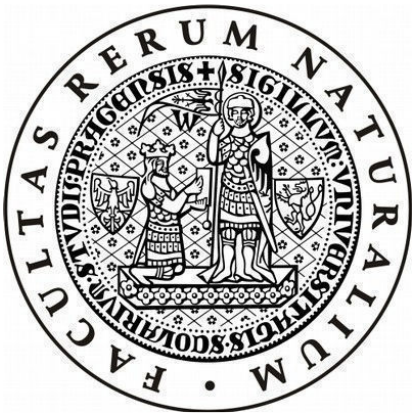
CHARLES UNIVERSITY IN PRAGUE
Faculty of Science
Institute for Environmental Studies
Benátská-2, Prague-2, 12801, Czech Republic



Source Apportionment of Sub-Micron Prague Aerosols from Combined Particle Number Size Distribution and Gaseous Composition Data by Bilinear Positive Matrix Factorization

Devraj THIMMAIAH, M.Sc.

Ph.D. Thesis

Supervisor: RNDr. Jan Hovorka, Ph.D.

Prague, February 2009

LIST OF CONTENTS

Contd…

LIST OF CONTENTS

Contd…

LIST OF CONTENTS

Contd…

LIST OF CONTENTS

*****

ACKNOWLEDGEMENTS

I hereby, would like to express my heartfelt sincere thanks and deep gratitude of appreciation to my supervisor RNDr. Jan Hovorka, Ph.D. (Institute for Environmental Studies, Charles University in Prague, Czech Republic) for his guidance and co-operation during my study period and to my external scientific consultant Prof. Philip. K. Hopke, (Department of Chemical and Biomolecular Engineering, Clarkson University, Potsdam, NY, USA) for providing me the necessary basic training to run the PMF2 program. His continuity of support in terms of sharing his vast research experience is greatly acknowledged.

I also take this opportunity to thank, Doc. RNDr. Martin Braniš, CSc. (previously, Head of the Institute) for his support during my research work and stay.

My continued list of appreciation and indebtedness are also with Doc. Ing. Jan Frouz, CSc. (Head of the Institute), Prof. RNDr. Karel Pivnička, DrSc.; Dr. Ing. Luboš Matějíček; Ms. Lenka Kupcová (school secretary), and Ms. Magdalena Cuříkova (study department) who have all extended their support during various stages of my stay. I would like to thank all the employees of our Institute whom, I have not mentioned here.

A special word of appreciation goes to Mgr. Martin Civiš, and RNDr. Martin Jankovský and to my other colleagues for their help, support and good companionship during my stay.

My profound eternal respects are towards my parents, my brother, my sister and my near relatives who all stood by me, by their moral support. Special thanks, also to my friends in Czech Republic and in India for their support.

I am indebted to the Czech Government Scholarship under India-Czech bilateral study program which ensured the path of my Ph.D. study. Financial support from the Grant Agency of Charles University (GAUK-PřF-49707/2007) is highly acknowledged.

DECLARATION

I affirm that neither this thesis, nor any of the results presented within has been submitted for the purpose of obtaining the title of Ph.D., or any other title, at another Institute.

Devraj THIMMAIAH, M.Sc.

Institute for Environmental Studies,
Faculty of Science, Charles University in Prague,
Benátská-2, Prague-2, Prague-128 01, Czech Republic.
E-mail: devraj151078@yahoo.com

In Prague, February 2009.

LIST OF TABLES

LIST OF FIGURES

Contd…

LIST OF FIGURES

Contd…

LIST OF FIGURES

*****

LIST OF ABBREVIATIONS

CMB	:	Chemical Mass Balance
CPF	:	Conditional Probability Function
DL	:	Detection Limit
EPA	:	Environmental Protection Agency
FA	:	Factor Analysis
HC	:	Hydrocarbons
INI	:	Initiation file
LPM	:	Laser Precipitation Monitor
ME	:	Multilinear Engine
NAAQS	:	National Ambient Air Quality Standard
NMHC	:	Non-Methane Hydrocarbons
PM	:	Particulate Matter
PMF	:	Positive Matrix Factorization
PMF2	:	Bilinear Positive Matrix Factorization
PNC	:	Particle Number Concentration
PSCF	:	Potential Source Contribution Function
RA	:	Regression Analysis
RH	:	Relative Humidity
SD	:	Standard Deviation
SMPS	:	Scanning Mobility Particle Sizer
T / Temp.	:	Temperature
THC	:	Total Hydrocarbons
UV-A	:	Ultraviolet solar radiation, wavelength 320 to 400 nm
UV-B	:	Ultraviolet solar radiation, wavelength 290 to 320 nm
WD	:	Wind Direction
WS	:	Wind Speed
#	:	Number
Hrs	:	Hours

CHAPTER-I
INTRODUCTION

CHAPTER-I
INTRODUCTION

Airborne particulate matter is one of the most important constituent for assessing air quality measured in a given urban airshed because of its adverse health and environmental impacts. The difficulties arise primarily from problems with pollution measurement and transport, identification of sources, estimation of emission rates, physical transport of substances, and physical and chemical transformation processes occurring during their transport (Hopke *et al*., 1999). A source apportionment or receptor model helps to address some of these problems.

Receptor models utilize information about the pollutants at a particular site to identify their sources and to develop a plan of effective air quality management (Hopke, 1985, 1991). There are generally two kinds of models attempting to explain variability of the air quality changes. They are dispersion and receptor models. While the first requires precise information on possible sources and air flow parameters in the airshed, the second, receptor model utilize information about the pollutants at a particular site to identify their sources and consecutively to develop a plan of effective air quality management (Michael *et al*., 2000). Receptor model application for source apportionment of Particulate Matter (PM) data have been used and developed over the last 25 years.

Receptor models include Chemical Mass Balance (CMB), Factor Analysis (FA) and time series analyses. The CMB model requires knowledge about source profiles/source signatures. Often, neither the number of sources nor their chemical profiles are known precisely. There is no comprehensive database of such source signatures available in the Czech Republic at present. Therefore, the FA will be carried out. Positive Matrix Factorization (PMF), a variant of FA technique has been applied to many studies inside and outside the U.S. and has proven to have reproduced satisfactory result. Further, the promulgation of the 1997 National Ambient Air Quality Standard (NAAQS) in the U.S. has further underlined the importance of receptor models. The U.S. Environmental Protection Agency (EPA) guidelines requires all the state and local air quality planning agencies to revise their plans for improving air quality and in reducing

the particulate level concentrations to meet the prescribed standard levels (EPA, 1986). The U.S. EPA has explicitly approved the use of receptor models in the planning process along with traditional dispersion models. Thus, receptor models have now become an accepted part of the regulatory process for air quality management.

Regarding the Prague area (428.6 km^2, about 1.21 million inhabitants), there are only a few applications of PM_{10} receptor modeling studies carried out at Prague cross-roads and at small settlement in Benešov, Czech Republic by Hovorka *et al*. (1996; 1999; 2001 and 2002). The first attempt to receptor modeling study in Czech air quality was carried out by Pinto *et al.* (1998) at Teplice and Prachatice, Czech Republic.

Receptor model studies provide information on the following objectives:-

- High concentration of PM at the sampling point in excedance of the standard requires research and analysis to investigate the possible sources of PM and PM precursors.
- To know whether the sampling and analysis setup is adequate to identify potential sources in the area.
- To identify potential sources and meteorological conditions to assist policy makers and modelers in developing control strategies.
- To identify main contributions to the PM so that appropriate control on PM and their precursors can be developed and implemented.
- To know how well current emission inventories and dispersion models represent the ambient conditions to model future control scenarios and the effect on PM concentrations.

In this study, the sources of sub-micron Prague aerosols have been apportioned using bilinear Positive Matrix Factorization (PMF2).

1.1. PMF Defined

PMF, an acronym for "Positive Matrix Factorization" is a variant of Factor Analysis technique with non-negative factor elements (Paatero, 1996; PMF-Guide). PMF can be applicable for the usual 2-dimensional matrices and for 3-way arrays. 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 (Paatero, 1999). In this study, the bilinear PMF2 model is used in apportioning the sub-micron Prague aerosols.

The PMF model is described in detail by Paatero and Tapper (1993; 1994) and by Paatero (1997a). In PMF any matrix X of data of dimension *n* rows and *m* columns, where *n* and *m* are the number of samples and the number of species, respectively, can be factorized into two matrices, namely G(*n* x *p*) and F(*p* x *m*), and the unexplained part of X, E, where *p* represents the number of factors extracted.

The 2-dimensional factor analytic "bilinear" model is explained by,

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

Where,

X= Measured matrix, G and F = Factor matrices to be determined, usually the elements of G and F are constrained to non-negative values only. E = Residual matrix, i.e. the unexplained part of X.

Or in index notation by,

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

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 (minimum sum of squares) for a given p is found. Q is defined as

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

Where,

σ_{ij} is the uncertainty of the j^{th} 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}.

In Robust mode, the PMF algorithm attempts to minimize Q_{Robust} (as in Eq. 4) rather than Q (as in Eq. 3) and Q_{Robust} is now referred to as Q_{True}. When 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 down weight their influence on the PMF solution. Most PMF analyzes of PM data give a value of α=4.0 (Reff *et al*., 2007).

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

Where,

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

1.2. 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, 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 is not found in literature.

1.3. PMF Analysis

Numerous procedural decisions must be made and algorithmic parameters selected when analyzing PM data with PMF. However, few publications document enough of these details for readers to evaluate, reproduce or compare results between different studies. Few studies, document why some species were used and others not used in the modeling, how the number of factors was selected, or how much uncertainty exists in the solutions (Reff *et al*., 2007). Numerous studies, recent and past have performed source apportionment of ambient PM using the PMF receptor model (selected publications are listed in Chapter-II, Literature Review). Despite the extensive use of PMF, there exists considerable variation in the procedures followed and decisions made to apportion sources of ambient PM using PMF.

The PMF modeling may be divided into three broad steps:

1. Preparing data to be modeled.
2. Processing the data with PMF to develop a feasible and robust solution.
3. Interpreting the solutions.

Specific decisions such as the creation of data uncertainties, selection of the best number of factors, and treatment of outliers need to be made when carrying out these steps.

The Fig. I.1, lists the general work flow for PMF analysis of ambient PM species data. Procedures relating to *Data preparation step* involves choosing the relevant species, handling missing and below detection limits data, handling poor or unknown data quality, whether or not? redundant species for example sulfur or sulfate or both need to be included? (Reff *et al*., 2007). PMF *Analysis step* involves selecting the PMF parameters-robust mode, setting the outliers distance (α), number of factors (*p*), the error model by which PMF calculates uncertainties at each iteration of the program, inputting the data and the uncertainty matrix created. After successful PMF program run to check the goodness of model fit, by comparing the species concentrations predicted by PMF to the original measurements by apportioning to the total PM mass measurements (Predicted conc. vs. Measured conc.), to examine the distributions of scaled residuals (e_{ij}/σ_{ij}) to ensure that scaled residuals for most species in their data sets lie between certain limits, typically -2 and +2 (Liu *et al*., 2003a; Wang and Shooter, 2005). In spite of non-negativity, there can exist a multiple number of F and G matrices that all produce the same minimum *Q* (Henry, 1987). This existence of possible solutions is referred to as rotational freedom. In PMF2 program, the common method of rotating solutions is done by adjusting the parameter FPEAK, multiple values of FPEAK are explored and the resulting *Q* values, F and G matrices, and scaled residuals are examined to select the optimum solution. The FPEAK values explored lie between -1.0 and +1.0 (Paatero *et al*., 2005).

Another method of inducing rotations in PMF2 program is to use the "FKey" and "GKey" matrices. Fkey and Gkey allow the user to specify whether values in the F and G matrices should be zero and how strongly the constraint should be applied. The S*ource interpretation and Analysis step* involves- source identification by looking at the time series of contributions patterns (G matrix) and comparing the profiles (F matrix) with

literature database and looking at the common strategy to explain the profiles. The daily, weekly, seasonal and yearly analysis of source contributions can aid in interpretation.

Additionally, *Auxillary analyses*, for example, local wind trajectories are often analyzed in conjunction with PMF analyses of PM, helps to identify the location of specific sources by using Conditional Probability Function (CPF), Potential Source contribution Function (PSCF), cluster analyses and residence time analysis. Regression Analyses (RA) of source contributions deduced by PMF with meteorological data and gaseous time series data (NO_x, O_3, SO_2, CO, Non Methane Hydrocarbon (NMHC) and CH_4) obtained in conjunction with the PM samples can aid in identifying the sources deduced by PMF. A review of existing methods for receptor modeling of ambient PM data using PMF is discussed elsewhere (Reff *et al*., 2007).

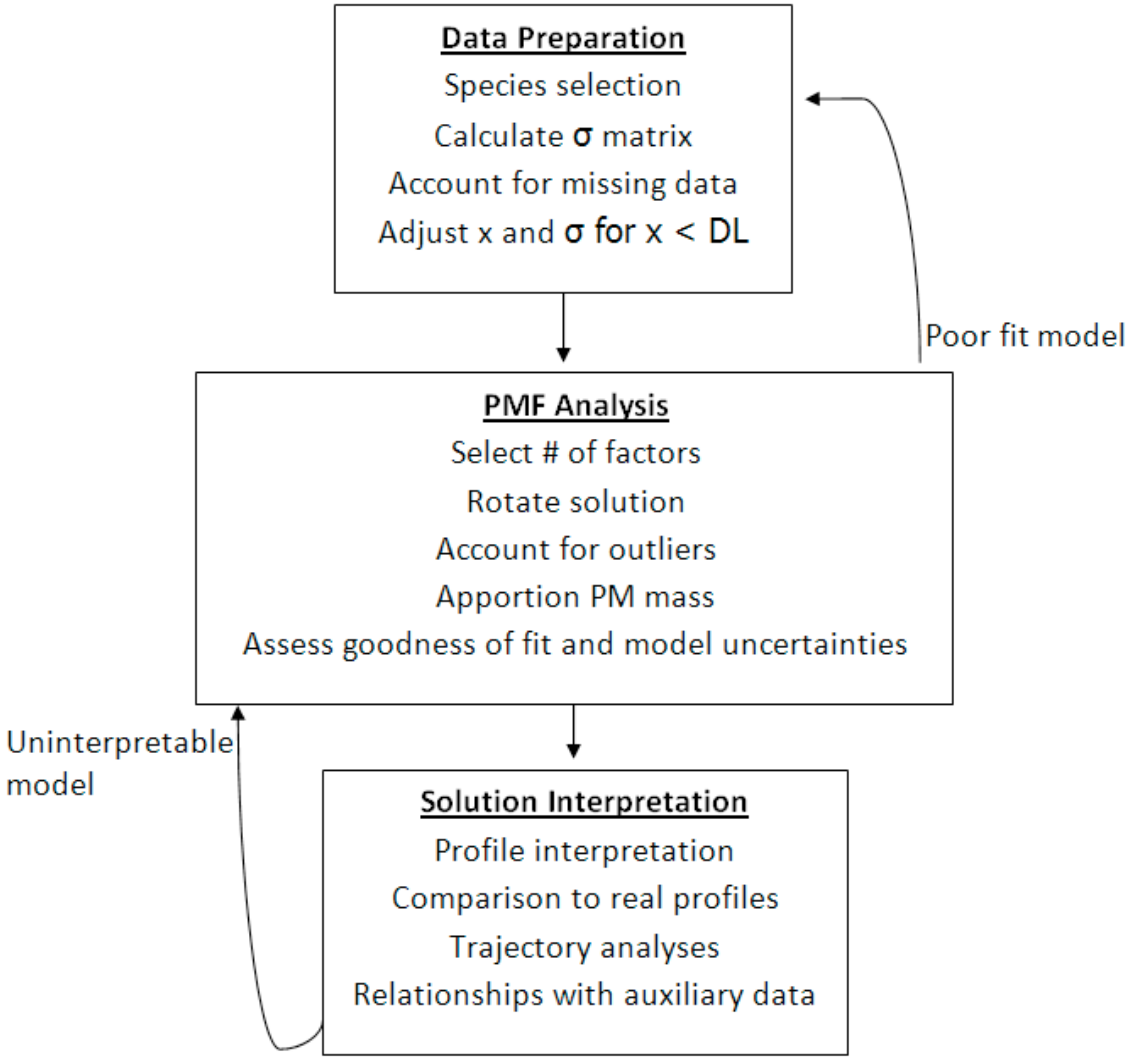


Figure I.1. General work flow for PMF analyses (Source Reff *et al*., 2007).

1.4. Apportioning Total PM (w)

There are two methods of apportioning total PM (w) [where, 'w' is replaced by mass/number concentration as per study objectives]. Firstly, to use total PM (w) as species along with the other particulate species in the data matrix and subsequently analyzed by PMF, then the PMF will apportion total PM (w) to each factors just as it apportions to other particulate species. This can be considered as an example of double counting as all the species used in the PMF analysis are already contained in total PM

(w). Secondly, to exclude total PM (w) from the data matrix and to regress PMF factor contributions onto the total PM (w) as shown in Eq. (5), Reff *et al*., (2007).

$$PM_i = \sum_{k=1}^{p} g_{ik} a_k \quad (5)$$

Where,

PM_i is the total PM (w) measurement from sample i, and a_k is the regression coefficient for factor k resulting from regressing the factor contributions (g_{ik}) onto PM_i. Negative values of a_k indicate that too many factors have been selected in the PMF model. If so, then the PMF analysis is redone with fewer factors, and the PMF factor contributions are regressed onto the PM (w) measurements.

1.5. Study Organization

- Literature survey on the subject is covered in **Chapter-II** to a limited extent.
- Information with respect to the receptor site, the two sampling period datasets comprising, particle number concentrations, meteorological and gaseous data, which formed the basis of materials required to be input as data matrix to the PMF2 program is discussed in **Chapter-III**. The methods involved in the preparation of the data and uncertainty matrix are also discussed.
- The **Chapter-IV** deals with Results and Discussion of the PMF2 deduced solutions.
- The study Conclusion is given in **Chapter-V** followed by the References and Annex.

CHAPTER-II
LITERATURE REVIEW

CHAPTER-II
LITERATURE REVIEW

Receptor modeling based on aerosol composition data obtained at a sampling site is often considered a reliable way to provide information regarding source characteristics (Hopke, 1985, 1991; Gordon, 1988). Recently, there have been significant improvements in multivariate models. One of those developments, Positive Matrix Factorization (PMF), a variant of factor analysis that constrains factor loadings and factor scores to non-negative values provides a flexible modeling approach that can effectively use the information in the data. The PMF as a factor analytic tool in receptor modeling studies and superior to the customary Principal Component Analysis (PCA) technique was first introduced by Paatero and Tapper, 1994. It is shown that the solutions by PMF are usually different than any solutions produced by the customary Factor Analysis (FA) i.e., PCA followed by rotations. Usually PMF produces better fit to the data than FA. Further, the result of PMF is guaranteed to be non-negative. The goodness of PMF are: it does not generate noise factors, allows handling of missing values and outliers without undue loss of information, and it is expected to generate minimum variance factors because it performs an optimally weighted least squares fit thus suitable for many environmental applications (Paatero and Tapper, 1994; Paatero, 1997a). Several studies have shown that PMF is a powerful tool for receptor modeling of airborne PM and to successfully assess particle source contributions (Liu *et al.,* 2003a; Ramadan *et al.,* 2000; Qin *et al.,* 2002). As previously reported by Paatero, 2003, the estimated uncertainties of each chemical species that had scaled residuals larger than ±2 were increased in order to reduce the weights and consequently reduce their residuals (Morishita *et al.,* 2006). Since its introduction, it has been continuously modified and subjected to lot of improvements in its ease of application and understanding.

2.1. Recent PMF Studies

In the last decade lot of aerosol researchers have successfully applied PMF to identify and apportion gaseous and ambient PM sources in numerous locations around the

world. The research paper, place of research and the investigating authors list is mentioned chronologically according to publication year. Finland (Anttila *et al.,* 1995 and Yli-Tuomi *et al.,* 1996), Alaska (Polissar *et al.,* 1998), Alert, Northwest Territories and Canada (Hopke *et al.,* 1998 and 1999), Hong Kong (Lee *et al.,* 1999), Narragansett, RI (Huang *et al.,* 1999), Arctic (Xie *et al.,* 1999), Thailand (Chueinta *et al.,* 2000), Phoenix, AZ (Ramadan *et al.,* 2000), Mace Head, Ireland (Huang *et al.,* 2001), Northeastern US (Washington, DC, Brigantine, NJ and Underhill, VT) (Song *et al.,* 2001), Vermont (Polissar *et al.,* 2001 and Poirot *et al.,* 2001) Brigantine National Wildlife Refuge, NJ (Lee *et al.,* 2002), Potsdam and Stockton, NY (Liu *et al.,* 2003a), Atlanta (Kim *et al.,* 2003a, 2003c), Northwest U.S. city Spokane, WA (Kim *et al.,* 2003b), Houston, Texas (Buzcu *et al.,* 2003), Western United states- (IMPROVE) program (Liu *et al.,* 2003b), Dundee, U.K. (Qin and Oduyemi, 2003), Eastern United States (Lapina and Paterson, 2004), Hanoi, Vietnam (Hien *et al.,* 2004), Santiago, Chile (Jorquera and Rappenglück, 2004), Toronto, Canada (Owega *et al.,* 2004), New York city, NY (Li *et al.,* 2004), Seattle, WA (Kim *et al.,* 2004e), Beijing, China (Song *et al.,* 2004), New York city (Ito *et al.,* 2004), Pittsburg (Zhou *et al.,* 2004), Washington DC (Kim and Hopke, 2004a, b), Atlanta (Kim *et al.,* 2004c), Seattle UNMIX and PMF (Kim *et al.,* 2004d), San Gorgonio, CA (Zhao and Hopke, 2004), Hotspot area in Dhaka, Bangladesh (Begum *et al.,* 2005), Washington DC (Begum *et al.,* 2006), Hanoi, Vietnam (Hien *et al.,* 2005), Central Chile (Hedberg *et al.,* 2005), Pittsburg (Zhou *et al.,* 2005), Bondville, ILon (Kim *et al.,* 2005a), St. Louis (Kim *et al.,* 2005b), Auckland, New Zealand (Wang and Shooter, 2005), Baltimore (Ogulei *et al.,* 2006), Detroit, MI (Morishita *et al.,* 2006), Houston, Texas (Xie and Berkowitz, 2006), Lake Champlain Basin, VT (Gao *et al.,* 2006), Beijing, China (Song *et al.,* 2006), NETL Pittsburg site (Eatough *et al.,* 2006), Hong Kong (Yuan *et al.,* 2006), Pittsburg supersite (Pekney *et al.,* 2006a (Part 1) and Bein *et al.,* 2006 (Part 2), Alsasua, Spain (Zabalza *et al.,* 2006), San Jose STN sites (Hwang and Hopke, 2006), St. Louis, MO (Lee and Hopke, 2006a), St. Louis-Midwest Supersite (Lee *et al.,* 2006b), Pittsburg (Pekney *et al.,* 2006b), New York city (Qin *et al.,* 2006), Indianapolis, IN (Zhao and Hopke, 2006a), Mammoth Cave National park Improve site in Kentucky (Zhao and Hopke 2006b), United States-Canada trade bridge (Ogulei *et al.,* 2007a), West coastal IMPROVE site (Hwang and Hopke,

2007), Southwestern Oregon USEPA STN in Chicago, Illinois (Rizzo and Scheff, 2007), Detroit, MI (Gildemeister *et al.,* 2007), Los Angeles (Brown *et al.,* 2007), Rochester, NY (Ogulei *et al.,* 2007b).

The PMF has also been accepted and approved by EPA that has led to the development of graphical interface PMF 1.1 multilinear engine version available for free download from US EPA official website in collaboration with Dr. Pentti Paatero, University of Helsinki, Finland.

CHAPTER-III
MATERIALS AND METHODS

CHAPTER-III
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, ^oC), 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:

3.1. Receptor Site

The receptor site was a well-equipped rooftop sampling station (at height about 25m above street level, 225m ASL) belonging to the Institute for Environmental Studies, Charles University (latitude-50^o 4' 17.46" N; longitude-14^o 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 is shown in Fig. III.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 100m 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. 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 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.

3.2. Sampling Periods and Parameters Measured

In this study, the two periods of sampling data measured at the receptor site were selected for the PMF2 Analysis. The sampling data set one was obtained between 17th February and 10th April 2007 (weekdays: 37 and weekends: 16, total: 53 days). The sampling data set two was obtained between 7th and 23rd January 2008 (weekdays: 13 and weekends: 04, total: 17 days). The parameters measured, the instrumentation used and the time resolution of data obtained are summarized in Table 3.1 and Table 3.2 respectively. The brief principle of operation of Scanning Mobility Particle Sizer (SMPS) is given in ***Section 3.3.*** The principle of gaseous measurements is given in parenthesis and the descriptive method of measurement is given in the subsequent ***Section 3.4***.

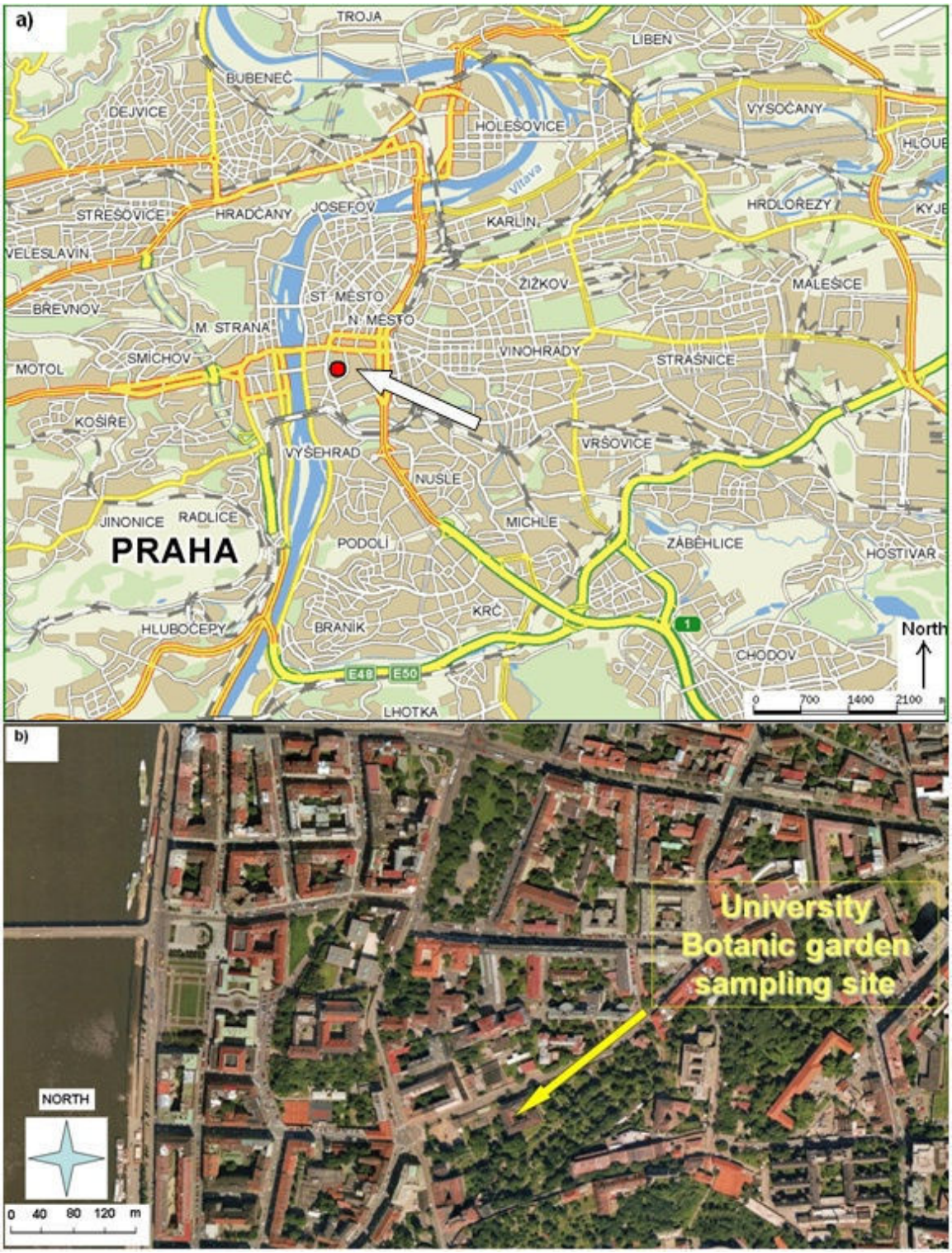


Figure III.1. (a) Prague (Praha) city and measurement site location (red circle); **(b)** Photo map of the measurement site.

Table 3.1. Summary of the sampling data set one (17th February to 10th April 2007).

Parameters	Instrumentation	Integration Time (minutes)
Particle Size	SMPS-3936L25	
14.6-736.5 nm	(TSI Inc., MN, USA)	10
Meteorological Data	Young vane type monitor	
WD, WS, T, RH, UV-A and UV-B	(Model 05103v, Fondriest Environmental Inc., USA)	15
Rain Intensity	Laser Precipitation Monitor (LPM Disdrometer, Adolf Thies Clima Inc., Germany)	01
Gaseous Components	***Horiba 360 series Analyzers:***	
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)	15
O_3	APOA (Ultra-violet-absorption method, NDUV)	
CH_4, NMHC and THC	APHA (Flame ionization detection method (FID) with selective combustion)	

Table 3.2. Summary of the sampling data set two (7th to 23rd January 2008).

Parameters	Instrumentation	Integration Time (minutes)
Particle Size	SMPS-3936L25	
14.6-736.5 nm	(TSI Inc., MN, USA)	05
Meteorological Data	Young vane type monitor	
WD, WS, T, RH, UV-A and UV-B	(Model 05103v, Fondriest Environmental Inc., USA)	05
Rain Intensity	Laser Precipitation Monitor (LPM Disdrometer, Adolf Thies Clima Inc., Germany)	01
Gaseous Components	***Horiba 360 series Analyzers:***	
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)	05
O_3	APOA (Ultra-violet-absorption method, NDUV)	
CH_4, NMHC and THC	APHA (Flame ionization detection method (FID) with selective combustion)	

3.3. PARTICLE SIZE MEASUREMENTS

3.3.1. The SMPS System

The Scanning Mobility Particle Sizer classifier (SMPS 3936L25, TSI Inc., USA) fitted with medium impactor type (Impactor nozzle diameter=0.0508 cm, aerosol flow=0.3 l/min and sheath flow=3.0 l/min) operated in the size range from 14.6 to 736.5 nm on the basis of their mobility diameter. The SMPS measures size distribution of particles using an electrical mobility detection technique. It uses a bipolar charger in the Electrostatic Classifier to charge the particles to a known charge distribution. The particles are then classified according to their ability to traverse an electrical field, and counted with Condensation Particle Counter (CPC 3025A, TSI Inc., USA). The Kr^{85} radioactive source aerosol neutralizer (3077/A, TSI Inc., USA) was used to neutralize polydisperse aerosols. Detailed operation principle of SMPS system can be found elsewhere (www.tsi.com).

3.4.GASEOUS MEASURMENTS PRINCIPLE

A brief discussion of gaseous measurements is outlined in the subsequent sections, necessary to understand the measurement principle. A detailed overview of the operation principle can be found elsewhere (www.horiba.com).

3.4.1. Ambient CO Monitor (APMA-360)

Horiba APMA monitor uses cross-flow modulated non-dispersive infrared absorption (NDIR) method to measure CO with good stability and high sensitivity (5 ppm). It uses AS-type interference-compensating detector, and a flowing reference gas. The reference gas is generated by flowing sample gas through an oxidation process where an oxidizing catalyst burns CO to CO_2. This results in accurate measurements and eliminates the interference from other elements. In its principle, fixed amount of sample gas and reference gas are injected alternately into the measurement cell by the solenoid

valve and the modulation signal is generated. By means of subtraction the actual signal obtained, the CO is determined. The monitor is approved by U.S. EPA for atmospheric measurements.

3.4.2. Ambient NO_x ($NO + NO_2$) Monitor (APNA-360)

Horiba APNA-360 monitor operates using a cross-flow modulated reduced pressure chemiluminescence (CLD) method. It measures atmospheric NO_x, NO_2 and NO by a combination of dual cross-flow modulation type chemiluminescence principle and adopting the subtraction method. It offers continuous long term measurements with high sensitivity (0.1 ppm) and good stability. It is approved by the U.S. EPA and offers sensitive, accurate and precise atmospheric pollution monitoring data. It is compact and no supplemental gas is required. The unit includes a reference gas generator, an ozone-source drier unit, an ozone decomposer and a sampling pump. The CLD principle is outlined below:

The chemiluminescence method uses the reaction of NO with O_3:

$$NO + O_3 \rightarrow NO_2 + O_2$$
$$NO + O_3 \rightarrow NO_2^* + O_2$$
$$NO_2^* \rightarrow NO_2 + hv$$

A portion of NO_2 generated by the outcome of reaction becomes NO_2^*. Chemiluminescence is generated in the range between 600 and 3000 nm when excited molecules NO_2^* returns to ground state. The light intensity is proportional to the concentration of NO molecules, by which the NO concentration of the sample is obtained. A reducing convertor changes NO_2 to NO, and is also measured. In other words, the NO_2 concentration can be obtained by the difference between the NO_x concentration measured when sample gas is directed through a convertor and the NO concentration measured when the gas is not run through the convertor.

3.4.3. Ambient SO_2 Monitor (APSA-360)

Horiba APSA-360 monitor operates on UV-fluorescence principle. It offers high sensitivity (0.05 ppm) and stability with minimum influence from moisture due to its unique, new fluorescent chamber design. The built in aromatic-hydrocarbon cutter with selective transmission reduces interference components. It requires no supplemental gas. It is approved by the U.S. EPA for atmospheric measurement data. The SO_2 molecules in the sample are excited by Ultra-Violet (UV) radiation; later, they emit characteristic fluorescence in the range of 220-240 nm. This fluorescence is measured and the SO_2 concentration is obtained by the changes in the intensity of the fluorescence. The reactive mechanism is given below:

$SO_2 + hv1 \rightarrow SO_2^*$ (SO_2 absorbs energy hvl by UV-radiation to form excited SO_2^*)
$SO_2^* \rightarrow SO + (O)$ (hv2 released when excited molecules return to ground state)
$SO_2^* \rightarrow SO_2 + hv2$ (decomposition by the light emitted from the excited molecules)
$SO_2^* + M \rightarrow SO_2 + M$ (quenching, i.e., energy lost on collision with other molecules)

3.4.4. Ambient O_3 Monitor (APOA-360)

Horiba APOA monitor uses cross-flow modulation type ultra-violet-absorption method (NDUV) in conjunction with comparative calculation method. It is approved by the U.S. EPA and permits continuous measurement with great stability and sensitivity (0.1 ppm). The heated deozoniser provides reference gas by decomposing the O_3 found in sample gas. This reduces the interference and makes it insensitive to great changes in moisture content. In its working principle, the O_3 absorbs UV-rays in the area of 254 nm. Measurements are done continuously, with alternate injections of the sample gas and the reference gas controlled by solenoid valve. A comparative calculation method compensates for all fluctuations in the mercury vapor light source and in the detector providing continuous measurements.

3.4.5. Ambient THC (CH_4 and NMHC) Monitor (APHA-360)

Horiba APHA monitor uses flame ionization detection (FID) method with selective combustion, with single detector method to provide continuous measurements of Total Hydrocarbons (THC), NMHC and CH_4 with high sensitivity (5 ppm). A catalytic unit for generating reference gas and auxiliary combustion air is all included within the unit for selective combustion (i.e., NMHC chopper). The supplemental gas Hydrogen is required. The FID principle in combination with selective combustion system, utilize the ionization that occurs as a result of high-temperature energy from combustion at the tip of the burner jet when carbon compounds are introduced into the H_2 flame. The H_2 flame is located between the electrodes and when electrical voltage is applied between these electrodes, a minute ion current proportional to the hydrocarbon (HC) concentration is produced. This current is amplified giving voltage readout for THC. To measure CH_4 the sample gas is passed through selective catalytic combustion (i.e., NMHC chopper) which oxidizes NMHC, without oxidizing CH_4, thus concentration of CH_4 is measured (denoted by 'A'). Let 'B' represents THC concentration without passing through selective catalytic combustion. Thus, by simple subtraction, (B-A) gives the concentration of NMHC. Final adjustments to calculation of concentration value using a relative sensitivity correction function coefficient; '*k*' is done as shown below:

CH_4 concentration = A

NMHC concentration = k (B-A)

THC concentration = A + k (B-A)

3.5. SAMPLING PERIOD DESCRIPTION:

3.5.1. Sampling Period One

The data selected for the study purpose was obtained between 17th February, 2007 and 10th April, 2007. It is still winter period during this time with low temperatures. The

lowest temperature recorded was -2.1^{o}C and the highest 18.4^{o}C with an average of 6.6 ± 4.0^{o}C (Standard Deviation-SD). Relative humidity was between 21 and 85.4 % (avg. RH 63.7 %). Wind speed was between 0.20 and 8.3 m/s (avg. WS 1.6 m/s). Compared to the previous years, this period was the warmest winter. Wet precipitation occurred on the following dates, 16th, 19th, 22nd, 25-28th February 2007; 1st -4th, 8th, 10th, 17th, 18th, 20-24th March, 2007 and 3rd April, 2007). In general, the residential and space heating systems were believed to be operated on most days and some days without any heating required. The vehicular traffic had no major changes and remained the same except during the weekends and state holidays. Few sunny days were also reported and active UV radiations which influence the ozone photochemistry and formation of secondary particles. The time-series plot of meteorological and gaseous data for sampling period-I, including rain intensity is given in Fig. III.2 with original time-resolution data. The temporal contours of SMPS (14.6-736.5 nm) particle number counts for sampling period-I is given in Fig. III.3 (see ***Section 3.6*** for PNC data worthiness).

The above data (except rain intensity) was reduced to hourly time integrates by arithmetic averaging and the data matrix (Observed/Measured matrix) was prepared for selected species, later to be input to the PMF2 software. The general procedures related to data matrix preparation and the selected species are discussed in subsequent sections. The directionality analysis of species with respect to the wind directions, were calculated using the Conditional Probability Function (CPF) plots. The use of CPF method is defined in ***Section 3.7***. Though there was 5-min time shift in the time-resolution of gaseous data as compared to the SMPS data obtained (see, Table 3.1) it was assumed not to have much change and uniformly synchronized and reduced to hourly concentrations for the convenience of data handling and matrix preparation.

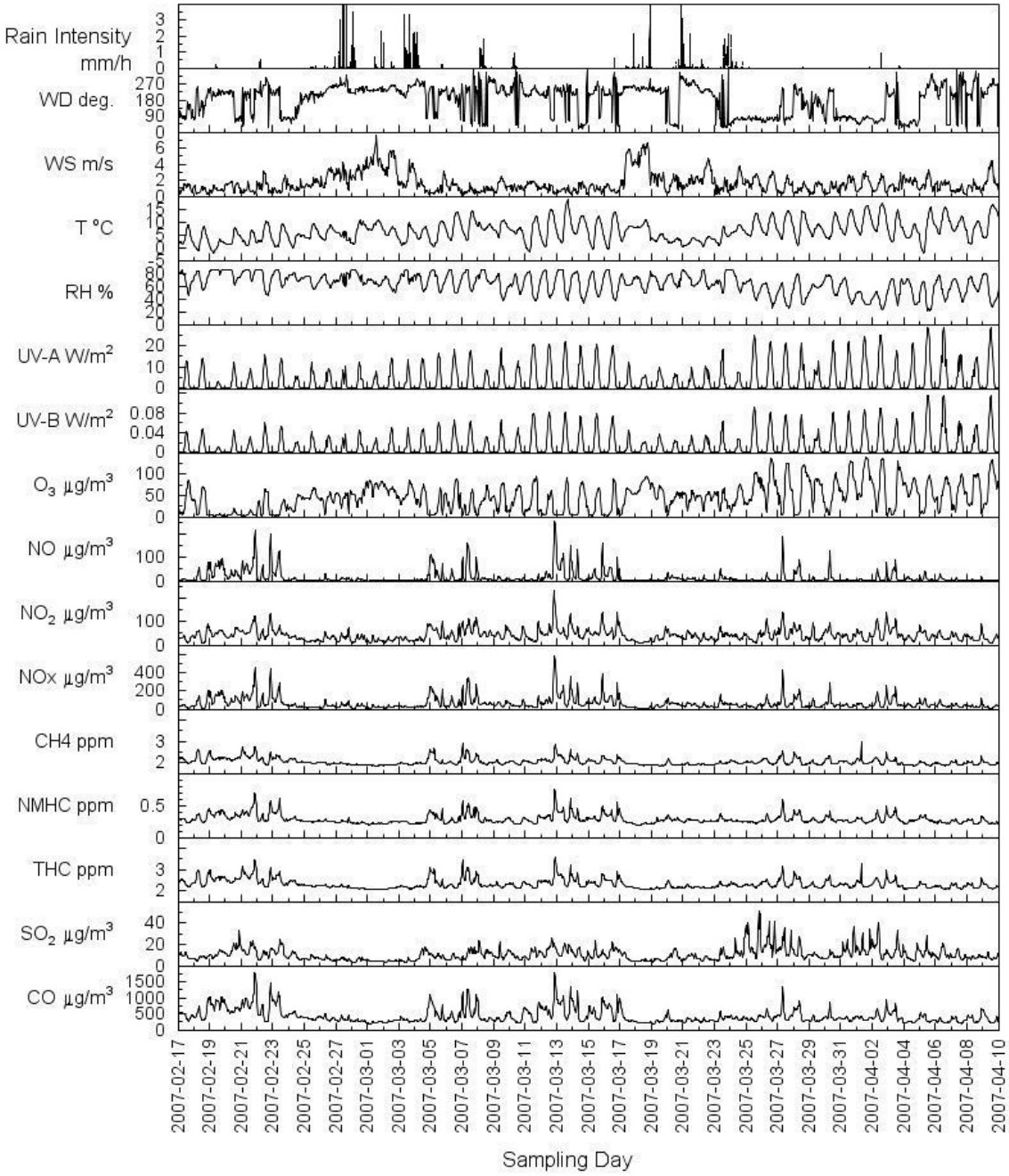


Figure III.2. Time-series of meteorological and gaseous data for sampling period-I, including rain intensity.

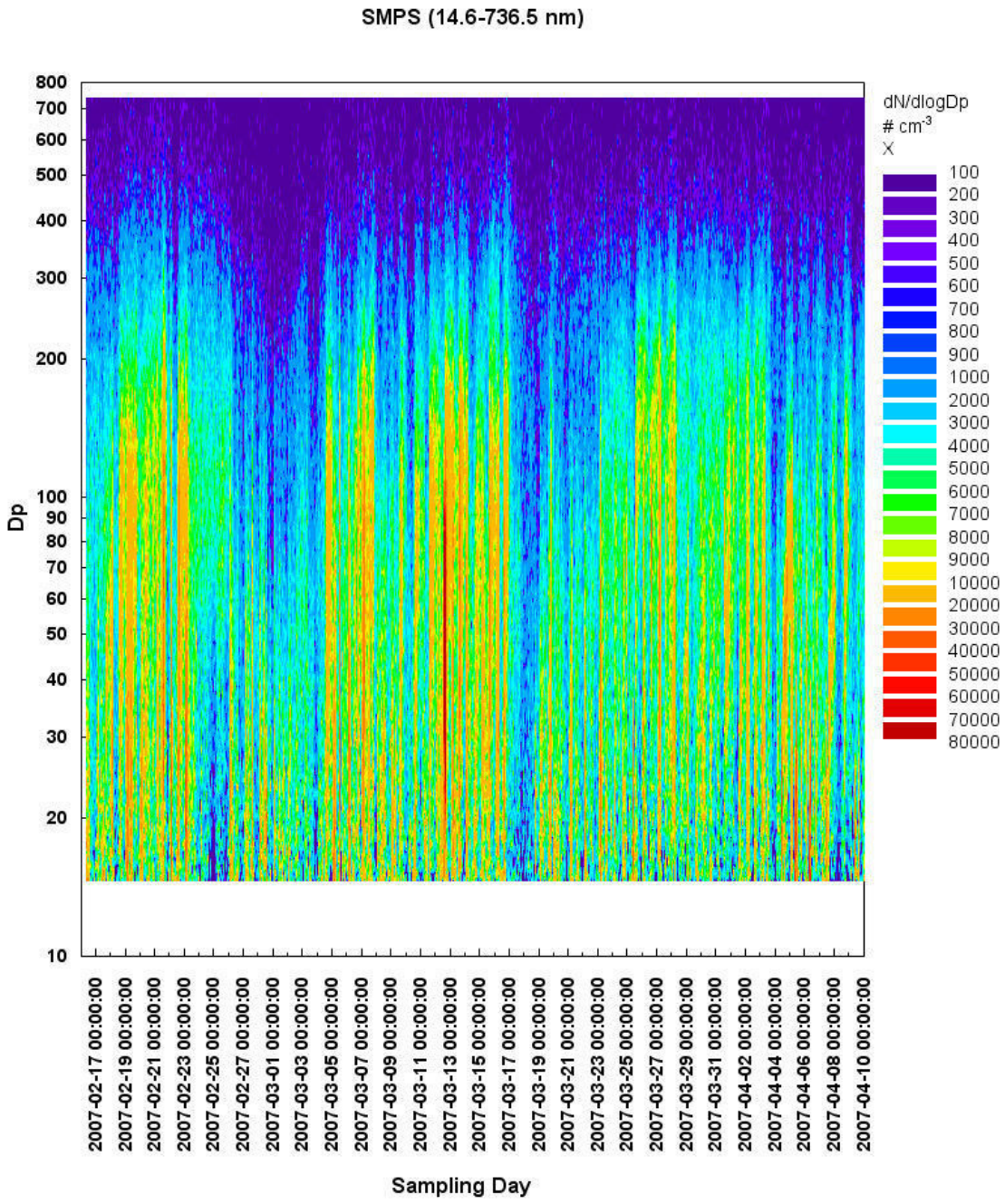


Figure III.3. Temporal contours of SMPS particle number counts for sampling period-I.

3.5.2. Sampling Period Two

The data selected for the study purpose was obtained between 7th and 23rd January 2008. 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). Relative humidity was between 52 and 85% (avg. RH 76.4 ± 7.6%). Wind speed 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 January, 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. The time-series plot of meteorological and gaseous data for sampling period-II, including rain intensity is given in Fig. III.4 with original time-resolution data. The temporal contours of SMPS (14.6-736.5 nm) particle number counts for sampling period-II is given in Fig. III.5 (see ***Section 3.6*** for PNC data worthiness).

Similarly, as described earlier in ***Section 3.5.1***, for the sampling period-II parameters, the hourly averages were recalculated after thoroughly checking for any outliers in the datasheet, outliers if found were replaced by the mean value of the preceding and the next observed value. However, only 15-20 outlying data points were found and were replaced with the mean value. The rain intensity values were not touched and 1-min time resolution information was retained. The directionality analysis of species with respect to their wind directions was done by CPF plots and the method is defined in ***Section 3.7***. The general procedures related to the selection of species and data matrix preparation are also discussed in subsequent sections.

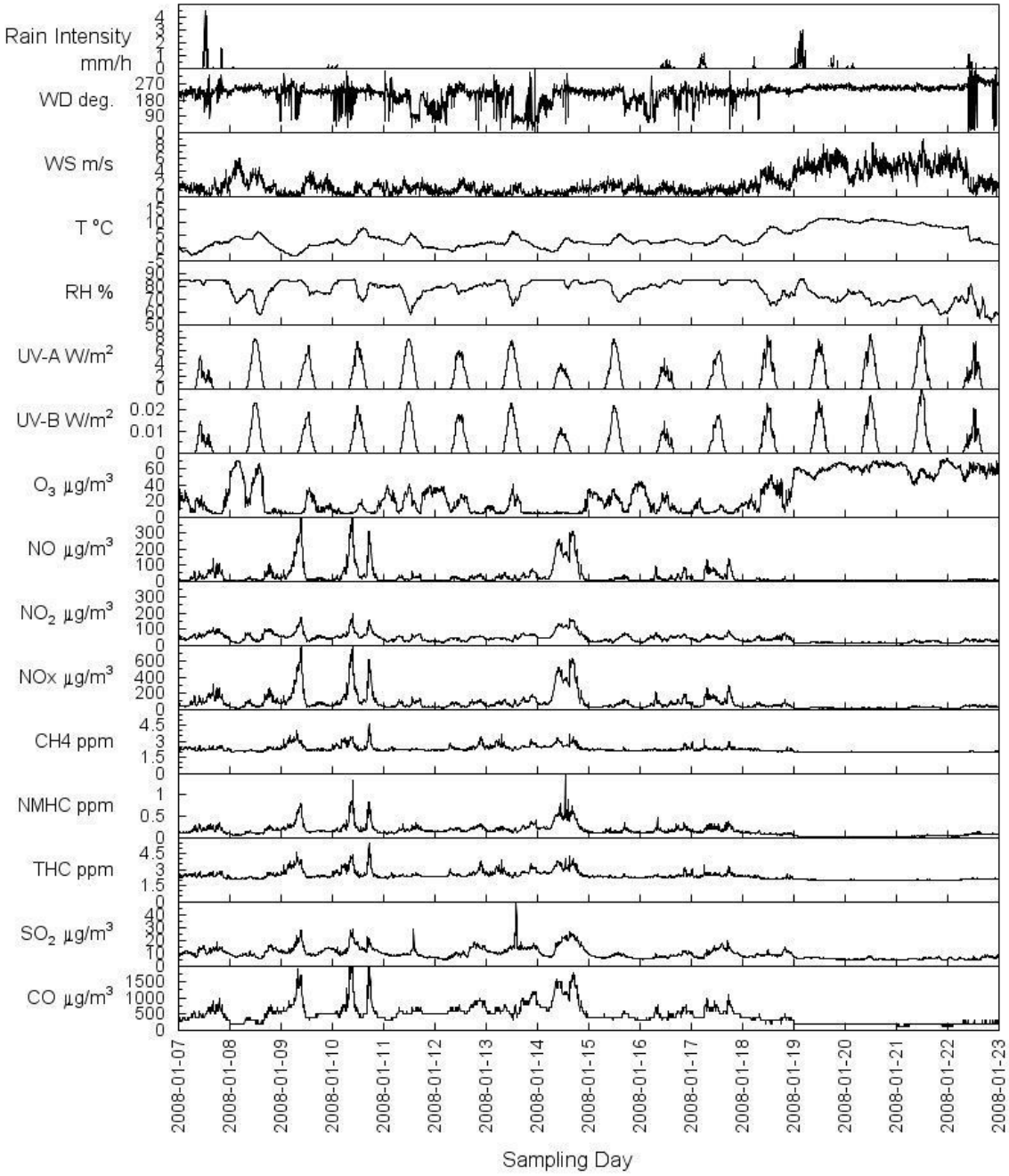


Figure III.4. Time-series of meteorological and gaseous data for sampling period-II, including rain intensity.

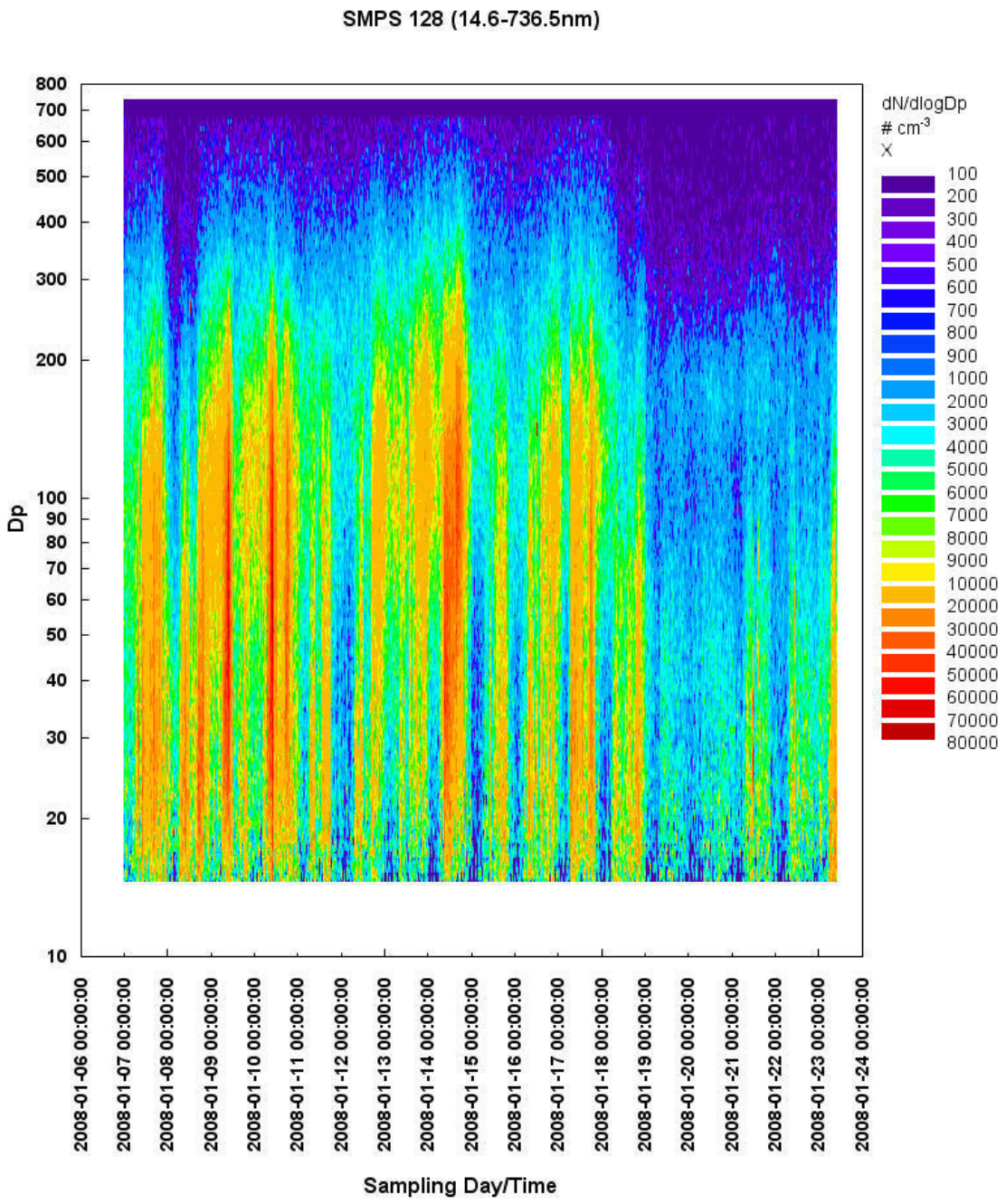


Figure III.5. Temporal contours of SMPS particle number counts for sampling period-II.

3.6. Worthiness of PNC data to fit PMF Model

Receptor models require 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 distributions caused by particle growth will deform the model solutions and make the sources unidentifiable (Zhou *et al*., 2005). Thus the temporal contours of PNC for each measurement days were drawn to check if the regional nucleation events occurred?. In a regional nucleation event, the newly formed particles continue growing. Particle growth events are confined to limited time intervals and size ranges from above 10 nm up to accumulation mode size. These particle growth events were observed at many other places, the experimental and theoretical studies on this phenomenon can be found elsewhere (Kulmala *et al*., 2004). Temporal contours of PNC for every single day of the two measurement periods were analyzed and no nucleation events were found. The Figs. III.3 and III.5 shows the temporal contours of SMPS PNC for the two measurement periods, i.e., sampling periods-I and II respectively. The individual day's temporal contours drawn are not shown here, as they number into total 70 graphs. Further, the particle size distributions vary significantly, i.e., seasonally and also depending on atmospheric processes. The environmental processes are dynamic and vary within the same season in short periods. The PMF analysis requires stationary or quasistationary size distributions measured at the receptor site. In a short period, the 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. Thus, a considerable approximation has been arrived in mindset by assuming such changes as minor.

3.7. Conditional Probability Function (CPF)

A conditional probability function (Ashbaugh *et al.,* 1985; Kim *et al.,* 2004f) was calculated by coupling the ambient gaseous components data to the wind direction values as measured at the receptor site. Similarly, wind direction values were assigned to the

source contributions of the factors deduced by PMF2 program (see Chapter-IV, Results and Discussions). The CPF is defined by Eq. (6).

$$\boldsymbol{CPF = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \qquad n_{\Delta\theta} > n_c} \tag{6}$$

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 contributions from each source, and $n_{\Delta\theta}$ is the total number of wind occurrences in the same direction sector $\Delta\theta$. Fractional contributions are used to avoid the influence of atmospheric dilution on CPF results. 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 gaseous components species using CPF, and the wind roses for the two sampling periods, the sampling periods-I and II are as shown in Figs. III.6 and III.7 respectively.

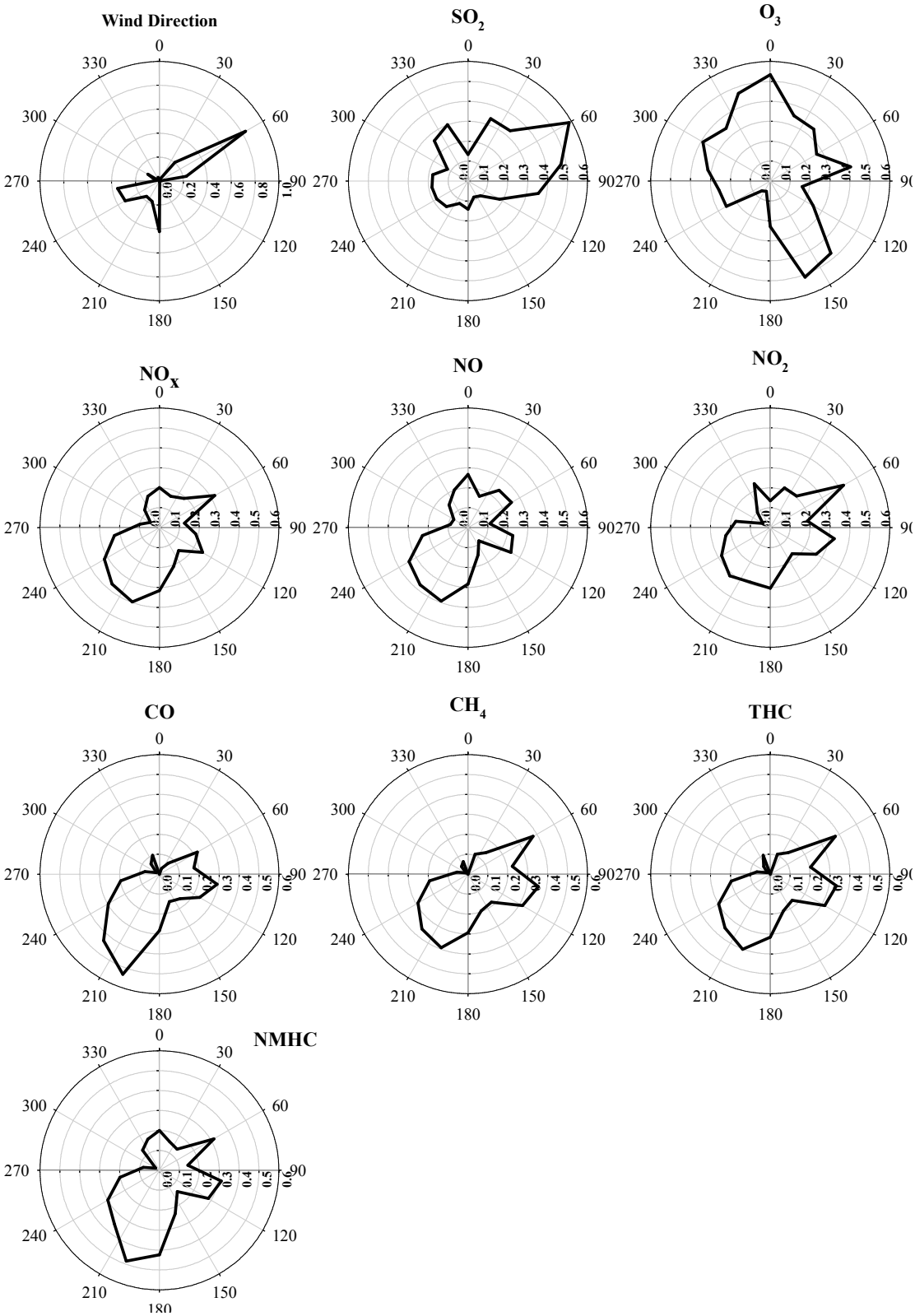


Figure III.6. Wind direction location of species using CPF for sampling period-I.

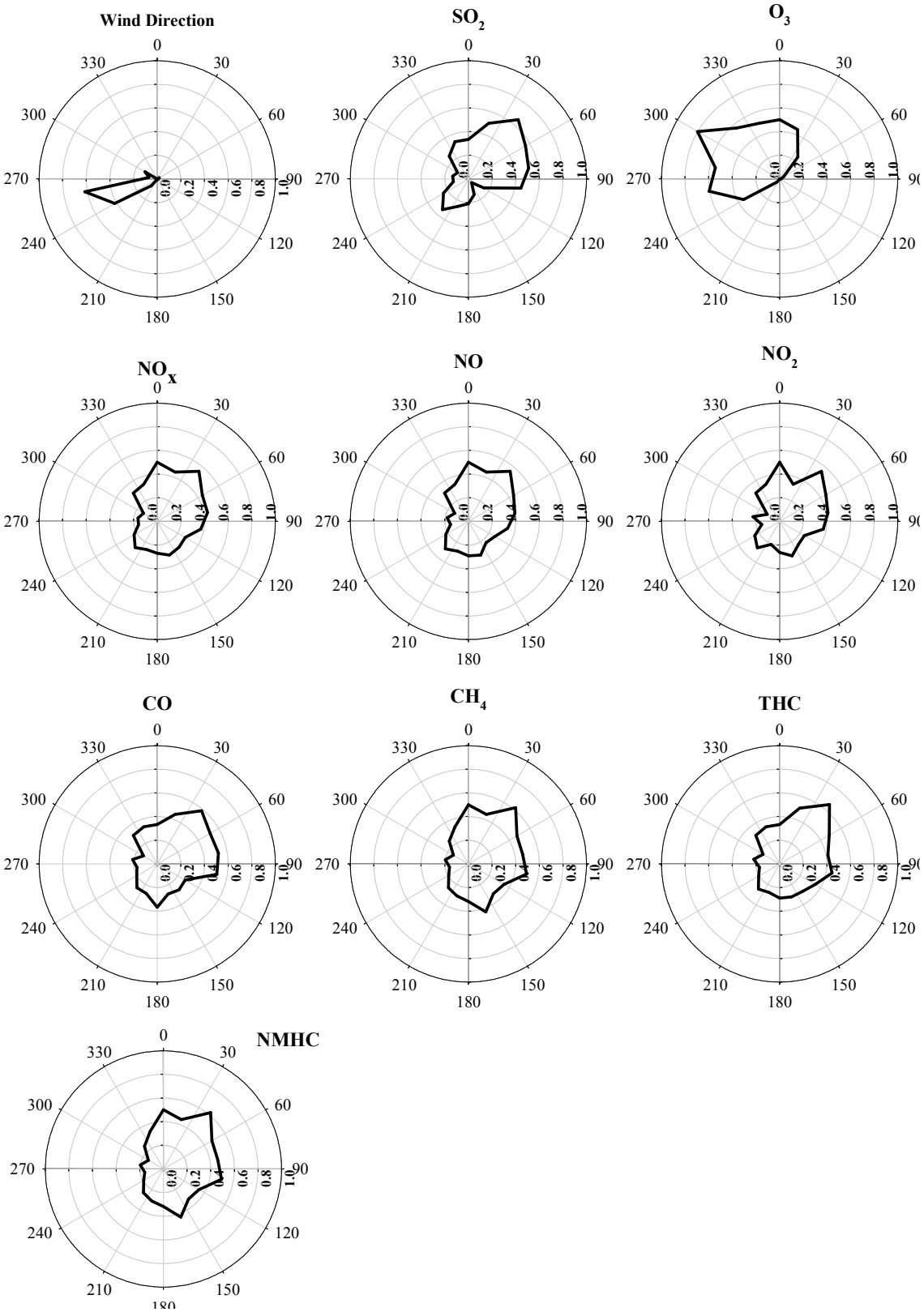


Figure III.7. Wind direction location of species using CPF for sampling period-II.

3.8. PMF2 Program Operation

To run PMF2, we need to input at least 2 matrices; one the data matrix and the second, the estimates of error known as error/uncertainty matrix. Both the matrices are to be saved as Comma Separated Variables (CSV). The methods used in preparation of the data matrix and uncertainty matrices are discussed in subsequent sections. A set of instructions for the program is saved in initiation file also called as the startup file, initiation file (INI). PFE32 text editor is used to make changes to the INI file. This INI file contains instructions for the PMF2 program which can be easily modified. All the above files should be in a single working directory. As example, the setting for the INI file for the sampling periods-I and II is given in Annex-I and II, respectively. The PMF2 program is run for different factor settings ranging from 3 to 10 (minimum 3 and maximum 14) with 5 repeats for individual factor settings (It is generally advisable to perform the analysis several times typically 5 to be certain that the same solution is obtained (Paatero, 1997a). The outlier distance (α) was set to 4.0. In PMF, the choice of number of factors is a compromise. Using too few factors will combine sources of different nature together. Using too many factors will make a real factor further dissociate into two or more non-existing sources.

The information from the residual matrix and the rotational matrix in PMF is useful in determining the number of factors and reduce the ambiguity due to manual judgment (Lee *et al*., 1999, Yakovleva *et al*., 1999). As one of the methods to determine the quality of the fit, the residuals e_{ij}, are examined. Typically the distributions of the residuals are plotted for each measured species. It is desirable to have symmetric distributions and to have all the residuals within +/- 3 standard deviations. Some researchers try to ensure that scaled residuals (e_{ij}/σ_{ij}) for most species in their data sets lie between certain limits, typically -2.0 and +2.0 (Liu *et al*., 2003a; Wang and Shooter, 2005). If the spread is large in residuals, then the number of factors should be reexamined. As an example, a histogram plot of residual analysis carried out for a selected species of column 28 is shown in Fig. III.10. The distributions with large spread might indicate that uncertainties are too high. There can be multiple combinations of

source profiles (F matrix) and source contributions (G matrix) that produce the same result. The source profiles and contribution plots were drawn for the minimum Q value obtained for individual runs. A suitable factor setting was chosen as increasing the number of factors did not show any credible additional sources. The scaled residuals showed symmetric distributions between +/- 3 standard deviations. Further, for the chosen factor setting (In this case, 5-Factors and 4-Factors for the sampling periods-I and II, respectively) the FPEAK analysis (Paatero, 1997a) was varied between -1.0 and +1.0 in steps of 0.2 and plot of Q as a function of FPEAK was drawn. The FPEAK setting with 0.0 was found to be the optimum with minimum Q values as shown in Figs. III.8 and III.9 for sampling periods-I and II, respectively. In PMF, 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. The use of FPEAK, Gkey and Fkey settings are described in introduction chapter, ***Section 1.2.*** The Fkey setting option was used for CO and SO_2 for sampling period-I as was showing up in all the factors deduced by the PMF2. However, the use of Fkey didn't show any improvement in the study. The Fkey setting was not used for sampling period-II as no species was showing up in places not necessary in the factor profiles which were physically interpreted as sources.

3.9. Data Matrix Preparation

Data point outliers if present may considerably distort the factor analysis approach leading to erroneous results and care was taken to remove the outliers. The Fig. III.11 (a, b) and Fig. III.12 (a, b) show before and after outliers were corrected for sampling periods-I and II, respectively. Unexpectedly high and low concentrations were identified by plotting graphs and the data points were replaced by the average concentrations of the preceding and the next sample. After successful replacement of outliers, the hourly integrates were recalculated from the original time-resolution of data obtained for the two measurement periods in MS excel spreadsheet by arithmetic averaging method. However, only few outlying data points were found and replaced by this method (only 20 and 15 replacements for sampling periods-I and II respectively).

Thus, there were 1279 and 384 hourly particle number size distributions in conjunction with the selected gaseous components for the sampling periods-I and II respectively. The particle number size distributions and gaseous components were synchronized according to 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 $\mu g/m^3$, CH_4 and NMHC were in ppm. Further, the original number of size bins in SMPS data file covering the size range between 14.6 to 736.5 nm were reduced to 35 columns by summing sets of three consecutive size bins as discussed below in ***Section 3.9.1***. The resulting new midpoint diameters of the size bins for both measurement periods are shown in Table 3.3 along with the selected gaseous components (CO, NO_x, SO_2, O_3, CH_4 and NMHC). Thus an input data matrix of hourly particle number size distribution and selected gaseous components was created with matrix size dimensions 1279 x 41 [(rows) x (columns)] and 384 x 41 respectively. Each individual row corresponds to each sample measured and columns the species selected, in this case, the particle size distributions in conjunction with the selective gaseous components. The uncertainty matrices of same dimensions as those of data matrices were prepared and is discussed below in ***Section 3.10***.

3.9.1. Particle Number Concentrations (PNC)

The SMPS recorded particle number concentrations between 14.6 and 736.5 nm. This range was reduced by summing up three consecutive size bins in order to reduce the number of columns for easy data handling. The resulting midpoint diameters for the two sampling periods are summarized in Table 3.3.

Table 3.3. Particle size bins plus gaseous species for the two sampling periods.

Sl. No.	Size Bins plus Gases	Midpoint Diameter Sampling Dataset One	Midpoint Diameter Sampling Dataset Two
1	Bin 1	15.1	-
2	Bin 2	16.8	-
3	Bin 3	18.8	18.8

Contd…

Table 3.3. contd...

Sl. No.	Size Bins plus Gases	Midpoint Diameter Sampling Dataset One	Midpoint Diameter Sampling Dataset Two
4	Bin 4	20.9	20.9
5	Bin 5	23.3	23.3
6	Bin 6	25.9	25.9
7	Bin 7	28.9	28.9
8	Bin 8	32.2	32.2
9	Bin 9	35.9	35.9
10	Bin 10	40.0	40.0
11	Bin 11	44.5	44.5
12	Bin 12	49.6	49.6
13	Bin 13	55.2	55.2
14	Bin 14	61.5	61.5
15	Bin 15	68.5	68.5
16	Bin 16	76.4	76.4
17	Bin 17	85.1	85.1
18	Bin 18	94.7	94.7
19	Bin 19	105.5	105.5
20	Bin 20	117.6	117.6
21	Bin 21	131.0	131.0
22	Bin 22	145.9	145.9
23	Bin 23	162.5	162.5
24	Bin 24	181.1	181.1
25	Bin 25	201.7	201.7
26	Bin 26	224.7	224.7
27	Bin 27	250.3	250.3
28	Bin 28	278.8	278.8
29	Bin 29	310.6	310.6
30	Bin 30	346.0	346.0
31	Bin 31	385.4	385.4
32	Bin 32	429.4	429.4

Contd…

Table 3.3. contd…

Sl. No.	Size Bins plus Gases	Midpoint Diameter Sampling Dataset One	Midpoint Diameter Sampling Dataset Two
33	Bin 33	478.3	478.3
34	Bin 34	532.8	532.8
35	Bin 35	593.5	593.5
36	Bin 36	661.2	661.2
37	Bin 37	723.5	723.5
38	Gas-1	CO	CO
39	Gas-2	NO_x	NO_x
40	Gas-3	SO_2	SO_2
41	Gas-4	O_3	O_3
42	Gas-5	-	CH_4
43	Gas-6	-	NMHC

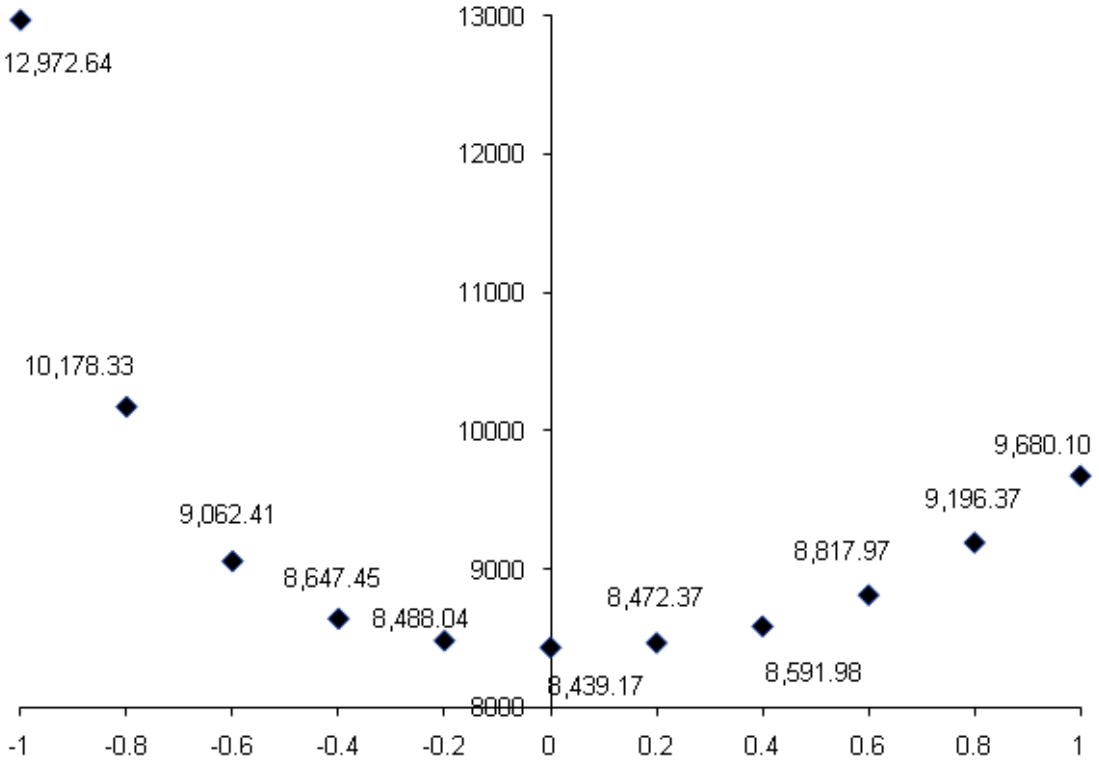


Figure III.8. Plot of Q as function of FPEAK b/w. -1.0 and +1.0 for sampling period-I.

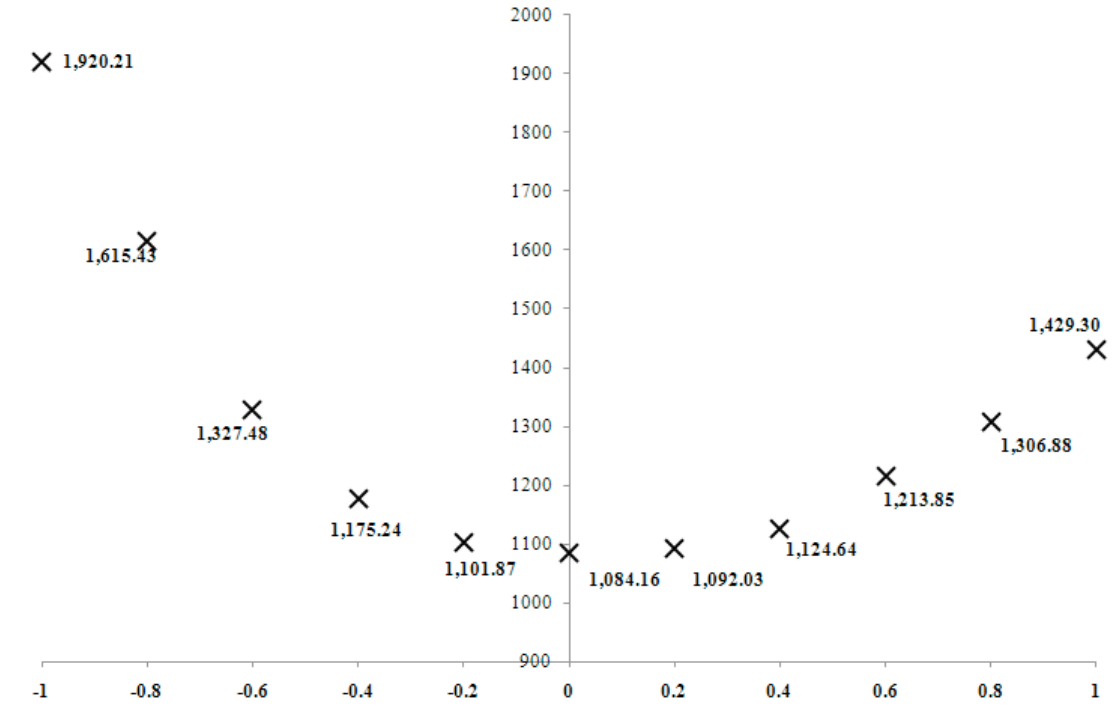


Figure III.9. Plot of Q as function of FPEAK b/w. -1.0 and +1.0 for sampling period-II.

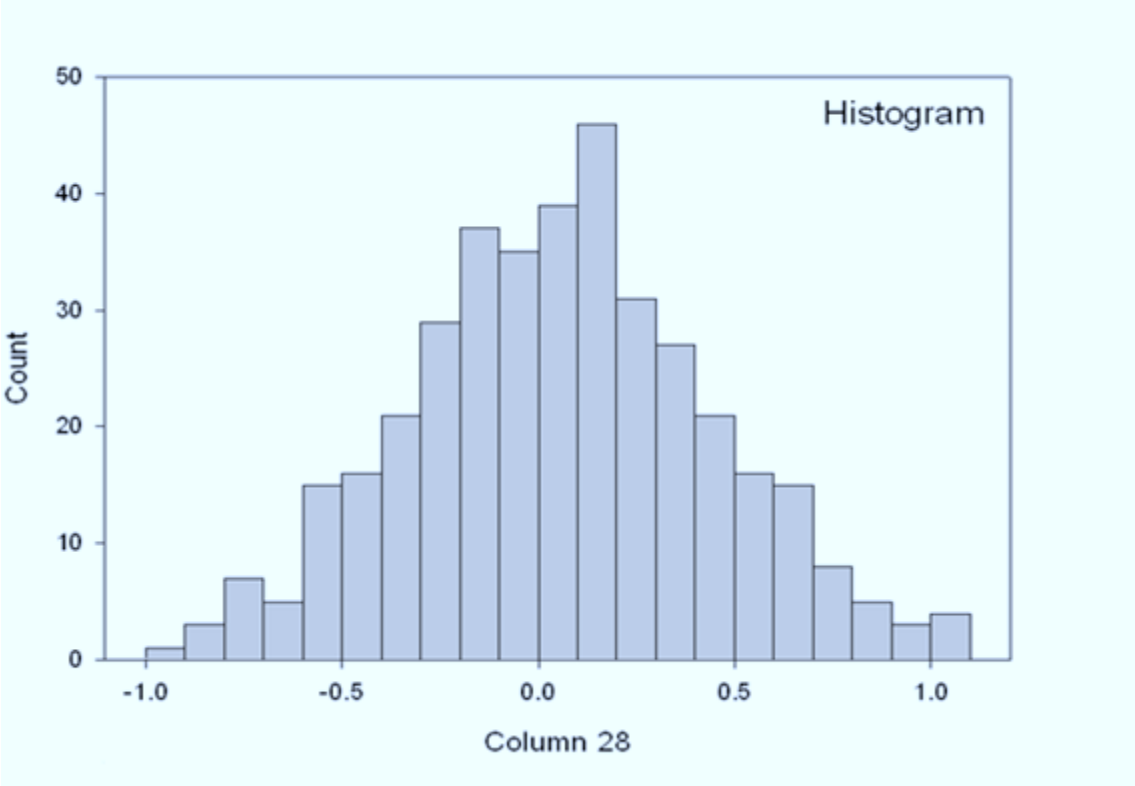


Figure III.10. An example of residual analysis, for the chosen species column.

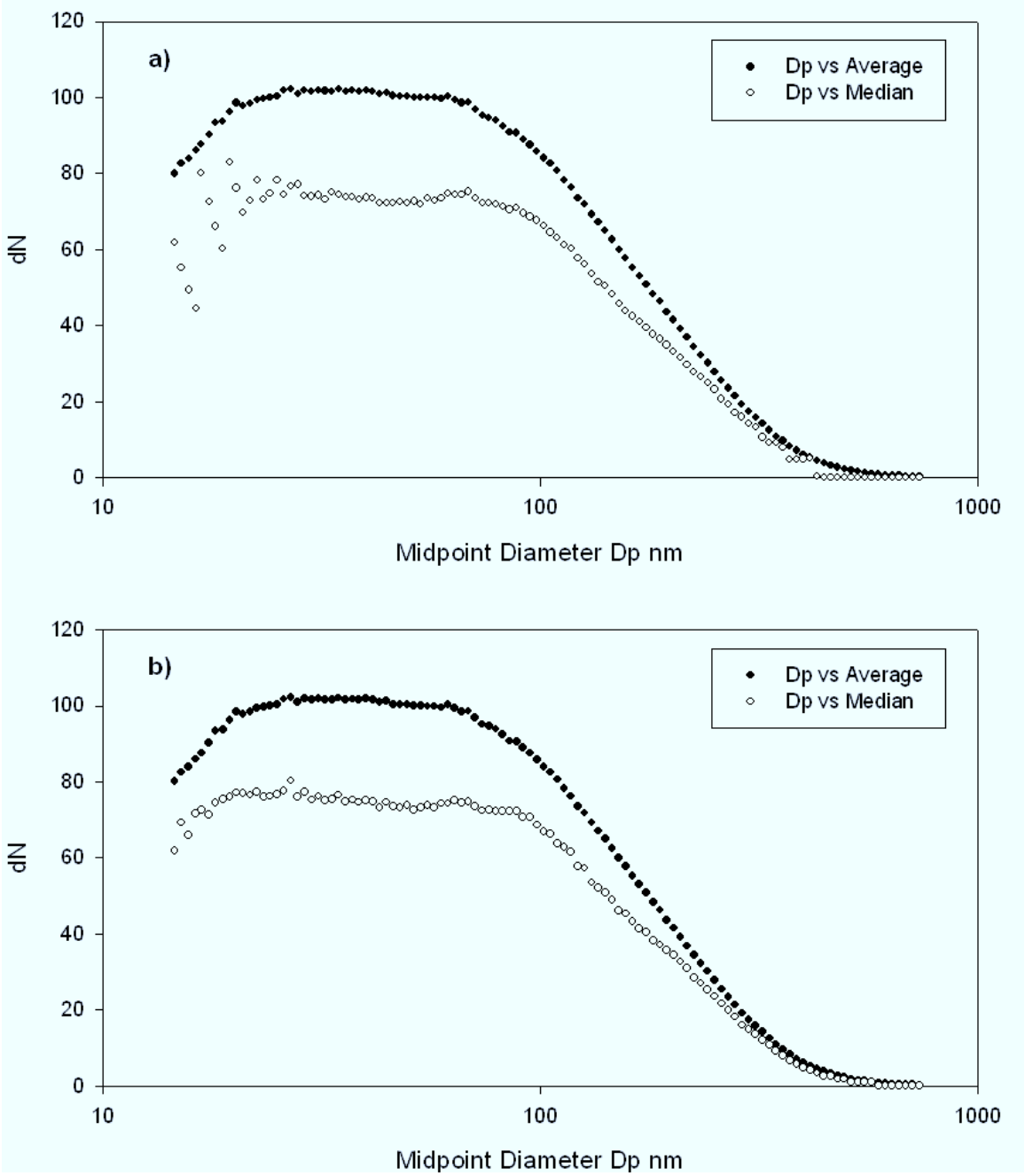


Figure III.11. (a) Before and **(b)** After outlier correction for sampling period-I.

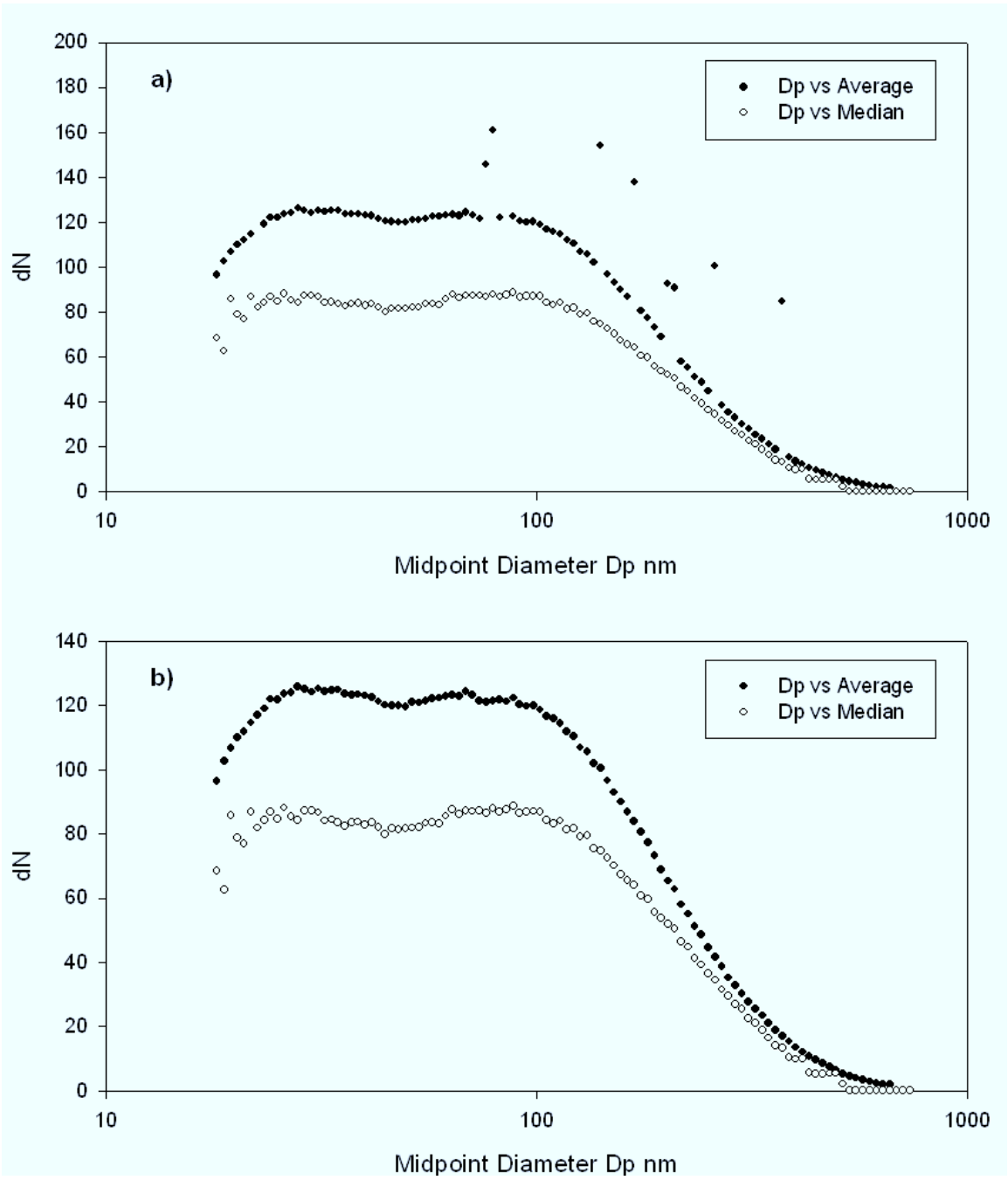


Figure III.12. (a) Before and **(b)** After outlier correction for sampling period-II.

3.10. Uncertainty Matrix Preparation

The error/uncertainty matrices with the same dimensions as with respect to the data matrices were 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. A review of uncertainty opted in many other studies can be found elsewhere (Reff *et al*., 2007).

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. (7) and gaseous PMF uncertainties are summarized in Table 3.4. 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.

$$\text{For particles,} \quad \boldsymbol{\sigma_{ij} = 1 + \sqrt{(x_{ij})} + 0.1 * (x_{ij})} \tag{7}$$

Table 3.4. 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})

3.11. Chemical Analysis

The best way for source apportionment studies is to have all the chemical speciation done and to know their concentration levels for the known mass of aerosol deposited on each filters. These are filter based measurements involving excess financial requirements and laborious time consuming processes for determining their chemical compositions. In the case of filter based measurements, the precision of results depends on the gravimetric measurements and the element instrument detection limits (DLs).

In this study, a new simple approach is being investigated to readily make use of the available real-time highly time resolved continuous particulate data recorded by SMPS instrument and to combine them along with the gaseous composition data in knowing their sources. This technique is feasible if the number of sources is expected to be relatively low. However, if there are many sources, it is advisable to have the chemical speciation data.

3.12. PMF2 Results

The subsequent sections list the number of factors deduced by PMF2 and the assigned source names (attribution of sources). The factors from PMF2 could be assigned to particle sources by examination of the number size distributions associated with the factors, the time frequency properties of each source contributions, and correlations of the contributions values with simultaneous gas phase measurements (O_3, NO_x, SO_2, CO, CH_4 and NMHC). Additionally, weekday/weekend effects were investigated. The CPF analysis was performed for each source deduced by PMF2 to ascertain the likely directions in which the sources are located. The comprehensive discussions with supporting materials are given in Chapter-IV, Results and Discussions.

3.12.1. Sampling Period One

The PMF2 analysis deduced 5 probable factors which were interpreted as sources in this study. They are mixed source-1, traffic, heating, mixed source-2 and O_3-secondary

particles. Out of 5 factors, 2 factors were assigned to be mixed sources-1 and 2. Often, some of the factors appear to be a mixture of several sources that cannot be further separated. Repetitive PMF2 run with different seeded values were performed with not much success. These two sources may be coming from background or long range transport. Similar difficulties in resolving the factors are encountered in other study (Zhou *et al*., 2005). The results from this study are published recently by Thimmaiah *et al*. (2008).

3.12.2. Sampling Period Two

The PMF2 analysis deduced 4 probable factors which were interpreted as sources in this study. 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. The results from this study are published recently by Thimmaiah *et al*. (2009).

The PMF2 deduced results for sampling periods-I and II; i.e., the source profiles, their time-series of source contributions, their attribution of source names, with the help of CPF analyses of deduced factor contributions by assigning to the wind direction measured at the receptor site and by additional analyses with respect to the diurnal pattern analyses of deduced factor contributions for determining their weekday/weekend contributions and their correlations with the actual measured gaseous species concentrations are discussed in Chapter-IV, Results and Discussions. The measured vs. predicted concentration plots showing good agreements are also shown.

CHAPTER-IV
RESULTS AND DISCUSSION

CHAPTER-IV
RESULTS AND DISCUSSION

4.1. SAMPLING PERIOD ONE

The PMF2 analysis deduced 5 probable factors, assigned to as sources in this study. The source profiles and source contributions of extracted 5 factors are shown in Figs. IV.1 and IV.2 respectively. The procedures related to PMF2 operation, species selection, and the steps involved in the preparation of data matrix and uncertainty matrix are discussed in Chapter-III, Materials and Methods.

The PMF2 deduced source contributions were assigned to receptor site wind direction measured for directionality analysis to known the likely locations of their source emissions as shown below in Fig. IV.3. The attribution of sources was possible by analyzing the CPF plots drawn for measured gaseous species (Fig. III.6) and the CPF plots of PMF2 deduced source contributions (Fig. IV.3). Additionally, the diurnal pattern weekday/weekend analysis of the 5 factor contributions deduced by PMF2 (Fig. IV.4) and the measured gaseous species (Figs. IV.5-IV.7) aided in the identification process.

The 5 PMF2 deduced factors (F1-F5) were assigned to as mixed source-1, traffic, heating, mixed source-2, and O_3-secondary particles, respectively. Out of 5 sources, 2 sources were assigned to as mixed sources-1 and 2. Repetitive PMF2 run with different seeded value failed to resolve these two mixed sources. Often, some of the factors appear to be a mixture of several sources that cannot be further separated (Zhou *et al*., 2005). These two sources may be coming from background or long range transport not known due to lack of chemical tracers.

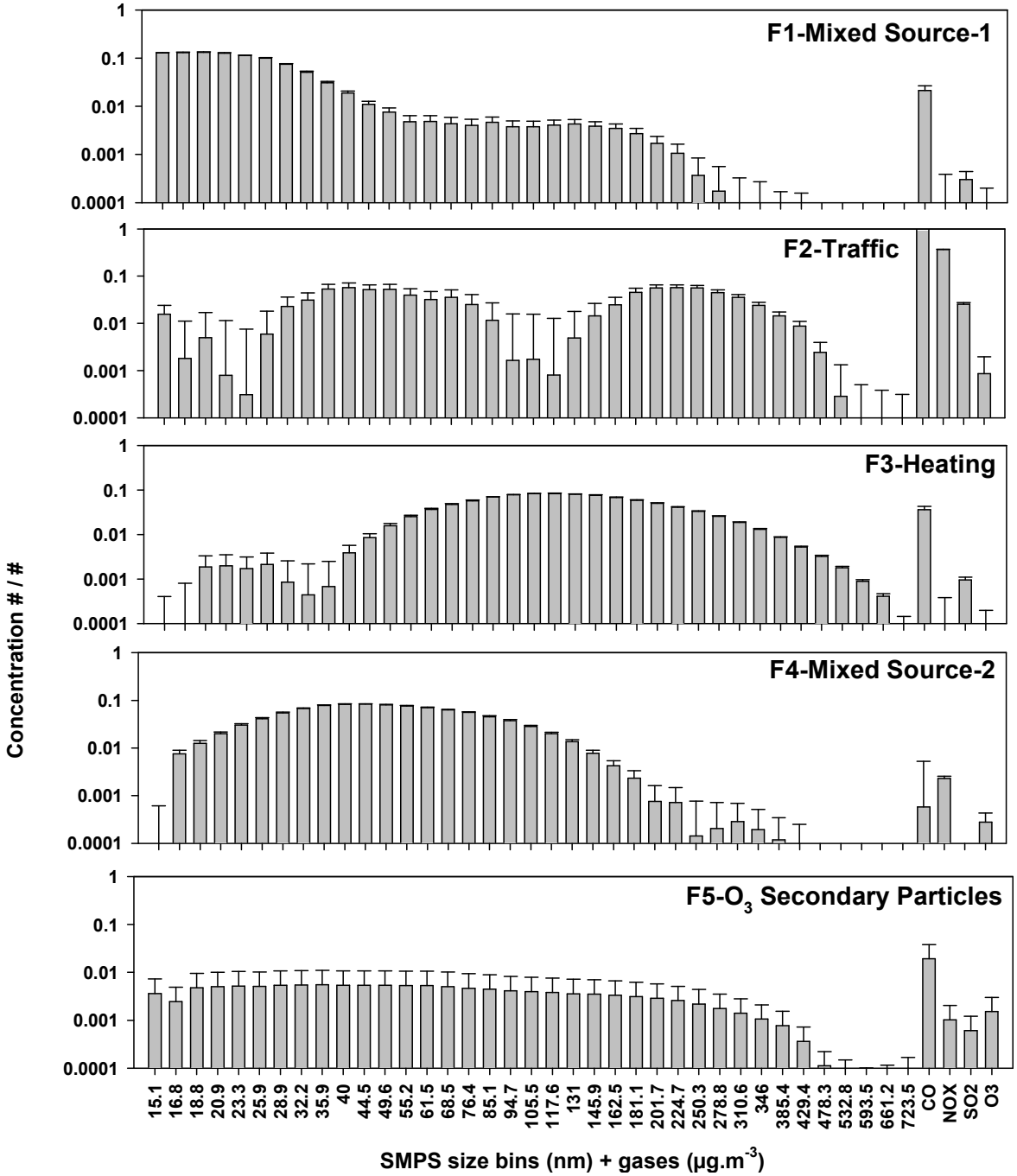


Figure IV.1. Source profiles of 5 factors deduced by PMF2 for sampling period-I.

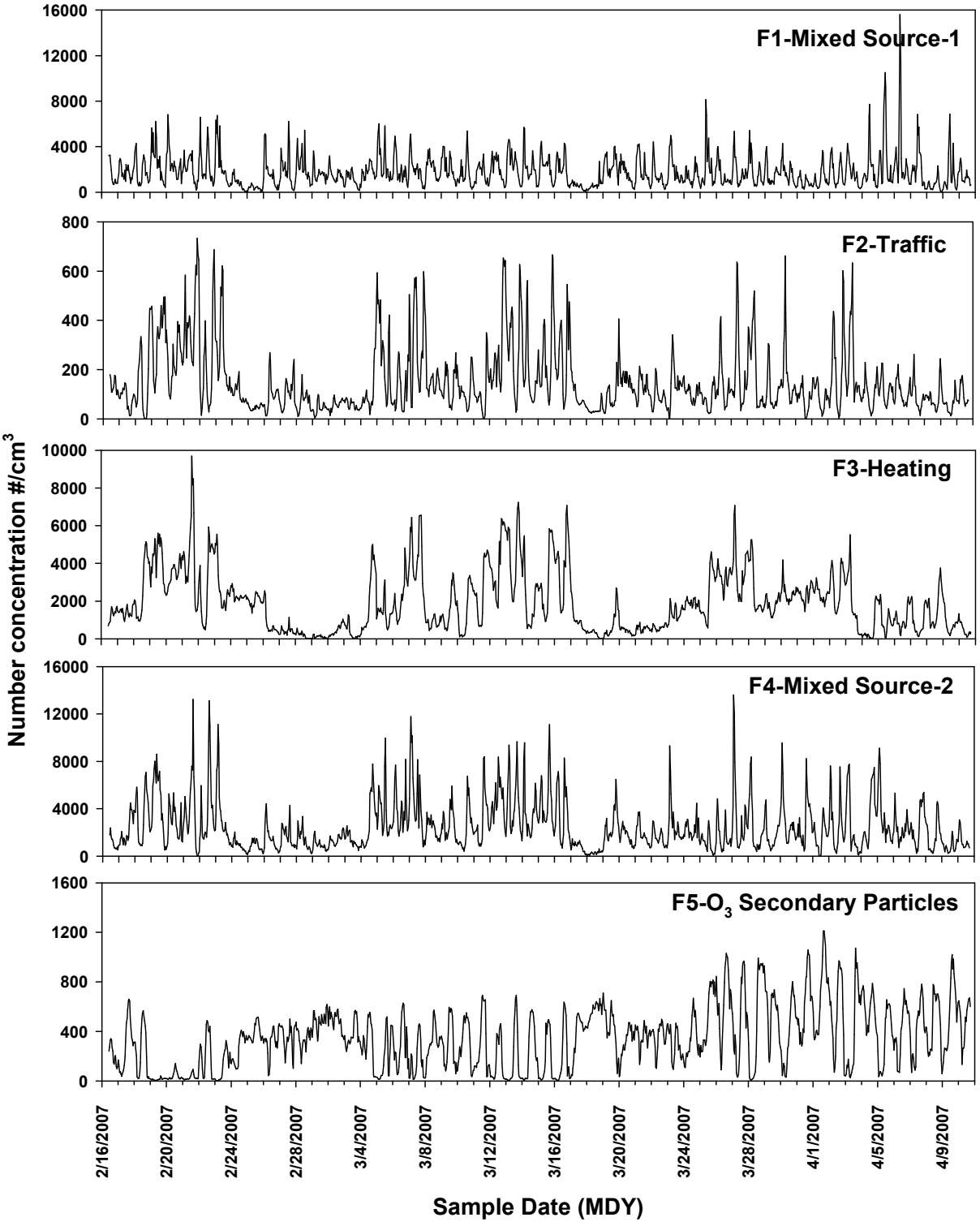


Figure IV.2. Source contributions of 5 factors deduced by PMF2 for sampling period-I.

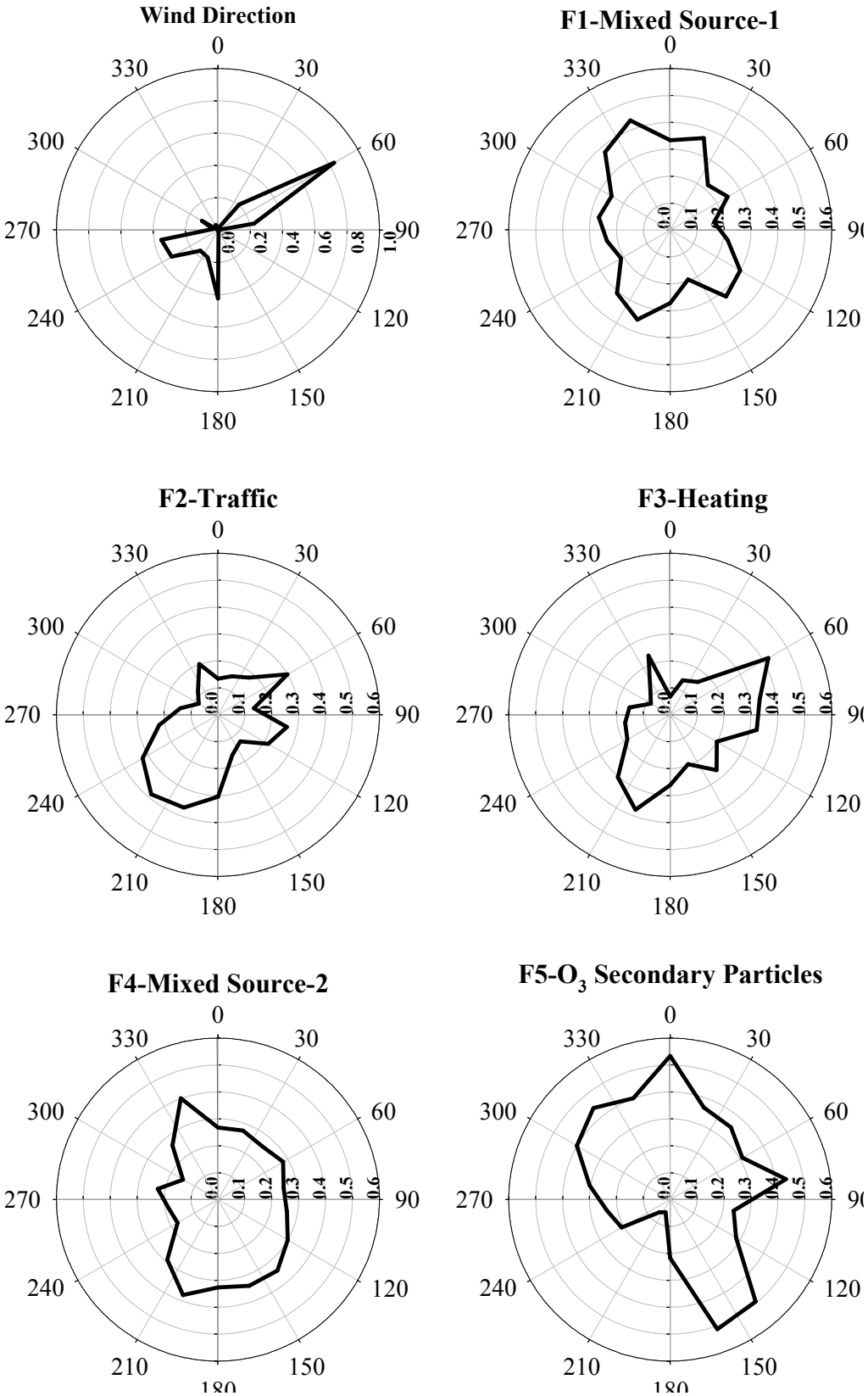


Figure IV.3. Wind direction locations of 5 factors using CPF for sampling period-I.

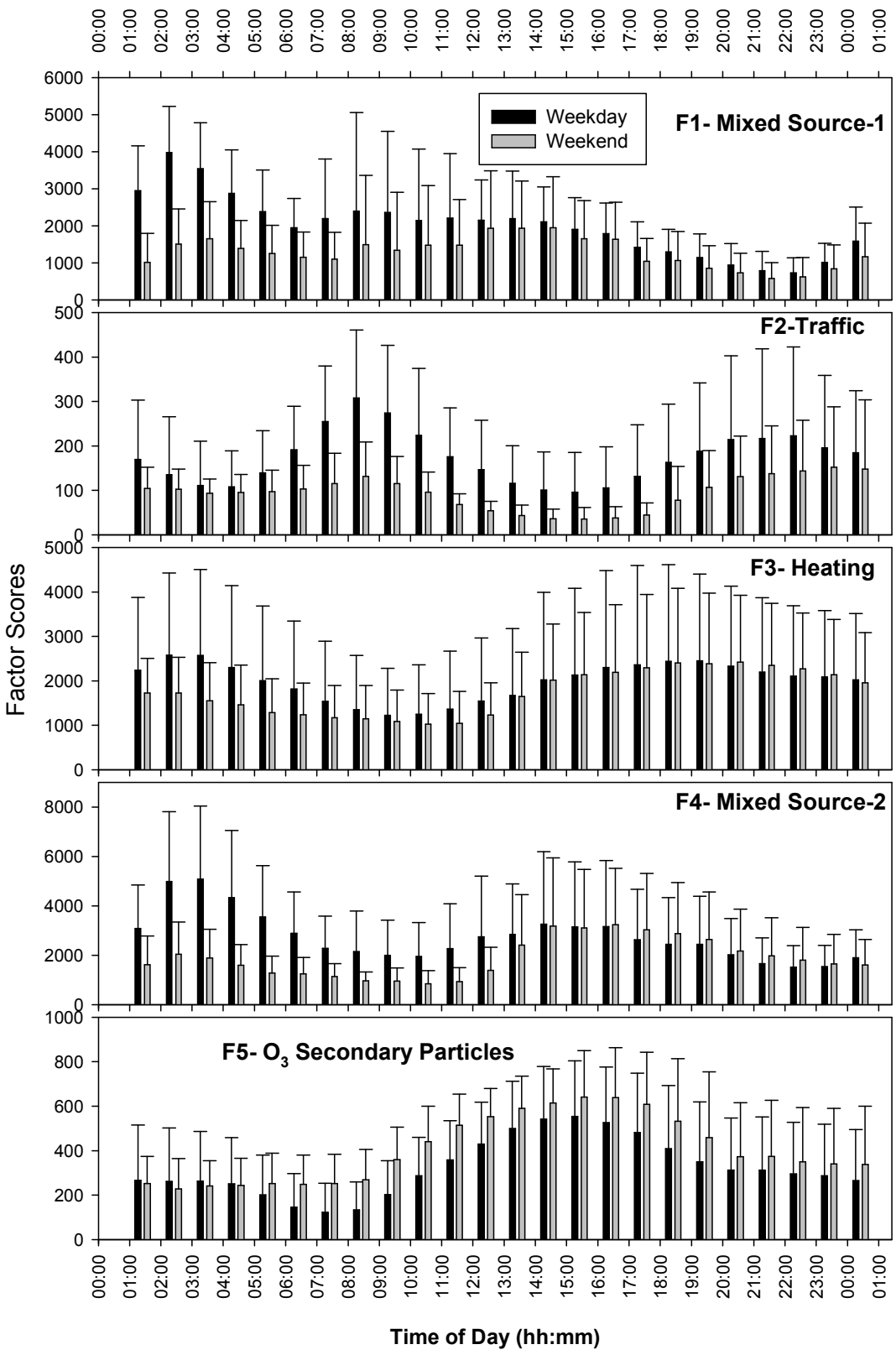


Figure IV.4. Weekday/weekend diurnal pattern of 5 factors deduced by PMF2 for sampling period-I.

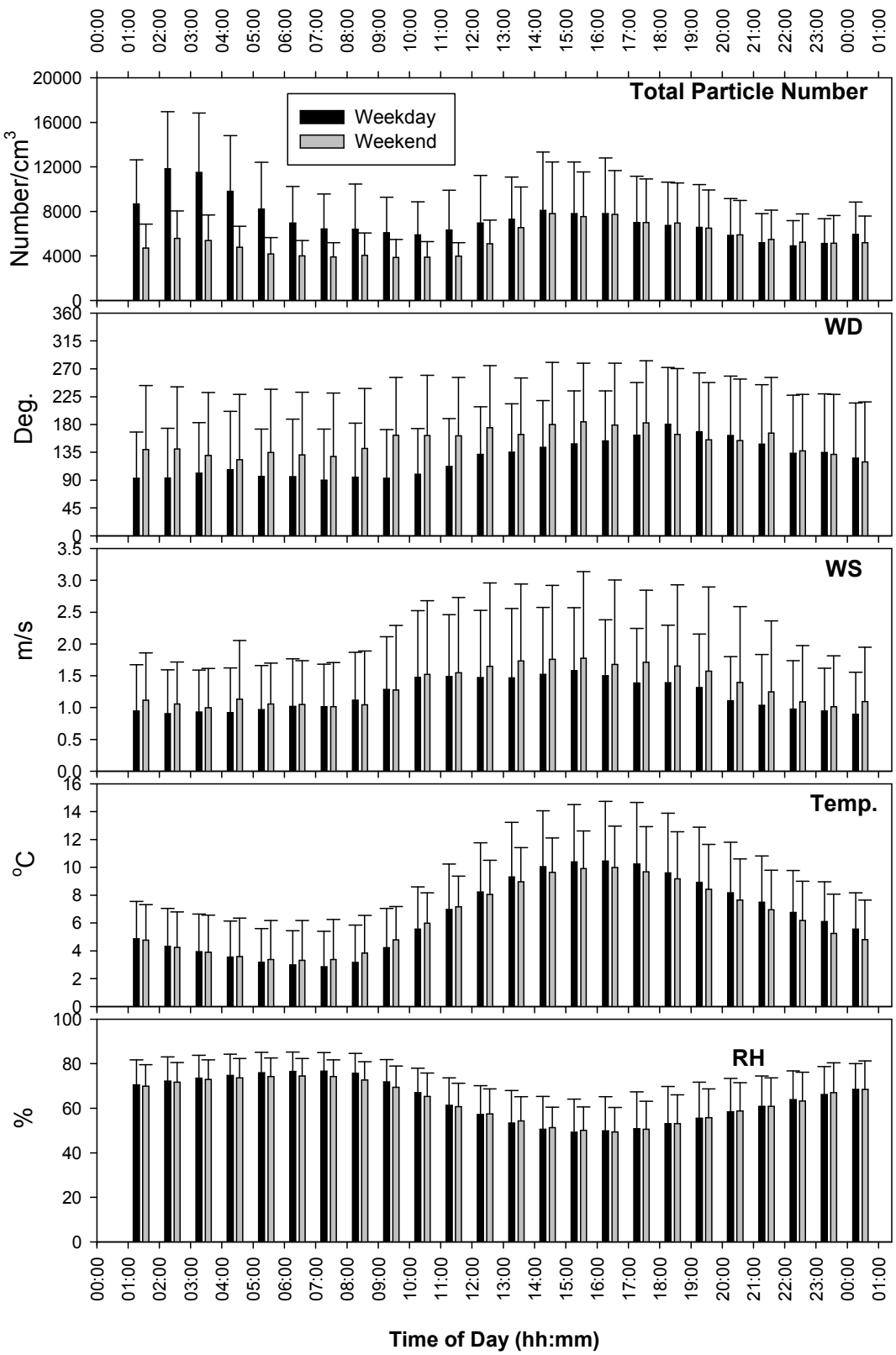


Figure IV.5. Weekday/weekend diurnal pattern of total PNC, WD, WS, Temp. and RH for sampling period-I.

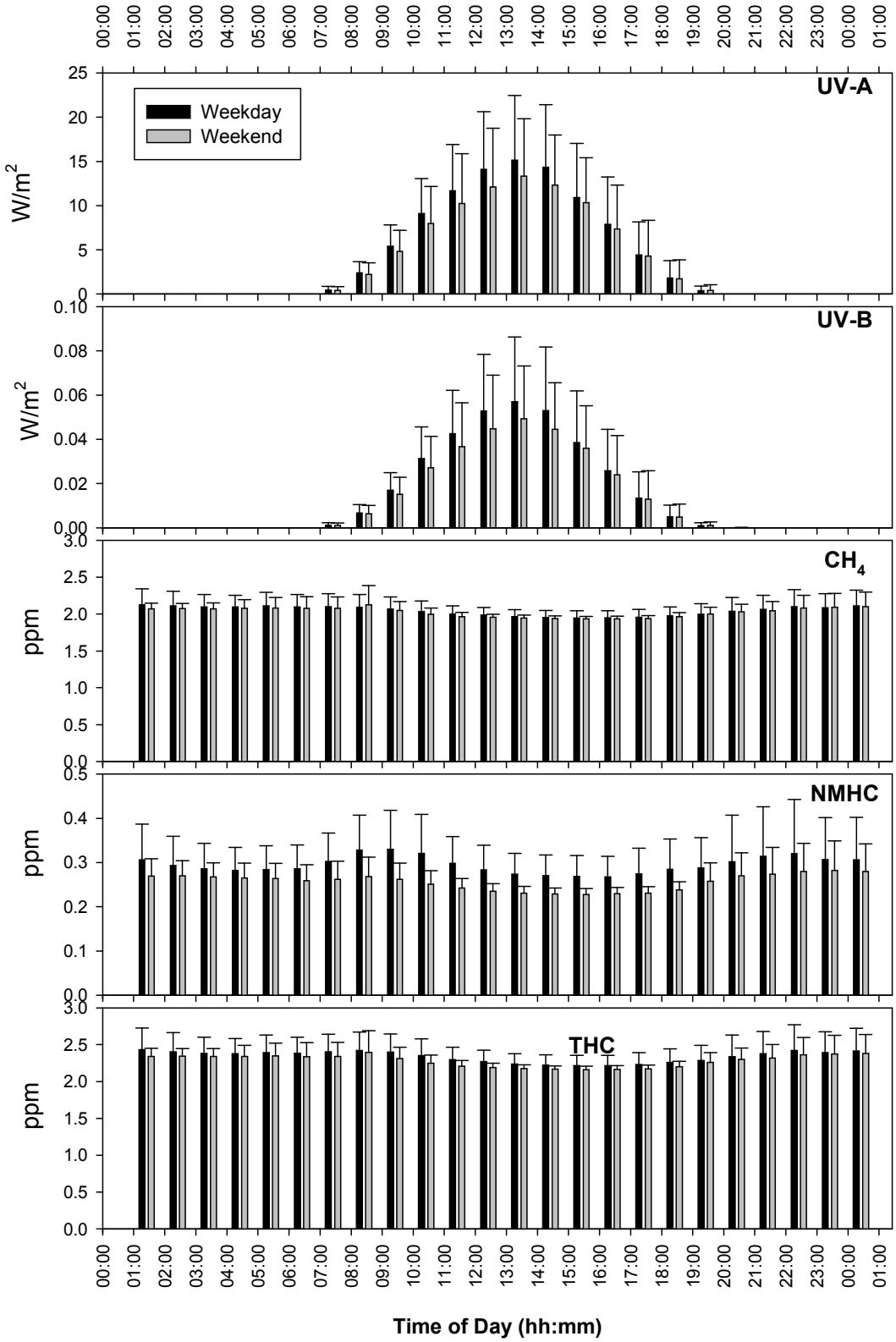


Figure IV.6. Weekday/weekend diurnal pattern of UV-A, UV-B, CH_4, NMHC and THC for sampling period-I.

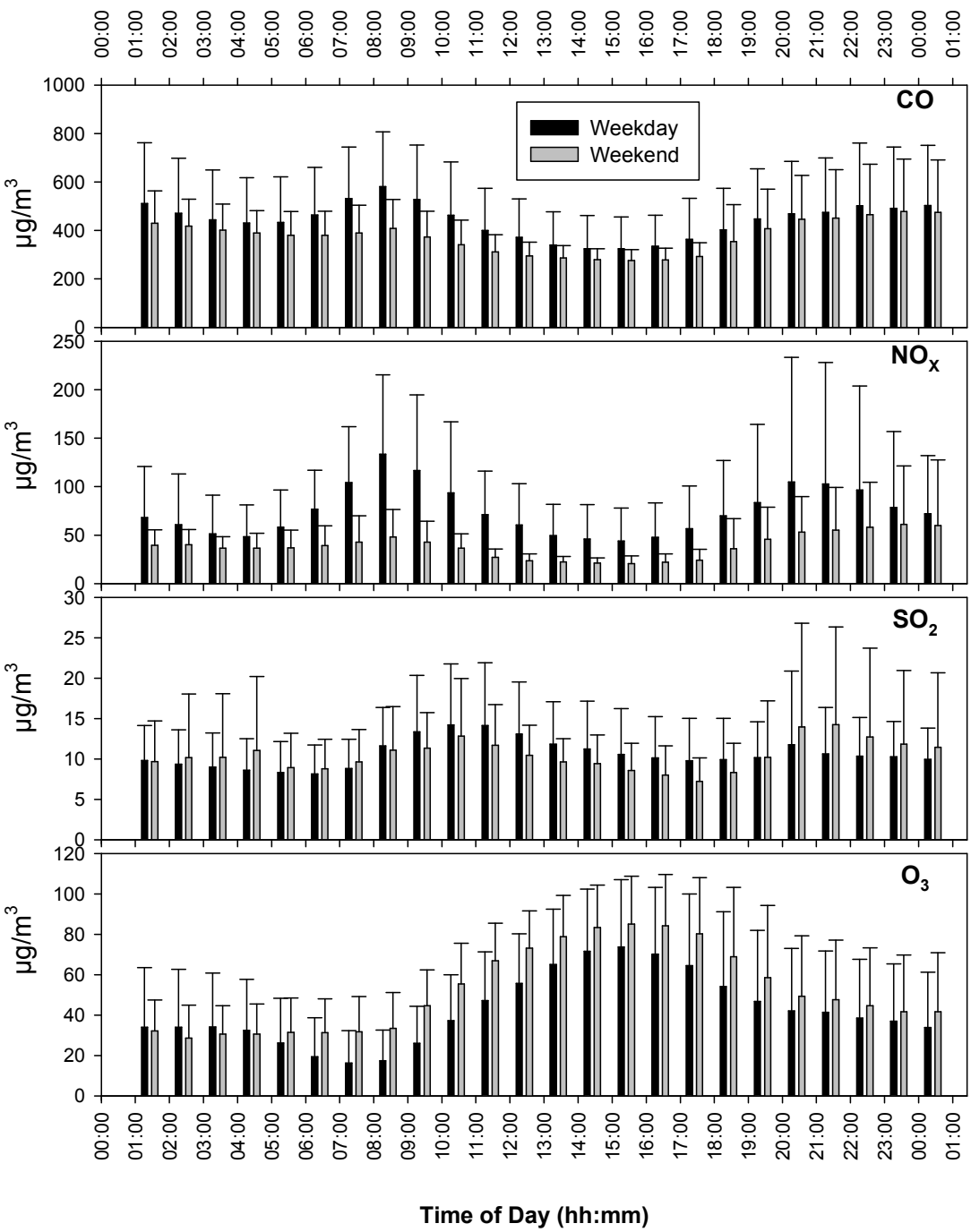


Figure IV.7. Weekday/weekend diurnal pattern of CO, NO_x, SO_2 and O_3 for sampling period-I.

4.2. ATTRIBUTION OF SOURCES (SAMPLING PERIOD-I)

4.2.1. F1-Profile (Mixed Source-1)

The Factor 1 profile (F1) shows high fractional abundance in the lower midpoint particle diameters and gradually decreasing towards the higher particle diameters (Fig IV.1). The CO and SO_2 gaseous species are present and absence of NO_x and O_3 is noted. The F1-CPF plot (Fig. IV.3) shows peaks in all the WD sectors. The diurnal weekday/weekend pattern analysis of F1 contributions show similar patterns during weekdays and on weekends but with low contributions (Fig. IV.4). The diurnal pattern observed maps the diurnal pattern of total PNC measured as seen in Fig. IV.5. This factor seems to be a mixture of two or more sources and repetitive PMF2 run with different seeded value failed to resolve them separately, and referred to as mixed source-1 in this study.

4.2.2. F2-Profile (Traffic)

The Factor 2 profile (F2) shows distinct bimodal size distribution pattern (Fig. IV.1). The gaseous species of CO, NO_x, and SO_2 are present. The F2-CPF plot (Fig. IV.3) shows similar peaks as seen in the case of NO_x-CPF (Fig. III.6). The diurnal weekday/weekend pattern of F2 contributions (Fig. IV.4) shows the existence of morning (06:00- 10:00 Hrs.) and evening rush hour (17:00-22:00 Hrs.) peaks with agreeable low contributions during the weekends. The diurnal pattern observed for F2 contributions (Fig. IV.4) maps the diurnal pattern of NO_x measured as seen in Fig. IV.7.

The hourly averaged traffic density studies by Hovorka *et al*. (2005a) at the botanical garden intersection showed highest hourly morning and evening traffic volume peaks between 8:00-9:00 Hrs and 17:00-18:00 Hrs respectively. In general, temporal pattern peaks in car traffic study for Prague, 2003 is between 7:00-9:00 Hrs and 15:00-18:30 Hrs (source www.udi.cz).

4.2.3. F3-Profile (Heating)

The Factor 3 profile (F3) shows high fractional abundance in the size range around 90 nm midpoint diameter (Fig. IV.1). The gaseous species CO and SO_2 are also seen. The weekday/weekend F3 diurnal patterns (Fig IV.4) show high contributions during early morning hours and at late evening hours when ambient temperatures are the lowest as can be seen in weekday/weekend diurnal plot for Temp. in Fig. IV.5. The heating boiler belonging to the Charles university faculty building and hospital boiler are in close proximity to the receptor site and significant contributions may arise from residential/office boilers in the vicinity. Though, the weekday/weekend diurnal patterns show identical behavior, the late evening weekend contributions were higher denoting office space heating turned on to keep it ready for people starting to work, beginning Mondays.

The findings from this study, match with the study by Hovorka et al. (2005a) for natural gas burning by botanical garden boiler and hospital heating boiler showing diurnal pattern of natural gas burn up at peak between 05:30-10:00 Hrs (botanical garden boiler) and 04:00-07:00 Hrs (hospital heating boiler), and for residential boilers between 04:00-06:00 Hrs and 18:00-22:00 Hrs. Further, the source profiles of heating (Fig. IV.1) shows the dominant contribution to the particle number concentrations in the size range around 90 nm midpoint diameter and match with the experimental study carried out by Hovorka *et al.* (2005b) for particles expelled by botanical garden boiler chimney showing monomodal size distribution with maximum mobility diameter around 90 nm as shown in Fig. IV.8 (Source: Hovorka *et al*., 2005b).

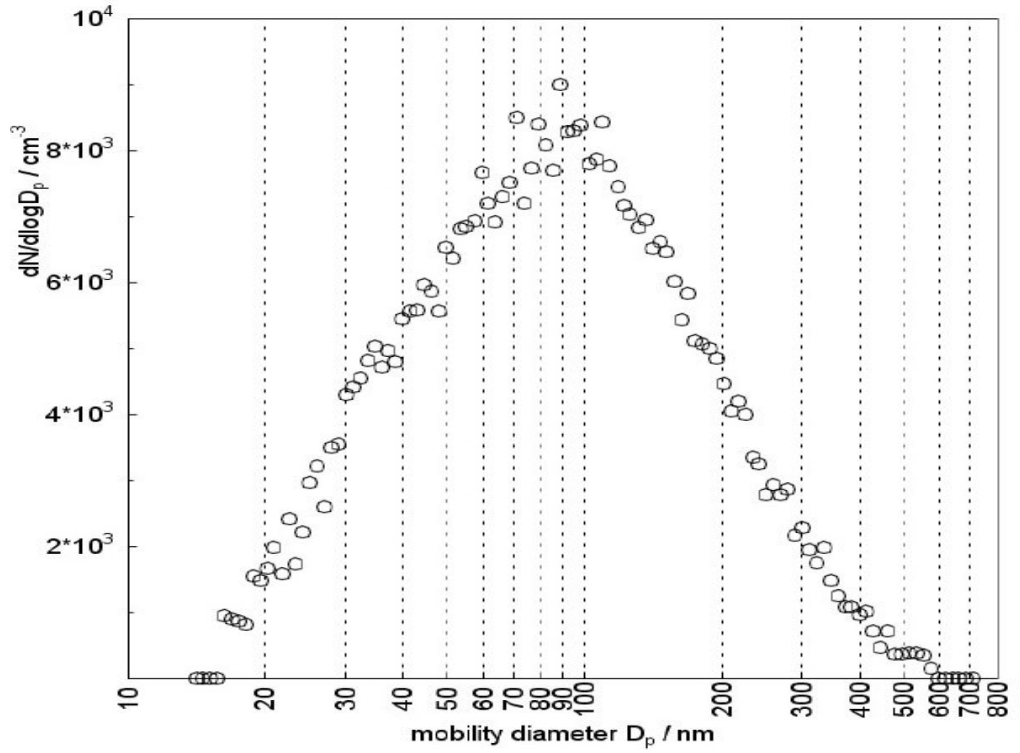


Figure IV.8. Monomodal particle size distributions from stack of heating boiler chimney, operated on natural gas showing GMD of 84.5 nm.

4.2.4. F4-Profile (Mixed Source-2)

The Factor 4 profile (F4) shows monomodal size distribution pattern (Fig. IV.1) with maximum mobility diameter around 50 nm size. The gaseous species, CO, NO_x and O_3 is seen. The F4 time-series contributions (IV.2) show similar pattern as observed in the case of F2 time-series contributions assigned to as traffic source. However, the weekday/weekend diurnal F4 source contributions (Fig. IV.4) doesn't show bimodal distribution pattern as seen in the case of traffic emissions with morning and evening rush hour peaks, and show morning peaks around 3:30 Hrs and afternoon peak around 14:30 Hrs. The diurnal pattern observed maps the diurnal pattern of total PNC measured as seen in Fig. IV.5, showing existence of early morning weekday peak and increased evening contributions after 16:30 Hrs during the weekends. Like F1-contribtuions, the F4-contributions seems to be a mixture of two or more sources and repetitive PMF2 run with

different seeded value failed to resolve them separately, and referred to as mixed source-2 in this study.

4.2.5. F5-Profile (O_3-Secondary Particles)

The Factor 5 source contributions (Fig. IV.2) show identical pattern as observed for measured O_3 time-series (Fig. III.2). Further, the F5-CPF plot (Fig. IV.3) and measured O_3-CPF plot (Fig. III.6) looks similar. The Weekday/weekend diurnal pattern of F5 source contributions (Fig. IV.4) and O_3 weekday/weekend diurnal pattern (Fig. IV.7) show peaks around midday, clearly showing the influence of UV-A, B radiations in their formation irrespective of weekdays or weekends in the presence of ozone. The diurnal patterns of UV-A, B (Fig. IV.6) show maximum solar radiance at peak during midday. In the case of O_3 (Fig. IV.7) and F5 (Fig. IV.4), the weekend contributions were higher than during the weekdays this phenomenon is referred in the literature as Ozone Weekend Effect (OWE).

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 VOCs are lower than on usual weekdays. While earlier, OWE was known as Sunday Effect-SE (Altshuler *et al*., 1995; Chinkin *et al*., 2003**)**. During weekends, the sum of NO_x ($NO+NO_2$) shifts towards lower NO and higher NO_2 concentrations. The NO_x emissions are lower because of lower heavy-duty diesel truck activity and slightly lower for VOCs on weekends. In rural settings, emissions from cars are highest on Fridays and Sundays (Altshuler *et al*., 1995**;** Suppan and Graf, 2000). Usually, higher ozone concentrations are observed on weekends in urban downwind areas (Marr and Harley, 2002). Detailed studies of this phenomenon indicate the primary cause of high O_3 on weekends is due to reduction in local NO_x emissions on weekends in VOC-limited chemical regime, in contrast, the lower O_3 on weekends in other locations maybe due to NO_x reductions in a NO_x-limited regime (Heuss *et al*., 2003).

The average weekday/weekend contributions of 5 deduced sources along with the total average PNC and the weekday/weekend averages of meteorological data and gaseous species measured at the receptor site are shown in Fig. IV.9.

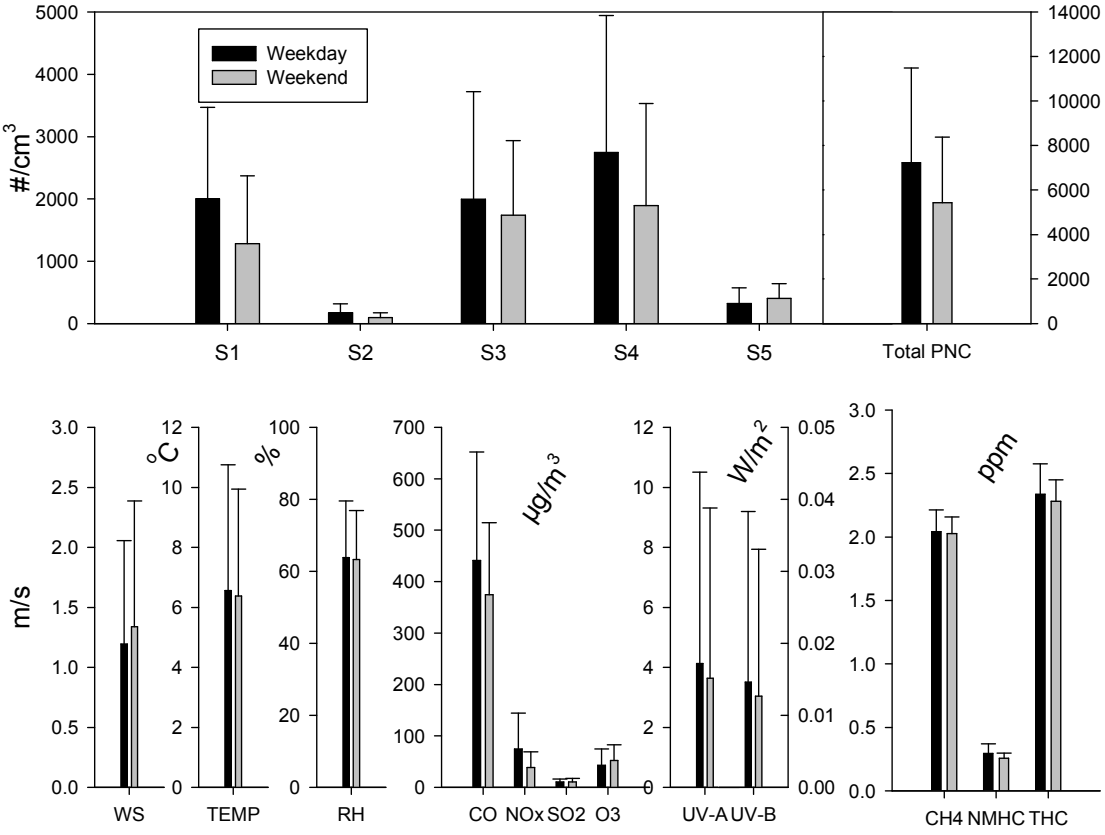


Figure IV.9. Weekday/weekend averages of 5 sources, total PNC, meteorological data and gaseous species.

4.2.6. Regression Analyses

Regression analyses between the 5 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 [6686 (100%)] for F1-mixed source-1, F2-traffic, F3-heating, F4-mixed source-2 and F5, O_3-secondary particles were 1787 (26.5%), 150 (2.5%), 1917 (28.7%), 2489 (37.0%) and 348 (5.2%), respectively. The linear regression correlation between the predicted vs. measured total PNC was excellent in agreement with R^2= 0.99 (see Fig. IV.10).

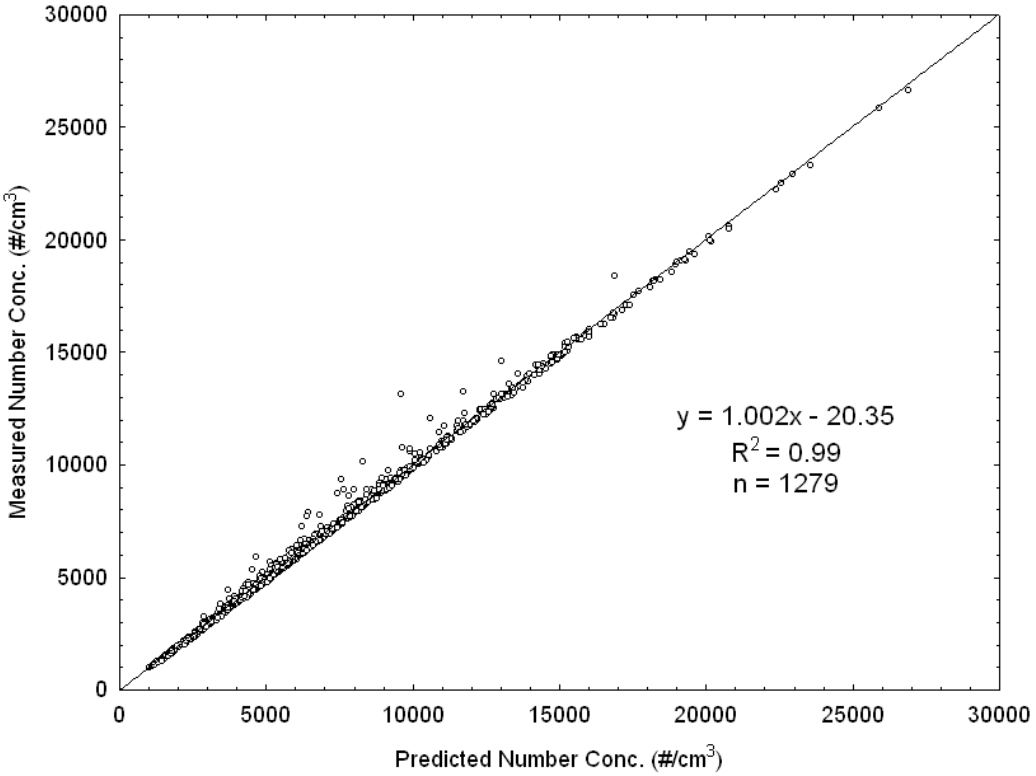


Figure IV.10. Regression between predicted and measured particle number concentration for sampling period-I.

4.3. SAMPLING PERIOD TWO

4.3.1. 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. IV.11 and IV.12, respectively. The following observations were made in the four factors source profiles and their contributions:

4.3.1.1. 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.

4.3.1.2. 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.

4.3.1.3. 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-

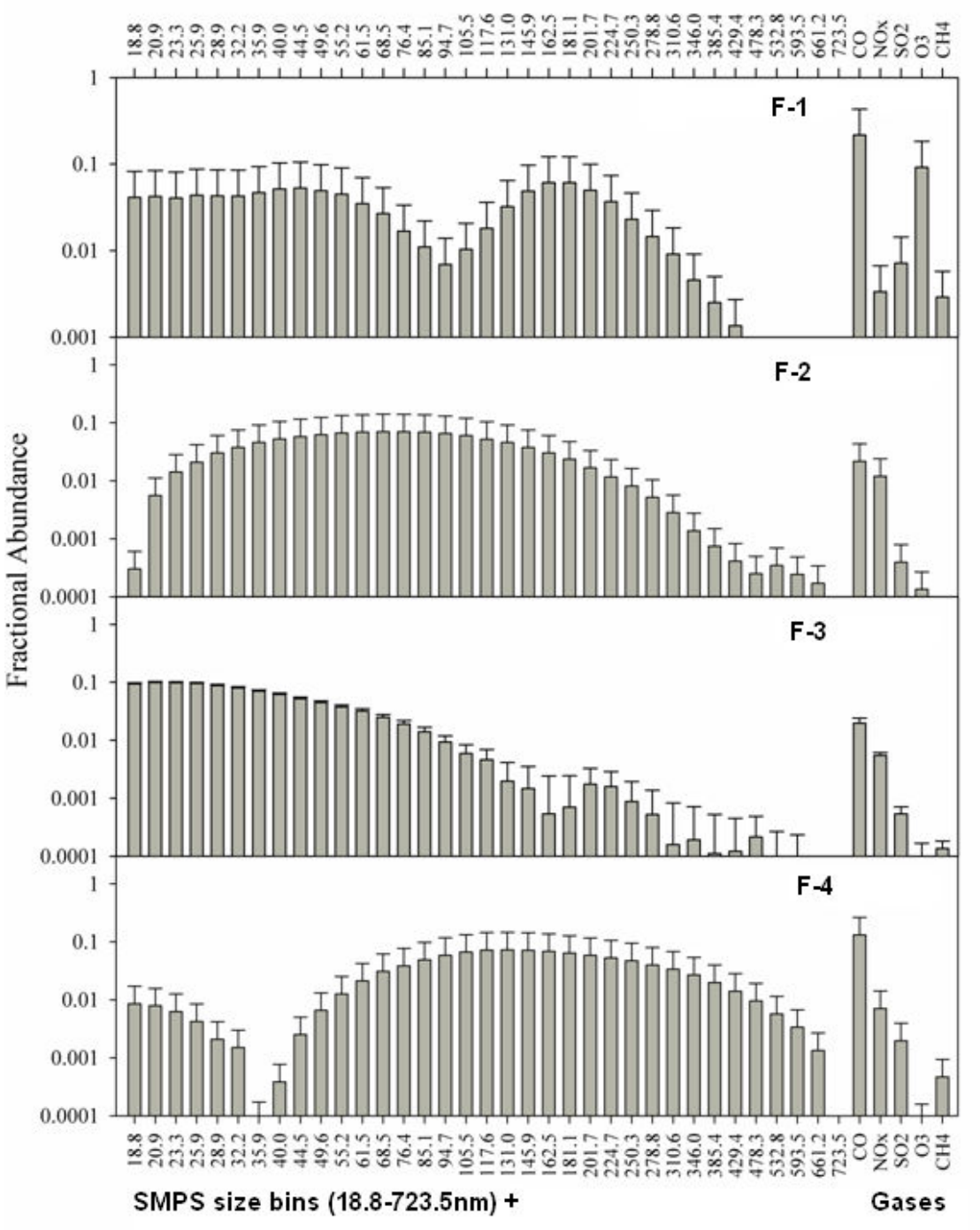


Figure IV.11. Source profiles of 4-factors deduced by PMF2 for sampling period-II.

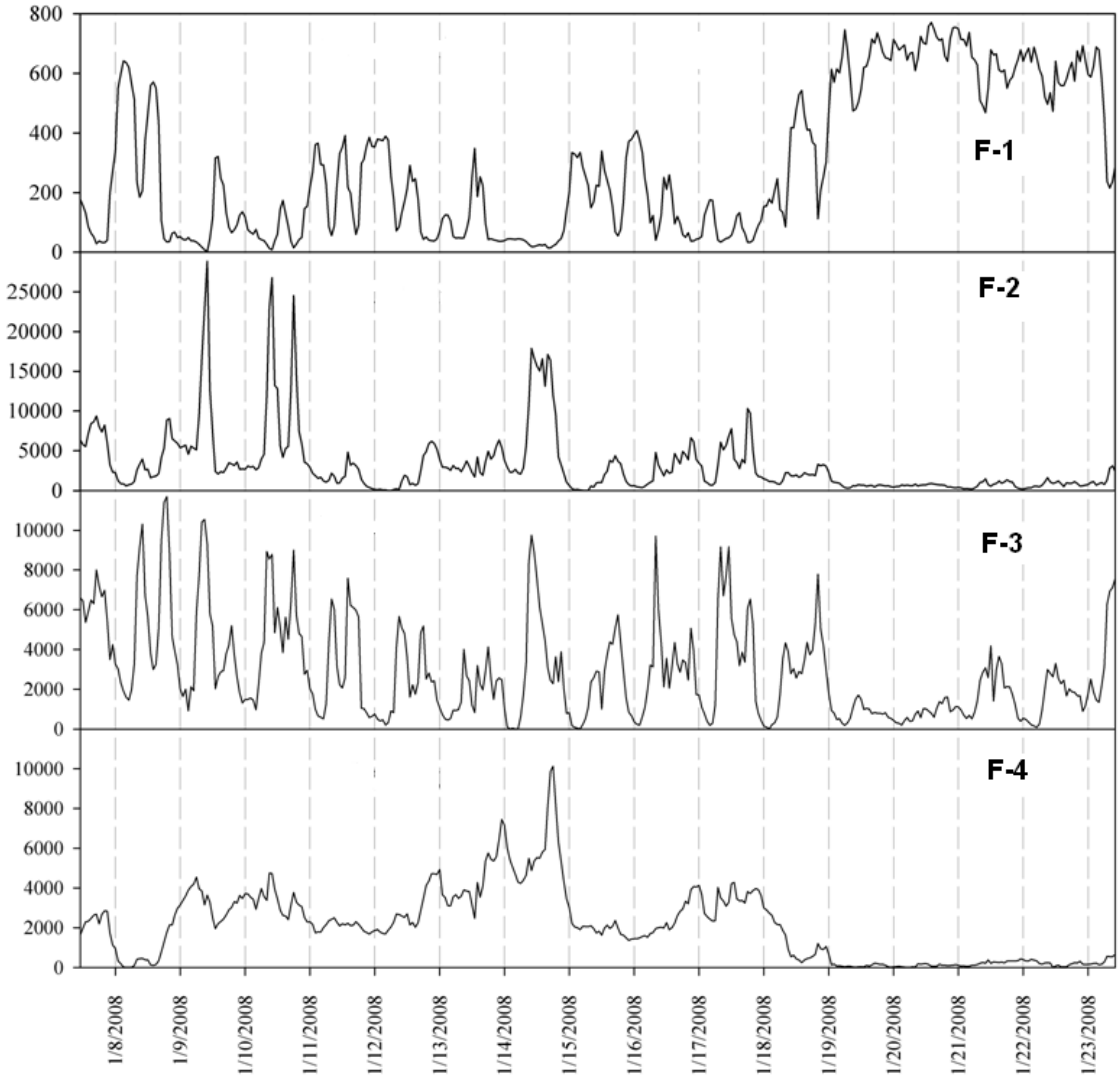


Figure IV.12. Source contributions of 4-factors deduced by PMF2 for sampling period-II.

-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.

4.3.1.4. 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.

4.4. 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. IV.13. 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. IV.13. From Fig. IV.13 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. IV.12) versus the species time-series (Fig. IV.13) it clearly shows that F1 is following the pattern of O_3 levels; F2, following NO_x levels; F3, seems to 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. IV.12), 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.

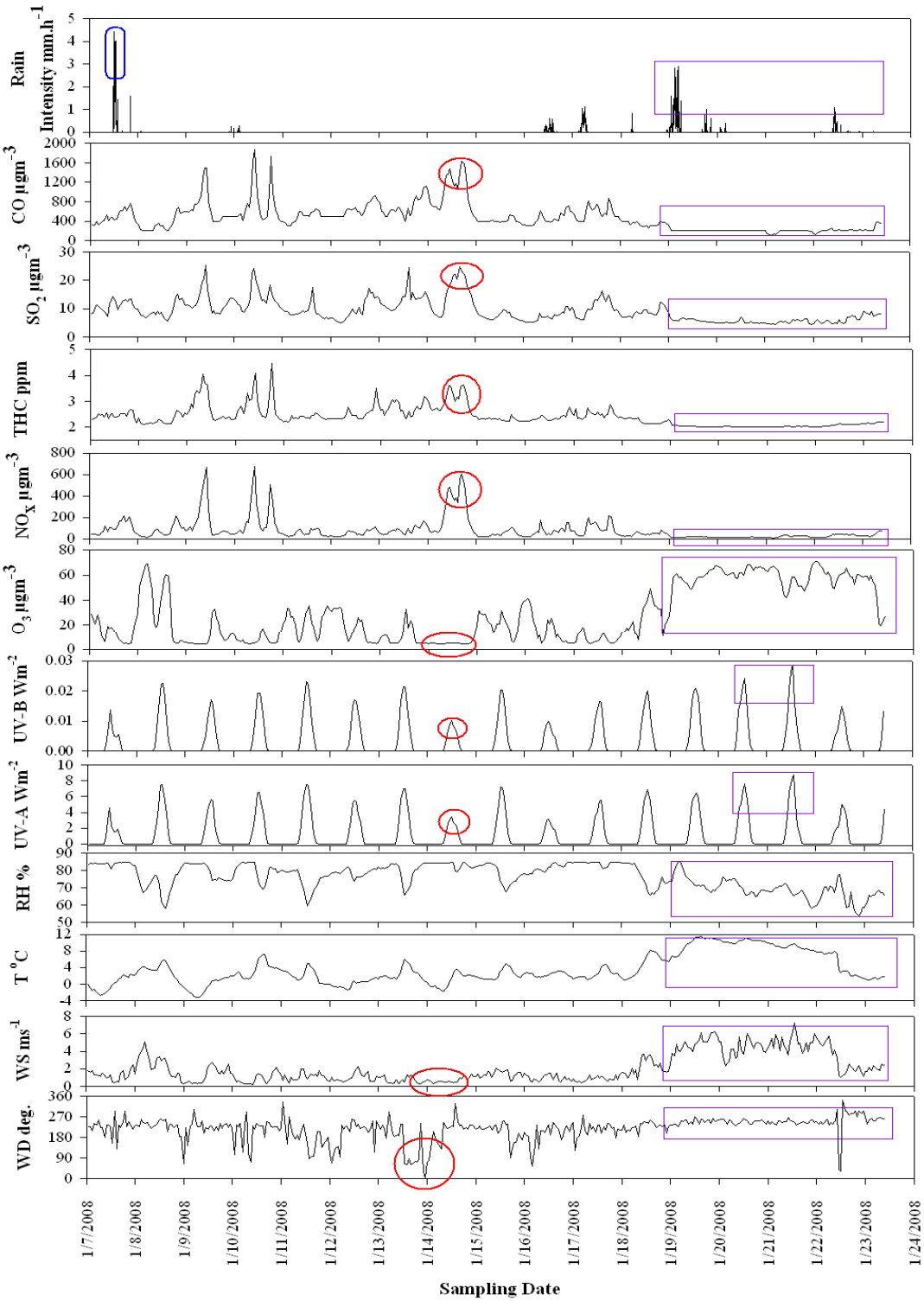


Figure IV.13. Time-series of species-gaseous concentration, meteorology and precipitation intensity for sampling period-II.

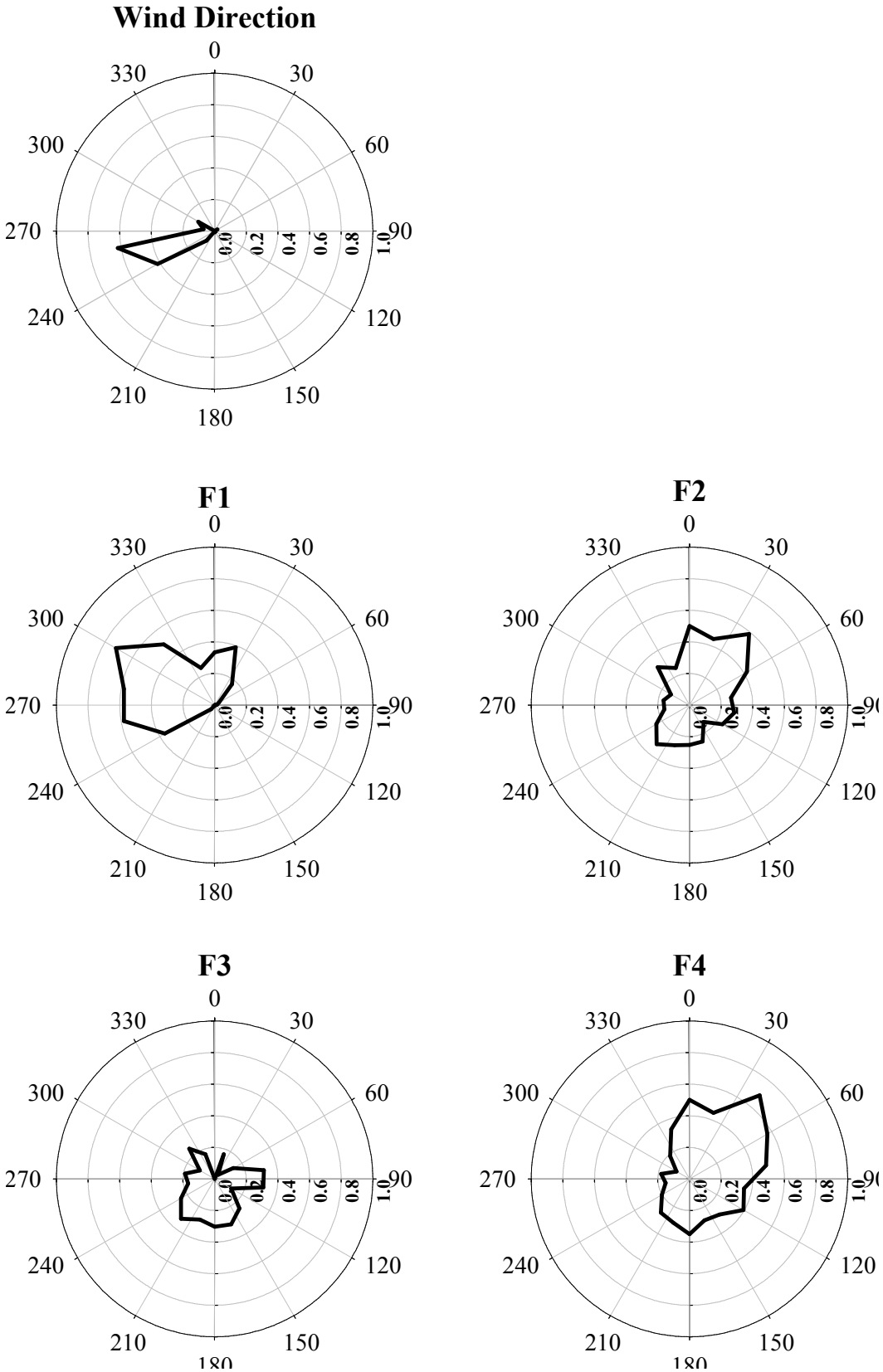


Figure IV.14. Wind direction locations of four factors using CPF for sampling period-II.

4.5. CPF Analyses

The wind direction locations of individual species and four factors using CPF are shown in Fig. III.7 and Fig. IV.14, respectively. From Fig. IV.14, 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. IV.14), the F1 CPF is identical to the O_3 CPF (Fig. III.7) 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 (see Fig. III.7). 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.

Because of the multiplicity of sources/processes that may contribute to the emissions of NO_x, 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, NO_x, CO, SO_2, CH_4 and NMHC (see supplemental Figs. S1-S4) and the correlation between ambient CH_4 and ambient temperature were done (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. IV.15 shows the 24-hour diurnal plots for factors F1 to F4 on weekdays and weekend.

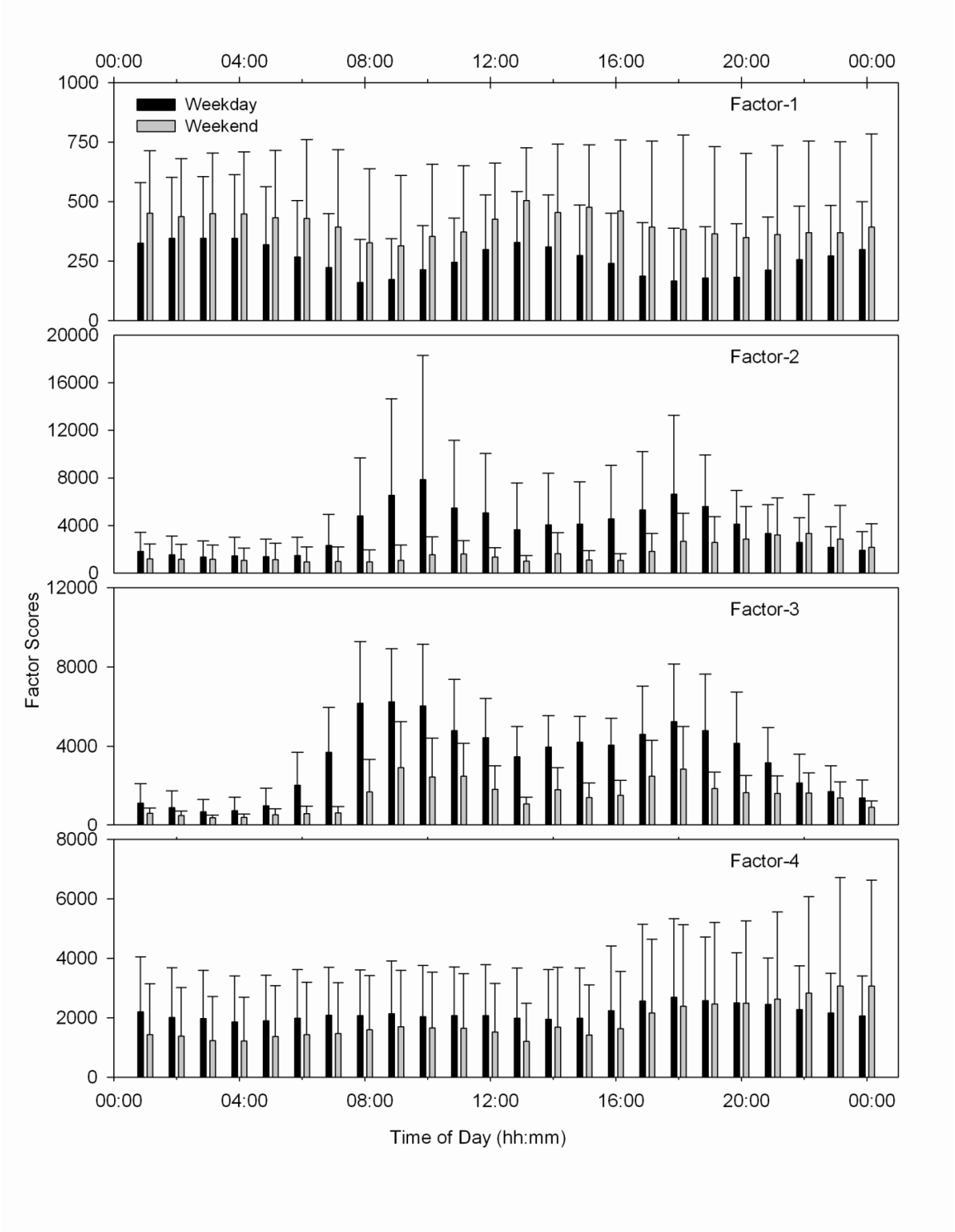


Figure 1V.15. Diurnal pattern of four factors on weekdays and weekend for sampling period-II.

4.6. ATTRIBUTION OF SOURCES (SAMPLING PERIOD-II)

4.6.1. F1-Ozone Rich, Transported Ozone/Ozone Precursors

The F1-CPF (Fig. IV.14) and species O_3-CPF (Fig. III.7) were seen to be similar, and the scatter plots between the F1 contributions and the ambient O_3 concentrations showed good correlation (see Fig. S1a). 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 and weekend with peak values at noon (see Fig. IV.15). 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 19^{th} -23^{rd} Jan 2008 and also on 8^{th} Jan 2008 (see Fig. IV.13). 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^o-275^o (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*. (2007b) in Rochester, NY.

4.6.1.1. High Ozone Concentration Periods

The species time series (see Fig. IV.13) 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 concentration 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. IV.16). 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).

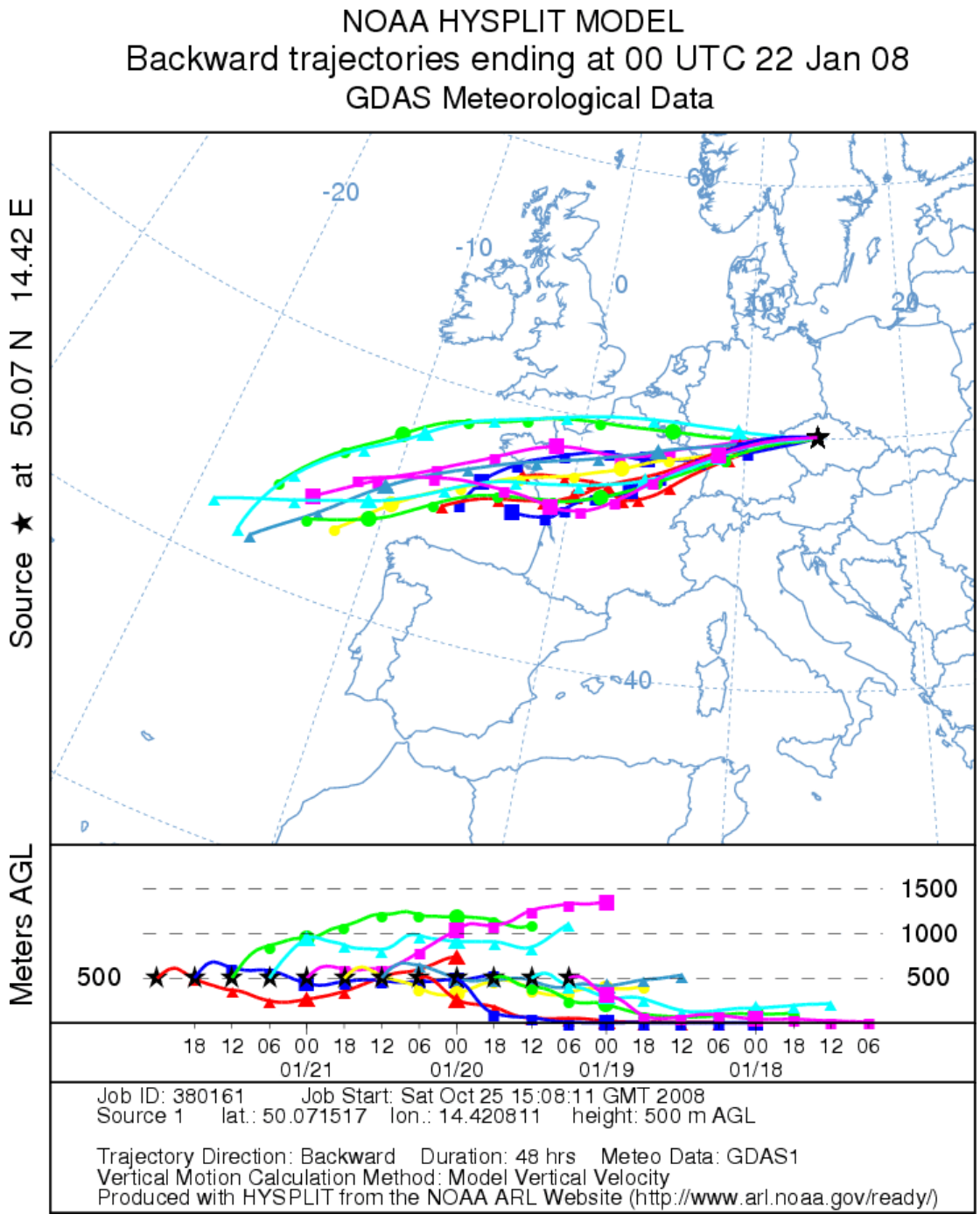


Figure IV.16. Back trajectories ending at 00 UTC 22 January 2008 for sampling period-II.

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. IV.16) shows that the trajectories at higher levels had south-western direction. 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.

4.6.2. F2-NO_x 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 (see Fig. III.7). 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. IV.15. 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.

4.6.3. 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. IV.15) 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. IV.13). 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. IV.11).

4.6.4. 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^oC and decrease to near zero for temperatures above 8^oC. 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 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. IV.15). 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.

The average weekday/weekend contributions of 4 deduced sources along with the total average PNC and the weekday/weekend averages of meteorological data and gaseous species measured at the receptor site are shown in Fig. IV.17.

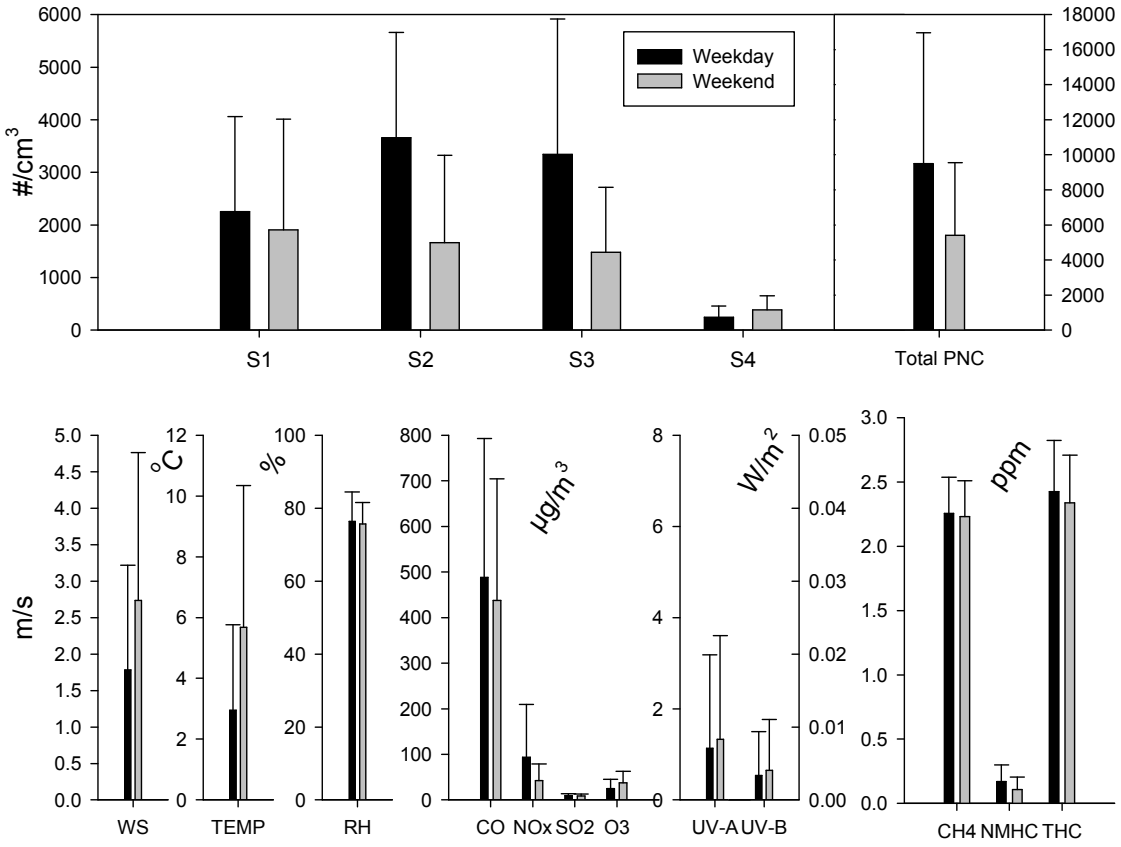


Figure IV.17. Weekday/weekend averages of 4 sources, total PNC, meteorological data and gaseous species.

4.6.5. 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 with R^2= 0.99 (see Fig. IV.18).

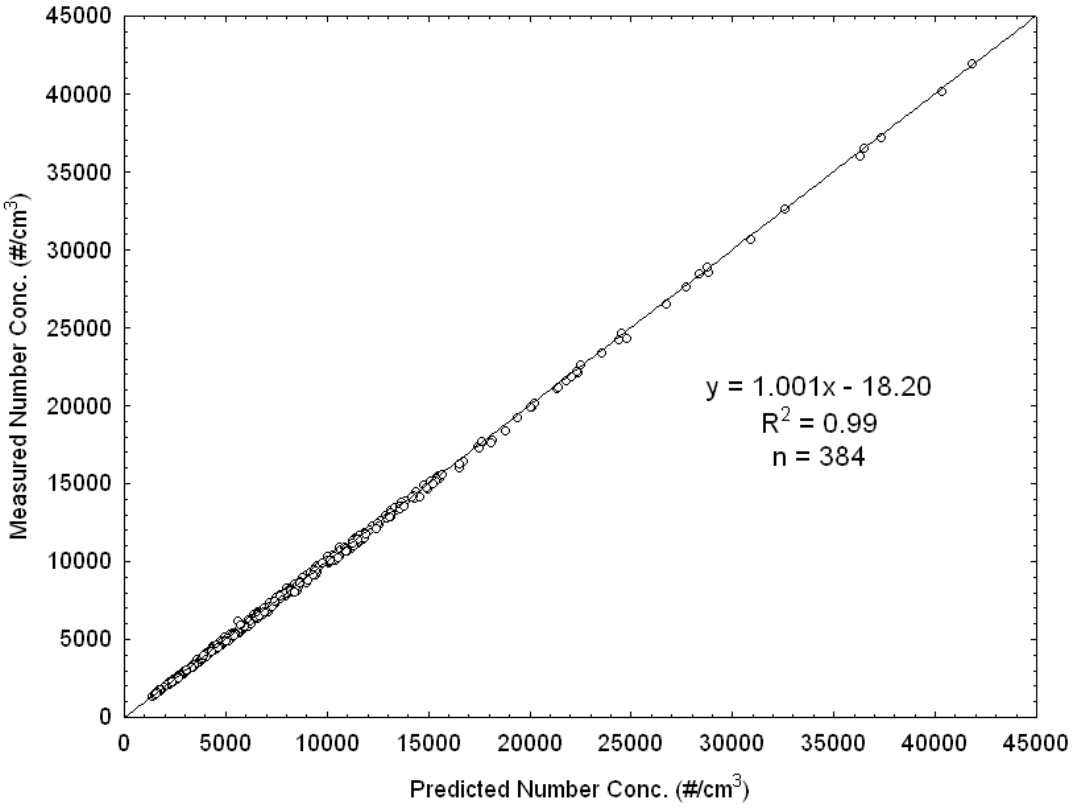


Figure IV.18. Regression between predicted and measured particle number concentration for sampling period-II.

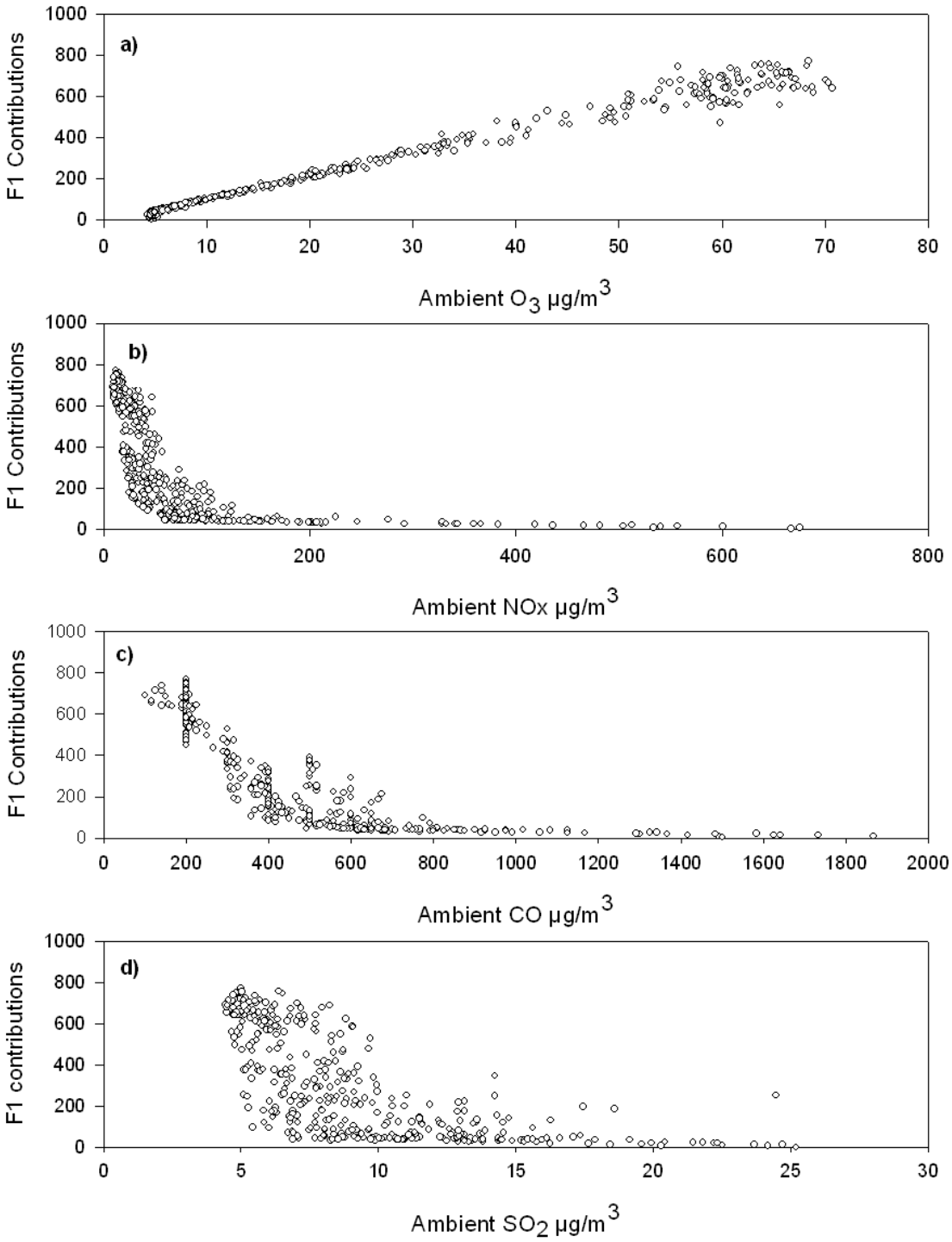


Figure S1 (a, b, c and d). F1 source contributions vs. O_3, NO_x, CO and SO_2.

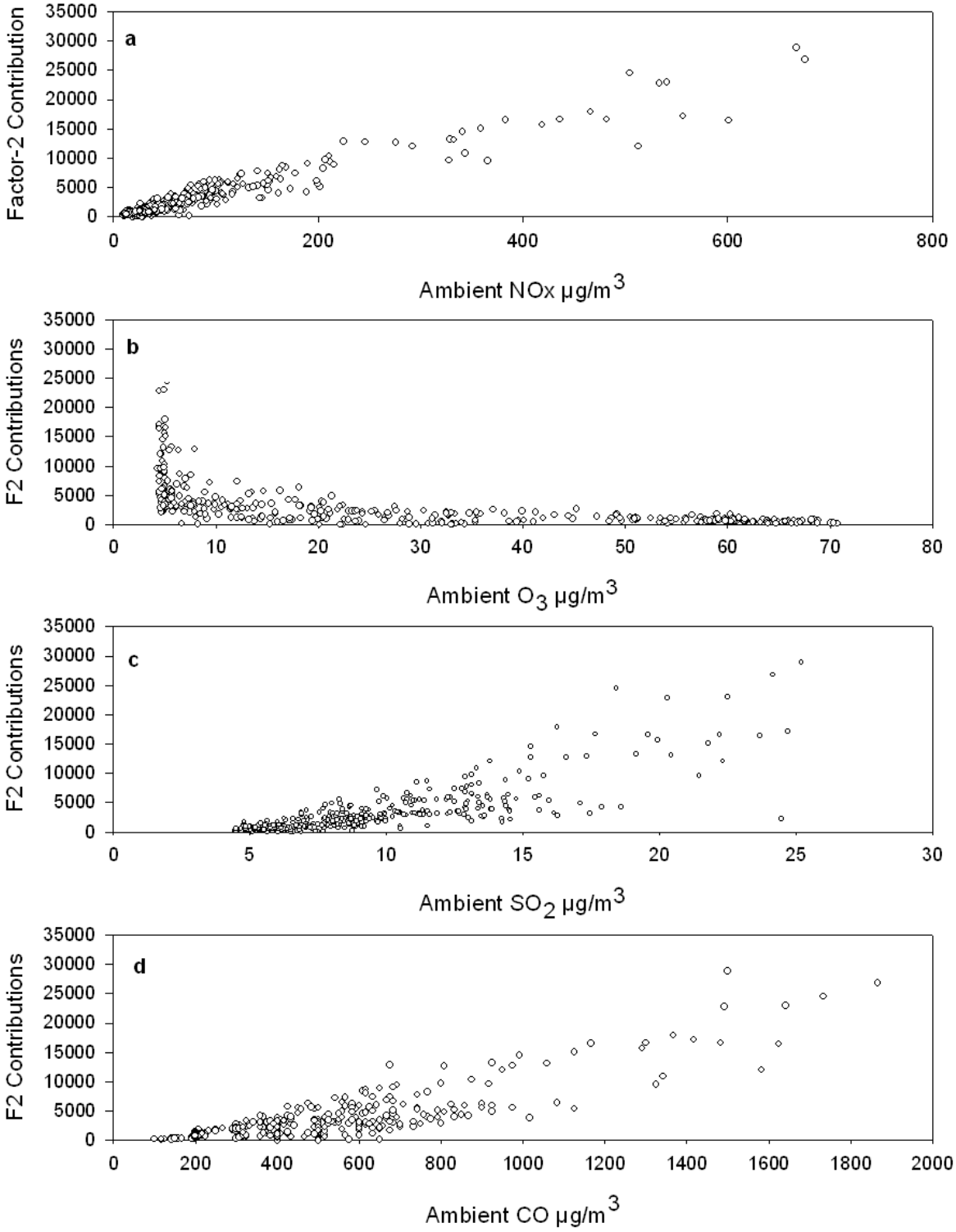


Figure S2 (a, b, c and d). F2 source contributions vs. NO_x, O_3, SO_2 and CO.

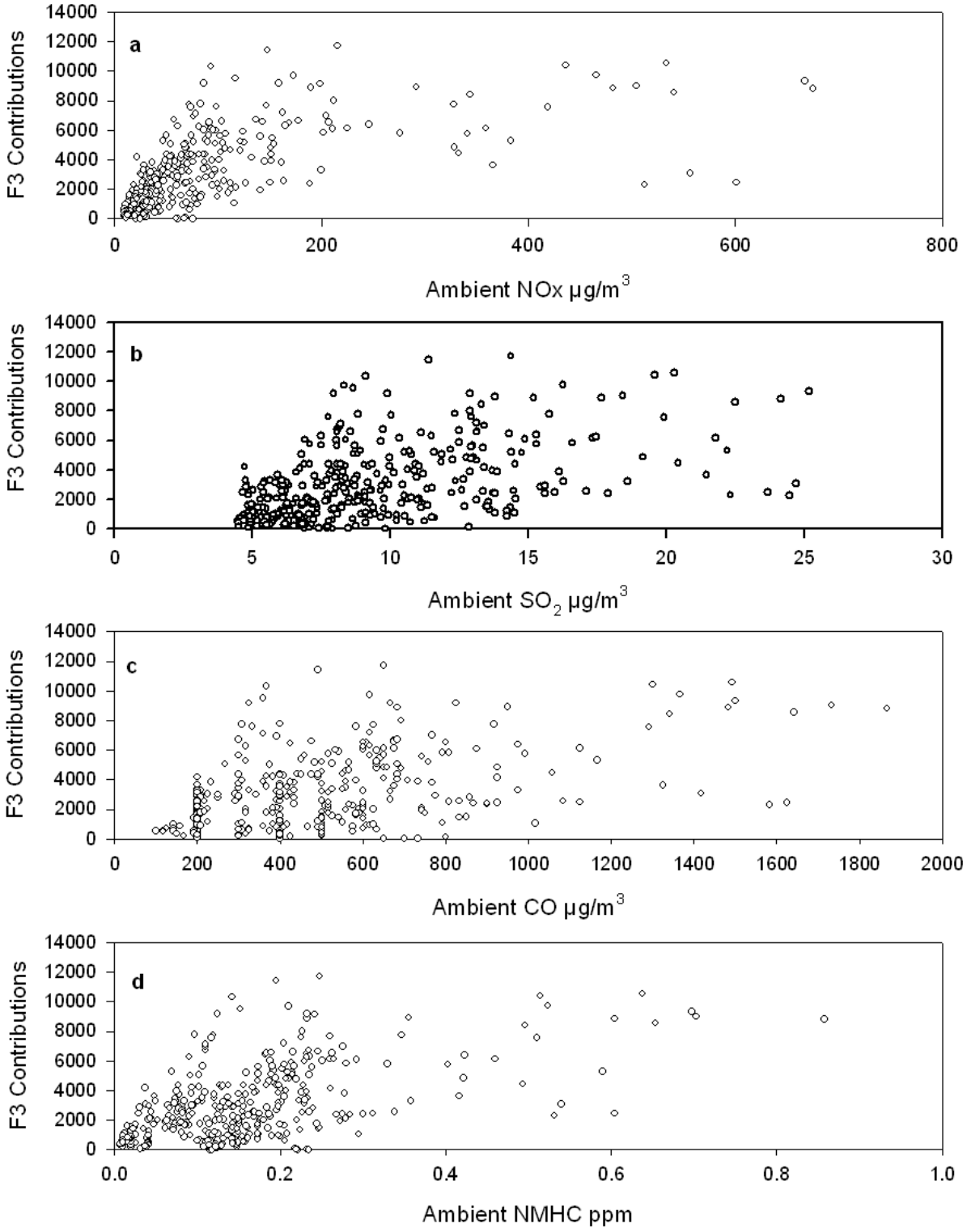


Figure S3 (a, b, c and d). F3 source contributions vs. NO_x, SO_2, CO and NMHC.

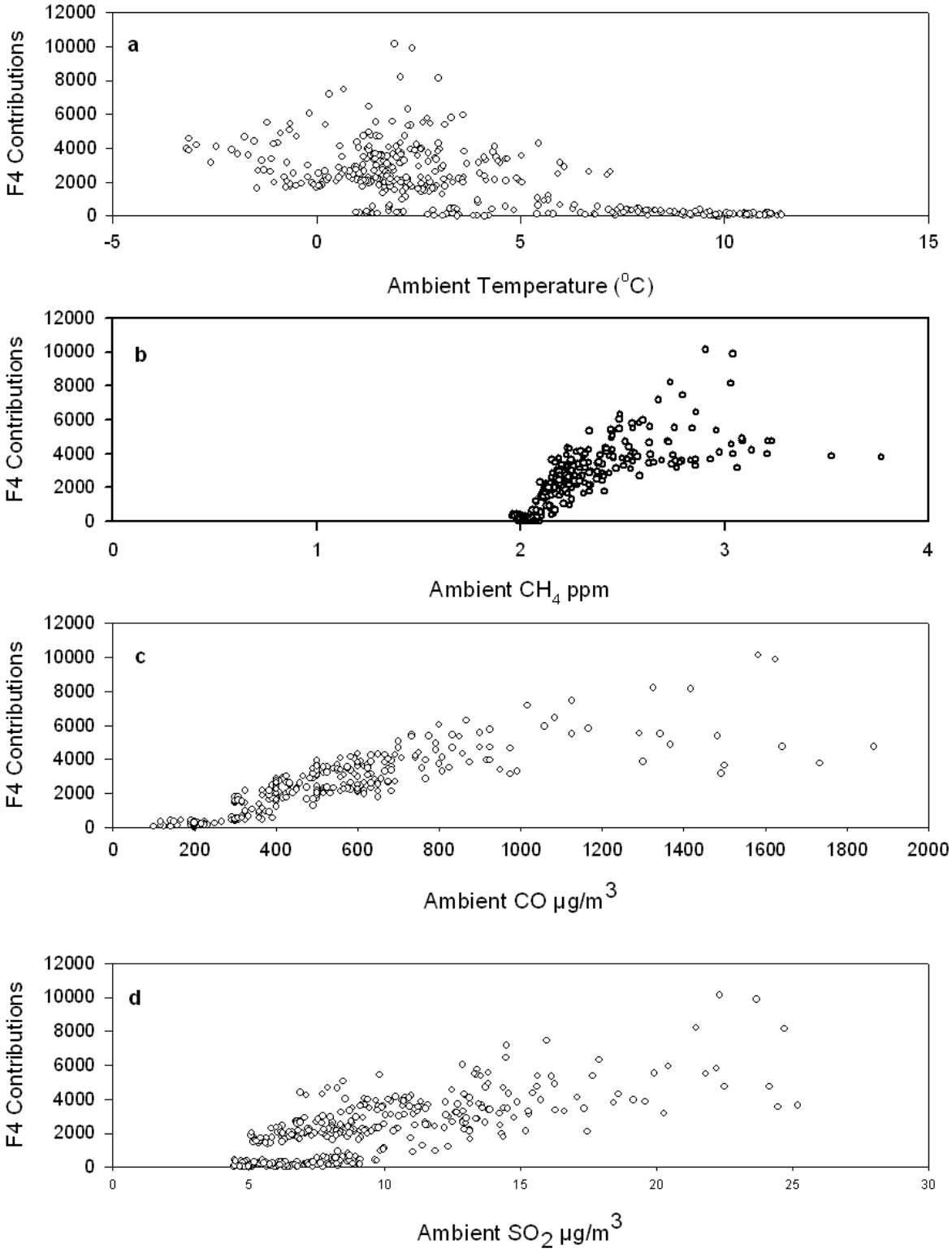


Figure S4 (a, b, c, and d). F4 source contributions vs. Temp, CH_4, CO and SO_2.

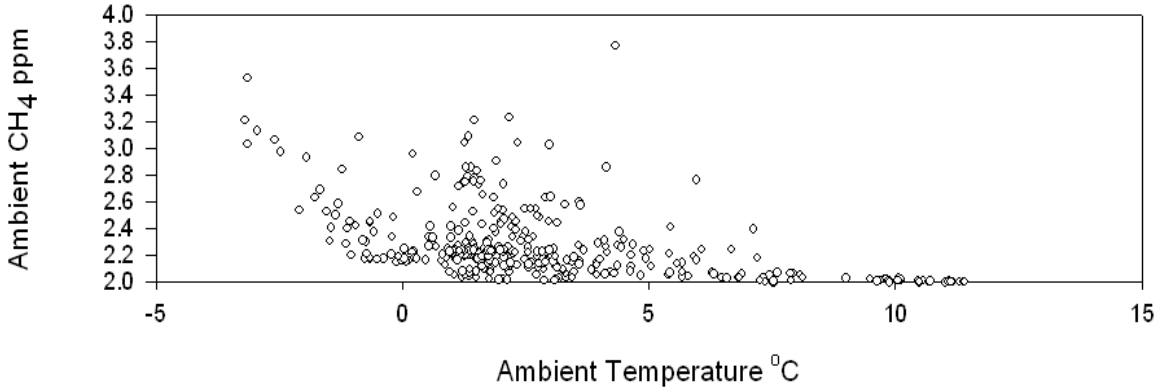


Figure S5. Ambient CH_4 vs. ambient Temp.

CHAPTER-V
CONCLUSION

CHAPTER-V
CONCLUSION

The combined particle number size distributions and readily available gaseous concentration data were used to identify winter sub-micron particle sources in the urban atmosphere of Prague, the capital of Czech Republic by applying bilinear Positive Matrix Factorization (PMF2). The ambient Particle Number Concentrations (PNC) were obtained using a Scanning Mobility Particle Sizer (SMPS) in the size range between 14.6 and 736.5 nm (midpoint diameters) along with the ambient gaseous concentrations of CO, SO_2, NO_x (NO + NO_2), O_3, CH_4, and Non Methane Hydrocarbons (NMHC) at the receptor site. The meteorological data concerning wind direction, wind speed, temperature, relative humidity and the UV-A and UV-B solar radiance were also obtained. The Conditional Probability Function (CPF) plots were used for directionality analyses, in determining the likely locations of the source emissions. In addition, diurnal patterns of factor contributions, the correlations of the factor contributions with gaseous pollutants (O_3, NO_x, SO_2, CO, CH_4 and NMHC) were used to assist in the interpretation of the sources.

For the sampling period-I, the PMF2 deduced 5 factor solution. They were assigned to as mixed source-1, traffic, heating, mixed source-2, and O_3-secondary particles, respectively. Out of 5 sources, 2 sources were assigned to as mixed sources-1 and 2 with CPF directionality from all sectors. Repetitive PMF2 run with different seeded value failed to resolve these two mixed sources. Often, some of the factors appear to be a mixture of several sources that cannot be further separated (Zhou *et al*., 2005). These two sources may be coming from background or long range transport not known due to lack of chemical tracers.

For the sampling period-II, the PMF2 analysis, identified 4 possible sources. 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.

REFERENCES

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.

Anttila, P., Paatero, P., Tapper, U. and Jarvinen, O. (1995). Source Identification of Bulk Wet Deposition in Finland by Positive Matrix Factorization. *Atmos. Environ.* 29:1705-1718.

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.

Begum, B.A., Biswas, S.K., Kim, E., Hopke, P.K. and Khaliquzzaman, M. (2005). Investigation of Sources of Atmospheric Aerosol at a Hot Spot Area in Dhaka, Bangladesh. *J. Air & Waste Manage. Assoc.* 55:227-240.

Begum, B.A., Hopke, P.K., Zhao, W. (2006). Source Identification of Fine Particles in Washington DC, by Expanded Factor Analysis Modeling. *Environ. Sci. Technol.* 39:1129-1137.

Bein, K.J., Zhao, Y.J., Pekney, N.J., Davidson, C.I., Johnston, M.V. and Wexler, A.S. (2006). Identification of Sources of Atmospheric PM at the Pittsburgh Supersite - Part II: Quantitative Comparisons of Single Particle, Particle Number, and Particle Mass Measurements. *Atmos. Environ.* 40:Suppl. 2, S424-S444.

Brown, S.G., Frankel, A. and Hafner, H.R. (2007). Source Apportionment of VOCs in the Los Angeles Area using Positive Matrix Factorization. *Atmos. Environ.* 41:227-237.

Buzcu, B., Fraser, M.P., Kulkarni, P. and Chellam, S. (2003). Source Identification and Apportionment of Fine Particulate Matter in Houston, TX, using Positive Matrix Factorization. *Environ. Eng. Sci.* 20:533-545.

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 at Urban and Suburban Residential Areas in Thailand by Positive Matrix Factorization. *Atmos. Environ.* 34:3319-3329.

Contd…

Eatough, D.J., Anderson, R.R., Martello, D.V., Modey, W.K., Mangelson, N.E. (2006). Apportionment of Ambient Primary and Secondary PM2.5 during a 2001 Summer Intensive Study at the NETL Pittsburgh Site using PMF2 and EPA UNMIX. *Aerosol Sci. Technol.* 40:925-940.

EPA (1986). PM10 SIP Development Guideline, Report No. EPA450/2-86-001. U.S. Environmental Protection Agency, Research Triangle Park, NC.

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(7): 802-815.

Gao, N., Gildemeister, A.E., Krumhansl, K., Lafferty, K., Hopke, P.K., Kim, E. and Poirot, R.L. (2006). Sources of Fine Particulate Species in Ambient Air Over Lake Champlain Basin, VT. *J. Air & Waste Manage. Assoc.* 56:1607-1620.

Gildemeister, A.E., Hopke, P.K. and Kim, E. (2007). Sources of Fine Urban Particulate Matter in Detroit, MI. *Chemosphere*, doi:10.1016/j.chemosphere.2007.04.027.

Gordon, G.E. (1988). Receptor Models. *Environ. Sci. Technol.* 22:1132–1142.

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 PM10 and Arsenic Concentrations in Central Chile using Positive Matrix Factorization. *Atmos. Environ.* 39:549-561.

Henry, R.C. (1987). Current Factor Analysis Receptor Models are ill Poised. *Atmos. Environ.* 21:1815-1820.

Heuss, J.M., Kahlbaum, D.F. and Wolff, G.T. (2003). Weekday/Weekend Ozone Differences: What can we learn from them? *J. Air & Waste Manage. Assoc.* 53(7):772-788.

Hien, P.D., Bac, V.T. and Thinh, N.T.H. (2004). PMF Receptor Modeling of Fine and Coarse PM10 in Air Masses Governing Monsoon Conditions in Hanoi, Northern Vietnam. *Atmos. Environ.* 38:189-201.

Contd…

Hien, P.D., Bac, V.T. and Thinh, N.T.H. (2005). Investigation of Sulfate and Nitrate Formation on Mineral Dust Particles by Receptor Modeling. *Atmos. Environ.* 39:7231-7239.

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.

Hopke, P.K., Xie, Y., Paatero, P., Barrie, L.A. and Li, S.M. (1998). Multiway Analysis of Airborne Particle Composition Data. *J. Aerosol Sci.* Suppl. 1, 29:S515-S516.

Hopke, P.K., Xie, Y.L. and Paatero, P. (1999). Mixed Multiway Analysis of Airborne Particle Composition Data. *J. Chemom.* 13:343-352.

Hovorka, J., Keohane, B., and Marshall, G.B. (1996). Elemental and Stable Lead Isotopic Composition of PM10 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 PM10 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 PM10 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, Czech Republic.

Hovorka, J., Braniš, M. and Schwarz, J. (2005a). Diurnal Variation of Submicron PM Size Distributions and Gaseous Pollutant Concentrations during a Year 2004 in Prague, Pollution Source Estimation. AAAR Conference, Particulate matter supersite program and related studies, 7–11 February, Atlanta, Georgia, USA.

Hovorka, J., Thimmaiah, D. and Civiš, M. (2005b). Size Distribution of Submicron Aerosols from the Stack Gas of Heating Boiler Chimney Operating on Natural Gas (English translation). In Czech: "Velikostní Distribuce Submikronových Částic Aerosolu ve Spalinách Kotle na Zemní Plyn". Conference Proceedings of Czech Aerosol Society, 15 November, Prague, Czech Republic, Pp: 31-32. ISBN 80-86186-13-X.

Contd…

Huang, S., Rahn, K.A. and Arimoto, R. (1999). Testing and Optimizing Two Factor-Analysis Techniques on Aerosol at Narragansett, Rhode Island. *Atmos. Environ.* 33:2169-2185.

Huang, S., Arimoto, R. and Rahn, K.A. (2001). Sources and Source Variations for Aerosol at Mace Head, Ireland. *Atmos. Environ.* 35:1421-1437.

Hwang, I. and Hopke, P.K. (2006). Comparison of Source Apportionments of Fine Particulate Matter at Two San Jose Speciation Trends Network Sites. *J. Air & Waste Manage. Assoc.* 56:1287-1300.

Hwang, I. and Hopke, P.K. (2007). Estimation of Source Apportionment and Potential Source Locations of PM2.5 at a West Coastal IMPROVE Site. *Atmos. Environ.* 41:506-518.

Ito, K., Xue, N. and Thurston, G. (2004). Spatial Variation of PM2.5 Chemical Species and Source-Apportioned Mass Concentrations in New York City. *Atmos. Environ.* 38:5269-5282.

Jorquera, H. and Rappenglück, B. (2004). Receptor Modeling of Ambient VOC at Santiago, Chile. *Atmos. Environ.* 38:4243-4263.

Kim, E., Hopke, P.K. and Edgerton, E.S. (2003a). Source Identification of Atlanta Aerosol by Positive Matrix Factorization. *J. Air & Waste Manage. Assoc.* 53:731-739.

Kim, E., Larson, T.V., Hopke, P.K., Slaughter, C., Sheppard, L.E. and Claiborn, C. (2003b). Source Identification of PM2.5 in an Arid Northwest U.S. City by Positive Matrix Factorization. *Atmos. Res.* 66:291-305.

Kim, E., Hopke, P.K., Paatero, P. and Edgerton, E.S. (2003c). Incorporation of Parametric Factors into Multilinear Receptor Model Studies of Atlanta Aerosol. *Atmos. Environ.* 37:5009-5021.

Kim, E. and Hopke, P.K. (2004a) [See correction i.e., 2004b]. Source Apportionment of Fine Particles in Washington, DC, Utilizing Temperature-Resolved Carbon Fractions. *J. Air & Waste Manage. Assoc.* 54:773-785.

Kim, E. and Hopke, P.K. (2004b). Source Apportionment of Fine Particles in Washington, DC, Utilizing Temperature-Resolved Carbon Fractions. *J. Air & Waste Manage. Assoc.* 54:1011-1012.

Contd…

Kim, E., Hopke, P.K. and Edgerton, E.S. (2004c). Improving Source Identification of Atlanta Aerosol using Temperature Resolved Carbon Fractions in Positive Matrix Factorization. *Atmos. Environ.* 38:3349-3362.

Kim, E., Hopke, P.K., Larson, T.V. and Covert, D.S. (2004d). Analysis of Ambient Particle Size Distributions using UNMIX and Positive Matrix Factorization. *Environ. Sci. Technol.* 38(1):202-209.

Kim, E., Hopke, P.K., Larson, T.V., Maykut, N.N. and Lewtas, J. (2004e). Factor Analysis of Seattle Fine Particles. *Aerosol Sci. Technol.* 38(7):724-738.

Kim, E., Hopke, P.K., Larson, T.V. and Covert, D.S. (2004f). Analysis of Ambient Particle Size Distributions using UNMIX and Positive Matrix Factorization. *Environ. Sci. Technol.* 38(1):202-209.

Kim, E., Hopke, P.K., Kenski, D.M. and Koerber M. (2005a). Sources of Fine Particles in a Rural Midwestern US Area. *Environ. Sci. Technol.* 39:4953-4960.

Kim, E., Hopke, P.K., Pinto, J.P. and Wilson, W.E. (2005b). Spatial Variability of Fine Particle Mass, Components, and Source Contributions during the Regional Air Pollution Study in St. Louis. *Environ. Sci. Technol.* 39:4172-4179.

Kulmala, M., Vehkamaki, H., Petaja, T., Dal Maso, M., Lauri, A., Kerminen, V.M., Birmili, W. and McMurry, P.H. (2004). Formation and Growth Rates of Ultrafine Atmospheric Particles: A Review of Observations. *J. Aerosol Sci.* 35:143-176.

Lapina, K. and Paterson, K.G. (2004). Assessing Source Characteristics of PM2.5 in the Eastern United States using Positive Matrix Factorization. *J. Air & Waste Manage. Assoc.* 54:1170-1174.

Lee, E., Chan, C.K. and Paatero, P. (1999). Application of Positive Matrix Factorization in Source Apportionment of Particulate Pollutants in Hong Kong. *Atmos. Environ.* 33:3201-3212.

Lee, J.H., Yoshida, Y., Turpin, B.J., Hopke, P.K., Poirot, R.L., Lioy, P.J. and Oxley, J.C. (2002). Identification of Sources Contributing to Mid-Atlantic Regional Aerosol. *J. Air & Waste Manage. Assoc.* 52:1186-1205.

Lee, J.H. and Hopke, P.K. (2006a). Apportioning Sources of PM2.5 in St. Louis, MO using Speciation Trends Network Data. *Atmos. Environ.* 40:Suppl. 2, S360-S377.

Lee, J.H., Hopke, P.K. and Turner, J.R. (2006b). Source Identification of Airborne PM2.5 at the St. Louis-Midwest Supersite. *J. Geophys. Res.-Atmos.* 111:D10S10, doi:10.1029/2005J-D006329.

Contd…

Li, Z., Hopke, P.K., Husain, L., Qureshi, S., Dutkiewicz, V.A., Schwab, J.J., Drewnick, F., Demerjian, K.L. (2004). Sources of Fine Particle Composition in New York City. *Atmos. Environ.* 38:6521-6529.

Liu, W., Hopke, P.K., Han, Y.J., Yi, S.M., Holsen, T.M., Cybart, S., Kozlowski, K. and Milligan, M. (2003a). Application of Receptor Modeling to Atmospheric Constituents at Potsdam and Stockton, NY. *Atmos. Environ.* 37:4997-5007.

Liu, W., Hopke, P.K. and VanCuren, R.A. (2003b). Origins of Fine Aerosol Mass in the Western United States using Positive Matrix Factorization. *J. Geophys. Res.* 108(D23):4716. Doi:10.1029/2003JD003678.

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). Spectral Analysis of Weekday-Weekend Differences in Ambient Ozone, Nitrogen Oxide, and Non-Methane Hydrocarbon Time Series in California. *Atmos. Environ.* 36(14):2327-2335.

Michael, R.A., Gulen, G., Jack, B. and Ilhan, O. (2000). Receptor Modeling for Elemental Source Contributions to Fine Aerosols in New York State. *J. Air & Waste Manage. Assoc.* 50:881-887.

Morishita, M., Keeler, G.J., Wagner, J.G. and Harkema, J.R. (2006). Source Identification of Ambient PM2.5 during Summer Inhalation Exposure Studies in Detroit, MI. *Atmos. Environ.* 40:3823-3834.

Ogulei, D., Hopke, P.K., Zhou, L., Pancras, J.P., Nair, N. and Ondov, J.M. (2006). Source Apportionment of Baltimore Aerosol from Combined Size Distribution and Chemical Composition Data. *Atmos. Environ.* 40:S396-S410.

Ogulei, D., Hopke, P.K., Ferro, A.R., Jaques, P.A. (2007a). Factor Analysis of Submicron Particle Size Distributions near a Major United States-Canada Trade Bridge. *J. Air & Waste Manage. Assoc.* 57:190-203.

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

Contd…

Owega, S., Khan, B.U.Z., D'Souza, R., Evans, G.J., Fila, M. and Jervis, R.E. (2004). Receptor Modeling of Toronto PM2.5 Characterized by Aerosol Laser Ablation Mass Spectrometry. *Environ. Sci. Technol.* 38:5712-5720.

Paatero, P. and Tapper, U. (1993). Analysis of Different Models of Factor Analysis as Least Square Fit Problems. *Chemom. Intell. Lab. Syst.* 18:183-194.

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

Paatero, P. (1996). User's Guide for Positive Matrix Factorization Programs PMF2.exe and PMF3.exe, University of Helsinki, Helsinki, Finland.

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. Stat.* 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. (2003). User's Guide for Positive Matrix Factorization Programs PMF2 and PMF3, Part 1: Tutorial. University of Helsinki, Helsinki, Finland.

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.

Pekney, N.J., Davidson, C.I., Bein, K.J., Wexler, A.S. and Johnston, M.V. (2006a). Identification of Sources of Atmospheric PM at the Pittsburgh Supersite, Part I: Single Particle Analysis and Filter-based Positive Matrix Factorization. *Atmos. Environ.* 40:Suppl. 2, S411-S423.

Pekney, N.J., Davidson, C.I., Robinson, A., Zhou, L.M., Hopke, P.K., Eatough, D. and Rogge, W.F. (2006b). Major Source Categories for PM2.5 in Pittsburgh using PMF and UNMIX. *Aerosol Sci. Technol.* 40 (10):910-924.

Contd…

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.

Poirot, R.L., Wishinski, P.R., Hopke, P.K. and Polissar, A.V. (2001). Comparative Application of Multiple Receptor Methods to Identify Aerosol Sources in Northern Vermont. *Environ. Sci. Technol.* 35:4622-4636.

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, D15:19045-19057.

Polissar, A.V., Hopke, P.K. and Poirot, R.L. (2001). Atmospheric Aerosol Over Vermont: Chemical Composition and Sources. *Environ. Sci. Technol.* 35:4604-4621.

Qin, Y., Oduyemi, K. and Chan, L.Y. (2002). Comparative Testing of PMF and CFA Models. *Chemom. Intell. Lab. Syst.* 61:75–87.

Qin, Y. and Oduyemi, K. (2003). Atmospheric Aerosol Source Identification and Estimates of Source Contributions to Air Pollution in Dundee, UK. *Atmos. Environ.* 37:1799-1809.

Qin, Y.J., Kim, E. and Hopke, P.K. (2006). The Concentrations and Sources of PM2.5 in Metropolitan New York City. *Atmos. Environ.* 40:Suppl. 2, S312-S332.

Ramadan, Z., Song, X.H. and Hopke, P.K. (2000). Identification of Sources of Phoenix Aerosol by Positive Matrix Factorization. *J. Air & Waste Manage. Assoc.* 50:1308-1320.

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(2):146-154.

Rizzo, M.J. and Scheff, P.A. (2007). Fine Particulate Source Apportionment using Data from the USEPA Speciation Trends Network in Chicago, Illinois: Comparison of Two Source Apportionment Models. *Atmos. Environ.* doi:10.1016/j.atmosenv. 2007.03.055.

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.

Song, X.H., Polissar, A.V. and Hopke, P.K. (2001). Sources of Fine Particle Composition in the Northeastern US. *Atmos. Environ.* 35:5277-5286.

Contd…

Song, Y., Tang, X.Y., Xie, S.D., Zhang, Y.H., Wei, Y.J., Zhang, M.S., Zeng, L.M. and Lu, S.H. (2004). Source Apportionment of PM2.5 in Beijing in 2004. *J. Hazard. Mater.* 146:124-130.

Song, Y., Zhang, Y.H., Xie, S.D., Zeng, L.M., Zheng, M., Salmon, L.G., Shao, M., and Slanina, S. (2006). Source Apportionment of PM2.5 in Beijing by Positive Matrix Factorization. *Atmos. Environ.* 40(8):1526-1537 and 40:7661-7662.

Suppan, P. and Graf, J. (2000). The Impact of an Airport on Regional Air Quality at Munich, Germany. *Int. J. Environ. Pollut.* 14:375-381.

Thimmaiah, D., Hovorka, J., Civiš, M. and Hopke, P.K. (2008). Source Apportionment of Sub-micron Prague Aerosol from Combined Particle Number Size Distribution and Gaseous Composition Data by Bilinear Positive Matrix Factorization. *Acta Univ. Carol.–Environmentalica*, 22:81-110.

Thimmaiah, D., Hovorka, J. and Hopke, P.K. (2009). Source Apportionment of Winter Submicron Prague Aerosols from Combined Particle Number Size Distribution and Gaseous Composition Data. *Aerosol and Air Qual. Res.* Xx:xxx-xxx **(**In Press).

Wang, H. and Shooter, D. (2005). Source Apportionment of Fine and Coarse Atmospheric Particles in Auckland, New Zealand. *Sci. of the Total Environ.* 340:189-198.

Xie, Y.-L., Hopke, P.K., Paatero, P. Barrie, L.A. and Li, S.-M. (1999). Identification of Source Nature and Seasonal Variations of Arctic Aerosol by Positive Matrix Factorization. *J. Atmos. Sci.* 56:249-260.

Xie, Y.L. and Berkowitz, C.M. (2006). The Use of Positive Matrix Factorization with Conditional Probability Functions in Air Quality Studies: An Application to Hydrocarbon Emissions in Houston, Texas. *Atmos. Environ.* 40:3070-3091.

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

Yli-Tuomi, T., Paatero, P. and Raunemaa, T. (1996). The Soil Factor in Rautavaara Aerosol in Positive Matrix Factorization Solutions with 2 to 8 Factors. *J. Aerosol Sci.* 27:S671-S672.

Yuan, Z.B., Lau, A.K.H., Zhang, H.Y., Yu, J.Z., Louie, P.K.K. and Fung, J.C.H. (2006). Identification and Spatiotemporal Variations of Dominant PM10 Sources over Hong Kong. *Atmos. Environ.* 40:1803-1815.

Contd..

Zabalza, J., Ogulei, D., Hopke, P.K., Lee, J.H., Hwang, I., Querol, X., Alastuey, A. and Santamaria, J.M. (2006). Concentration and Sources of PM10 and its Constituents in Alsasua, Spain. *Water, Air, Soil Pollut.* 174:385-404.doi:10.1007/s11270-006-9136-8.

Zhao, W.X. and Hopke, P.K. (2004). Source Apportionment for Ambient Particles in the San Gorgonio Wilderness. *Atmos. Environ.* 38:5901-5910.

Zhao, W.X. and Hopke, P.K. (2006a). Source Investigation for Ambient PM2.5 in Indianapolis, IN. *Aerosol Sci. Technol.* 40:898-909.

Zhao, W.X. and Hopke, P.K. (2006b). Source Identification for Fine Aerosols in Mammoth Cave National Park. *Atmos. Res.* (4):309-322.

Zhou, L.M., Kim, E., Hopke, P.K., Stanier, C.O. and Pandis, S. (2004). Advanced Factor Analysis on Pittsburgh Particle Size-Distribution Data. *Aerosol Sci. Technol.* Suppl. 1. 38:118-132.

Zhou, L,M., Kim, E., Hopke, P.K., Stainer, C. and Pandis, S.N. (2005). Mining Airborne Particulate Size Distribution Data by Positive Matrix Factorization. *J. Geophys. Res.-Atmos.* 110:D07S19, doi:10.1029/2004JD004707.

WEBSITE

1. www.tsi.com
2. www.horiba.com
3. www.udi.cz

ANNEX

PMF2 Initiation file (INI file) for Sampling Period-I

```
##PMF2 .ini file for: Sampling Period-I
## Monitor code M: if M>1, PMF2 writes output every Mth step
## For finding errors, use M<1 to output debug information
##      M       PMF2 version number
       1         4.2
## Dimensions: Rows, Columns, Factors. Number of "Repeats"
                    1279        41        5         5
##   "FPEAK"  (>0.0 for large values and zeroes on F side)
   0.0
## Mode(T:robust, F:non-robust)  Outlier-distance       (T=True F=False)
          T                     4.000
## Codes C1 C2 C3 for X_std-dev, Errormodel EM=[-10 ... -14]
       10.0000    1.0000    0.1000     -14
## G Background fit:  Components   Pullup_strength
                          0                 0.0000
## Pseudorandom numbers:   Seed     Initially skipped
                                  7       0
## Iteration control table for 3 levels of limit repulsion "lims"
##  "lims"    Chi2_test  Ministeps_required  Max_cumul_count
    1.00000   0.50000          5          100
    0.30000   0.50000          5          150
    0.00300   0.30000          5          200
## Table of FORMATs, with reference numbers from 50 to 59
## Number  Format_text(max 40 chars)
     50   "(A)                                "
     51   "((1X,5G13.5E2))                    "
     52   "((1X,10F8.3))                      "
     53   "((1X,20(I3,:' ')))                 "
     54   "((1X,150(G12.5E1,:' ')))           "
     55   "((1X,180(F9.4,:' ')))              "
     56   "(1X,A)                             "
     57   "((1X,150(G13.5E2,:' ')))           "
     58   "((1X,350(F4.3,:' ')))              "
     59   "((1X,600(I2,:' ')))                "
## Table of file properties, with reference numbers from 30 to 39
## Num- In  Opening  Max-rec File-name(max 40 chars)
## ber  T/F status   length
   30   T "OLD    "  2000  "FpeakDatamatrix.csv    "
   31   T "OLD    "  2000  "FpeakErrormatrix.csv   "
   32   T "OLD    "  2000  "G8.txt                   "
   33   T "OLD    "  2000  "F8.txt                   "
   34   F "UNKNOWN"  2000  "PMF34.DAT               "
   35   F "UNKNOWN"  2000  "MISC8-Fkeypeak.TXT     "
   36   F "UNKNOWN"  2000  "G8-Fkeypeak.TXT         "
   37   F "UNKNOWN"  2000  "F8-Fkeypeak.TXT         "
   38   F "UNKNOWN"  2000  "TEMP-Fkeypeak.TXT      "
```

```
   39   F "UNKNOWN"  2000  "$.DAT                          "
## Input/output definitions for 21 matrices
##  ===HEADING=====  ========MATRIX==========      default HEADING
##  --IN---- --OUT-   -----IN------   ---OUT--       for each matrix
## FIL(R)FMT FIL FMT FIL(R)(C)FMT(T) FIL FMT(T) ------max 40 chars----...
    0 F  50  38  56  30 F     0 F   38  57 F  "X (data matr)         "
    0 F  50  38  56  31 F     0 F   38  57 F  "X_std-dev /T (constant)"
    0 F  50   0  56   0 F     0 F    0  57 F  "X_std-dev /U (sqrt)    "
    0 F  50   0  56   0 F     0 F    0  57 F  "X_std-dev /V (proport) "
    0 F  50   0  56  32 T  F   0 F    0  57 F  "Factor G(orig.)        "
    0 F  50   0  56  33 T  F   0 F    0  57 F  "Factor F(orig.)        "
    0 F  50   0  56   0 F     0 F    0  53 F  "Key (factor G)         "
    0 F  50   0  56   0 F     0 F    0  59 F  "Key (factor F)         "
    0 F  50   0  56   0 F     0 F    0  52 F  "Rotation commands      "
    0 F  50   0  56               36  57 F  "Computed Factor G Q=   "
    0 F  50   0  56               37  57 F  "Computed Factor F      "
    0 F  50  35  56               35  57 F  "Computed std-dev of G  "
    0 F  50  35  56               35  57 F  "Computed std-dev of F  "
    0 F  50   0  56                0  57 F  "G_explained_variation  "
    0 F  50  35  56               35  58 F  "F_explained_variation  "
    0 F  50   0  56                0  57 F  "Residual matrix X-GF    "
    0 F  50  35  56               35  57 F  "Scaled resid. (X-GF)/S "
    0 F  50   0  56                0  57 F  "Robustized residual    "
    0 F  50  35  56               35  55 F  "Rotation estimates.  Q="
    0 F  50   0  56                0  55 F  "Computed X_std-dev     "
    0 F  50   0  56                0  55 F  "Background coefficients"
## If Repeats>1, for input matrices, select (R)=T or (C)=T or none
##    (R)=T: read(generate) again   (C)=T,"chain": use computed G or F
##    none, i.e.(R)=F,(C)=F: use same value as in first task
## (T)=T: Matrix should be read/written in Transposed shape
##
## Normalization of factor vectors before output. Select one of:
##   None   MaxG=1   Sum|G|=1 Mean|G|=1  MaxF=1 Sum|F|=1 Mean|F|=1
       T       F       F       F       F       F       F
## Special/read layout for X (and for X_std-dev on following line)
## Values-to-read (0: no special) #-of-X11  incr-to-X12  incr-to-X21
                      0       0       0       0
                      0       0       0       0
## A priori linear constraints for factors, file name: (not yet available)
   "none                          "
## Optional parameter lines (insert more lines if needed)
## (FIL#4 = this file)    (FIL#24 = .log file)
## After next 2 lines, you may include matrices to be read with FIL=4
## but observe maximum line length = 120 characters in this file
## and maximum line length = 255 characters in the .log file
    41*0
    37*0 4 1*0 3 1*0
    37*0 4 3*0
    41*0
    37*0 4 3*0
```

PMF2 Initiation file (INI file) for Sampling Period-II

```
##PMF2 .ini file for: Sampling period-II
 ## Monitor code M: if M>1, PMF2 writes output every Mth step
 ## For finding errors, use M<1 to output debug information
 ##      M       PMF2 version number
        1         4.2
 ## Dimensions: Rows, Columns, Factors. Number of "Repeats"
                    384        41        4         5
 ##   "FPEAK"  (>0.0 for large values and zeroes on F side)
    0.0
 ## Mode(T:robust, F:non-robust)  Outlier-distance       (T=True F=False)
           T                    4.000
 ## Codes C1 C2 C3 for X_std-dev, Errormodel EM=[-10 ... -14]
       10.0000    1.0000    0.1000     -14
 ## G Background fit:  Components   Pullup_strength
                          0                0.0000
 ## Pseudorandom numbers:   Seed     Initially skipped
                                  7       0
 ## Iteration control table for 3 levels of limit repulsion "lims"
 ##  "lims"    Chi2_test  Ministeps_required  Max_cumul_count
     1.00000   0.50000          5          100
     0.30000   0.50000          5          150
     0.00300   0.30000          5          200
 ## Table of FORMATs, with reference numbers from 50 to 59
 ## Number  Format_text(max 40 chars)
      50   "(A)                                "
      51   "((1X,5G13.5E2))                      "
      52   "((1X,10F8.3))                      "
      53   "((1X,20(I3,:' ')))                 "
      54   "((1X,150(G12.5E1,:' ')))             "
      55   "((1X,180(F9.4,:' ')))               "
      56   "(1X,A)                            "
      57   "((1X,150(G13.5E2,:' ')))             "
      58   "((1X,350(F4.3,:' ')))               "
      59   "((1X,600(I2,:' ')))                "
 ## Table of file properties, with reference numbers from 30 to 39
 ## Num- In  Opening  Max-rec File-name(max 40 chars)
 ## ber  T/F status   length
    30   T "OLD    "  2000  "CH4NMHCDataMatrix.csv                        "
    31   T "OLD    "  2000  "CH4NMHCErrorMatrix.csv                     "
    32   T "OLD    "  2000  "PMF32.DAT                          "
    33   T "OLD    "  2000  "PMF33.DAT                          "
    34   F "UNKNOWN"  2000  "PMF34.DAT                              "
    35   F "UNKNOWN"  2000  "MISC8.TXT                              "
    36   F "UNKNOWN"  2000  "G8.TXT                           "
    37   F "UNKNOWN"  2000  "F8.TXT                           "
    38   F "UNKNOWN"  2000  "TEMP.TXT                               "
```

```
   39   F "UNKNOWN"  2000  "$.DAT                              "
## Input/output definitions for 21 matrices
##  ===HEADING=====  ========MATRIX==========      default HEADING
##  --IN---- --OUT-   -----IN------   ---OUT--       for each matrix
## FIL(R)FMT FIL FMT FIL(R)(C)FMT(T) FIL FMT(T) ------max 40 chars----...
    0 F  50  38  56  30 F     0 F   38  57 F  "X (data matr)         "
    0 F  50  38  56  31 F     0 F   38  57 F  "X_std-dev /T (constant)"
    0 F  50   0  56   0 F     0 F    0  57 F  "X_std-dev /U (sqrt)    "
    0 F  50   0  56   0 F     0 F    0  57 F  "X_std-dev /V (proport) "
    0 F  50   0  56   0 F  T   0 F    0  57 F  "Factor G(orig.)        "
    0 F  50   0  56   0 F  T   0 F    0  57 F  "Factor F(orig.)        "
    0 F  50   0  56   0 F     0 F    0  53 F  "Key (factor G)         "
    0 F  50   0  56   0 F     0 F    0  59 F  "Key (factor F)         "
    0 F  50   0  56   0 F     0 F    0  52 F  "Rotation commands      "
    0 F  50   0  56              36  57 F  "Computed Factor G Q=   "
    0 F  50   0  56              37  57 F  "Computed Factor F      "
    0 F  50  35  56               35  57 F  "Computed std-dev of G  "
    0 F  50  35  56               35  57 F  "Computed std-dev of F  "
    0 F  50   0  56               0  57 F  "G_explained_variation  "
    0 F  50  35  56               35  58 F  "F_explained_variation  "
    0 F  50   0  56               0  57 F  "Residual matrix X-GF    "
    0 F  50  35  56               35  57 F  "Scaled resid. (X-GF)/S "
    0 F  50   0  56               0  57 F  "Robustized residual    "
    0 F  50  35  56               35  55 F  "Rotation estimates.  Q="
    0 F  50   0  56               0  55 F  "Computed X_std-dev     "
    0 F  50   0  56               0  55 F  "Background coefficients"
## If Repeats>1, for input matrices, select (R)=T or (C)=T or none
##    (R)=T: read(generate) again   (C)=T,"chain": use computed G or F
##    none, i.e.(R)=F,(C)=F: use same value as in first task
## (T)=T: Matrix should be read/written in Transposed shape
##
## Normalization of factor vectors before output. Select one of:
##   None   MaxG=1   Sum|G|=1 Mean|G|=1  MaxF=1 Sum|F|=1 Mean|F|=1
       T       F       F       F       F       F       F
## Special/read layout for X (and for X_std-dev on following line)
## Values-to-read (0: no special) #-of-X11  incr-to-X12  incr-to-X21
                      0       0       0       0
                      0       0       0       0
## A priori linear constraints for factors, file name: (not yet available)
   "none                              "
## Optional parameter lines (insert more lines if needed)
## (FIL#4 = this file)    (FIL#24 = .log file)
## After next 2 lines, you may include matrices to be read with FIL=4
## but observe maximum line length = 120 characters in this file
## and maximum line length = 255 characters in the .log file
```