Source Apportionment of Winter Submicron Prague Aerosols from Combined Particle Number Size Distribution and Gaseous Composition Data

Devraj Thimmaiah[1*], Jan Hovorka[1], Philip K. Hopke[2]

[1] *Institute for Environmental Studies, Faculty of Science, Charles University in Prague, Benátská-2, Prague 12801, Czech Republic*

[2] *Department of Chemical and Biomolecular Engineering, Clarkson University, Potsdam, New York 13699, USA*

Abstract

Combined particle number size distributions and readily available gaseous concentration data were used to identify winter particle sources in the urban atmosphere of Prague, the capital of Czech Republic by applying bilinear Positive Matrix Factorization (PMF2). Ambient Particle Number Concentration (PNC) were obtained using a Scanning Mobility Particle Sizer (SMPS) in the size range between 18.8 and 723.5 nm (midpoint diameters) along with the ambient gaseous concentrations of CO, SO_2, NO_x ($NO + NO_2$), O_3, CH_4, Non Methane Hydrocarbons (NMHC) and Total Hydrocarbons (THC) at the receptor site with 5-minute time resolution. Meteorological data including wind direction, wind speed, temperature, relative humidity and solar the UV-A,B radiation were also recorded in 5-minute intervals to assist in the interpretation and to support the use of Conditional Probability Function (CPF) in determining the directionality of the sources. For convenient data handling and matrix preparation, all of the data obtained were reduced to hourly average concentrations. This study has identified and apportioned four major sources. They are ozone-rich, (transported ozone/ozone precursors, mixed down from above boundary layer associated with high wind speed and temperature), NO_x-rich (diesel emissions), traffic and local heating.

Keywords: Receptor modeling; Particle size distribution; Positive Matrix Factorization; Prague.

INTRODUCTION

* Corresponding author. Tel.: +420-221 951 894; Fax.: +420-224 914 803.

E-mail addresses: devraj151078@gmail.com; evraj151078@yahoo.com

Epidemiological studies have shown association between ambient particulate matter concentrations and mortality (Dockery *et al.*, 1993; Dockery and Brunekreef, 1996; Dockery

and Schwartz, 1996; Schwartz and Dockery, 1992a and 1992b). The need to measure fine and ultrafine aerosols has increased substantially over recent years because of their significant role in epidemiological studies showing relationships with mortality, morbidity and increase in cardiovascular hospital admissions and respiratory diseases (Schwartz *et al.*, 1996; Neas *et al.*, 1999; Korrick *et al.*, 1998; Peters *et al.*, 2000). While past studies have shown correlations between PM_{10} (Particulate Matter with aerodynamic diameter (d_{ae}) < 10 μm) or $PM_{2.5}$ (Fine Particles (FP) i.e., with d_{ae} < 2.5 μm), the role of Sub-Micron (SM i.e., d_{ae} < 1.0 μm) and Ultra Fine Particles (UFP i.e., d_{ae} < 0.1 μm) has not been extensively investigated. Studies for the years 1995-2001 in Erfurt, Germany on particle size distribution data obtained by combination of differential mobility analyzer and an optical laser aerosol spectrometer to provide daily means of particle number concentrations in three UFP size categories (0.01-0.03 μm, 0.03-0.05 μm, 0.05-0.10 μm) showed an increase in the number of daily deaths in association with increased UFP number concentrations and the magnitude of associations of particle measures with daily mortality, changed with changing air pollution levels and that UFP may be relevant (Stolzel *et al.*, 2004).

Particle Number Concentrations (PNCs) are also relevant with respect to epidemiological estimates of health effects. The processes/sources that emit these particles and their PNC in the specified size categories are of interest. Though the US EPA regulations recommend Particle Mass Concentrations (PMC) obtained by Federal Reference Method (FRM)/Federal Equivalent Method (FEM) samplers, based on filter gravimetry as acceptable into the Federal Register Database (FRD). In Europe, majority of the ambient monitoring networks utilize automated continuous monitors because of their ease in operation and long intervals between maintenance.

As an example, the PMC of Diesel Particulate Matter (DPM) is of less importance while compared with PNC. The DPM consists of fine particles with high number of UFP and are highly respirable, have large surface area where organics adsorb easily (including Polycyclic Aromatic Hydrocarbons (PAHs) and nitro-PAHs, sulfate, nitrate, metals and other trace elements) and can cause acute irritation and neurophysiological, respiratory and asthma-like and other allergic symptoms, may contribute to cardiopulmonary morbidity and mortality and lung cancer (Wichmann, 2007). If this is the case, there is still threat to human health even at mass concentrations well below the present ambient air quality standards. Hence, the PNC and the readily available gaseous composition data were combined to resolve easily identifiable sources contributing to air pollution at the receptor site by receptor modeling methods. Similar source apportionment studies were conducted in Erfurt by Yue *et al.* (2008), as such measurements are cost-effective when compared to the cost of determining the chemical composition of airborne PM. Thus, PNC measurements and source apportionment

methods can provide a cost-effective approach to provide useful information for air quality management.

Receptor modeling is an application of multivariate statistical methods to the identification and quantitative apportionment of air pollutants to their sources and to develop a plan of effective air quality management (Hopke, 1985 and 1991). The receptor model includes Chemical Mass Balance (CMB), Factor Analysis (FA), time series and regression analyses. The CMB model requires knowledge about source profiles/source signatures. There is no comprehensive database of such source signatures available in the Czech Republic at present. Therefore, the FA approach is suitable. The FA approach includes the traditional Principal Component Analysis (PCA) and a new variant of FA, Positive Matrix Factorization (PMF), which is better suited to environmental applications than PCA (Paatero and Tapper, 1994). Unlike PCA, PMF utilizes the known (or estimated) errors of the data matrix and imposes non-negativity constraints on the resulting factors.

PMF is a recent development in the class of data analysis techniques called factor analysis in which the fundamental problem is to resolve the identities and contributions in an unknown mixture. PMF has been extensively used for source apportionment of ambient particulate matter, where the goal is to resolve the mixture of sources that contribute to PM samples (Reff *et al.*, 2007). The PMF approach has been implemented in programs called PMF2 (Paatero, 1997a) and PMF3 (Paatero, 1997b), respectively. The applicability has been extended to arbitrary multilinear models in 1999 by means of the program "Multilinear Engine" ME-2 (Paatero, 1999). In this study, the bilinear PMF2 model is used in apportioning the sub-micron Prague aerosols.

MATERIALS AND METHODS

The materials, the measured sub-micron PNC in conjunction with the gaseous components data (CO, NO_x, SO_2, O_3, CH_4 and Non Methane Hydrocarbons-NMHC) that served as the input species of data matrix for the PMF2 program analysis, were obtained at the same receptor site location. The meteorological data including wind direction (WD, deg.), wind speed (WS, m/s), temperature (T, ^{o}C), relative humidity (RH, %), global radiations-UV-A (W/m^2) and UV-B (W/m^2) were also measured simultaneously. Additionally wet precipitation was recorded (Rain Intensity, RI in mm/h). A brief description of the receptor site is outlined below:

Receptor Site
The receptor site was a well-equipped rooftop sampling station (at height about 25 m above street level, 225 m ASL) belonging to the Institute for Environmental Studies, Charles University (latitude-50°4'17.46" N; longitude-14°25'14.92" E). It is situated inside the university botanical garden (area 0.035 km^2). Although the sampling station is positioned in the Prague city center, the site is shielded from direct sources of pollution and there are no street canyon conditions that

might affect the sampling conditions. The Prague city map and the monitoring site location along with the photo map of measurement site are shown in Figs. 1(a) and (b) respectively. With respect to the measurement site, much of the Prague city is built towards west, north, northeast and southward directions. The built area between the western and southern sector is less while compared to the other sectors. About 100 m to the west from the measurement site is a major

traffic intersection. The heating boiler chimney of the institute building and the heating boiler belonging to the Charles University Hospital are in close proximity to the measurement site towards north and north eastern sectors.

Prague, the capital of Czech Republic is known for its historic architectural monuments and sculptures and is the most popular attractive destination for tourists during all seasons. About 1.2 million people reside in Prague that is located in central Europe.

Fig. 1. (a) Prague (Praha) city map and measurement site location (red circle); **(b)** Photo map of the measurement site.

The Czech Republic shares borders with Germany, Austria, Slovakia and Poland. Prague is one of the most polluted regions within the country and the center for most traffic emissions. Regarding the Prague area (428.6 km^2), there are only few applications of receptor modeling studies conducted primarily for PM_{10} at a cross-roads in Prague and at small settlement in Benešov, Czech Republic by Hovorka *et al*. (1996; 1999; 2001; 2002). The first receptor modeling study of Czech air quality was done by Pinto *et al*. (1998) at Teplice and Prachatice, Czech Republic.

Sampling Period and Parameters Measured

The data selected for the study purpose was obtained between 7th and 23rd Jan 2008 (weekdays: 13 and weekends: 4, total: 17 days). The parameters, the instrumentation used and time resolution of the data obtained are summarized in Table 1 and the principle of gaseous measurement is given in parenthesis and the descriptive method can be found elsewhere (www.horiba.com). It is winter period with low temperatures. The lowest temperature recorded was -3.3^{o}C and the highest 11.5^{o}C with an average of 3.5 ± 3.6^{o}C (SD). RH was between 52 and 85% (avg. RH 76.4 ± 7.6%). WS was between 0.10 and 8.9 m/s (avg. WS 2.0 ± 1.7 m/s). Wet precipitation occurred on the following dates, 7th-10th, 16-20th, 22nd-23rd Jan 2008 lasting from a few minutes to a few hours on some days. Residential and space heating systems were operated on most days. The vehicular traffic had no major day-to-day changes and remained relatively constant except during the

weekends. A few sunny days were also reported and UV radiation that influenced ozone photochemistry was measured.

PMF Model

The bilinear PMF is the most widely used receptor model to analyze PM for source apportionment studies. The mathematical model in matrix form is given by

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

where, X= measured matrix. G and F are factor matrices to be determined; usually the elements of G and F are constrained to non-negative values only. E is residual matrix, i.e. the unexplained part of X. Or in component notation,

$$X_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \tag{2}$$

where x_{ij} is the concentration of species *j* measured on sample *i*, *p* is the number of factors contributing to the samples, f_{kj} is the concentration of species *j* in factor profile *k*, g_{ik} is the relative contribution of factor *k* to sample *i*, and e_{ij} is error of the PMF model for the species *j* measured on sample *i*. Now, the goal is to find g_{ik}, f_{kj} and *p* that best reproduce x_{ij}. The values of g_{ik} and f_{kj} are adjusted until a minimum value of *Q* for a given *p* is found. *Q* is defined as

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{e_{ij}}{\sigma_{ij}} \right)^2 \tag{3}$$

where σ_{ij} is the uncertainty of the jth species concentration in sample i, n is the number of samples, and m is the number of species. In PMF, the solutions are obtained using a weighted least squares fit, where the known uncertainties for the values x_{ij} are used for determining the weights of the residuals e_{ij}.

Numerous procedural decisions must be made and algorithmic parameters selected when analyzing PM data with PMF. A review of receptor modeling existing methods for ambient PM using PMF is given by Reff *et al.* (2007). In the robust mode, the PMF algorithm attempts to minimize Q_{Robust} (Eq. (4)) rather than Q (Eq. (3)) and Q_{Robust} is referred to as Q_{True}. When the robust mode is used, the uncertainties of measurements whose scaled residuals (e_{ij}/σ_{ij} in Eq. (3)) that are greater than the parameter called the outlier distance (α) are increased to downweight their influence on the PMF solution. Most PMF analyzes of PM data using a value of $\alpha = 4.0$ (Reff *et al.*, 2007).

Table 1. Summary of the sampling data obtained at the receptor site.

Parameters	Instrumentation	Integration Time (minutes)
Particle Size	SMPS-3936L25	05
14.6-736.5 nm	(TSI Inc., MN, USA)	
Meteorological Data	Young vane type monitor	05
WD, WS, T, RH, UV-A and UV-B	(Model 05103v, Fondriest Environmental Inc., USA)	
Rain Intensity	Laser Precipitation Monitor	01
	(LPM Disdrometer, Adolf Thies Clima Inc., Germany)	
Gaseous Components	***Horiba 360 series Analyzers:***	05
CO	APMA (Cross Flow Modulated Infrared Absorption technology, NDIR)	
NO$_x$ (NO+NO$_2$)	APNA (Cross flow modulation reduced pressure chemiluminescence, CLD)	
SO$_2$	APSA (UV fluorescence)	
O$_3$	APOA (Ultra-violet-absorption method, NDU)	
CH$_4$, NMHC and THC	APHA (Flame Ionization Detection method (FID) with selective combustion)	

$$Q_{Robust} = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{e_{ij}}{h_{ij}\sigma_{ij}} \right)^2 \qquad (4)$$

$$h_{ij} = 1 \qquad \qquad for \ \left| e_{ij}/\sigma_{ij} \right| \leq \alpha$$

$$h_{ij} = \left| e_{ij}/\sigma_{ij} \right|/\alpha \qquad for \ \left| e_{ij}/\sigma_{ij} \right| > \alpha$$

PMF Parameters-FPEAK, FKey and GKey

One of the key features of PMF2 is, the rotations can be controlled by using the parameter FPEAK (Paatero *et al*., 2002), when used forces rows and columns of the F and G matrices to be added and/or subtracted from each other depending on the sign of the FPEAK value. Typically, the PMF solutions for multiple values of FPEAK are explored, and the resulting Q values, F and G matrices, and scaled residuals (e_{ij} / σ_{ij}) are examined to select the optimum solution. The values of FPEAK selected lie between -1.0 and +1.0. (Hedberg *et al.*, 2005; Han *et al.*, 2006). Recently, Paatero *et al*. (2005) proposed a method of finding optimum values of FPEAK by plotting source contributions from the PMF analysis (G matrix columns) against each other and adjusting FPEAK until so-called "edges" in the plots become parallel to the plot axis.

Another method of inducing rotations when using PMF2 is through the use of the Fkey and Gkey matrices. Fkey and Gkey allow the user to specify whether the values in the F and G matrices should be zero and how strongly that constraint should be applied (Reff *et al*., 2007). If specific values of profiles or time series are known to be zero, then it is possible to force the solution toward zero for those values

through appropriate settings of Fkey and Gkey values, but the use of Gkey has not been previously reported.

Data Matrix Preparation

The highly time-resolved (5-min interval) SMPS sampling data obtained starting from 7th Jan 2008 (10:00 a.m.) until 23rd Jan 2008 (10:00 a.m.) were thoroughly checked for outliers, if found, the outlying data points were replaced by the average concentration of the preceding and the next sample value; only 15 outlying data points found were replaced. The hourly concentrations of particle number size distribution were recalculated and thus, there were 16 × 24 hours = 384 hourly measurements. The original number of size bins in SMPS data file were reduced to 35 columns by summing sets of three consecutive size bins to reduce the uncertainty in the number concentrations in each resulting bin while retaining the critical size information for resolving different source types. The midpoint diameters of the resulting size bins were 18.8, 20.9, 23.3, 25.9, 28.9, 32.2, 35.9, 40.0, 44.5, 49.6, 55.2, 61.5, 68.5, 76.4, 85.1, 94.7, 105.5, 117.6, 131.0, 145.9, 162.5, 181.1, 201.7, 224.7, 250.3, 278.8, 310.6, 346.0, 385.4, 429.4, 478.3, 532.8, 593.5, 661.2 and 723.5 nm respectively. The 41 columns included 35 size bins covering the size range between 18.8 and 723.5 nm plus the 6 gaseous components (CO, NO_x, SO_2, O_3, CH_4 and NMHC). The particle number size distribution and gaseous components were synchronized according to the sampling date and time. For uniformity, the units of number concentrations over all the size bins were

reported in Number/cm^3 and the CO, NO$_x$, SO$_2$, and O$_3$ concentrations were in µg/m^3, CH$_4$ and NMHC were in ppm. A 384 × 41 [(rows) × (columns)] input data matrix was created containing hourly concentrations of particle number size distribution and selected gaseous components (CO, NO$_x$, SO$_2$, O$_3$, CH$_4$ and NMHC).

Uncertainty Matrix Preparation

The error/uncertainty matrix with the same dimension 384 × 41 was prepared. The measured value is denoted by x_{ij}, and its associated uncertainty by σ_{ij}. The PMF solution depends on the uncertainties for each of the data values. Polissar *et al*. (1998), Yakovleva *et al*. (1999) and Chueinta *et al*. (2000) have suggested some approaches for the concentration values and their associated error estimates.

In this study, the error matrix for particle number concentration and gaseous composition was obtained by assuming Poisson errors for particle counts. The error estimations are based on certain assumptions of sampling errors and may vary from study-to-study. The PMF uncertainty for the particles is given in Eq. (5) and gaseous PMF uncertainties are summarized in Table 2. For gases, 10% of the measured value plus selective fraction was used (based on instrument detection limit and lowest measured value) to calculate the uncertainty value for each data point measured.

$$\text{For particles,} \quad \sigma_{ij} = 1 + \sqrt{(x_{ij})} + 0.1 \times (x_{ij})$$

(5)

Table 2. Gaseous PMF uncertainty (σ_{ij}).

Gases	Uncertainty σ_{ij}
CO	0.5 (100) + 0.1 (x_{ij})
NO$_x$	0.5 (0.1) + 0.1 (x_{ij})
SO$_2$	0.5 (2.7) + 0.1 (x_{ij})
O$_3$	0.5 (1.0) + 0.1 (x_{ij})
CH$_4$	0.1 (1.0) + 0.1 (x_{ij})
NMHC	0.01 (1.0) + 0.1 (x_{ij})

PMF2 Program Operation

The data matrix and uncertainty matrix prepared were input to the PMF2 program. The outlier distance (α) was set to 4.0. The PMF2 program was run for different factor settings ranging from 3 to 10 factors with 5 random starts for each individual factor setting. The source profiles and source contributions plots were examined for the minimum Q value obtained for individual runs. The PMF2 with 4 factor setting was chosen as increasing the number of factors did not show any credible additional sources. For the 4 factor solution, FPEAK was varied between - 1.0 and + 1.0 in steps of 0.2. The FPEAK setting with 0.0 was chosen. The scaled residuals showed symmetric distributions within ± 3 standard deviations. Fkey was not used in this study as no unwanted species was present in the factors deduced by PMF2, that were physically interpreted as sources.

RESULTS AND DISCUSSION

PMF2 Source Profiles and Contributions

The PMF2 factor analysis deduced probable four factor solution, their source profiles and

time-series of source contribution plots are shown in Figs. 2 and 3, respectively. The following observations were made in the four factors source profiles and their contributions:
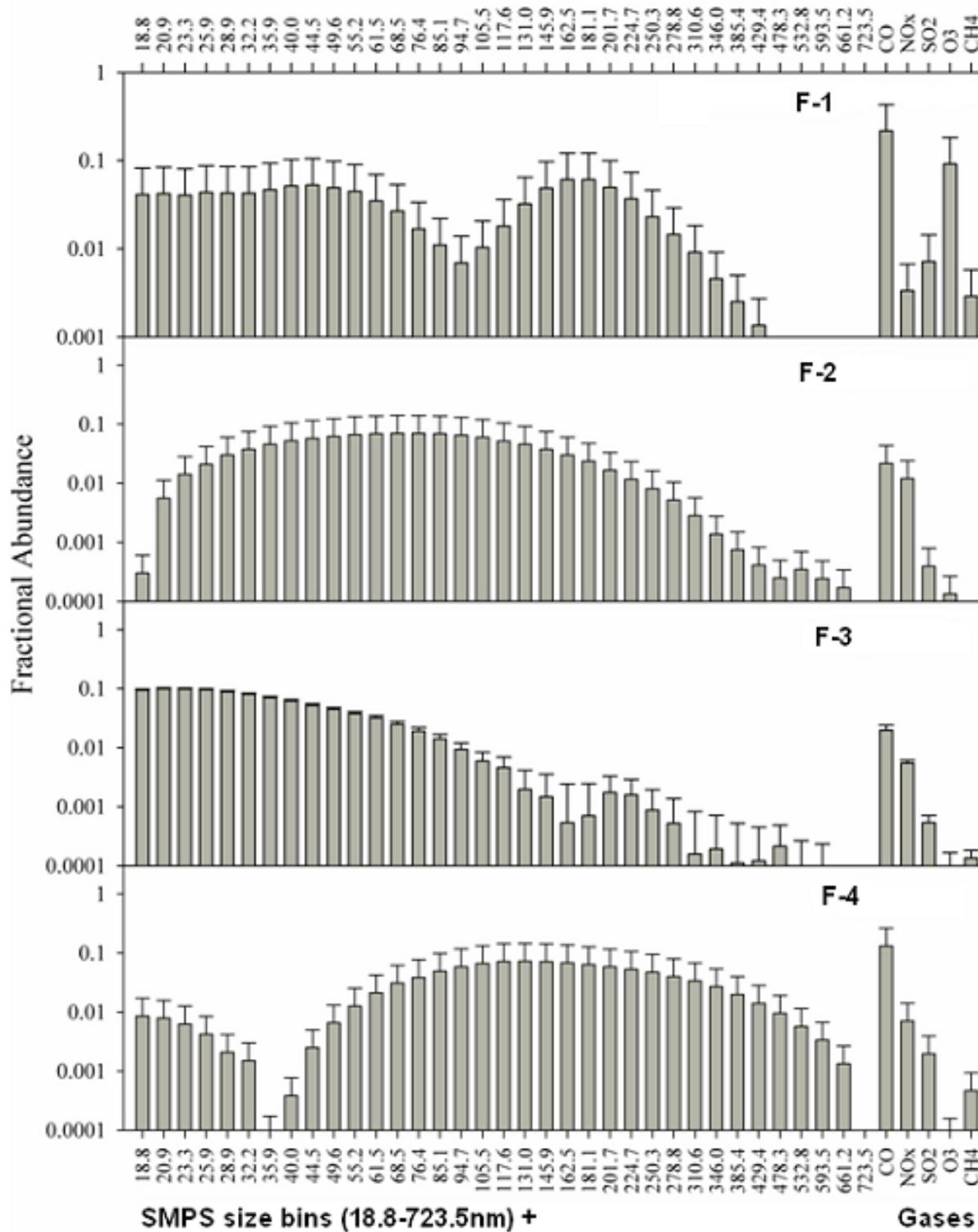

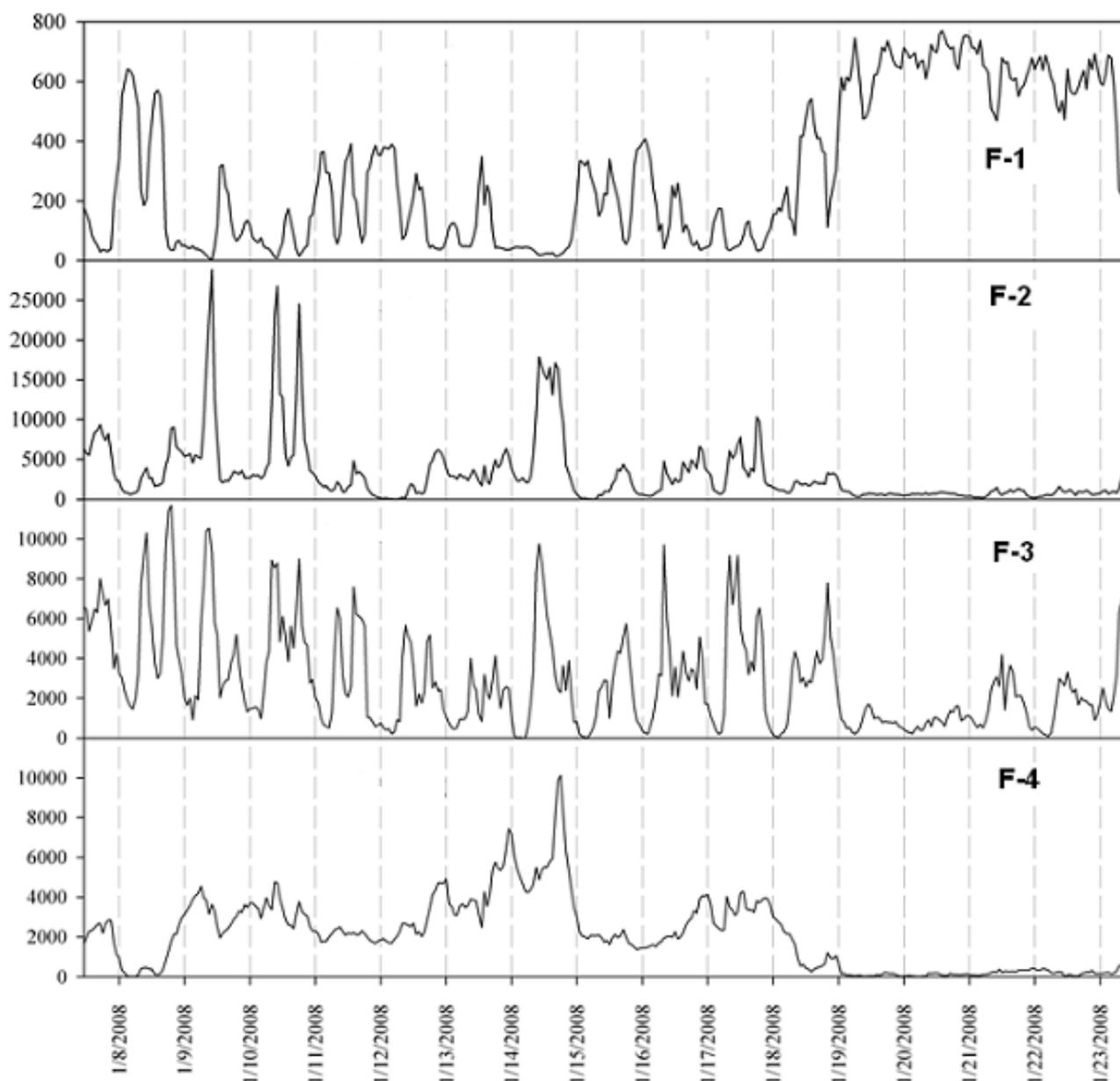**Fig. 2.** Source profiles of 4-factors deduced by PMF2.

Fig. 3. Source contributions of 4-factors deduced by PMF2.

Factor-1 Profile (F1)

The source profile of F1 shows bimodal particle size distribution, with high O_3 content, while compared against the profiles of F2, F3 and F4. Similarity in CO was seen with the profile of F4. The NO_x, CH_4 and slightly high SO_2 are seen. The F1 time-series contributions show a significant increase in contributions during the last 5 days of the sampling campaign period.

Factor-2 Profile (F2)

The source profile of F2 shows single mode particle size distribution, with their fractional abundance decreasing in the top midpoint diameter size range, CH_4 is absent, CO and

NO_x found in almost equal amounts, SO_2 found similar to F3 but slightly lower and O_3 traces. The F2 time-series contributions show low contribution towards the last 5 days of the sampling campaign period, contrary to the observations made in the case of F1 time-series contributions.

Factor-3 Profile (F3)

The source profile of F3 shows increased fractional abundance in the lower midpoint diameter size range and gradually lowering towards the higher midpoint diameter size bins. The variability is also high in these size ranges. The gaseous component CO was distinctly higher than NO_x and SO_2. The O_3 is absent and little CH_4 was observed. The F3 time-series contributions show suppressed contributions towards the last 5 days of the sampling campaign period. In contrast to both F1 and F2 time-series factor contributions, it's neither too high nor too low as compared to the previous days contributions but showed suppressed activity, but also shows continuity in its existence. When individual day time-series are looked at, it shows each day with two distinct peaks in most cases, and show diurnal variability nature of this particular source contribution.

Factor-4 Profile (F4)

The source profile of F4 shows distinct bimodal size distribution, with abundance in larger size fractions. High CO, followed by NO_x, SO_2 and CH_4 and absence of O_3 is observed. The CH_4 is in significant fractional abundance as compared to F2 and F3 and quite similar to F1 profile. The F4 time-series contributions show low contributions towards the last 5 days of the sampling campaign period. When looked at time-series for previous days, the contribution lines rarely touchdown to y-axis 0 value and show continuity in existence except during the last 5 days.

Gaseous Concentrations and Meteorological Data

The time-series of species gaseous concentrations, meteorology and precipitation intensity recorded at the receptor site is discussed and plots shown in Fig. 4. Just before the start of the sampling campaign, there was wet precipitation and also on the first day as shown by marked region (vertical round edged rectangle) in Fig. 4.

From Fig. 4 marked region (circles) it can be seen that the WD is considerably changed and WS reduced and increase in concentration levels of NO_x, Total Hydrocarbons (THC), SO_2 and CO emissions and reduced UV-A, UV-B radiation and lowest O_3 levels and no precipitation events recorded. Whereas, from the marked region (rectangles) it can be seen that during the end of the campaign period (last 5 days) the WD is steady, with high WS, T, UV-A, UV-B radiation and O_3 but low in RH, NO_x, THC, SO_2 and CO and with precipitation events being recorded.

Now, by visual comparison between 4 factor source contributions time-series (see Fig. 3) versus the species time-series (Fig. 4) it clearly shows that F1 is following the pattern of O_3 levels; F2, following NO_x levels; F3, seems to
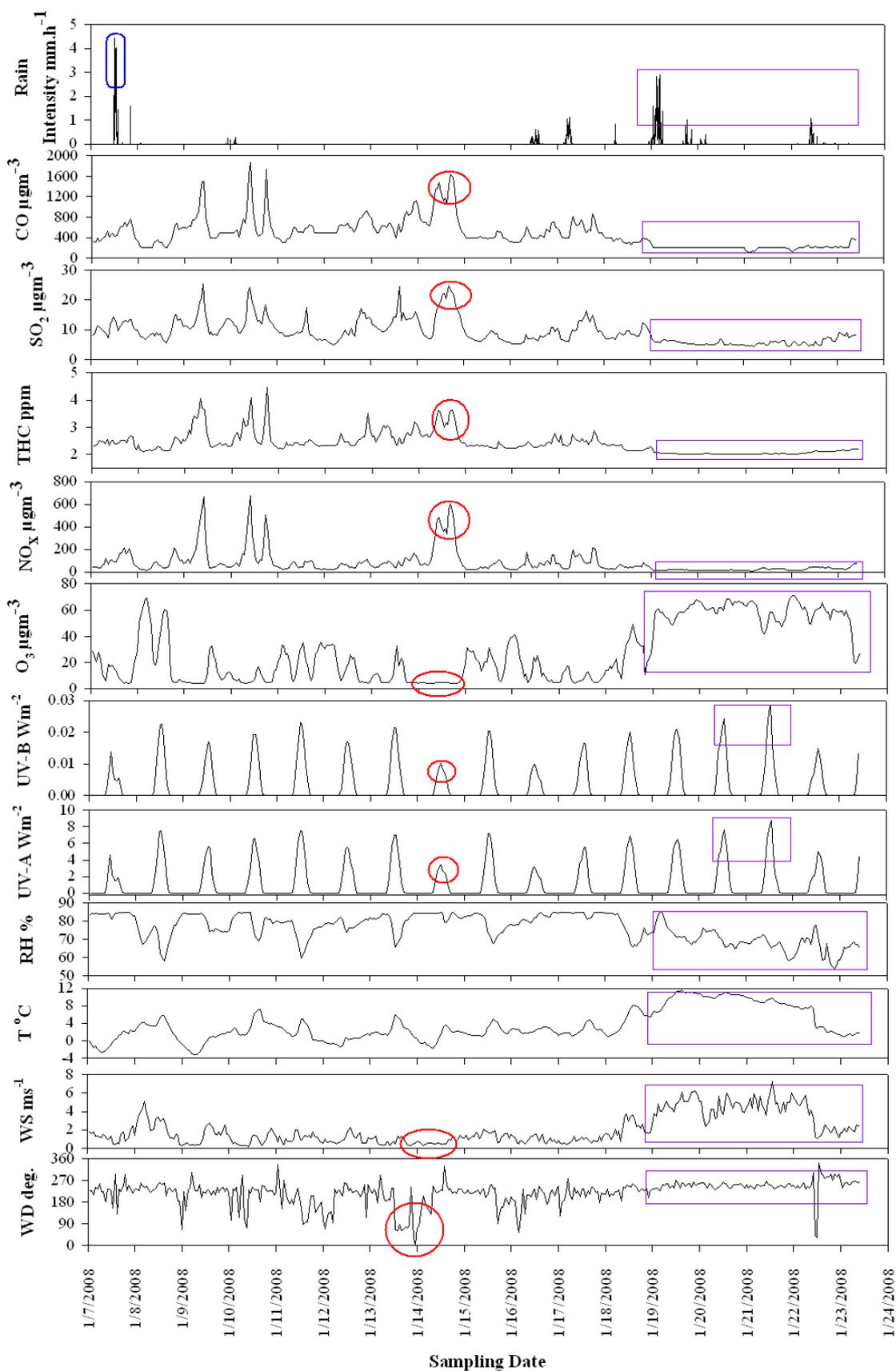
Fig. 4. Time-series of species-gaseous conc., meteorology and precipitation intensity.

be a combination of CO, SO_2 and NO_x; and F4, following THC. During the days 13th to 15th Jan 2008, sudden fall in F1 contributions and high peaks in the case of F2, F3 and F4 source contributions (see Fig. 3), has something to do with the change in the WD and reduced photoactivity as can be seen from reduced UV-A and UV-B global radiation. From the receptor site location these changed WD sectors point towards the north, northeast, and east directions where much of the Prague city is built, and the increased NO_x, THC, SO_2, CO is pointing to some local source emissions from these sectors not too far from the receptor site. Hence, it was considered important to know the species directionalities with respect to wind direction, in order to help in identifying their emission sources.

Conditional Probability Function (CPF)

In order to assist in the interpretation of factors, source contribution values and the species values were associated with wind directions using the conditional probability function (CPF) (Ashbaugh *et al.*, 1985; Kim *et al.*, 2004). The point of interest here is to know, when and from what directions the wind brings in majority of pollutants to the receptor site even while blowing at lower WS and less frequency of occurrences from these sectors. The CPF is defined by

$$CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \qquad n_{\Delta\theta} > n_c \qquad (6)$$

Where, $m_{\Delta\theta}$ is the number of occurrences in the wind direction sector $\Delta\theta$ that exceeds the threshold, defined as the upper 25th percentile of the fractional contribution from each source, and $n_{\Delta\theta}$ is the total number of wind occurrences in the same wind direction sector $\Delta\theta$. When $n_{\Delta\theta}$ is below 10 ($n_c = 10$) the CPF value is set to zero. Calm wind periods (< 1 m/s) were excluded from the CPF analysis. The sources are likely to be located in the direction sectors with high CPF values.

The wind direction locations of individual species and four factors using CPF are shown in Figs. 5 and 6, respectively. From Fig. 5, the wind direction CPF shows that the majority of the wind is blown from western and southwestern sectors. Usually it's more clean air being brought by the southwesterly winds, and also in these regions the built area of Prague is less as compared to the north, northeast and eastern regions. Looking at the four factors CPF (see Fig. 6), the F1 CPF is identical to the O_3 CPF and shows that F1 is associated with ozone. The F2 and F4 CPF plots show maxima in the northeast sectors, similar to the maxima for NO_x, SO_2, CO and THC. The F3 CPF surrounded all sectors except between 330° to 360° and more distinctly pointing at 225° similar to species CPF plots of SO_2 and NO_x. The gaseous species CPF may provide direct information regarding the processes that emits them. For example CH_4 may represent emissions from residential/office space heating with natural gas. NMHC can also be an indicator of hydrocarbon emissions related to vehicular traffic. Incomplete combustion of gas, oil, kerosene, wood or charcoal due to malfunctioning of heating appliances produce

CO emissions and can be an indicator of emissions due to space heating. Cold starts and idling of cars also produce CO. SO_2 (burning of sulfurous coal, oil and high sulfur diesel oil) and NO_x emissions arise when fossil fuels are burned.
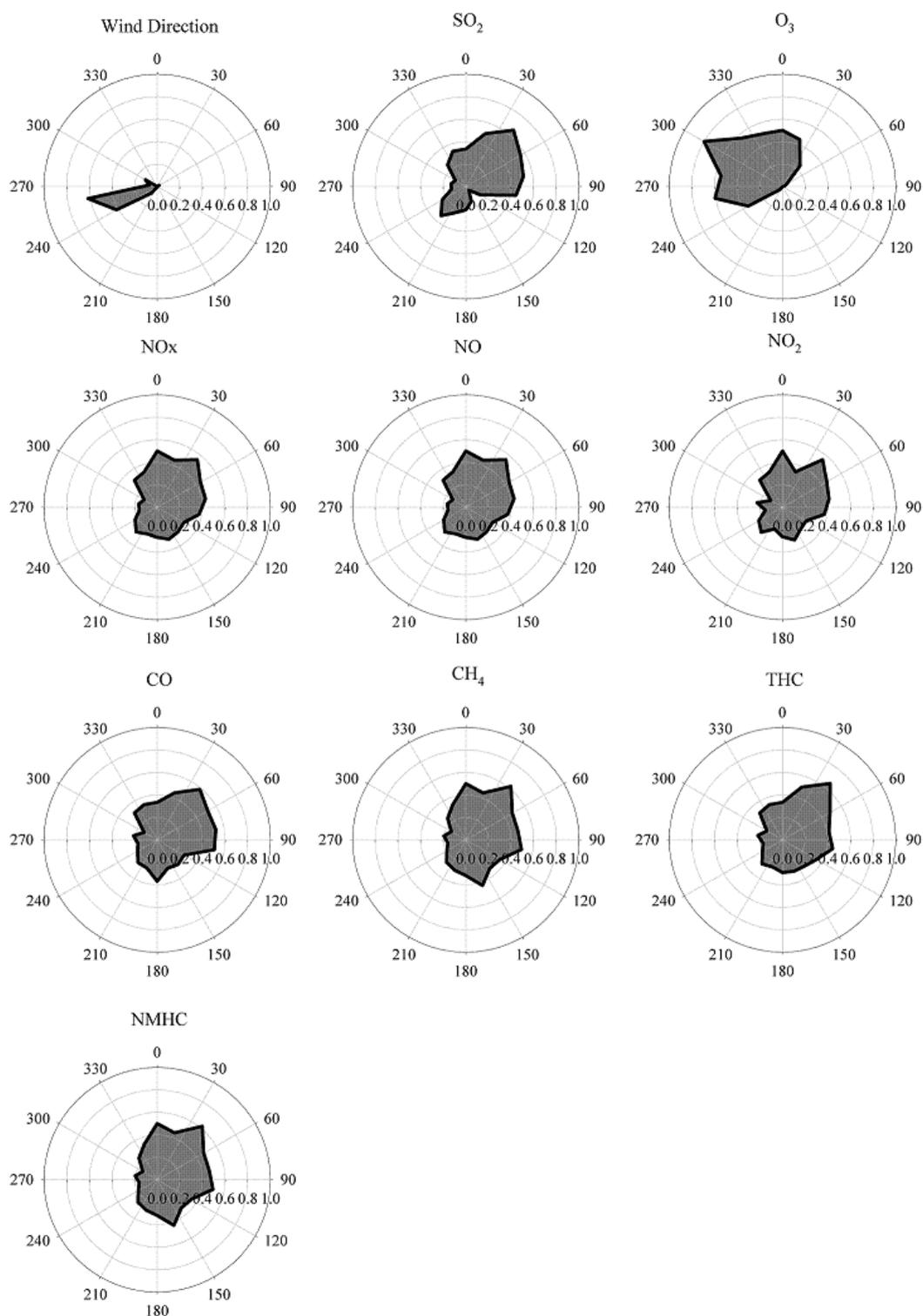

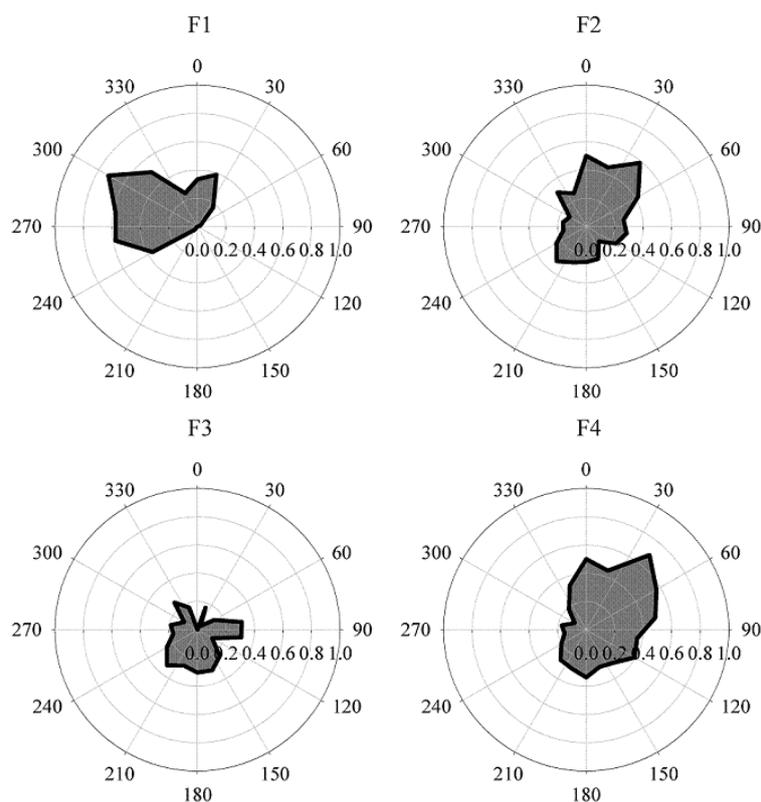**Fig. 5.** Wind direction locations of species using CPF.

Fig. 6. Wind direction locations of four factors using CPF.

Because of the multiplicity of sources/processes that may contribute to the emissions of NOx, SO$_2$, CO, CH$_4$ and NMHC, the standalone four factor CPF is not adequate to identify the sources based on their directionality analyses. It is also necessary to understand the correlations between each factor against the selected gaseous species. The correlation of four factors with gaseous pollutants O$_3$, NOx, CO, SO$_2$, CH$_4$ and NMHC (see supplemental Figs. S1-S4) and the correlation between ambient CH$_4$ and ambient temperature were calculated (see supplemental Fig. S5). Secondly, as an additional aid, the diurnal variability on weekdays and weekend may help to resolve and label the sources. Thus, the four factors deduced by PMF2 were

subjected to 24-hour diurnal pattern analyses on weekdays and weekend. Fig. 7 shows the 24-hour diurnal plots for factors F1 to F4 on weekdays and weekend.

Attribution of Sources
F1-Ozone Rich, Transported Ozone/Ozone precursors

The F1-CPF and species O$_3$-CPF were seen to be similar, and the scatter plots between the F1 contributions and the ambient O$_3$ concentrations showed good correlation (see Fig. S1(a)). The F1 contributions scatter with ambient NO$_x$, CO and SO$_2$ showed no other strong correlation (see Figs. S1(b), S1(c), and S1(d)). The 24-hour diurnal pattern analyses for F1 showed similar patterns for weekdays
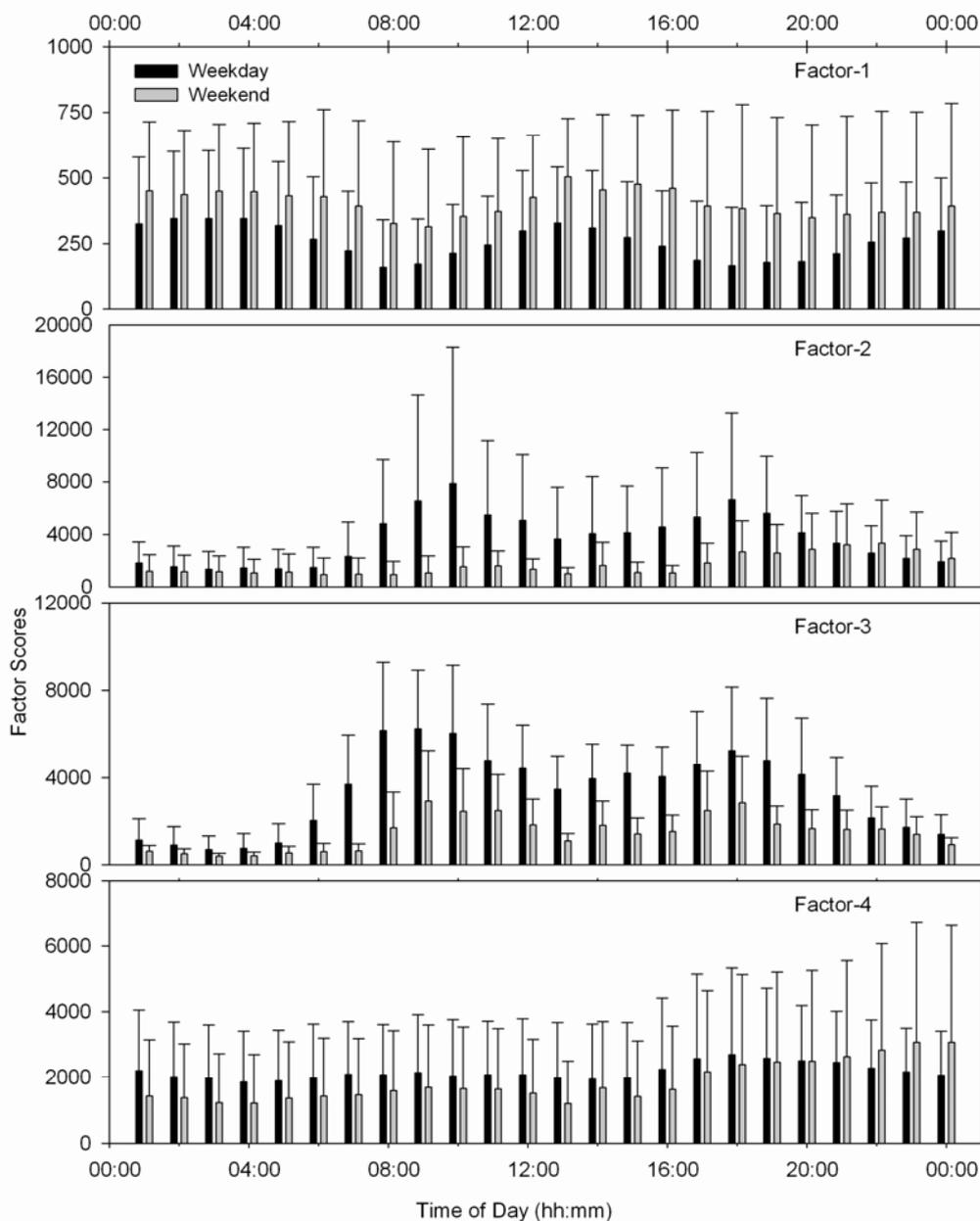
Fig. 7. 24-hour diurnal pattern of four factors on weekdays and weekend.

and weekend with peak values at noon (see Fig. 7). However, the weekend contributions were higher as compared to weekdays. This phenomenon is referred in literature as weekend effect/Ozone Weekend Effect (OWE) (Marr and Harley, 2002; Fujita *et al*., 2003). Higher weekend O_3 concentrations than on weekdays were first observed in 1970s and since then many studies have reported and supported the existence of such differences. During weekends, the ozone precursor emissions of NO_x and Volatile Organic Carbons (VOCs) are lower than on usual weekdays. This effect was earlier known as Sunday Effect-SE (Altshuler *et al*., 1995; Chinkin *et al*., 2003).

In general, such noon peaks are observed in the case of UV-A and UV-B radiations that are at their maximum at noon time. Increased O_3 concentrations were observed during the last 5 days 19th -23rd Jan 2008 and also on 8th Jan 2008 (see Fig. 4). These events were associated with a particular WD, with high WS and T and low RH. Thus, it is likely that the ozone/ozone precursors may be transported from rural regions and mixed down from above the boundary layer in to the areas. During the high O_3 days, the WD was between 225°-275° (SW-W sector) and the study receptor site is in the urban downwind location and the high concentrations of ozone may be due to transported ozone precursors from rural regions. The locally emitted NO_x, THC, CO and SO_2 concentrations were lower in comparison to the other days due to reduced office/space heating induced by high ambient temperatures; additionally high WS facilitates good dispersion.

In winter, we often see the high O_3 days as ones when the air is cleaner, so more dispersion and less titration of O_3 by the locally emitted NO. Wet precipitation events preceded the high O_3 days. Precipitation events, led to reduction in fine particles and clear skies (last 5 days), high WS and T with frontal passage supports the assignment of F1 to be ozone-rich, similar to ozone-related factor observed by Ogulei *et al.* (2007) in Rochester, NY.

High Ozone Concentration Periods

The species time series (see Fig. 4) show similar patterns for CO, SO_2, NO_x, THC and opposite from those of the O_3. Similar observations were made by Markovic *et al.* (2008) in Belgrade. The influence of meteorological conditions on observed concentration levels reveals the role of high WS on the O_3 concentrtion levels. Meteorological conditons (T, RH, WD, WS and solar radiation) strongly influence ozone formations and destructions (Markovic and Markovic, 2005). In winter, high O_3 concentration days are generally marked by clear skies triggered by wet precipitation events leading to washout of particles, less scatter of sunlight due to low fine particle concentrations and increased sunlight availability, enhances photochemically formed ozone.The high O_3 concentration levels also seems to arise from a particular WD and preceding with precipitation events The high O_3 concentrations pointing to the phenomenon of ozone transport during episodic measurements needed investigation, hence air back trajectoreis for the high level concentration periods were analysed (see Fig. 8). Ambinet O_3 concentrations at a given location are effected by topography and by transport of O_3 and/or formations from its precursors from extraneous regions, although occasionally there may be intrusons from the stratosphere (Finlayson-Pitts and Pitts, 1999).

The calculation of backward trajectores better provides understanding of transport effects. The backward trajectores ending at 00 UTC 22 Jan 2008 for a duration of 48 hours (see Fig. 8) shows that the trajectories at higher levels had south-western direction. Recently, HYSPLIT simulation and analysis
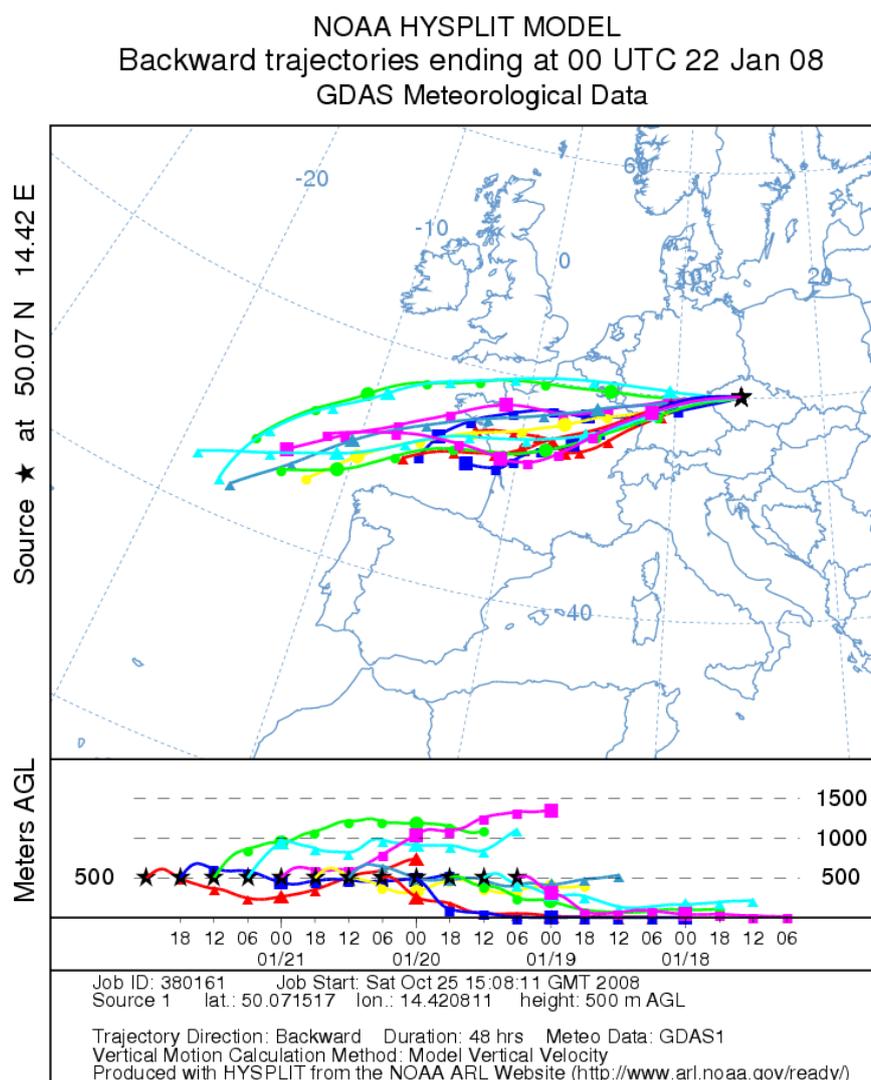
Fig. 8. Back trajectories ending at 00 UTC 22 Jan 2008.

was used in analyzing regional influence of PM on the city of Beijing, China (Xu *et al.*, 2008). In case of Belgrade urban air, Markovic *et al.* (2008) observed ozone transport form Adriatic area while in this study, strong westerly winds brings clean air from the Atlantic where ozone laden air is being transported with good disperisons and less titration from locally emitted NO and is a similar kind of ozone transport episode. Markovic *et al.* (2008) also observed ozone conentrations to be higher on weekends than

on weekdays, despite lower weekend ozone precursors emissions expected than during weekdays.

F2-NOx Rich Diesel Emissions

For F2, the directionality analyses of F2-CPF points in the same direction as the species CPF plots for NO_x, SO_2 and CO. The F2 contribution scatter plots shows good scatter with ambient NO_x, SO_2 and CO respectively (see Figs. S2(a), S2(c) and S2(d)). Further, the 24-hour diurnal pattern analyses show morning

and evening peaks and also higher contributions on weekdays than on weekends as seen in Fig. 7. However, during the late evening hours on weekends the contributions were higher than on weekdays suggesting activity at the end of weekends. This factor composition profile is high in NO_x and there is a strong correlation between its contributions and the NO_x concentrations. Given the size distribution with sizes in the 60 to 70 nm range and the strong relationship with NO_x, this factor appears to be related to diesel emissions. Schwarz *et al.* (2008) report that 16% are diesel vehicles, (62% Buses, 38% Trucks) and the direction is toward the center of the city where traffic would be most intense. The Sunday night – early Monday morning traffic may be delivery vehicles completing shipments to start the work week.

F3- Traffic, Spark Ignition Vehicles

For F3, the directionality analyses of F3-CPF showed that some contributions to arise from all sectors except between 330°-360°, with a small peak at 225°. The scatter analyses showed good F3 contribution correlations with ambient NO_x, SO_2 and NMHC (see Figs. S3(a), S3(b) and S3(d)). The 24-hour diurnal pattern analyses (see Fig. 7) show a strong bimodal distribution representing morning and evening rush hours. Weekend days show reduced levels of activity compared to weekdays and during the end of weekends show increased activity pointing to the fact that these represents increased vehicular traffic related activities when people start returning home back to the city center late at night on Sunday evenings

while people who are living at the periphery also start preparing to return back to the city center to start routine work starting on Mondays.

The center of the city is towards the north, northeast and east relative to the receptor site. However, traffic can be expected to occur in essentially all directions. The wind direction plot shows that the prevailing winds are dominated by winds from 225° to 275°. Steady southwesterly winds blew at the end of the campaign period (last 5 days, see Fig. 4). Thus, the traffic contributions from the north, northeast, and east will be reduced and the remaining traffic contributions are from southwestern sector highway vehicular emissions dominated by spark ignition vehicles given the size distribution (see Fig. 2).

F4-Local Heating Sources

For F4, the directionality analyses of F4-CPF showed strong influences from the north, northeast and east rather than from the prevailing southwesterly wind direction. The pattern is similar to the species CPF plots of CH_4, CO and NO_x. The F4 contributions scatter plots with the individual species showed quite different patterns than observed for the other factors. Fig. S4(a) shows that the F4 contributions decreased with increasing ambient temperature. The contributions were higher as the temperature nears $0^{o}C$ and decrease to near zero for temperatures above$8^{o}C$. From Fig. S4(b), ambient CH_4 seems to be well correlated with the F4 contributions as are those of CO and SO_2 (see Figs. S4(c) and S4(d)). In addition, Fig. S5
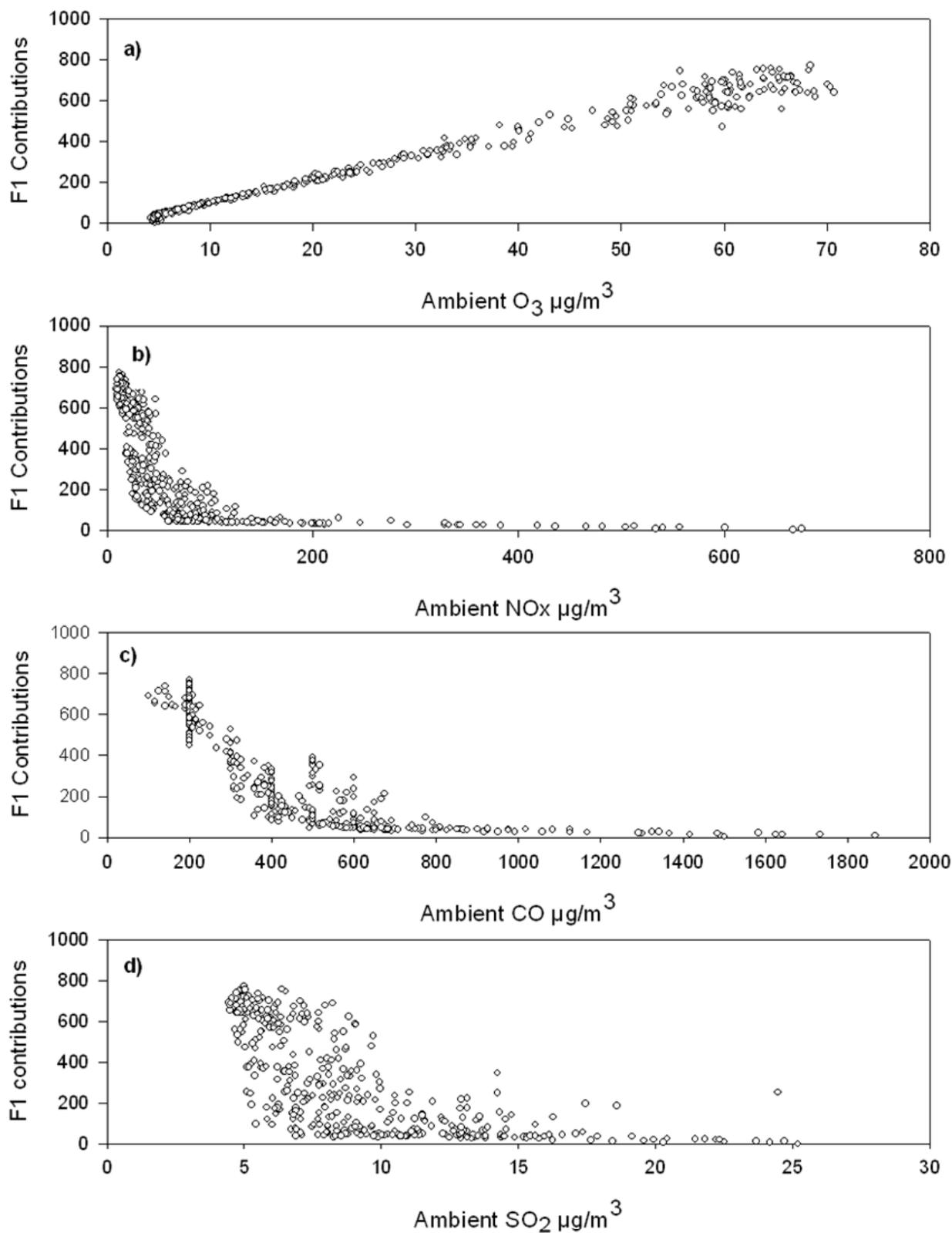
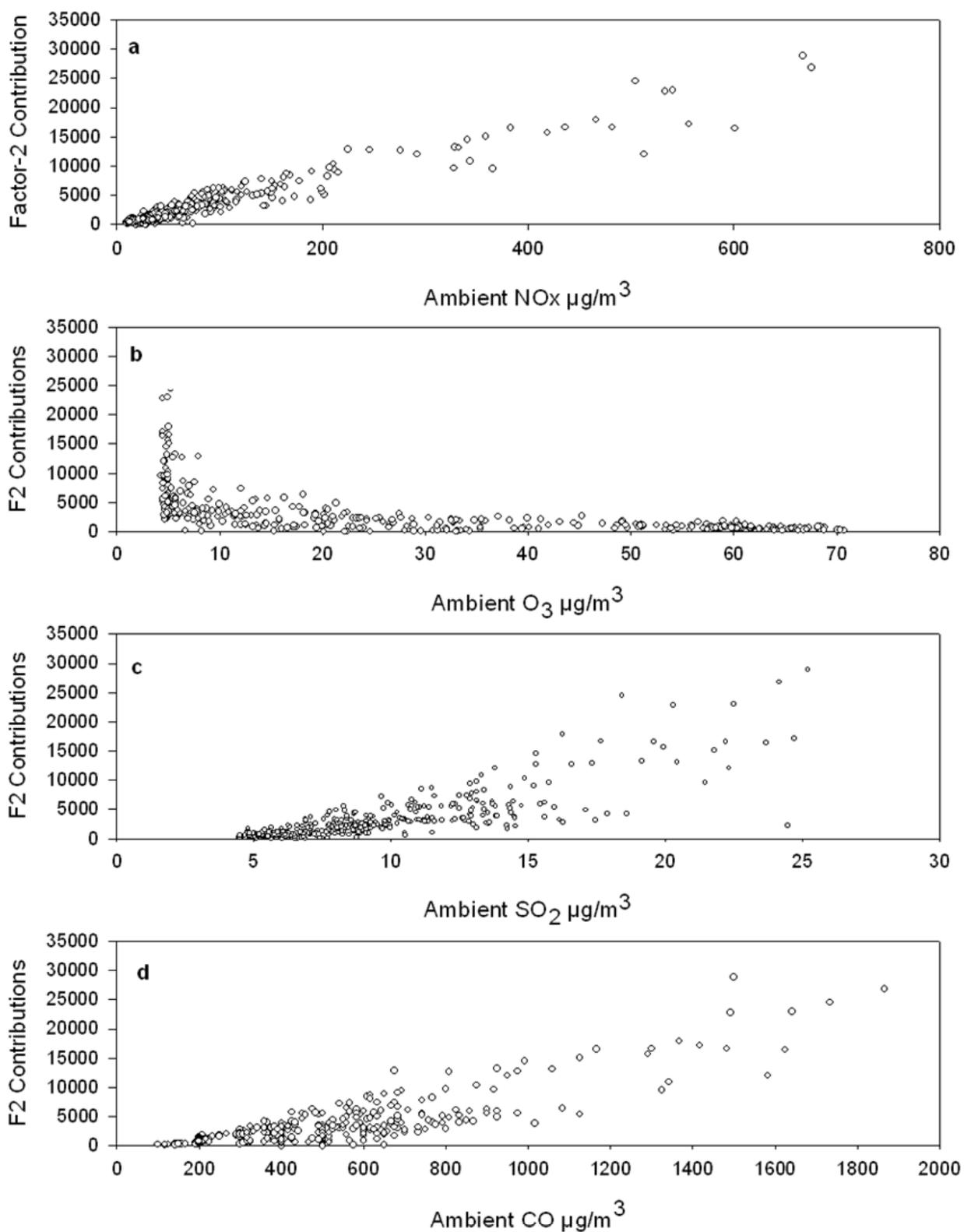**Fig. S1 (a, b, c and d).** F1 source contributions vs. O_3, NO_x, CO and SO_2.

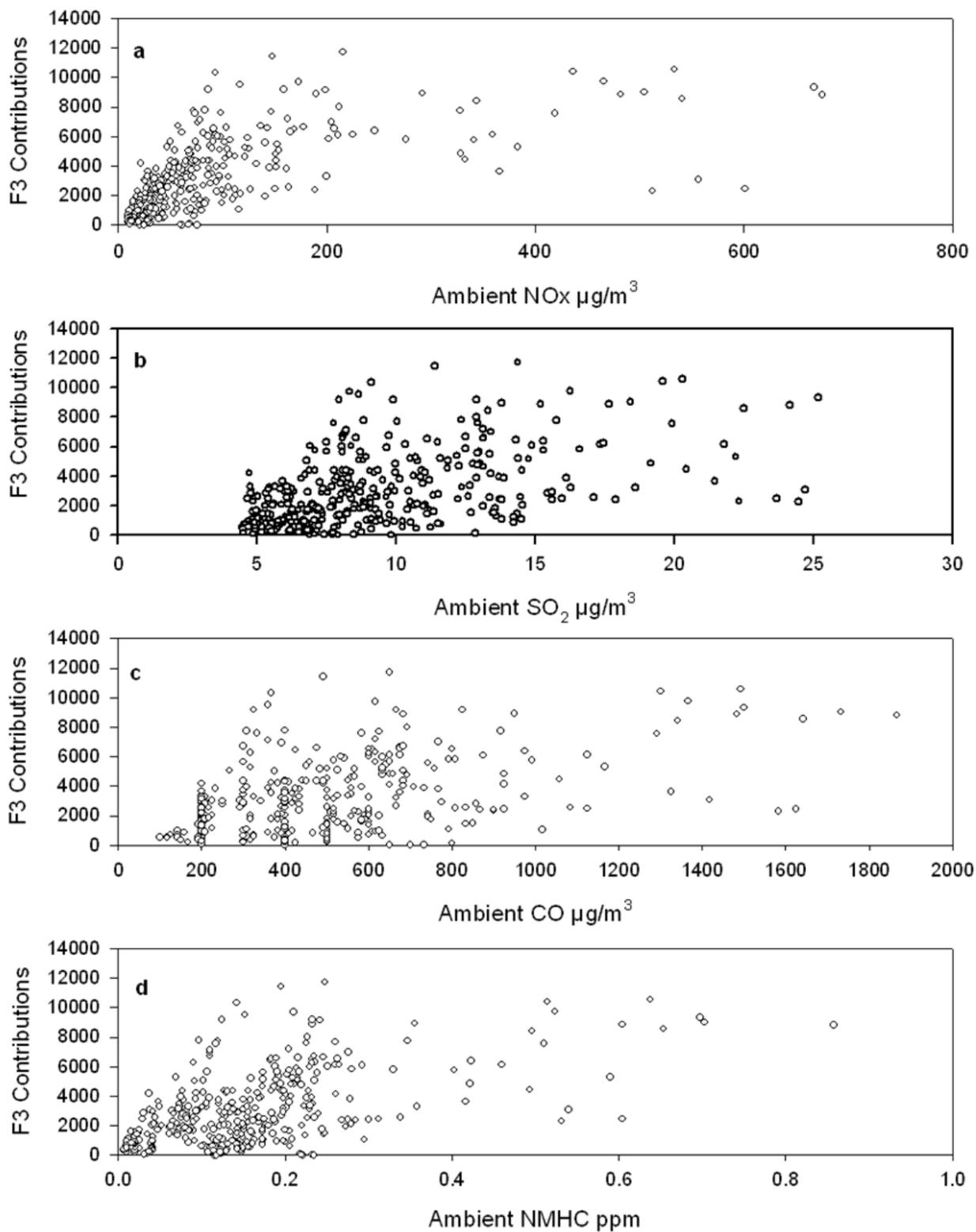**Fig. S2 (a, b, c and d).** F2 source contributions vs. NO_x, O_3, SO_2 and CO.

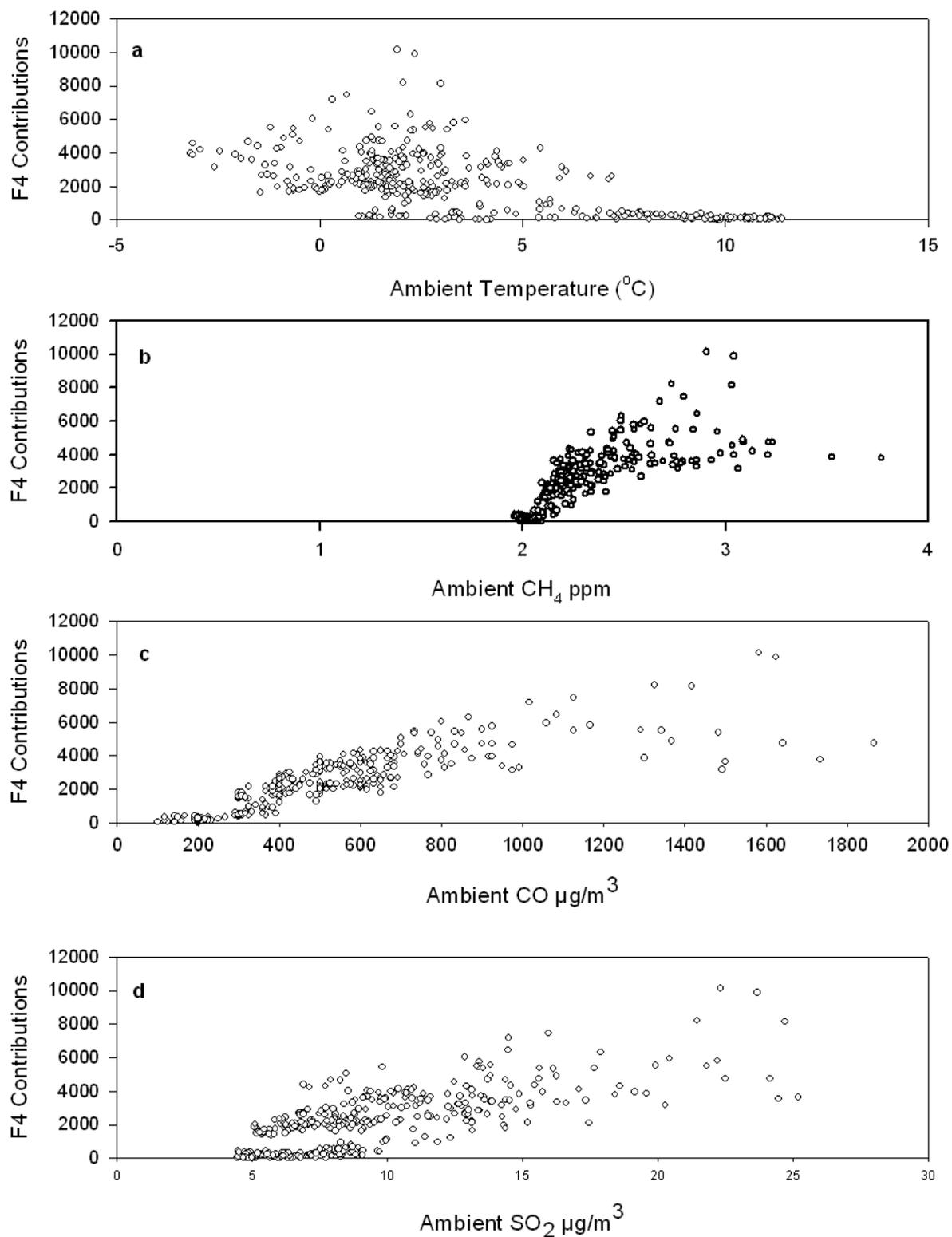**Fig. S3 (a, b, c and d).** F3 source contributions vs. NO_x, SO_2, CO and NMHC.

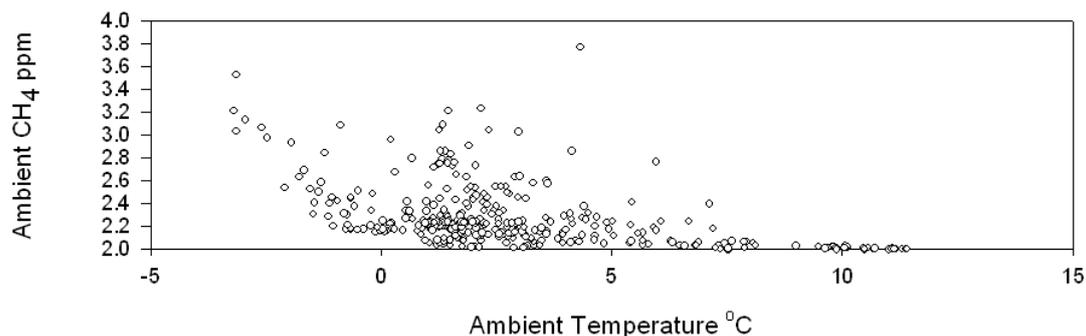**Fig. S4 (a, b, c, and d).** F4 source contributions vs. Temp, CH_4, CO and SO_2.

Fig. S5. Ambient CH_4 vs. ambient Temp.

shows that when the ambient temperature increases the ambient CH_4 levels drops and is highest when the ambient temp is lowest. This regulated emission of CH_4 points to office space/residential building heating by natural gas. Since the city is built more towards the north, northeast and east, the F4-CPF directionality is reasonable since the heating sources are high in these sectors. Considering the dominant southwesterly winds, emissions from this direction should be reduced by the prevailing winds. However, in close proximity to the receptor site at about 50-60 m distance, there are 2 heating boiler chimneys belonging to the institute building of the Charles University and the hospital. Thus, these specific sources may have significant contributions in the immediate vicinity particularly with the turbulence generated by the built environment in the vicinity of the monitoring site.

The diurnal pattern analyses of F4 contributions on weekdays and weekends add supportive information in labeling as heating sources (see Fig. 7). The diurnal pattern looks similar on both weekdays and on weekends (while contributions are lower on weekend). However, the evening contributions were high both during weekdays and weekend. During weekdays, it responded to increased home activities as people get back home after office hours starting from 17:00 hours. If this is the case, one should expect all time high during weekend when people are not working, but most people may keep away from their homes. However, during the end of the weekend, the contributions were higher than on weekdays between 22:00 hours and midnight. This shows the office space heating switched on to get it ready for people to start working from beginning of the week, Mondays.

Regression Analyses

Regression analyses between the factor contributions and total PNC were performed to calculate the contribution of each factor to the total PNC. The mean contribution of each factor to total number concentrations [8473 (100%)] for F1-ozone rich, F2-NO_x rich diesel emissions, F3-traffic and F4-heating sources were 294 (3.5%), 3206 (37.8%), 2895 (34.2%) and 2085 (24.6%) respectively. The linear regression correlation between the predicted vs. measured total PNC was excellent in agreement ($R^2 = 0.98$).

CONCLUSIONS

In this study, PMF2 analysis, based on the particle size distribution data along with the gaseous compositional data and meteorology, identified four possible sources for ambient winter period sub-micron particles in Prague, Czech Republic. The sources were identified as ozone-rich (transported ozone/ozone precursors, mixed down from above boundary layer associated with high wind speed and temperature), NO_x-rich (diesel emissions), traffic and heating sources. In addition, diurnal patterns of factor scores, the correlations of the factor contributions with gaseous pollutants (O_3, NO_x, SO_2, CO, CH_4 and NMHC) and CPF plots were used to assist in the interpretations of the sources.

ACKNOWLEDGEMENTS

REFERENCES

Altshuler, S.L., Arcado, T.D. and Lawson, D.R. (1995). Weekday vs. Weekend Ambient Ozone Concentrations - Discussion and Hypotheses with Focus on Northern California. *J. Air Waste Manage. Assoc.* 45: 967-972.

Ashbaugh, L.L., Malm, W.C. and Sadeh, W.Z. (1985). A Residence Time Probability Analysis of Sulfur Concentrations at Grand-Canyon-National-Park. *Atmos. Environ.* 19: 1263-1270.

Chinkin, L.R., Coe, D.L., Funk, T.H., Hafner, H.R., Roberts, P.T., Ryan, P.A. and Lawson, D.R. (2003). Weekday versus Weekend Activity Patterns for Ozone Precursor Emissions in California's South Coast Air Basin. *J. Air Waste Manage. Assoc.* 53: 829-843.

Chueinta, W., Hopke, P.K. and Paatero, P. (2000). Investigation of Sources of Atmospheric Aerosol Urban and Suburban Residential Areas in Thailand by Positive Matrix Factorization. *Atmos. Environ.* 34: 3319-3329.

Dockery, D.W., Pope, C.A., Xu, X.P., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. and Speizer, F.E. (1993). An Association between Air-Pollution and Mortality in 6 United-States Cities. *New Engl. J. Med.* 329: 1753-1759.

Dockery, D.W. and Brunekreef, B. (1996). Longitudinal Studies of Air Pollution Effects on Lung Function. *Am. J. Respir. Crit. Care Med.* 154: S250-256.

Dockery, D.W. and Schwartz, J. (1996). Particulate Air Pollution and Mortality-Reply. *Epidemiology.* 7: 213-214.

Finlayson-Pitts, B.J. and Pitts, J.N. (1999). *Chemistry of the Upper and Lower Atmosphere*, Academic Press, New York.

Fujita, E.M., Stockwell, W.R., Campbell, D.E., Keistar, R.E., and Lawson, D.R. (2003). Evolution of the Magnitude and Spatial

Extent of the Weekend Ozone Effect in California's South Coast Air Basin, 1981-2000. *J. Air Waste Manage. Assoc.* 53: 802-815.

Han, J., Moon, K., Lee, S., Kim, Y., Ryu, S., Cliff, S. and Yi, S. (2006). Size-Resolved Source Apportionment of Ambient Particles by Positive Matrix Factorization at Gosan Background Site in East Asia. *Atmos. Chem. Phys.* 6: 211-223.

Hedberg, E., Gidhagen, L. and Johansson, C. (2005). Source Contributions to PM_{10} and Arsenic Concentrations in Central Chile Using Positive Matrix Factorization. *Atmos. Environ.* 39: 549-561.

Hopke, P.K. (1985). *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons Inc., New York.

Hopke, P.K. (1991). An Introduction to Receptor Modeling. *Chemom. Intell. Lab. Syst.* 10: 21-43.

Hovorka, J., Keohane, B. and Marshall, G.B. (1996). Elemental and Stable Lead Isotopic Composition of PM_{10} Aerosols by ICP-MS. *Acta Univ. Carol.-Environmentalica.* 10: 63-70.

Hovorka, J. and Donkelaar, M. (1999). Elemental Composition and Stable Lead Isotopic Ratios of Wintertime PM_{10} Aerosols in Prague-Centre Crossroads, COST 319 Conference, Transport and Air Pollution, Technical University Graz, 76, IX/7, 39.

Hovorka, J., Braniš M. and Přibil R. (2001). Wintertime PM_{10} Elemental Composition and Source Apportionment in Prague and Benešov, Czech Republic. *J. Aerosol. Sci.*

32: S783.

Hovorka, J. (2002). Aerosol Source Apportionment Using by Multielemental and Isotopic Composition, Ph.D. Thesis, Institute for Environmental Studies, Charles University in Prague.

Kim, E., Hopke, P.K., Larson, T.V. and Covert, D.S. (2004). Analysis of Ambient Particle Size Distributions Using UNMIX and Positive Matrix Factorization. *Environ. Sci. Technol.* 38: 202-209.

Korrick, S.A., Neas, L.M., Dockery, D.W., Gold, D.R., Allen, G.A., Hill, L.B., Kimball, K.D., Rosner, B.A. and Speizer, F.E. (1998). Effects of Ozone and Other Pollutants on the Pulmonary Function of Adult Hikers. *Environ. Health Perspect.* 106: 93-99.

Markovic, D.M. and Markovic, A.M. (2005). The Relationship between Some Meteorological Parameters and the Tropospheric Concentrations of Ozone in the Urban Area of Belgrade. *J. Serb. Chem. Soc.* 70: 1478–1495.

Markovic, D.M., Markovic, D.A., Jovanovic, A., Lazic, L. and Mijic, Z. (2008). Determination of O_3, NO_2, SO_2, CO and PM_{10} Measured in Belgrade Urban Area. *Environ. Monit. Assess.* 145: 349-359.

Marr, L.C. and Harley, R.A. (2002). Modeling the Effect of Weekday-Weekend Differences in Motor Vehicles Emissions on Photochemical Air Pollution in Central California. *Environ. Sci. Technol.* 36: 4099-4106.

Neas, L.M., Schwartz, J. and Dockery, D. (1999). A Case-Crossover Analysis of Air Pollution and Mortality in Philadelphia.

Environ. Health Perspect. 107: 629-631.

Ogulei, D., Hopke, P.K., Chalupa, D.C. and Utell, M.J. (2007). Modeling Source Contributions to Submicron Particle Number Concentrations Measured in Rochester, New York. *Aerosol Sci. Technol.* 41: 179-201.

Paatero, P. and Tapper, U. (1994). Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estiamtes of Data Values. *Envirometrics.* 5: 111-126.

Paatero, P. (1997a). Least Squares Formulation of Robust Non-Negative Factor Analysis. *Chemom. Intell. Lab. Syst.* 37: 23-35.

Paatero, P. (1997b). A Weighted Non-Negative Least Squares Algorithm for Three-Way 'PARAFAC' Factor Analysis, *Chemom. Intell. Lab. Syst.* 38: 223-242.

Paatero, P. (1999). The Multilinear Engine-A Table-Driven Least Squares Program for Solving Multilinear Problems, Including the n-way Parallel Factor Analysis Model. *J. Comput. Graph. Statist.* 8: 854-888.

Paatero, P., Hopke, P.K., Song, X.H., and Ramadan, Z. (2002). Understanding and Controlling Rotations in Factor Analytic Models. *Chemom. Intell. Lab. Syst.* 38: 253-264.

Paatero, P., Hopke, P.K., Begum, B.A. and Biswas, S.K. (2005). A Graphical Diagnostic Method for Assessing the Rotation in Factor Analytical Models of Atmosheric Pollution. *Atmos. Environ.* 39: 193-201.

Peters, A., Liu, E., Verrier, R.L., Schwartz, J., Gold, D.R., Mittleman, M., Baliff, J., Oh, J.A., Allen, G., Monahan, K. and Dockery, D.W. (2000). Air Pollution and Incidence of Cardiac Arrhythmia. *Epidemiology.* 11: 11-17.

Pinto, J.P., Stevens, R.K., Willis, R.D., Kellogg, R., Mamane, Y., Novak, J., Šantroch, J., Beneš, I., Leniček, J. and Bureš, V. (1998). Czech Air Quality Monitoring and Receptor Modeling Study. *Environ. Sci. Technol.* 32: 843-854.

Polissar, A.V., Hopke, P.K., Paatero, P., Malm, W.C. and Sisler, J.F. (1998). Atmospheric Aerosol over Alaska, 2. Elemental Composition and Sources. *J. Geophys. Res.* 103: 19045-19057.

Reff, A., Eberly, S.I. and Bhave, P.V. (2007). Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods. *J. Air Waste Manage. Assoc.* 57: 146-154.

Schwartz, J. and Dockery, D.W. (1992a). Particulate Air-Pollution and Daily Mortality in Steubenville, Ohio. *Am. J. Epidemiol.* 135: 12-19.

Schwartz, J. and Dockery, D.W. (1992b). Increased Mortality in Philadelphia Associated with Daily Air-Pollution Concentrations. *Am. Rev. Respir. Dis.* 145: 600-604.

Schwartz, J., Dockery, D.W. and Neas, L.M. (1996). Is Daily Mortality Associated Specifically with Fine Particles? *J. Air Waste Manage. Assoc.* 46: 927-939.

Schwarz, J., Chi, X., Maenhaut, W., Civiš, M., Hovorka, J. and Smolík, J. (2008). Elemental and Organic Carbon in Atmospheric Aerosols at Downtown and

Suburban Sites in Prague. *Atmos. Res.* 90: 287-302.

Stolzel, M., Peters, A., Cyrys, J., Pitz, M., Wolke, G., Kreyling, W., Heinrich, J. and Wichmann, H.E. (2004). Particulate Matter in Several Size Classes and Daily Mortality in Erfurt, Germany. *Epidemiology*. 15: S59-59.

Wichmann, H.E. (2007). Diesel Exhaust Particles. *Inhal. Toxicol*. 19: 241-244.

Xu, X., Barsha, N.-A-F. and Li, J. (2008). Analyzing Regional Influence of Particulate Matter on the City of Beijing, China. *Aerosol Air Qual. Res.* 8: 78-93.

Yakovleva, E., Hopke, P.K. and Wallace, L. (1999). Receptor Modeling Assessment of PTEAM Data. *Environ. Sci. Technol.* 33: 3645-3652.

Yue, W., Stolzel, M., Cyrys, J., Pitz, M., Heinrich, J., Kreyling, W.G., Wichmann, H.E., Peters, A., Wang, S. and Hopke, P.K. (2008). Source Apportionment of Ambient Fine Particle Size Distribution Using Positive Matrix Factorization in Erfurt, Germany. *Sci. Total Environ*. 398: 133-144.