philip K. Hopke

Center for Air Resources Engineering and Science and Department of Chemical Engineering, Clarkson University, Potsdam, New York, USA

Charles Stanier[2] and Spyros N. Pandis

Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[1]   Airborne particulate size distribution data acquired in Pittsburgh from July 2001 to June 2002 were analyzed as a bilinear receptor model solved by positive matrix factorization (PMF). The data were obtained from two scanning mobility particle spectrometers and an aerodynamic particle sampler with a temporal resolution of 15 min. Each sample contained 165 size bins from 0.003 to 2.5 μm. Particle growth periods in nucleation events were identified, and the data in these intervals were excluded from this study so that the size distribution profiles associated with the factors could be regarded as sufficiently constant to satisfy the assumptions of the receptor model. The values for each set of five consecutive size bins were averaged to produce 33 new size intervals. Analyses were made on monthly data sets to ensure that the changes in the size distributions from the source to the receptor site could be regarded as constant. The factors from PMF could be assigned to particle sources by examination of the number size distributions associated with the factors, the time frequency properties of the contribution of each source (Fourier analysis of source contribution values), and the correlations of the contribution values with simultaneous gas phase measurements (O_3, NO, NO_2, SO_2, CO) and particle composition data (sulfate, nitrate, organic carbon/elemental carbon). Seasonal trends and weekday/weekend effects were investigated. Conditional probability function analyses were performed for each source to ascertain the likely directions in which the sources were located. Five factors were separated. Two factors, local traffic and nucleation, are clear sources, but each of the other factors appears to be a mixture of several sources that cannot be further separated.

Citation: Zhou, L., E. Kim, P. K. Hopke, C. Stanier, and S. N. Pandis (2005), Mining airborne particulate size distribution data by positive matrix factorization, *J. Geophys. Res.*, *110*, D07S19, doi:10.1029/2004JD004707.

1. Introduction

[2]   There have been many studies that find relationships between elevated morbidity and mortality and higher particulate matter (PM) concentrations [*Dockery and Pope*, 1994; *Pope et al.*, 1995; *Brunekreef et al.*, 1995; *van Bree and Cassee*, 2000]. In order to develop an effective control strategy for airborne particles, the relationship between the sources and receptor concentrations needs to be understood.

[3]   Positive matrix factorization (PMF) is a powerful tool for solving receptor models with aerosol composition data and has been used successfully in identifying the sources of the airborne particles in many studies [*Xie et al.*, 1999; *Lee et al.*, 1999; *Ramadan et al.*, 2000; *Chueinta et al.*, 2000; *Polissar et al.*, 2001; *Song et al.*, 2001]. Recently, the aerosol size distribution data have been analyzed by principal component analysis [*Ruuskanen et al.*, 2001; *Wahlin et al.*, 2001] and PMF [*Kim et al.*, 2004].

[4]   It is obvious that particles in different size ranges have unique characteristics and one always needs to consider classifying the size range before detailed analyses. A three-modal structure was widely used to describe continental particle number size distribution, including a nucleation mode at 10–20 nm, an Aitken mode at 40–80 nm, and an accumulation mode at 100–300 nm [*Whitby*, 1978; *Raes et al.*, 1997; *Hussein et al.*, 2004]. Lognormal distribution functions representing the three modes were used to parameterize size distribution, and the modes fit by the lognormal distribution functions were thought to be associated with particle sources and processes during the transport [*Birmili et al.*, 2001]. Using constant modal distribution functions was also proved possible if particle growth processes from nucleation mode to Aitken or even accumulation mode were

---

not considered [*Mäkelä et al.*, 2000]. It is worth mentioning that for size distributions with a stable structure, using variant modal distributions instead of constant ones tends to be an overfitting, which means that the small random variations of the modes and even measurement errors may be fit and, consequently, the features of the modes may be smeared.

[5] The Pittsburgh supersite was located in a strong anthropogenic source region, and the size distribution measured there cannot be fit well by the aforementioned three-modal distribution [*Stanier et al.*, 2004c]. Thus some other methods need to be applied for classifying the size ranges.

[6] Previously, size distribution data from July 2001 measured by the Pittsburgh Air Quality Study (PAQS) have been analyzed to identify the particle sources by PMF [*Zhou et al.*, 2004]. Five sources were identified: secondary aerosol (from distant sources), local stationary combustion, remote traffic (probably from the interstate highway 1 mile away), local traffic (from the street and minor roads close to the receptor site), and nucleation, with decreasing sizes. The PMF method has successfully divided the particles into the five classes on the basis of both their size ranges and temporal behavior without prior assumptions such as log-normal distribution and the number of modes. The PMF analysis requires stationary or quasistationary size distributions measured at the receptor site. In a short period such as one month, conditions that might affect particle size changes such as photochemical activity or temperatures may be taken as relatively constant, and the change of the particle size distribution can be thought to be sufficiently constant to permit the PMF analyses.

[7] Our recent study has indicated linear relationships between size distribution data and chemical composition data simultaneously measured at the Pittsburgh supersite [*Zhou et al.*, 2005], and this is a direct proof of the stationarity of the size distribution.

[8] In this study, a larger data set from PAQS, containing 1 year of size distribution data from July 2001 to June 2002, has been analyzed for source identification. Data mining is a process to discover patterns and relationships in data using various tools. The tools used in this study, including PMF, will be introduced in the next sections. Each factor found by PMF can be thought of as a pattern that represents the variations of a source or a group of sources. The possible sources associated with each factor (or pattern) will be investigated.

[9] Over a full year the atmospheric processes influencing the size distributions vary significantly, and our previous efforts in directly analyzing the full year data set proved to be inappropriate [*Zhou et al.*, 2003]. Therefore this analysis will be performed on a month-by-month basis.

[10] In our previous work [*Zhou et al.*, 2004], the days with extensive nucleation events, especially those events followed by particle growth, were excluded, and this leads to an incomplete description of the summer situation at the Pittsburgh area. In this study, a special method has been designed to remove only the data representing particle growth events, and thus a more thorough and complete investigation will be made.

2. Description of the Data

[11] The receptor site was located in Schenley Park, Pittsburgh (latitude 40.4395°, longitude −79.9405°). An
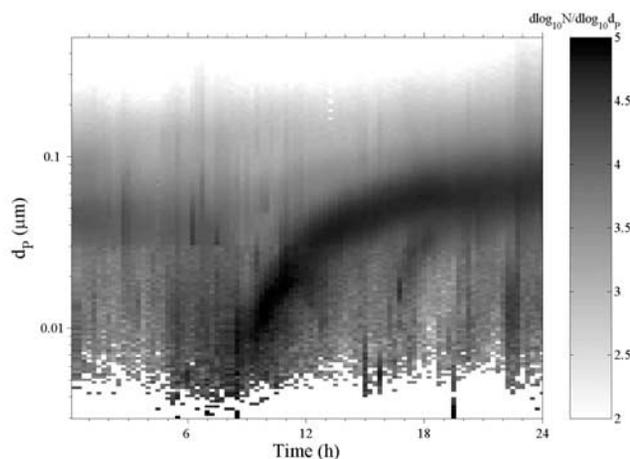

Figure 1. Typical particle growth on 2 July 2001.

overall summary and preliminary results for PAQS were given elsewhere [*Wittig et al.*, 2003]. The size distribution data were obtained from July 2001 to June 2002. Samples were collected and measured continuously every 15 min. The data were from two scanning mobility particle spectrometers (SMPS) and an aerodynamic particle sampler (APS). Above 583 nm, the data used in this study represent the electrical mobility diameter inferred from aerodynamic mobility and estimated density [*Khlystov et al.*, 2004]. The ratio of APS diameter and SMPS diameter is around 1.3. The samples were collected at 25% relative humidity, and "dry" particle distributions were obtained [*Stanier et al.*, 2004a].

[12] The gas phase concentrations (O_3, NO, NO_x, SO_2, and CO, with a 10 min resolution), particle mass and composition ($PM_{2.5}$, sulfate, and nitrate, with 10 min time resolution; organic carbon (OC) and elemental carbon (EC), with 2 to 4 hour resolution), and meteorological conditions (wind direction, wind speed, etc., with 10 min resolution) were measured at the same time and location as the size distribution data. A detailed description of the instrument and sampling methods can be found elsewhere [*Wittig et al.*, 2003; *Stanier et al.*, 2004b]. Each size distribution sample contained 165 geometrically equal sized intervals covering the particle size range of 0.003–2.5 μm.

3. Exclusion of Data Representing Particle Growth

[13] Nucleation events observed at the Pittsburgh site were classified as regional and short-lived by *Stanier et al.* [2004b]. In a regional nucleation event, it is often observed that the newly formed particles continue growing. Particle growth events are confined to limited time intervals and size ranges from over 10 nm up to accumulation mode size. A typical growth event after the nucleation is shown in Figure 1, where the number concentrations on 2 July 2001 are shown. These particle growth events were also observed at many other places, and the experimental and theoretical studies on this phenomenon were recently reviewed by *Kulmala et al.* [2004]. Discussions of the mechanisms of these events are beyond the scope of this study.
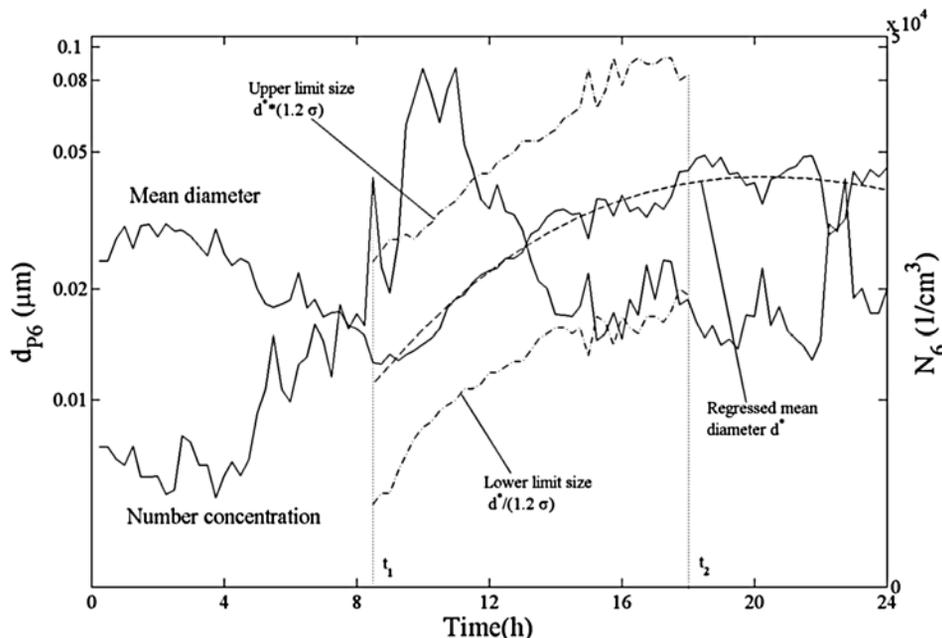
Figure 2. Number concentration and mean size variations on 2 July 2001.

[14] Receptor models require that the property being apportioned be stationary or, in this case, that there must be constant size distribution profiles associated with all factors. The temporal variations of the size distribution caused by particle growth events will deform the model solutions and make the sources unidentifiable. Particle growth during a regional nucleation event was identified, and the data representing the growth were removed by the following method. The number-concentration-weighted mean diameter is defined by the following equation:

$$\log d_{av} = \sum_{d_p} N(d_p) \log d_p / \sum_{d_p} N(d_p). \qquad (1)$$

In addition to the variation of number concentration dN/dt, the number-concentration-weighted diameter can also be used to investigate the nucleation event. If there is a nucleation event, the number concentration increases sharply, and the mean diameter should simultaneously drop sharply since the newly formed particles have the smallest sizes and the largest number concentrations.

[15] Since many particle growth events occurred in the size range above 6 nm, the total number concentration and mean diameter above 6 nm, N_6 and $d_{p6,av}$, were used. Here, N_6 is the total number concentration over 6 nm; $d_{p6,av}$ is computed by equation (1) for all particles larger than 6 nm. The beginning of particle growth is defined quantitatively as the point at which no significant decrease of mean diameter occurs when the number concentration has a sharp increase and the mean diameter rises gradually afterward. In Figure 2, t_1 is found to be the start of particle growth. The mean diameter $d_{p6,av}$ does not change with time when particle growth becomes stable. A polynomial regression of $\log d_{p6,av}$ against time

was made to estimate the time after which the variation in mean diameter is sufficiently small, as shown in Figure 2. When the first derivative of the regressed value is below a criterion value, such as at t_2 in Figure 2, the growth can be considered to have terminated.

[16] By assuming the number concentration distribution versus $d_{p6,av}$ to be lognormal (during the particle growth after the nucleation, the number concentration distribution over 6 nm is usually unimodal), we define the data, within $\pm 1.2\sigma$ from the regressed $d_{p6,av}$ and between t_1 and t_2, representing particle growth, where σ is the standard deviation of the lognormal distribution for each time interval. Figure 3 shows the result of this definition for 2 July 2001. The data thus defined were treated as missing values, which means replacing these values by the mean values and
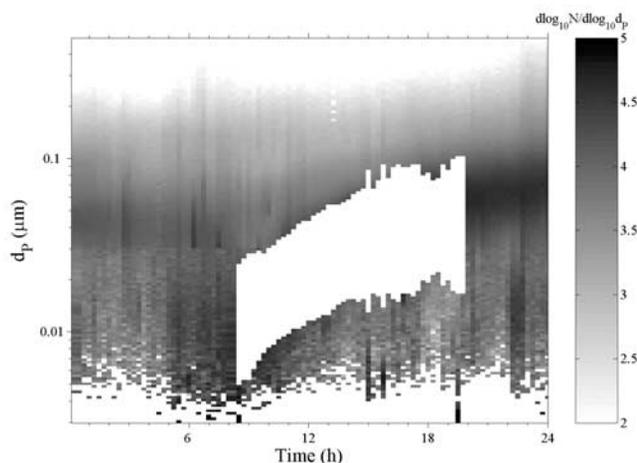
Figure 3. Definition of particle growth zone for 2 July 2001.

Table 1. Days With a Particle Growth Zone Defined (Data by This Definition Excluded in PMF Analysis)

Month	Days
July 2001	2, 11, 14, 15, 24, and 30 July
Aug. 2001	7, 11, 14, and 18 Aug.
Sept. 2001	1, 5, 11, 14, 16, and 26 Sept.
Oct. 2001	7, 10, 11, 15, 18, 20, 24, 25, 29, and 30 Oct.
Nov. 2001	3, 4, 8, 9, 10, and 22 Nov.
Dec. 2001	None
Jan. 2002	25 and 26 Jan.
Feb. 2002	13, 14, and 25 Feb.
March 2002	6, 7, 9, 15, 23, 24, and 29 March
April 2002	2, 10, 11, 12, 16, 17, 18, and 19 April
May 2002	4, 5, 10, 11, 15, 16, 17, 23, 24, 26, 27, 28, 29, 30, and 31 May
June 2002	2, 3, 4, 5, 8, 9, 21, 22, 23, 24, and 29 June

assigning them high uncertainties (low weights), such as 10 times their concentrations. Thus these data will not influence the PMF result and will not be reflected in the factors obtained by PMF. The data before and after this operation were inspected visually, and sometimes the parameters were adjusted manually to properly define the particle growth zone. If a regional nucleation event is followed with no particle growth, the corresponding data were not processed by this method. The days processed by this method are listed in Table 1. From this table, it can be

seen that this kind of particle growth phenomenon rarely happened in winter.

4. PMF and Other Tools for Mining the Size Distribution Data

4.1. PMF

[17] A detailed introduction to PMF can be found elsewhere [*Paatero*, 1997]. In this study, PMF2 (two-way PMF) is used to solve the following two-way receptor model:

$$X = GF + E \tag{2}$$

or, in the elemental form,

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}, \tag{3}$$

where X is the matrix of observed data and the element x_{ij} is the number concentration value of the ith sample at the jth size bin. G and F are the source contributions and size distribution profiles, respectively, of the sources that are unknown and to be estimated from the analysis. To be specific, g_{ik} is the concentration of particles from the kth source associated with the ith sample and f_{kj} is the size distribution associated with kth source. E is a matrix of residuals.

Table 2. Description of the Whole Year Data (Data Representing Particle Growth Excluded)

Size, μm	Number Concentration, number cm^{-3}							Mean Volume Concentration, μm^3 cm^{-3}
	Mean, × 10^3	SD,[a] × 10^3	Min	Max, × 10^3	Median, × 10^3	Lower 25%, × 10^3	Upper 25%, × 10^3	
0.0032	21.5	53	0	1550	7.8	0	23	0.00038
0.0039	11.2	21	0	496	5.7	0	13	0.00034
0.0046	8.8	12.7	0	241	5.4	2.4	10	0.00045
0.0055	8.4	10.1	0	306	5.9	3.2	9.7	0.00074
0.0066	8.8	9.7	0	667	6.9	3.9	10	0.0013
0.0079	9.9	14	0	2110	8.3	4.9	12	0.0026
0.0095	11.2	9.1	0	406	10.1	5.8	13	0.0050
0.0113	12.6	9.6	0	304	11.5	6.7	15	0.0096
0.0136	13.7	10.4	0	405	12.6	7.2	16	0.018
0.0163	14.3	12.5	0	688	12.0	6.9	17	0.030
0.0195	14.6	12.7	0	481	12.3	7.2	17	0.056
0.0233	15.0	13.3	0	371	12.3	7	17	0.10
0.0279	15.5	15.1	0	1170	12.7	7.5	18	0.18
0.0334	16.1	13.9	0.6	444	13.4	7.8	20	0.31
0.0400	16.2	13.2	0.44	224	13.7	7.7	19	0.54
0.048	15.6	12.0	0	150	13.6	7.6	19	0.90
0.057	14.6	10.8	0	131	12.6	7.2	17	1.4
0.069	13.0	9.6	0	124	11.2	6.7	15	2.2
0.082	11.2	8.1	0	118	9.8	6.0	13	3.2
0.098	9.5	6.7	0	96.9	8.5	5.1	11	4.7
0.118	7.8	5.2	0	51.8	6.9	4.1	9.3	6.6
0.141	6.03	3.87	0	38.0	5.3	3.3	7.3	8.8
0.169	4.43	2.74	0	23.5	3.8	2.5	5.7	11
0.20	3.11	1.90	0	19.2	2.6	1.7	4.1	13
0.24	2.09	1.36	0.32	19.0	1.7	1.1	2.7	15
0.29	1.31	0.99	0	11.0	0.98	0.62	1.7	16.5
0.35	0.74	0.68	0	5.9	0.51	0.29	0.96	16
0.41	0.37	0.41	0	4.1	0.22	0.12	0.45	14
0.50	0.16	0.21	0.43	2.6	0.089	0.045	0.19	11
0.63	0.087	0.092	1.81	0.84	0.056	0.027	0.109	11
0.90	0.01	0.010	0.23	0.21	0.0075	0.0037	0.12	3.8
1.29	0.0021	0.0018	0.1	0.054	0.00190281	0.0012	0	2.3
1.84	0.0011	0.0009	0	0.036	0.0009	0.0007	0.0013	3.6

[a]SD, standard deviation.
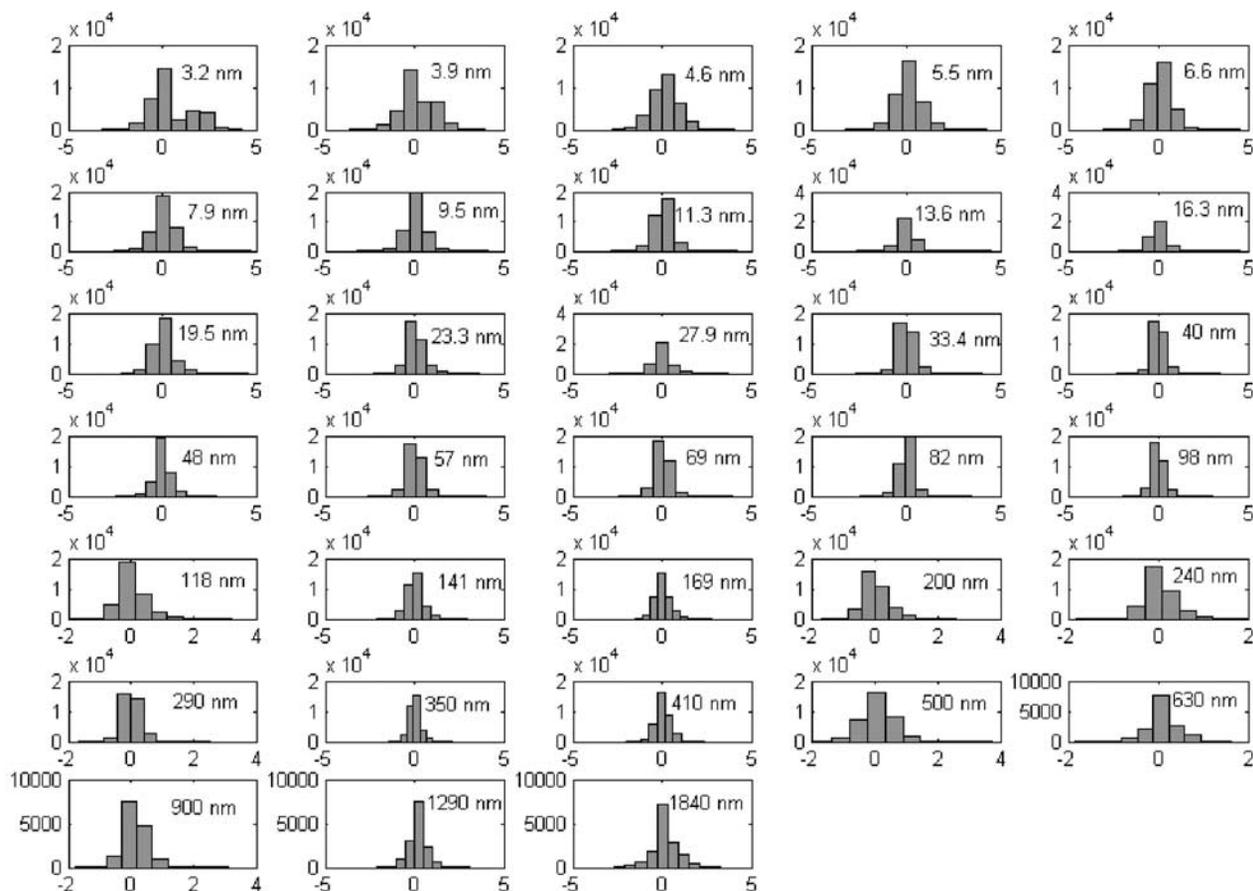
Figure 4. Histograms of the scaled regression residuals. The horizontal axes indicate the scaled residual, and the vertical axes indicate the frequencies.

[18] The residual sum of squares (Q) is minimized by finding the optimal F and G values as indicated in equations (2) and (3):

$$Q = \left\| \frac{(X - GF)}{s} \right\|_{F,G}^2 = \sum_i \sum_j \left(\frac{e_{ij}}{s_{ij}} \right)^2. \tag{4}$$

[19] Rotations are obtained by setting an FPEAK value [*Paatero et al.*, 2002]. When the FPEAK value is positive, the following additional term is included in the object function Q:

$$Q^P = \beta^2 \left(\sum_{k=1}^p \sum_{j=1}^n f_{kj} \right)^2, \tag{5}$$

where β^2 corresponds to the FPEAK value. The term defined above attempts to pull the sum of all the elements of F toward zero and makes the program do elementary transformations for F and G by subtracting the F vectors from each other and adding corresponding G vectors to obtain a more physically realistic solution. Recently, a method for solving rotational ambiguities by detecting edges in G space have been developed, and the best solution is obtained without clearly defined edges in G space [*Paatero et al.*, 2005]. In this study, the FPEAK

value is set to 0.2 for each of the months to eliminate G space edges, and five similar factors were found for each month.

[20] The F values were normalized to make the sum of each size interval equal to 1 for each factor. The G values were then rescaled correspondingly. The uncertainties were estimated with the method described by *Zhou et al.* [2004]. To smooth the size distribution data and minimize the error caused by the discontinuity between instruments, every five consecutive size bins were combined into one, and 33 new size intervals were produced from the original 165 size bins. The analysis of volume size distribution data does not provide much more information since it generates similar source contributions as number size distribution analyses do [*Zhou et al.*, 2004], and hence it will not be performed in this study. During the period between October 2001 and April 2002 the APS was not working, and only 29 new size intervals were produced for these months. This also caused incomplete information on volume size distribution. Table 2 summarizes the data set after the aforementioned treatment.

4.2. Correlation Analyses

[21] Correlations of some compositions and gases with source contributions were investigated. Using only gas data as an input to PMF will not bring much more information, and both data sets need to be averaged to 30 min. The 15 min resolution has made possible a more thorough time series
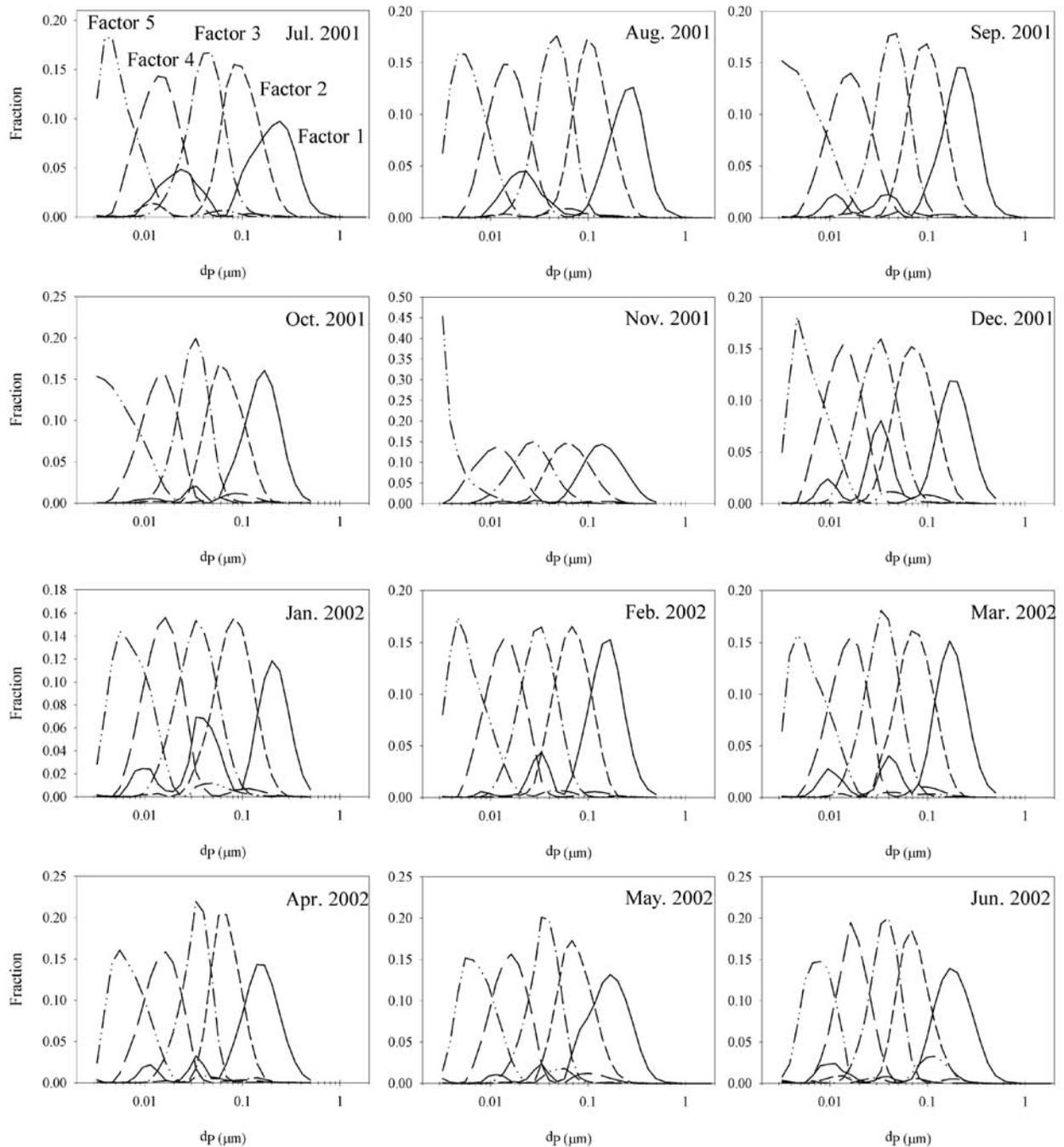
**Figure 5.** Size distribution profiles for each month.

analysis than 30 min resolution, such as in the time of day effect.

4.3. Conditional Probability Function (CPF)

[22] A conditional probability function [*Ashbaugh et al.*, 1985; *Kim et al.*, 2004] was calculated using the source contributions obtained by PMF2 and wind direction values by the following equation:

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}}, \quad (6)$$

where $m_{\Delta\theta}$ is the number of occurrences in the direction sector that exceeds the threshold, upper 25th percentile of the fractional contribution from each source, and $n_{\Delta\theta}$ is the total number of wind occurrences in the same direction sector. Fractional contributions are used to avoid the influence of atmospheric dilution on CPF results. The angular width of direction sector is set as 10°, and thus there are 36 directional sectors. Those samples corresponding to a wind speed below 1.0 m s^{-1} are excluded from this study, and two thirds of the total samples were excluded. The sources are thought to be located in the direction sectors
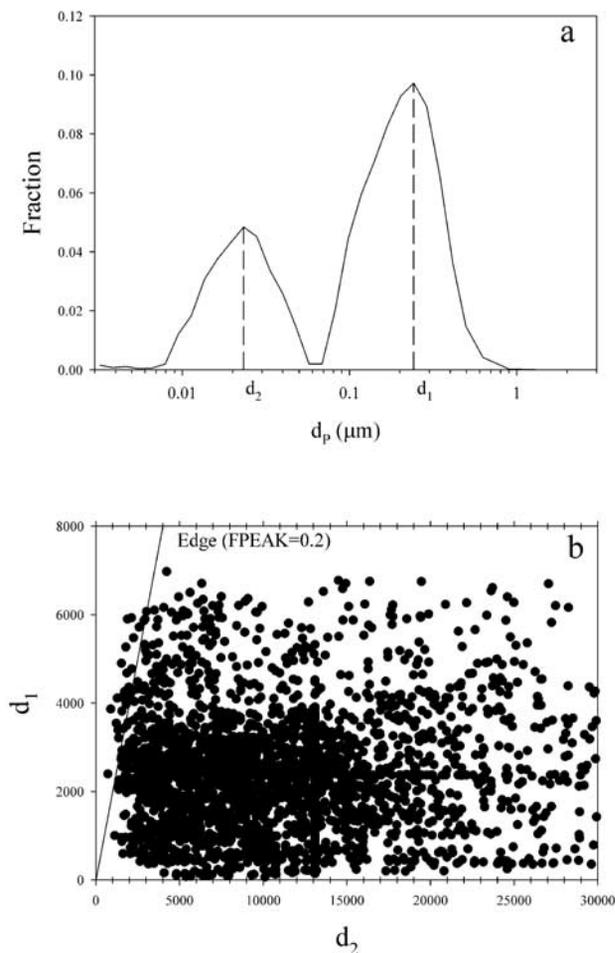
Figure 6. Discussion of the two modes of factor 1 in July 2001. (a) Size distribution of factor 1. (b) Scatterplot of the number concentration of the central size intervals of the two modes.

with high CPF values. When $n_{\Delta\theta}$ is below 10, the CPF value is set to zero.

4.4. Fourier Analysis

[23] The source contributions from each month are linked in a temporal sequence to form annual source contribution series, and Fourier analyses have been performed for these annual series. In the signal-processing field, it is well known that Fourier analysis has no temporal specificity. For example, if the Fourier analysis says that the series has a daily pattern (24 hour period), it cannot determine that the daily pattern exists at any specific time in the series. Therefore many time-frequency analysis techniques have been developed and used in combination with Fourier analyses. To address the problem in this study, the Fourier analysis was performed for the source contributions of each month. The correlations of the source contributions with the gas and particle composition data were calculated for each month.

5. Results and Discussion

[24] Table 2 describes the average number and volume size distribution of the full year of data. The volume distribution mode is located between 0.1 and 1 μm. The smallest particles

have the largest mean number concentration, but their median concentrations are not the highest. This situation may be caused by the frequent nucleation events that produce large numbers of new particles. Table 2 also indicates that the minimum number concentrations in all the size intervals are nearly zero, implying very low contributions from certain sources. The 12 months of data were summarized and described in detail elsewhere [*Stanier et al.*, 2004c]. Figure 4 indicates the histograms of the scaled regression residuals for all size intervals of all 12 months. Most of the residuals are within −2 and +2, and the distributions are symmetric and close to normal distributions. If the size distributions measured at the receptor are not stationary or quasistationary, then these good regressions cannot be obtained. The five factors are arranged in order of decreasing size from factor 1 to factor 5. These five factors are similar to the sources identified in our previous study [*Zhou et al.*, 2004]. Detailed discussion is provided as follows.

[25] Factor 1 has a number mode between 0.15 and 0.25 μm and a submode at 0.02 μm, as shown in Figure 5. Detecting edges in *F* space is a method to help resolve rotation problems [*Henry*, 2003]. For July 2001 the number concentrations of the two size intervals d_1 (240 nm) and d_2 (23 nm), the center size intervals of the two modes, were plotted in Figure 6. The edge corresponding to FPEAK = 0.2 was also plotted. When pulling down the submode by increasing the FPEAK value, the edge will rotate toward the d_1 axis, but it is clear that the edge cannot overlap with the d_1 axis, suggesting that the submode cannot be eliminated completely. Nevertheless, for the solutions with different rotations, the major mode and source contributions did not experience large changes, and our analysis and conclusions were not severely influenced.

[26] In Figure 7, the Fourier analysis shows a weak daily pattern for factor 1. Table 3 indicates that the daily pattern is clear in February and April of 2002 but not that clear in other months. Fourier analyses have also been applied for sulfate and nitrate data for each month. These analyses found that sulfate does not have daily patterns except in the summer months (it had weak daily patterns in July and August of 2001 and June of 2002) and nitrate usually has a strong diurnal pattern during the whole year except in November and December of 2001 and January and March of 2002. On the basis of these facts, one possible explanation for the periodicity of factor 1 is as follows: When the sulfate concentration is low and nitrate has a strong simultaneously diurnal pattern, such as the situation of February and April of 2002, the temporal behavior of factor 1 can be influenced by nitrate and shows observable daily patterns. Another possible reason is that some local emissions have been included in factor 1. However, the reason cannot be completely clarified within the current data sets.

[27] Factor 1 has a strong correlation with $PM_{2.5}$ mass over the entire period as indicated in Table 4 and is related to the major components of the $PM_{2.5}$ mass at the Pittsburgh supersite. In winter the correlation of nitrate with factor 1 increases, suggesting that more nitrate is included in factor 1.

[28] As indicated in Table 4, factor 1 usually has high correlations with sulfate. Secondary sulfate was formed by the oxidation of SO_2 through photochemical reactions.
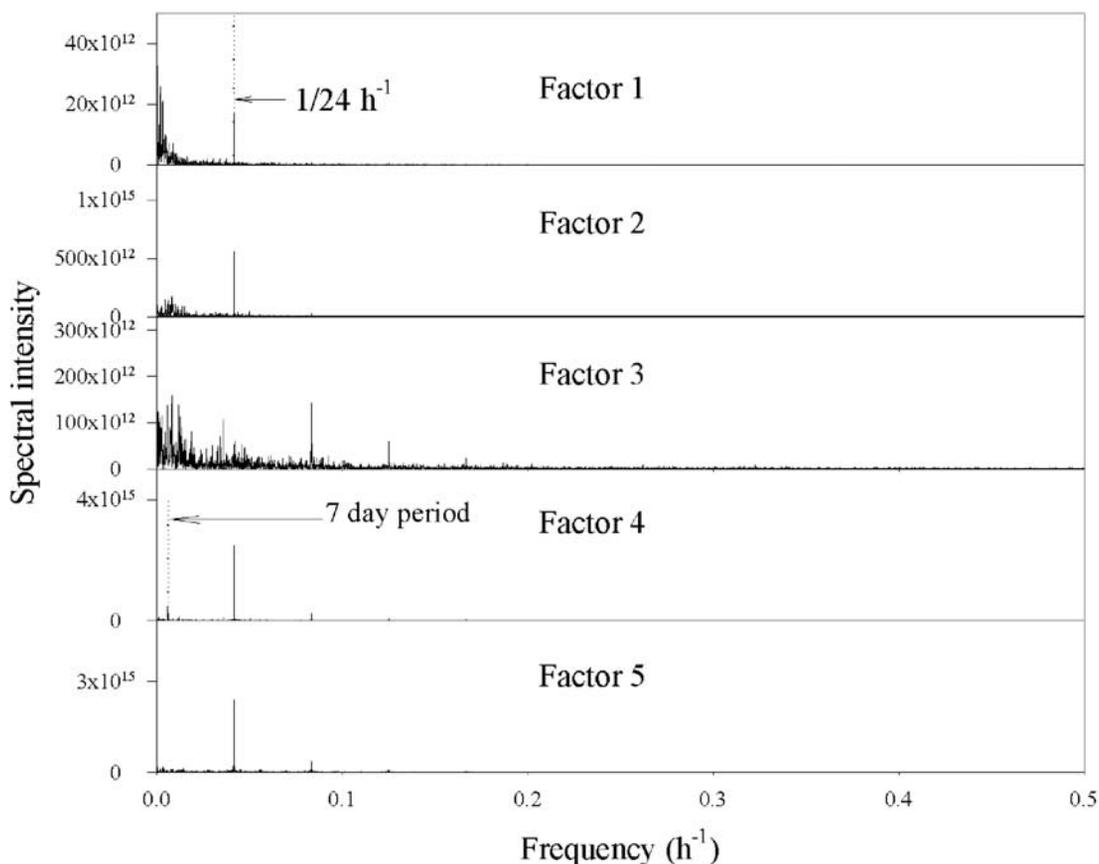
Figure 7. Fourier transformation for the annual contribution of each factor. (The powers are all nearly zero for all factors from 0.2 h^{-1} to 2 h^{-1}, the Nyquist frequency.)

The main SO_2 sources are coal-fired power plants hundreds of kilometers away. In winter the correlation of factor 1 with nitrate increases noticeably. The lower temperatures support ammonium nitrate in the solid phase, and the fraction of nitrate in $PM_{2.5}$ mass increases. The increase of the correlation coefficient with nitrate may explain the higher correlation of factor 1 with $PM_{2.5}$ than with sulfate during winter, and this is also consistent with the fact that there was much less sulfate in winter.

[29] Factor 1 has a higher correlation with OC than with EC in the summer and fall of 2001, but the correlations are similar for both species in the first half of 2002. The higher correlation with OC than with EC in the summer may be caused by secondary organic matter condensing onto pre-existing particles during the transport as well as more production of secondary OC in the summer. When primary OC dominates, the correlations with OC and with EC are similar since they are from the same sources and have similar temporal variations. In Figure 8 the CPF analysis shows that factor 1 is from the south. Table 5 summarizes the characteristics of factor 1 and other factors.

[30] Factor 1 typically includes sulfate, nitrate (in winter), and secondary organics as well as primary organics that have aged in the atmosphere and have grown from their original size. The particles of the submodes cannot be from distant places; otherwise, they would be depleted during the transport. The small modes of factor 1 are more likely to be from fresh emissions of some combustion sources south of

the site. Since the dominating direction of factor 1 is south, the particles of the small mode have the same variations as other large particles in factor 1, which may explain the correlation with CO. Factor 1 includes secondary, aged primary aerosol particles and also fresh primary particles from local combustion sources.

[31] The number mode of factor 2 is at $0.08 \sim 0.1$ μm in July, August, and September of 2001. In 2002 the number mode is at $0.06 \sim 0.07$ μm. In Figure 9 a daily pattern, caused by the reduction of mixing height at night, is clearly observed. The nocturnal increase of the source contribution, owing to the reduction of the mixing height, may be the

Table 3. Results of Monthly Fourier Analysis for Daily Patterns[a]

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
July 2001	no	weak	no	strong	strong
Aug. 2001	no	no	no	strong	strong
Sept. 2001	very weak	very weak	no	strong	strong
Oct. 2001	very weak	weak	no	strong	strong
Nov. 2001	no	very weak	no	strong	strong
Dec. 2001	no	no	no	strong	strong
Jan. 2002	no	no	no	strong	strong
Feb. 2002	weak	very weak	no	strong	strong
March 2002	no	no	no	strong	strong
April 2002	weak	weak	no	strong	strong
May 2002	no	weak	no	strong	strong
June 2002	no	weak	no	strong	–

[a]The strength of the daily patterns was decided by comparing the spectral intensity at 1/24 h^{-1} with other frequencies.

Table 4. Correlations r of the Source Contributions With Gas and Particle Composition Data[a]

	O_3	NO	NO_x	SO_2	CO	$PM_{2.5}$	Sulfate	Nitrate	OC	EC
					Factor 1					
July 2001	0.18	0.03	0.18	0.16	0.19	0.87	0.85	0.23	0.63	0.28
Aug. 2001	−0.02	0.20	0.29	0.07	0.42	0.75	0.54	0.29	0.56	0.42
Sept. 2001	0.10	0.26	0.34	0.21	0.38	0.80	0.78	0.46	−[b]	−[b]
Oct. 2001	0.03	0.32	0.41	0.27	0.51	0.73	0.75	0.61	0.67	0.56
Nov. 2001	−0.05	0.42	0.48	0.33	0.57	0.90	0.68	0.40	0.89	0.67
Dec. 2001	−0.47	0.46	0.45	0.14	0.61	0.81	0.47	0.54	0.74	0.76
Jan. 2002	−0.42	0.51	0.56	0.17	0.47	0.70	0.48	0.67	0.56	0.55
Feb. 2002	−0.53	0.70	0.75	0.34	0.76	0.91	0.61	0.54	0.82	0.76
March 2002	−0.50	0.43	0.57	0.30	0.59	0.88	0.62	0.57	0.65	0.69
April 2002	−0.33	0.44	0.52	0.18	0.52	0.73	0.58	0.45	0.51	0.43
May 02	−0.12	0.39	0.45	0.22	0.50	0.85	0.74	0.31	0.80	0.66
June 02	0.18	0.13	0.35	0.17	0.49	0.82	0.63	0.31	0.58	0.56
					Factor 2					
July 2001	−0.18	0.47	0.57	0.48	0.54	0.29	0.15	0.54	0.46	0.56
Aug. 2001	−0.15	0.34	0.48	0.32	0.35	0.25	0.23	0.31	0.31	0.39
Sept. 2001	−0.24	0.60	0.73	0.32	0.77	0.61	0.52	0.58	−[b]	−[b]
Oct. 2001	−0.46	0.66	0.78	0.46	0.56	0.46	0.23	0.61	0.43	0.58
Nov. 2001	−0.51	0.74	0.79	0.42	0.79	0.48	0.23	0.34	0.41	0.75
Dec. 2001	−0.46	0.53	0.53	0.36	0.67	0.81	0.47	0.40	0.80	0.81
Jan. 2002	−0.37	0.59	0.71	0.44	0.59	0.72	0.61	0.45	0.71	0.71
Feb. 2002	−0.61	0.71	0.78	0.58	0.70	0.79	0.56	0.38	0.66	0.71
March 2002	−0.50	0.53	0.68	0.59	0.61	0.70	0.57	0.43	0.53	0.61
April 2002	−0.41	0.50	0.59	0.48	0.54	0.49	0.46	0.34	0.36	0.34
May 2002	−0.29	0.60	0.68	0.47	0.64	0.65	0.41	0.38	0.64	0.71
June 2002	−0.21	0.36	0.58	0.27	0.61	0.45	0.19	0.52	0.50	0.58
					Factor 3					
July 2001	−0.24	0.37	0.40	0.23	0.28	−0.14	−0.23	0.23	−0.04	0.26
Aug. 2001	−0.27	0.32	0.44	0.18	0.31	−0.01	−0.11	0.25	0.14	0.35
Sept. 2001	−0.31	0.36	0.51	0.29	0.43	0.13	0.07	0.30	−[b]	−[b]
Oct. 2001	−0.24	0.39	0.45	0.24	0.22	0.13	0.02	0.28	0.08	0.27
Nov. 2001	−0.32	0.33	0.36	0.15	0.21	0.04	−0.15	0.17	−0.07	0.21
Dec. 2001	−0.25	0.20	0.18	0.34	0.24	0.34	0.13	0.31	0.27	0.30
Jan. 2002	−0.20	0.38	0.42	0.40	0.30	0.28	0.41	0.35	0.20	0.33
Feb. 2002	−0.23	0.16	0.20	0.45	0.08	0.11	0.19	0.10	−0.10	0.02
March 2002	−0.32	0.35	0.42	0.52	0.28	0.31	0.36	0.22	0.14	0.25
April 2002	−0.28	0.30	0.37	0.59	0.28	0.22	0.19	0.17	0.02	0.13
May 2002	−0.21	0.33	0.41	0.49	0.27	0.15	0.01	0.11	0.06	0.21
June 2002	−0.17	0.22	0.30	0.21	0.20	0.02	−0.05	0.20	0.04	0.17
					Factor 4					
July 2001	−0.10	0.18	0.19	0.06	0.01	−0.03	−0.06	0.09	−0.11	0.10
Aug. 2001	−0.13	0.23	0.21	0.07	−0.05	−0.20	−0.23	0.01	−0.17	0.03
Sept. 2001	−0.09	0.17	0.11	0.12	−0.10	−0.18	−0.20	−0.03	−[b]	−[b]
Oct. 2001	0.01	0.11	0.07	0.18	−0.03	−0.11	−0.03	0.02	−0.07	0.04
Nov. 2001	0.20	−0.17	−0.18	0.06	−0.23	−0.11	−0.17	−0.10	−0.16	−0.16
Dec. 2001	0.05	−0.04	−0.04	0.06	−0.09	−0.10	−0.20	0.06	−0.15	−0.02
Jan. 2002	0.05	0.16	0.14	0.16	0.07	−0.10	0.02	0.16	−0.15	0.03
Feb. 2002	0.11	−0.07	−0.08	0.13	−0.19	−0.23	−0.09	−0.10	−0.27	−0.16
March 2002	−0.10	0.14	0.15	0.21	0.06	0.10	0.12	0.14	0.01	0.10
April 2002	−0.03	0.08	0.07	0.34	−0.08	−0.05	−0.07	−0.03	−0.19	−0.02
May 2002	0.02	0.03	0.05	0.24	−0.08	−0.08	−0.11	−0.03	−0.17	−0.07
June 2002	−0.04	0.21	0.13	0.21	0.04	−0.05	−0.08	0.01	−0.10	0.01
					Factor 5					
July 2001	0.08	−0.13	−0.16	−0.02	−0.15	−0.15	−0.11	−0.19	−0.25	−0.18
Aug. 2001	0.12	−0.05	−0.12	0.03	−0.17	−0.22	−0.18	−0.19	−0.24	−0.17
Sept. 2001	0.09	−0.11	−0.20	0.03	−0.19	−0.24	−0.20	−0.24	−[b]	−[b]
Oct. 2001	0.17	−0.17	−0.25	0.03	−0.25	−0.27	−0.14	−0.27	−0.19	−0.12
Nov. 2001	0.17	−0.12	−0.15	0.13	−0.14	−0.11	−0.14	−0.07	−0.14	−0.17
Dec. 2001	0.27	−0.06	−0.08	−0.07	−0.10	−0.19	−0.21	−0.13	−0.15	−0.08
Jan. 2002	0.20	−0.08	−0.15	−0.10	−0.16	−0.28	−0.25	−0.21	−0.26	−0.17
Feb. 2002	0.35	−0.16	−0.21	−0.03	−0.20	−0.27	−0.20	−0.22	−0.22	−0.20
March 2002	0.21	−0.08	−0.14	−0.08	−0.11	−0.11	−0.10	−0.09	−0.02	−0.03
April 2002	0.13	−0.01	−0.05	0.16	−0.09	−0.03	0.00	−0.15	−0.08	0.02
May 2002	0.18	−0.11	−0.14	0.11	−0.14	−0.08	−0.11	−0.16	−0.14	−0.13
June 2002	−[c]	−[c]	−[c]	−[c]	−[c]	−[c]	−[c]	−[c]	−[c]	−[c]

[a]The source contributions and the species with 10 min resolution were averaged to 30 min; for OC and EC the source contributions were averaged to the corresponding OC/EC sampling period.

[b]OC or EC data are missing.

[c]Half of the data between 3 nm and 15 nm were missing during this month because the SMPS was not functioning properly.
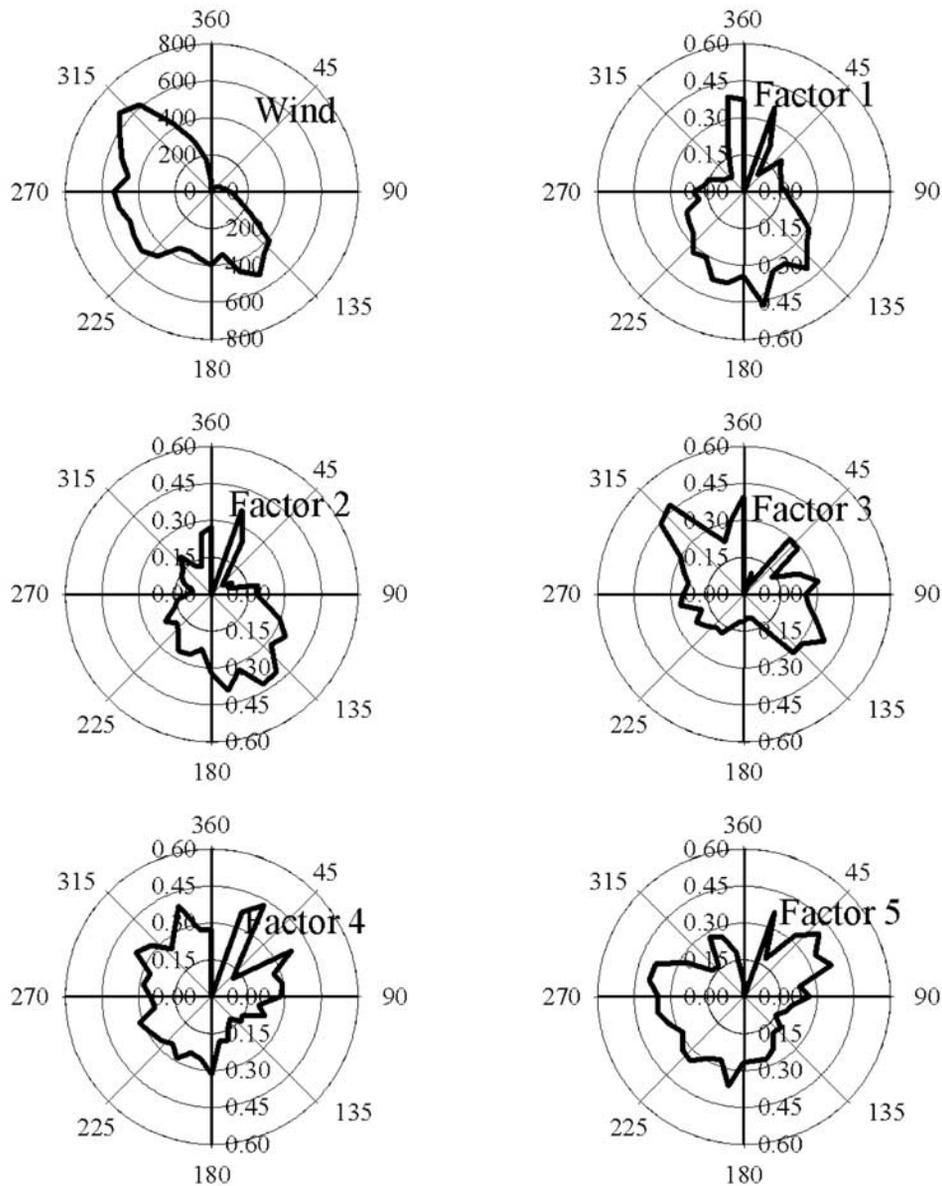
Figure 8. Wind profile and CPF of the whole year contribution of each factor.

cause of the negative correlation with ozone. Table 3 shows that the diurnal pattern of factor 2 does not appear in all months, indicating that this daily pattern is weak and is easily disturbed.

[32] The correlations between factor 2 and these gases, NO, NO_x, CO, and SO_2, suggest that the particles are from combustion sources and arrived at the receptor site accompanied by these gases. The strong correlation with SO_2 suggests coal burning, including coal power plants and steel mills. The correlation with CO can be explained by the emission from steel mills as well as a boiler 1 km away to the northwest of the site.

[33] Factor 2 has a weak correlation with OC in summer, and in winter the correlation becomes stronger. Like factor 1, the winter correlations with OC and with EC are similar, indicating primary OC sources. Factor 2 probably contains both emissions from wood burning and other local com-

bustion sources that are too similar in size to separate. The average contribution of factor 2 in the winter months (December 2001, January 2002, February 2002, and March 2002) is higher than that in the summer months (July 2001, August 2001, and September 2001) by 34%. In comparison, the average number contribution of factor 3 increases from summer to winter by only 17%. The higher increase of the factor 2 contribution may suggest an additional source from wood burning. As shown in Figure 9, factor 2 has no significant differences between weekdays and weekends. The CPF analysis shows it is also from the south to southeast.

[34] Factor 2 is thus assigned as stationary combustion, including emissions from local combustion sources. Probably, it may also include wood burning in winter. The emission from the boiler may also be included, except in summer when it seldom ran. The higher correlation of factor

Table 5. Summary of the Characteristics of All Factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Sources assigned	secondary and aged primary aerosol, fresh primary aerosol from local combustion sources	stationary combustion sources	remote Pittsburgh traffic, local point sources	local traffic	local nucleation
Size range	0.15 ~ 0.25 μm	0.08 ~ 0.1 μm in July, August and September 2001; 0.06 ~ 0.07 μm in other months	0.03 ~ 0.04 μm	15 nm	<10 nm
Diurnal pattern[a]	very weak	weak	no	strong	strong
Weekday/weekend difference	small	no	moderate	significant	no
Correlations with gas and particle composition data	no correlation with ozone; strong correlation with sulfate; correlations with other species	negative correlation with ozone; strong correlations with other gases; correlations with sulfate, nitrate and OC/EC	weak correlations with NO, NO_x and SO_2; no correlations with other species	no obvious correlations with any species	a weak correlation with ozone; no correlations with other species
Dominating direction by CPF	south	south and southeast	southeast and northwest	no clear dominating directions	no clear dominating directions

[a]Monthly results can be found in Table 4.

2 with SO_2 and lower correlation with sulfate suggests that the sources are located close to the receptor site, <100 km [*Zhou et al.*, 2004]. This is also consistent with results for a coal-fired power plant that was found to increase SO_2 close to the source but that produced little sulfate because of the short distance and time for oxidation [*Green et al.*, 2003].

[35] The number mode of factor 3 is at 0.04 μm in summer and 0.03 μm in the fall and winter. This factor has been associated with Pittsburgh traffic in a previous paper [*Zhou et al.*, 2004], but its behavior is a little puzzling here. Factor 3 is weakly correlated with NO, NO_x, and CO during all months. Figure 7 and Table 3 both indicate that it has no daily patterns. Table 6 also indicates that factor 3 has no correlation with the traffic flow on Interstate 376. One reason may be that the distance of transport weakened the daily pattern and the correlations, and meteorological conditions have more influence than the emissions. Another possible reason is the presence of some particles from local point sources in factor 3. Figure 7 indicates frequency peaks at $1/12\ h^{-1}$, $1/8\ h^{-1}$, and $1/6\ h^{-1}$ for factor 3. These are the harmonic frequencies of $1/24\ h^{-1}$, caused by a nonsinusoidal periodicity of 24 hours. Although these frequencies may suggest a hidden daily pattern, they only appeared in January and February of 2002, and we cannot conclude that factor 3 has daily patterns.

[36] Taking the periodicity of a series as "signal" and the nonperiodicity as "noise," then factor 3 has weak signals as just discussed, and noise is then sufficiently large to prevent a detectable 24 hour periodicity by Fourier analyses. The effect of the noise is random, and it affects any time of day to the same extent. Since there are a large number of samples for each time of weekday and weekend, we can expect a horizontal line with small fluctuations for the daily average in Figure 9. Thereafter, the concentration peak of factor 3 at morning rush hours as well as the weekday/weekend difference in Figure 9 suggests the influence of traffic.

[37] Although the number size distribution changes significantly within ~100 m of highways [*Zhu et al.*, 2002a, 2002b, 2004], the mode of the size distribution becomes stable at farther distances, and factor 3 is consistent with the unimodal size distributions found several kilometers downwind of the highway in Los Angeles [*Kim et al.*, 2002].

[38] Factor 3 is probably a collection of point sources and remote Pittsburgh traffic. These are particles produced in the city but not close to the measurement station, several kilometers away. The current techniques cannot separate these two source categories, and the source characteristics are not identified well.

[39] Factor 4 has its number mode at 15 nm. Figure 9 clearly shows the concentration peak during the morning rush hour for weekdays and shows no such peak for weekends. Sometimes particles in 10–20 nm were formed by nucleation followed with no detectable particle growth, and these data were not excluded by the method described in the previous section. These nucleation events keep the average concentration value of factor 4 from decreasing rapidly after 1000 local time (LT) to the afternoon. In Figure 7, Fourier analyses found two frequency peaks corresponding to a 7 day period and a 24 hour period. Factor 4 has no correlations with any gas and particle composition data, and this can be attributed to the small traffic flow around the site.
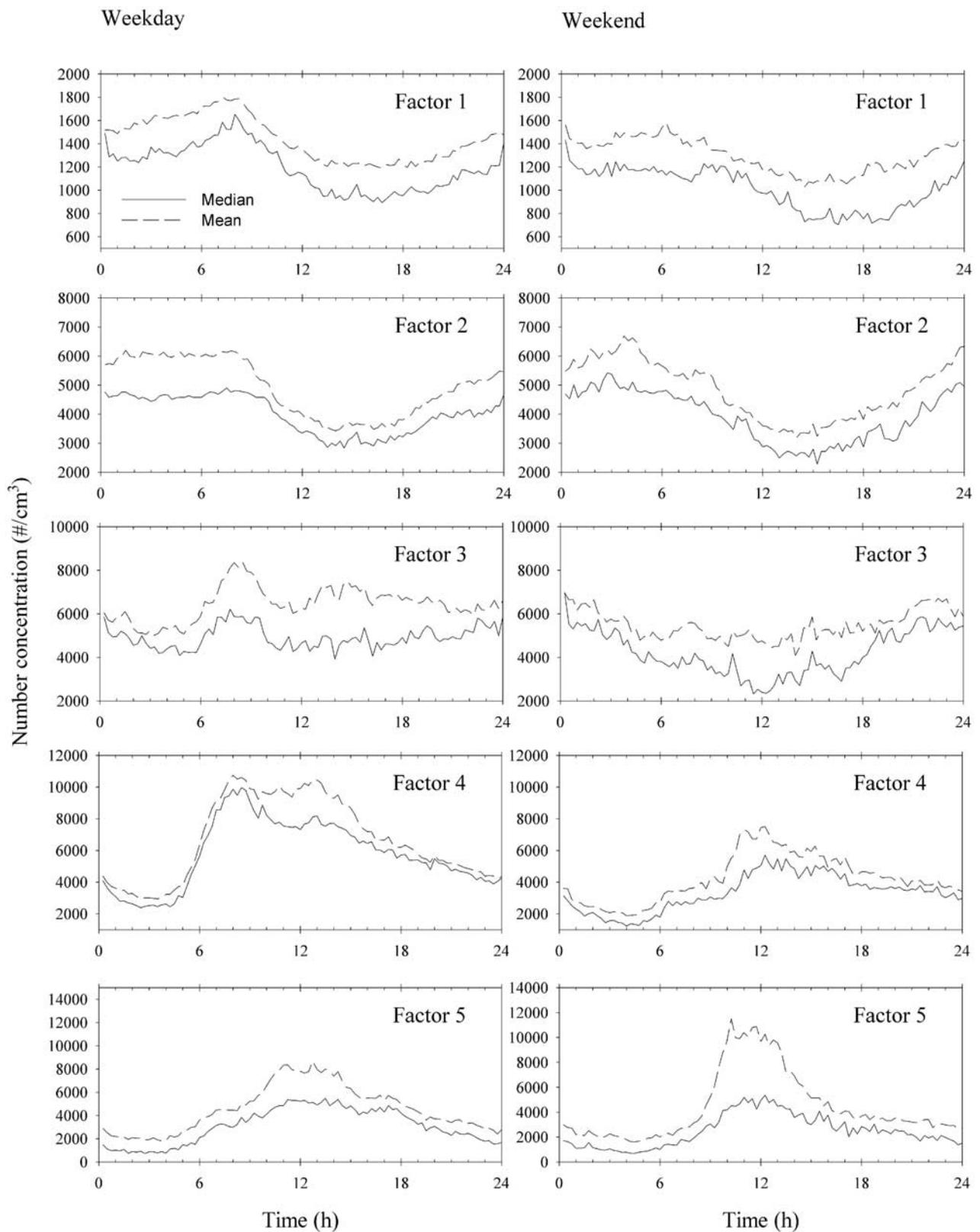
Figure 9. Daily average contribution from each factor for weekdays and weekends. The strength of the daily pattern can be determined by the spectral intensity at $1/24$ h^{-1} in Figure 6. The daily pattern strength is strong for factors 4 and 5, weak for factors 1 and 2, and nearly nothing for factor 3.

Table 6. Correlations of Gases and Source Contributions With Traffic Flow Volume on Schenley Drive, Forbes Avenue, and Interstate 376[a]

	NO	NO$_x$	CO	Factor 3	Factor 4
Schenley Drive[b]	0.056	−0.0024	−0.045	0.066	0.30
Forbes Avenue[b]	M0.021	−0.081	−0.145	−0.0069	0.27
Interstate 376[c]	−0.032	−0.125	−0.154	0.046	0.39

[a]Distance to the site: Schenley Drive, ∼200 m; Forbes Avenue, ∼600 m; and Interstate 376, ∼1600 m.

[b]For 1–31 January and 1–30 April 2002.

[c]For 11–31 January and 1–30 April 2002.

The traffic flow rate at three places (Schenley Drive, Forbes Avenue, and Interstate 376) near the site were not correlated with NO, NO$_x$, or CO as indicated in Table 6, indicating that the temporal variations of these gases do not show traffic patterns. The contribution of factor 4 has a positive correlation with the traffic flows. These facts support the conclusion that factor 4 is from local traffic while most NO, NO$_x$, or CO is not from local traffic.

[40] In 2002, from mid-February to mid-April, the roads near the station were closed in the early weekend mornings (0600–1200 LT) for motorless "buggy" practice. Figure 10 shows the contribution series of factor 4 and also the moving average of this series. The moving average time was chosen as 6 hours so that rapid variations were filtered. Usually, there are low emissions from local traffic on weekends, corresponding to low source contributions of factor 4. When the roads were closed for buggy practice, factor 4 shows the lowest contributions, as indicated in Figure 10. Factor 4 is thus assigned to be from local traffic. It may be from Forbes Avenue and other minor roads close to the measurement station within a distance of 1 km.

[41] Factor 5 represents particles smaller 10 nm from nucleation. These small particles are not involved in the particle growth event that we have discussed. A detailed study of nucleation events during PAQS was presented by *Stanier et al.* [2004b]. On average, the number mode is larger than 3 nm, and this may be caused by the nucleation happening upwind or happening at higher elevations with subsequent downward mixing of the particles. The similarity of data between the upwind (Florence) nucleation and Pittsburgh indicates that the nucleation is occurring upwind of Pittsburgh [*Stanier et al.*, 2004b]. Figure 7 shows a clear daily pattern, and Figure 9 indicates that local nucleation events are more active during the daytime, especially around noon. The similar pattern of mean and median values indicates that nucleation happened frequently in Pittsburgh, and this cannot be explained by occasional
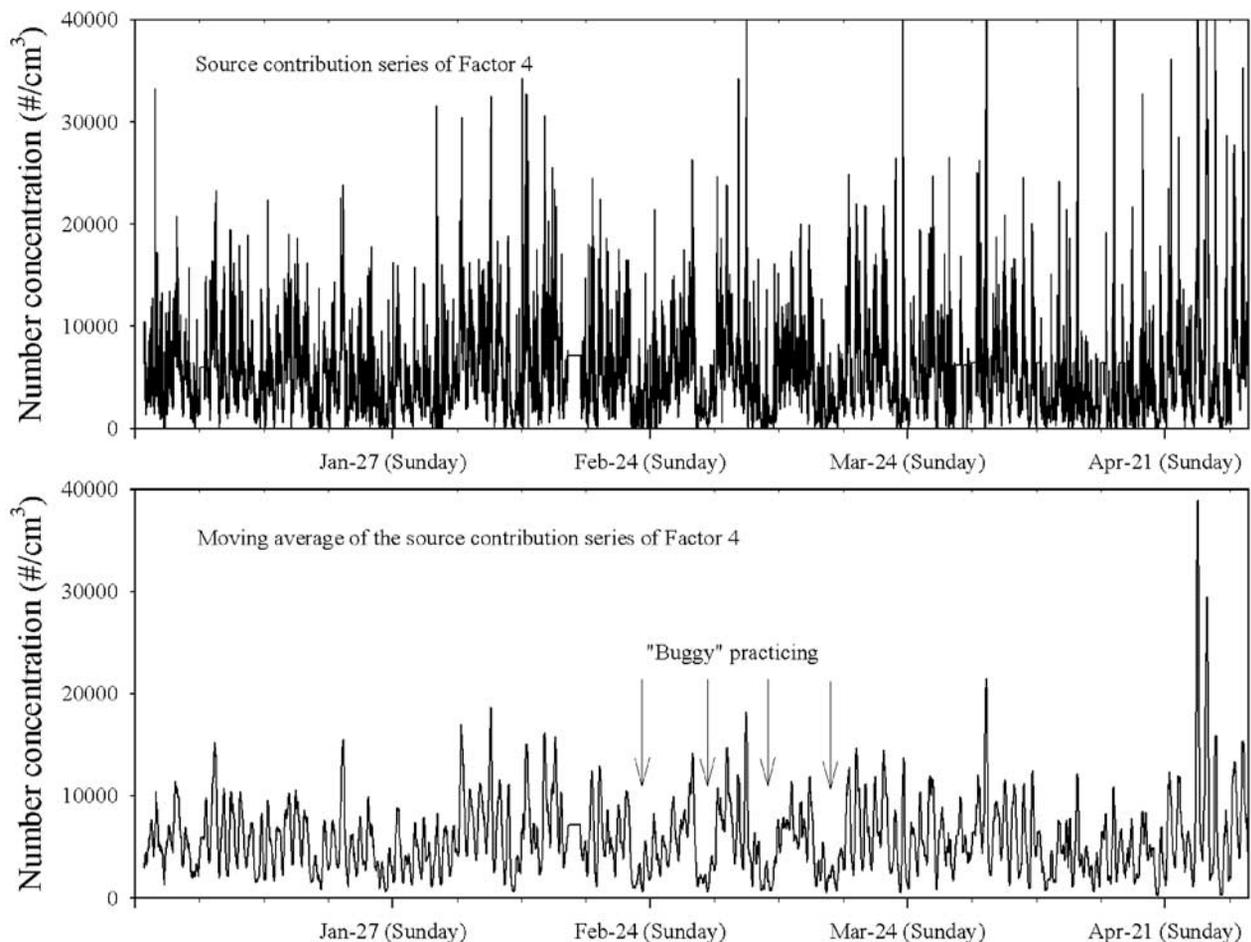
Figure 10. Source contribution series of factor 4 from 1 January to 30 April 2002.
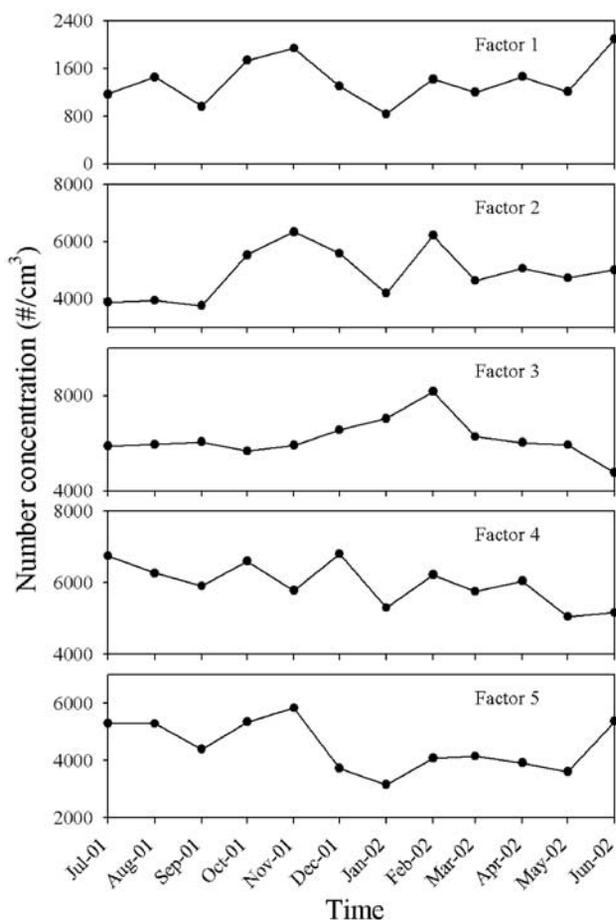
Figure 11. Monthly variations of average number contribution from each factor.

occurrences of nucleation with a large number of new particles produced.

[42] Figure 11 illustrates the variation of the monthly mean number contribution from each factor. The fluctuations are within a factor of 2. Factors 1 and 2 have similar seasonal trends, high in the fall and low in the winter. They all reach their highest concentrations in November 2001, and the SO_2 concentration is also the highest in that month. These high concentrations in November 2001 may be attributed to the dominant wind direction from the south, where more coal power plants are located. Figure 12 indicates the monthly volume contribution variations. The volume contribution is calculated from the number contribution and size distribution of each factor. For factor 1 the volume contribution only includes particles smaller than 0.5 μm for all months since the lack of APS data in some months prevents us from investigating the volume contribution over this size. The monthly variation of factor 1 is similar to $PM_{2.5}$ mass concentration. In the summer of 2001, particles from all sources seem to be larger than in other seasons.

[43] Table 7 summarizes the average contribution from each factor through the full year. The number contributions from factors 2, 3, 4, and 5 are approximately the same, and factor 1 contributes far fewer particles. The volume contribution is dominated by factors 1 and 2, and factor 5 contributes little to the total volume concentration. The

volume contribution of factor 1 is underestimated because of the missing APS data.

6. Conclusion

[44] Positive matrix factorization and other data-mining techniques have been applied for extracting source information with the full year size distribution data from the Pittsburgh Air Quality Study. The data representing particle growth after nucleation events were excluded. The analysis was performed for each month, and the same five factors were found for all months.

[45] The five factors found in this analysis represent five different size patterns. Each pattern is caused by a source or source group. This analysis has succeeded in separating and identifying local traffic and nucleation. The effect of this method is limited in separating sources in factors 1, 2, and 3. Factor 1 includes local sources besides secondary and aged primary aerosol; factor 2 includes power plants, but it cannot be separated further from other stationary combustion sources; factor 3 is probably Pittsburgh traffic, but the evidence is insufficient.

[46] For the purpose of source apportionment, the approach by itself has significant limitations compared with traditional receptor modeling based on chemical composi-
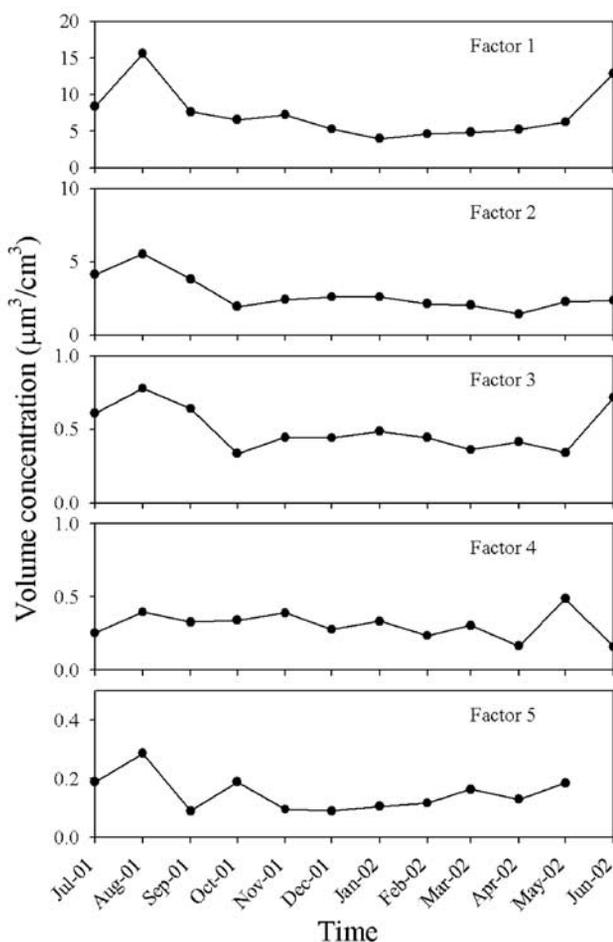
Figure 12. Monthly variations of average volume contribution from each factor.

Table 7. Average Contribution of All Factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Number concentration, number cm^{-3}	1402	4907	6182	5972	4514
Volume concentration, $\mu m^3\ cm^{-3}$	7.40	2.78	0.500	0.306	0.149[a]

[a]June 2002 is not included.

tion data. Comparison of the results of this approach with other chemical-composition-based analysis in the future will provide more information, especially in identifying the sources included in factors 1, 2, and 3. However, there is useful information that can be obtained from such analyses. Although the initial cost of the equipment is substantial, the operational costs are low, and they provide far more information than can be obtained from particle counts alone.

References

Ashbaugh, L. L., W. C. Malm, and W. Z. Sadeh (1985), A residence time probability analysis of sulfur concentrations at Grand Canyon National Park, *Atmos. Environ.*, 19, 1263–1270.

Birmili, W., A. Wiedensohler, J. Heintzenberg, and K. Lehmann (2001), Atmospheric particle number size distribution in central Europe: Statistical relations to air masses and meteorology, *J. Geophys. Res.*, 106, 32,005–32,018.

Brunekreef, B., D. W. Dockery, and M. Krzyzanowski (1995), Epidemiologic studies of short-term effects of low levels of major ambient air pollution components, *Environ. Health Perspect.*, 103, 3–13.

Chueinta, W., P. K. Hopke, and P. Paatero (2000), Investigation of sources of atmospheric aerosol at urban and suburban residential area in Thailand by positive matrix factorization, *Atmos. Environ.*, 34, 3319–3329.

Dockery, D. W., and C. A. Pope III (1994), Acute respiratory effects of particulate air pollution, *Annu. Rev. Public Health*, 15, 107–132.

Green, M., H. Kuhns, M. Pitchford, R. Dietz, L. Ashbaugh, and T. Watson (2003), Application of the tracer-aerosol gradient interpretive technique to sulfur attribution for the Big Bend Regional Aerosol and Visibility Observational Study, *J. Air Waste Manage. Assoc.*, 53, 586–595.

Henry, R. C. (2003), Multivariate receptor modeling by N-dimensional edge detection, *Chemom. Intell. Lab. Syst.*, 65, 179–189.

Hussein, T., A. Puustinen, P. P. Aalto, J. M. Makela, K. Hameri, and M. Kulmala (2004), Urban aerosol number size distributions, *Atmos. Chem. Phys.*, 4, 391–411.

Khlystov, A., C. Stanier, and S. N. Pandis (2004), An algorithm for combining electrical mobility and aerodynamic size distributions when measuring ambient aerosol, *Aerosol Sci. Technol.*, 38, 229–238.

Kim, E., P. K. Hopke, T. V. Larson, and D. S. Covert (2004), Analysis of ambient particle size distributions using UNMIX and positive matrix factorization, *Environ. Sci. Technol.*, 38, 202–209.

Kim, S., S. Shen, C. Sioutas, Y. Zhu, and W. C. Hinds (2002), Size distribution and diurnal and seasonal trends of ultrafine particles in source and receptor sites of the Los Angeles Basin, *J. Air Waste Manage. Assoc.*, 52, 297–307.

Kulmala, M., H. Vehkamaki, T. Petajda, M. Dal Maso, A. Lauri, V. M. Kerminen, W. Birmili, and P. H. McMurry (2004), Formation and growth rates of ultrafine atmospheric particles: A review of observations, *J. Aerosol. Sci.*, 35(2), 143–176.

Lee, E., C. K. Chun, and P. Paatero (1999), Application of positive matrix factorization in source apportionment of particulate pollutants, *Atmos. Environ.*, 33, 3201–3212.

Mäkelä, J. M., E. K. Koponen, P. Aalto, and M. Kulmala (2000), One-year data of submicron size modes of tropospheric background aerosol in southern Finland, *J. Aerosol. Sci.*, 31(5), 595–611.

Paatero, P. (1997), Least squares formulation of robust, non-negative factor analysis, *Chemom. Intell. Lab. Syst.*, 37, 23–55.

Paatero, P., P. K. Hopke, X. H. Song, and Z. Ramadan (2002), Understanding and controlling rotations in factor analytic models, *Chemom. Intell. Lab. Syst.*, 60, 253–264.

Paatero, P., P. K. Hopke, B. A. Begum, and S. K. Biswas (2005), A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution, *Atmos. Environ.*, 39, 193–201.

Polissar, A. V., P. K. Hopke, and R. L. Poirot (2001), Atmospheric aerosol over Vermont: Chemical composition and sources, *Environ. Sci. Technol.*, 35, 4604–4621.

Pope, C. A., III, D. W. Dockery, and J. Schwartz (1995), Review of epidemiologic evidence of health effects of particulate air pollution, *Inhalation Toxicol.*, 7, 1–18.

Raes, F., R. Van Dingenen, E. Cuegas, P. F. J. Van Velthoven, and J. M. Prospero (1997), Observations of aerosols in the free troposphere and marine boundary layer of the subtropical northeast Atlantic: Discussion of processes determining their size distribution, *J. Geophys. Res.*, 102, 21,315–21,318.

Ramadan, Z., X. H. Song, and P. K. Hopke (2000), Identification of sources of Phoenix aerosol by positive matrix factorization, *J. Air Waste Manage. Assoc.*, 50, 1308–1320.

Ruuskanen, J., et al. (2001), Concentrations of ultrafine, fine and PM$_{2.5}$ particles in three European cities, *Atmos. Environ.*, 35, 3729–3738.

Song, X. H., A. V. Polissar, and P. K. Hopke (2001), Source of fine particle composition in the northeastern U.S., *Atmos. Environ.*, 35, 5277–5286.

Stanier, C., A. Khlystov, W. R. Chan, M. Mandiro, and S. N. Pandis (2004a), A method for the in-situ measurement of fine aerosol water content of ambient aerosol: The Dry-Ambient Aerosol Spectrometer (DAASS), *Aerosol Sci. Technol.*, 38, Suppl. 1, 215–228.

Stanier, C., A. Khlystov, and S. N. Pandis (2004b), Nucleation events during the Pittsburgh Air Quality Study: Description and relation to key meteorological, gas phase, and aerosol parameters, *Aerosol Sci. Technol.*, 38, Suppl. 1, 253–264.

Stanier, C., A. Khlystov, and S. N. Pandis (2004c), Ambient aerosol size distributions and number concentrations measured during the Pittsburgh Air Quality Study (PAQS), *Atmos. Environ.*, 38, 3275–3284.

van Bree, L., and F. R. Cassee (2000), A critical review of potentially causative PM properties and mechanisms associated with health effects, *RIVM Rapp. 650010015*, Natl. Inst. of Public Health and the Environ., Bilthoven, Netherlands.

Wahlin, P., F. Palmgren, R. V. Dingenen, and F. Raes (2001), Experimental studies of ultrafine particles in streets and the relationship to traffic, *Atmos. Environ.*, 35, Suppl. 1, S63–S69.

Whitby, K. T. (1978), The physical characteristics of sulfur aerosols, *Atmos. Environ.*, 12, 135–159.

Wittig, B., N. Anderson, A. Y. Khlystov, S. N. Pandis, C. Davidson, and A. L. Robinson (2003), Pittsburgh Air Quality Study overview and preliminary scientific findings, *Atmos. Environ.*, 38, 3107–3125.

Xie, Y. L., P. K. Hopke, P. Paatero, L. A. Barrie, and S. M. Li (1999), Identification of source nature and seasonal variations of Arctic aerosol by positive matrix factorization, *J. Atmos. Sci.*, 56, 249–260.

Zhou, L., E. Kim, P. K. Hopke, C. Stanier, and S. N. Pandis (2003), Source apportionment using particle size distribution data from the Pittsburgh Air Quality Study (PAQS), paper presented at AAAR PM Conference, Am. Assoc. for Aerosol Res., Pittsburgh, April.

Zhou, L., E. Kim, P. K. Hopke, C. Stanier, and S. N. Pandis (2004), Advanced factor analysis on Pittsburgh particle size distribution data, *Aerosol Sci. Technol.*, 38, Suppl. 1, 118–132.

Zhou, L., P. K. Hopke, C. O. Stanier, S. N. Pandis, J. M. Ondov, and J. P. Pancras (2005), Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares (PLS) and positive matrix factorization (PMF), *J. Geophys. Res.*, D07S18, doi:10.1029/2004JD005050.

Zhu, Y., W. C. Hinds, S. Kim, S. Shen, and C. Sioutas (2002a), Study of ultrafine particles near a major highway with heavy-duty diesel traffic, *Atmos. Environ.*, 36, 4323–4335.

Zhu, Y. F., W. C. Hinds, S. Kim, and C. Sioutas (2002b), Concentration and size distribution of ultrafine particles near a major highway, *J. Air Waste Manage. Assoc.*, 52, 1032–1042.

Zhu, Y., W. C. Hinds, S. Shen, and C. Sioutas (2004), Seasonal trends of concentration and size distribution of ultrafine particles near major highways in Los Angeles, *Aerosol Sci. Technol.*, 38, Suppl. 1, 5–13.

————————————

P. K. Hopke and E. Kim, Center for Air Resources Engineering and Science, Clarkson University, P. O. Box 5708, Potsdam, NY 13699-5708, USA. (hopkepk@clarkson.edu)

S. N. Pandis, Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

C. Stanier, Department of Chemical and Biochemical Engineering, University of Iowa, 4122 Seamans Center, Iowa City, IA 52242, USA. (cstanier@engineering.uiowa.edu)

L. Zhou, Providence Engineering and Environmental Group LLC, 6160 Perkins Road, Suite 100, Baton Rouge, LA 70808, USA. (limingzhou@providencebr.com)