Dissertation Report

ON

Source Apportionment of Aerosol in Delhi using UNMIX Receptor Model

Under guidance and co-guidance of

Dr. Gazala Habib (Assisstant Professor, IIT Delhi)
Mr. K. S. Babu (Assisstant Professor,Thapar University, Patiala)

Submitted by

Ms. Arti Choudhary

Department of Biotechnology and Environmental Science
Thapar University, Patiala, 147004
&
Department of Civil Engineering
Indian Institute of Technology Delhi, New Delhi, 110016

2010-11

CERTIFICATE

This is to certify that the present work titled **"Source Apportionment of Aerosol in Delhi using UNMIX Receptor Model"**, is a bonafied record of work done by Ms Arti Choudhary for the fulfillment of the requirements for the award of degree of Masters of Technology from Thapar University, Patiala, during year 2010-11. She has worked under my supervision and has fulfilled the requirement for the submission of this report. The results presented in this work have not been submitted in part or full to any other university for awards of degree.

Dr. Gazala Habib

Department of Civil Engineering

Indian Institute of Technology Delhi

New Delhi-110016

K. S. Babu

Mr. K. S. Babu

Department of Biotechnology &

Environmental Science

Thapar University, Patiala

Head, D.B.T.E.S

Thapar University

Patiala, 147004

Dean of Academic Affairs

Thapar University

Patiala, 147004

Head
Dept. of Biotech & Env. Sciences
Thapar University,PATIALA-147 004 (Pb.)

ACKNOWLEDGEMENT

Above all, I praise God for giving me the courage and protection in all my day to day activities. Next, I would like to express my appreciation to the Department of Civil Engineering, IIT Delhi, for giving me the opportunity to work on this project.

I would like to forward my sincere and special thanks to Dr. Gazala Habib and Mr. K.S. Babu for her/his continuous support, encouragement and patience, who have given their invaluable guidance during my project work. Her knowledge, insight and constant motivation at each step of the project has been instrumental for my work.

I would like to forward my special thanks to Mr. Sanjay Gupta (Lab Superintendent) of Civil Engineering Department, Mr Jai Prakash, Mr Gaurav, Mr. Pravin Kumar Mutiyar, and Ms. Vijaya Singh (Research scholar), Civil Engineering Department who helped in better during this work.

I am very thankful to my friends, Anil JKurian, Tarun Arigela, and Raginee for their moral support and help.

I would like to take this opportunity to express my deep gratitude to my family for their love and support throughout my entire education.

Atti choudhary

By Atti Choudhary

[redacted]1004

ABSTRACT

PM1 (particles having aerodynamic diameter < 1.0 μm) concentrations were measured in previous study at a sampling site inside the Indian Institute of Technology (IIT) Delhi campus for 5 months from November- 2009 to March- 2010. In this study the source identification and quantification were done using UNMIX (Version 6) model. The input data included 15 species (Na, K, Cl, Mn, SO4, NO3, Ca, NH4Cd, Cr, Cu, Ni, Zn, Pb, Fe, and Mg) for 51 days.

The sources with eigenvalue >1 were finally selected by model as clear source and other with eigenvalue<1 were omitted by default. UNMIX provided the results in three different ways namely Analyze Run, Highlight Run Output, or Diagnostic Plots.

The three sources were identified by model as clear sources. For source profile 1 composition analysis showed high contribution from Mn, and Ca, but associated with large uncertainty at 95% CI. The factor loading analysis also revealed high loading of Ca (0.63), Mn (0.85) to source 1. The source profile 1 was partially contaminated with NH4 and SO4 as moderate to low contribution with low uncertainty at 95% CI was estimated. One of the plausible explanation could be the crustal soil particle were enveloped by sulphate aerosol forming shell around the dust core. The above analysis revealed that source profile 1 was primarily influence by Ca and Mn and can be interpreted as crustal soil. In addition to this the EF Crust values of individual elements in PM_1 size fractions were analysed, which varied 2-9 for Ca and 4-8 for Mn in different months. The EF crust less than 10 indicates these elements are predominantly from crustal soil confirming the model interpretation.

Source 2 highly loaded with elements K (0.84) and Na (0.90) can be interpreted as biomass burning. The evidence of open agricultural field burning and biomass fuel burning for space heating during winter have been reported in literature (Parmer et al., 2001; Nair et al., 2006, Chowdhury et al., 2007) for peninsular and Delhi region. Source 2 was validated using the K/Na ratio for different biomass burning reported in literature. The K/Na ratio varies 0.6 to 2.92 for various type of biomass burned. The ratio estimated in present study also varied from 0.98 to 1.86 lies in the range of values reported in literature indicting the strong influence of biomass burning during winter in Delhi region.

Sources profile 3 was significantly influenced by $SO4^{-2}$, $NO3^{-}$ and $NH4^{+}$ with large uncertainty, however, again the fraction apportionment of these species to source profile 3 indicated low variability. In factor loading analysis source 3 was significantly loaded with

$NO3^-$ (0.88), $SO4^{-2}$ (0.84) and $NH4^+$ (0.89) indicating the secondary aerosol formation mechanism as a source. The formation of $SO4^{-2}$ and $NO3^-$ in the atmosphere takes place by photochemical reaction (Chakroburty and Gupta, 2010, Saolapurkar and Sharma, 2006), and are termed as secondary aerosol. Fog during winter the during winter the photochemical reactions of precursor gases (SO2 and NOx) in the presence of UV radiation result in aerosol formation. Singhai (2010) reported high concentration of SO4 ions in densely foggy days of December, 2009 and January, 2010. Aerosol particles mediate the formation of fog in the atmosphere (Pandis et al., 1990; Seinfeld and Pandis,1998) through the preferred heterogeneous nucleation of water vapour on pollution particles, at high RH. Fog droplets further aid aerosol formation through aqueous-phase reactions of soluble gaseous precursors (e.g. SO2 and H2O2), leading to higher aerosol concentrations of species like sulfate on fog abatement, which then nucleate subsequent fog–smog–fog cycles (Pandis et al., 1990). Particulate inorganic matter would also contain soluble inorganic ions (Cl_, NO3-, NH4+) present in combustion aerosols, which could also mediate water uptake and fog formation (Mehta et al., 2009) hence, again enhance the sulphate aerosol formation.

The secondary aerosol formation is the major contributor (67%) to atmospheric aerosol loading over Delhi region followed by crustal soil (22%) and biomass burning (10%). The finding in this study has implications in regional climate study.

The model predicted element concentrations were correlated with observed concentration. The predicted Vs measured PM1 concentrations plot (r2=0.77) showed that the UNMIX model worked well and the predicted and measured PM1 mass concentrations are in good agreement. The $NO3^-$, K, $NH4^+$ $SO4^{-2}$ and Ca showed high correlation with r^2 value greater than 0.7, while Cl, Na, and Mn showed moderate correlation with R2 varied between 0.52-0.67.

TABLE and CONTENTS

LIST OF FIGURE

LIST OF TABLES

GLOSSERY

Coefficient of Correlation (r): A statistic representing how closely two ariables co-vary; they can vary from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation).

Collinearity: A situation in which a near-perfect linear relationship exists among some or all of the independent variables in a regression model; in practical terms, there is some degree of redundancy or overlap among the variables.

Confidence Interval (CI): CI for a population parameter is an interval with an associated probability p that is generated from a random sample of an underlying population such that if the sampling was repeated numerous times and the confidence interval recalculated from each sample according to the same method, a proportion p of the confidence intervals would contain the population parameter in question.

Covariance: A statistical measure of correlation of the fluctuations of two different quantities.

Edge: A line that defines the boundary of the relationship between two parameters on a scatter plot.

Factor analysis: A procedure for grouping data by similarity among variables (i.e., variables that are highly correlated are grouped).

Interquartile range (IQR): The difference between the 75th and 25th percentiles of a data set.

Method Detection Limit (MDL): The minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from the analysis of a sample in a given matrix containing the analyte.

Outlier: Data physically, spatially, or temporally inconsistent.

Particulate matter (PM): A generic term referring to liquid and/or solid particles suspended in the air.

R-squared (r2): Statistical measure of how well a regression line approximates real data points; an r2 of 1.0 (100%) indicates a perfect fit

CHAPTER-1

1. INTRODUCTION

In the past few years interest in air quality monitoring has increased, specifically pertaining to the identification of pollution sources and their quantification needed to implement air pollution control strategies. Since observing the quantity of various pollutants emitted from all potential pollution sources is virtually impossible, receptor models are used to analyze concentrations of pollutants or particles measured over time in order to gain insight concerning the unobserved pollution sources. Multivariate receptor modeling aims to identify the pollution sources and assess the amounts of pollution by resolving the measured mixture of chemical species into the contributions from the individual source types. The basic physical model comes from the laws of chemistry. The number of sources is the first problem encountered and need to be resolved. When the number and the composition of pollution sources are unknown, factor analytic approaches can been employed in order to identify pollution sources. As in the factor analysis models, the choice of the number of pollution sources (factors) used in receptor models is crucial, generally, the number of sources is chosen using one of many methods suggested in the literature. However, these methods often are not satisfying and in many papers the number of pollution sources is fixed on the basis of previous studies and/or specific assumptions made by the researcher. Once a model with known sources has been fitted, interest often lies in describing the composition of each pollution source and the amount of pollution emitted from each source. Such information is of great value when formulating and evaluating air quality policy and studying the climate change.

1.1 OBJECTIVE:

With the above background the following objectives were outlined for the study:

- Identification and Quantification of aerosol sources in Delhi
- Influence of winter season on aerosol formation and source distribution.

CHAPTER-2

2.0. LITERATURE REVIEW

2.1. SOURCE OF AEROSOL

Atmospheric aerosol comes from both natural and anthropogenic sources. Aerosols emitted from natural sources like forest fires and volcanic eruptions (Safai *et al.*, 1993) comprise a large fraction of atmospheric aerosol. However, the role of anthropogenic processes as the source of aerosol has been increasing continuously with the onset of the industrial revolution (Kim et al., 2002). Atmospheric aerosols comprise of organic carbon, black carbon ions, trace metals and organic compounds.

More than a dozen different types of motor vehicles release particles from incomplete combustion of fuel and lubricating oil, including gasoline-powered autos and light trucks with and without catalytic converters, diesel trucks and busses, railroad locomotives, aircraft, Ships, and off-highway construction equipment. Industrial fuel combustion, home heating and industrial manufacturing processes contribute to the particle emissions. In addition, there are dispersed area-wide sources that occur at the individual household or small business level, such as food cooking (e.g. meat charbroiling), and wood combustion in fireplaces and wood stoves, that can create substantial particle emissions when aggregated over a city having many millions of residents.

2.2. ATMOSPHERIC FINE AEROSOL CONCENTRATION

Most recently, fine particulate pollution gave rise to debate in numerous European countries as concentration limits were overstepped more frequently and more clearly than in the past. Urban areas were especially affected. Industry, traffic, agriculture and residential heating proof mainly responsible for high fine particulate emission levels. In Austria for instance, within the residential heating sector small-scale biomass combustion plants are responsible for about 88% of the PM_{10} emissions, most of them (about 86%) are poorly controlled old biomass combustion systems.

Compared to oil or natural gas, biomass contains a considerable amount of ash. This inevitably leads to fly ash emissions during combustion. Fly ash emissions of complete biomass combustion, which is usually achieved by modern and automatically controlled small-scale biomass combustion plants, consist mainly of inorganic aerosols and are related

to the PM_1 fraction (particulate matter with a particle size smaller 1.0 µm). Carbonaceous aerosols are a product of incomplete combustion, which is often the case in poorly controlled old biomass combustion systems or natural draught systems (like chimney stoves or tiled stoves), and consist of elementary carbon (soot) or condensed hydrocarbon compounds (organic aerosols). Moreover, old residential heating systems emit considerably higher PM_1 emissions than modern small-scale biomass combustion systems.

Fine particulate matter is the thoracic fraction out of total particulate matter, i.e. the fraction that can pass the larynx and reach the lung. There is strong epidemiologic indication that particulate matter in air has serious adverse health effects.

2.2.1. ATMOSPHERIC FINE AEROSOL (PM_1) CONCENTRATION IN INDIA

In India number of studies has been done on the $PM_{2.5}$ in various cities like source apportionment for Delhi, Mumbai, Kolkata, and Chandigarh were conducted in March 2001 to January 2002, using receptor-based chemical mass balance modelling. Five major sources of primary $PM_{2.5}$ were quantified: diesel exhaust, gasoline exhaust, road dust, coal combustion, and biomass combustion. Important trends in the seasonal and spatial patterns of the impact of these five sources were observed. Primary emissions from fossil fuel combustion (coal, diesel, and gasoline) contributed 25–33% of the $PM_{2.5}$ in Delhi, 21–36% in Mumbai, 37–57% in Kolkata, and 28% in Chandigarh. These figures can be compared to the biomass combustion contributing 7–20% in Delhi, 7–20% in Mumbai, 13–18% in Kolkata, and 8% in Chandigarh.

Another seasonal study was done in Delhi to examining atmospheric $PM_{2.5}$ concentrations was conducted over a 24-month (Jan 2007-Jan 2008) period at a suburban site (Jorhat) of India by using principal component analysis. Major source types to be identified and apportioned were crustal source, burning of coals, vehicular and traffic emissions, wood burning and secondary aerosol formation. Crustal source is the major contributor to $PM_{2.5}$ concentration followed by burning of coals.

In a study of Mumbai Airborne particulate matter in two size fractions $PM_{2.5\text{-}10}$ and $PM_{2.5}$ collected at a residential site in Vashi. These samples have been analysed for chemical composition using INAA and EDXRF. Varimax rotated factor analysis identified five major

sources contributing to coarse and fine particulate mass. FA-MLR technique is applied to apportion the sources. Source apportionment studies showed maximum contribution of the coarse fraction was from sea salt (35%) and crustal (25%) sources. A considerable amount of the mass was also contributed from industrial (14%), vehicular (10%) and fugitive emissions (7%). These results also showed the percentage contribution of soil, two-stroke emission with fugitive dust, industrial emission, motor vehicles and sea salt to the average fine mass concentration was 3%, 18%, 23%, 29% and 9%, respectively.

In India only few studies have been done which focus on fine particulate matter and their chemical constituents. A similar study carried out at different locations in Delhi [Srivastava et al., 2008] during winter 2005-06 using a high volume five stage cascade impactor including one submicron (<0.7μm) and one super micron stage (0.7-1.6μm) and together they reported as a fine size fraction of PM mass. In this study identified source were resuspended road dust, building material and vehicular emission.

One more study for fine particulate matter (PM_1) source apportionment was done in IITK campus. Still higher concentrations of PM_1 were observed indicating its transport by wind from far off emission sources. The identified sources from this study were road dust, vehicular emission, coal combustion, and secondary sources.

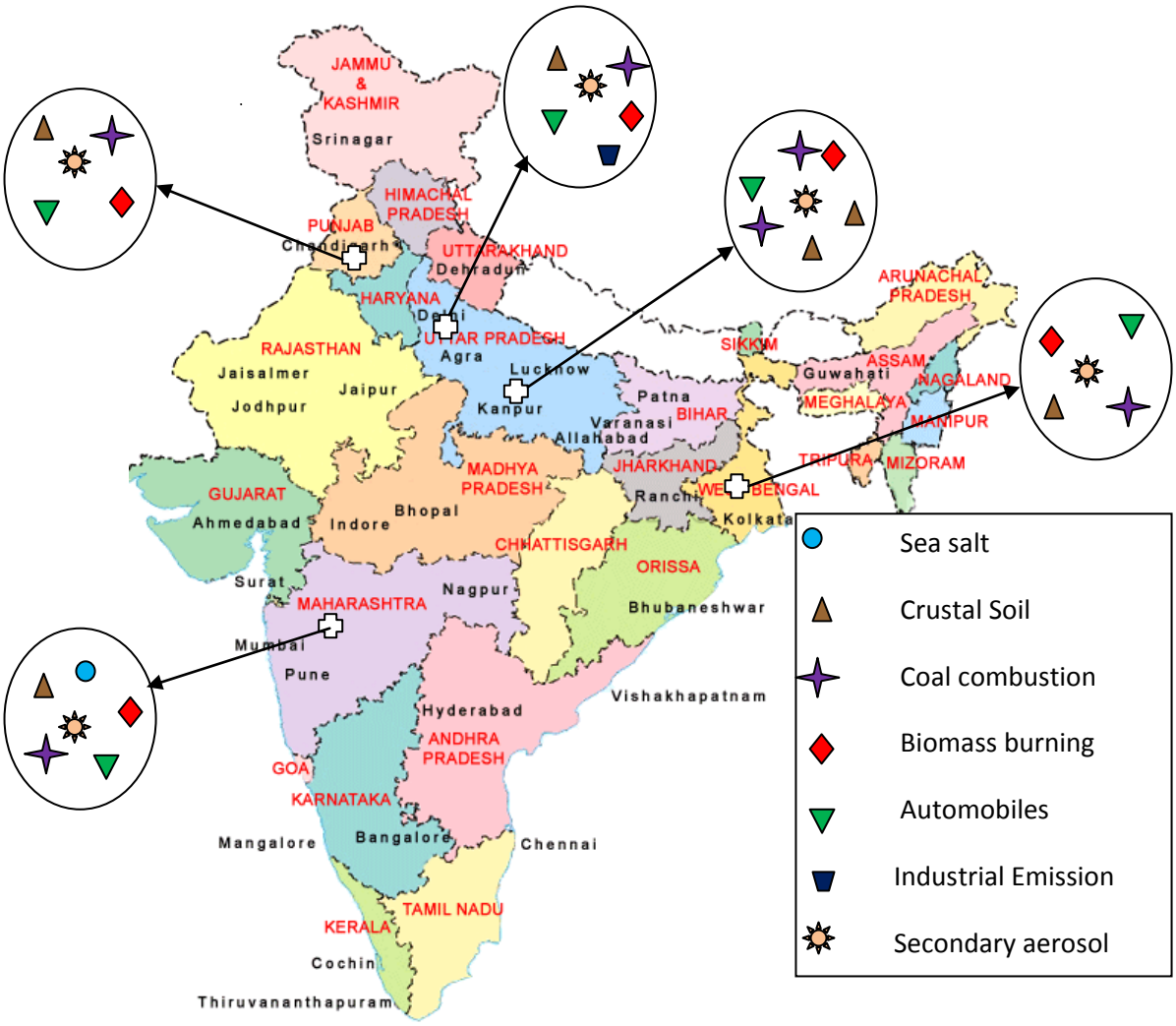

Figure 2.1: Source apportionment studies in Indian cities**.**

2.2.2. ATMOSPHERIC FINE AEROSOL CONCENTRATION IN WORLD REGION

Several research studies have been carried on $PM_{2.5}$ and very few studies have been carried out on PM_1 around the world. In most of the studies PM mass was collected for some time period in a particular season or in different seasons over a year followed by quantitative chemical speciation in terms of elements, anions, organics etc. to identify major sources and their contributions. In a related study carried out at Xi'an, China from Dec. 1 to Dec. 31, 2006 and the concentrations of 11 water-soluble ions (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}, F^-, Cl^-, Br^- NO_2^-, NO_3^- and SO_4^{2-}) and elemental carbon (EC) and organic carbon (OC) were determined in the filter samples to characterize the chemical composition of PM_1 over Xi'an, China during winter, here PM_1 level in winter in Xi'an was 149.7 $\mu g/m^3$, which was higher than

those observed in Hong Kong and Taipei. Ionic species and carbonaceous aerosol were dominant of PM_1, which attributed 46.0% and 27.5% to the total particle mass. SO_4^{2-}, NO_3^-, Cl^- and NH_4^+ were major ions, which occupied to 18.1%, 9.1%, 4.6% and 4.7% of PM_1 mass. The two secondary ionic species were in the form of NH_4HSO_4 and NH_4NO_3. OC concentration in Xi'an in winter is higher than those at Hong Kong and Taipei. High OC/EC ratios were observed in Xi'an, which may be due to more complicated sources of PM_1 carbon. Factor analysis for the eight carbon fractions demonstrated that coal combustion and gasoline exhaust, diesel exhaust, and biomass burning were three significant factors and explained 93% contributions of carbonaceous aerosol.

Another study carried out in the Tito Scalo—Southern Italy—from April 2006 to March 2007 (Rosa Caggiano et.al, 2009). In this period examined, PM_1 mean value is generally lower than those found in other anthropized sites. PCA results pointed out three probable PM_1 sources that are industrial emissions, traffic and re-suspension of soil dust.

In urban area of the city Kaohsiung, Taiwan (Lim and Lee; 2003), a study indicate that the ambient coarse (PM_1–PM_{10}) and submicron (PM_1) particles were being contributed by different sources. On average for the 8-month samples, 52% of the PM_{10} was made up of PM_1. The PM_1-PM_{10} ratio was observed to vary between summer and winter, it being higher in the summer (approximately 62% in summer and 48% in winter). Results from recent studies (J. Air Waste Manage. Assoc. 51 (2001) 489; Atmos. Environ. 36 (2002) 1911) at this same study site supported that combustion sources and secondary aerosols played significant roles in the formation of ambient submicron (PM_1) aerosol particles in the urban area.

The distribution of the solvent-extractable organic components in the fine (PM_1) and coarse ($PM_{1–10}$) fractions of airborne particulate was studied for the Algeria. That was done during October 2006 concurrently in a big industrial district, a busy urban area, and a forest national park located in Algiers, Boumerdes, and Blida, respectively, which are the three biggest provinces of Northern Algeria. Identified source were industrial, vehicular and biogenic emissions. Most particulate organics were associated to fine aerosols. The industrial district was more affected by air pollution than the urban zone, while the forest area had the cleanest atmosphere.

Table 2.1: Source Apportionment studies in world region

Reference	Source identified	Markers	Location
Zhenxing Shen et.al, 2009	Coal Combustion and Gasoline Exhaust, Diesel Exhaust, Biomass Burning	(OC2, OC3, OC4), (EC2, EC3), (EC1, OP)	China, Xi'an city
Rosa Caggiano et.al, 2010	Crustal origin, traffic and re-suspension of soil dust, , crustal origin, non-ferrous metal industry, chromium plating or fertilizer use, building material	(Al, Mg, Ti, Fe), (Pb, Zn, Mn), (Na, K, Ti), (Cd, Cr), (Cu, Ni), (Ca)	Tito Scalo—Southern Italy
Viana. M et.al; 2008	vehicular source, crustal source, sea-salt source, mixed industrial/fuel-oil combustion, secondary aerosol	(Fe/Ba/Zn/Cu), (Al/Si/Ca/Fe), (Na/Cl/Mg), ($V/Ni/SO_4^{-2}$), ($SO_4^{-2}/NO_3^-/NH_4^+$)	Europe
Arifa Lodhi, 2009	soil and road side dust, industrial emissions from Iron & Steel Industry, vehicular/traffic emission, secondary aerosols , Vehicular/Industrial Oil burning	(Al, Ca, K), (Co, Cr, Fe, Ni, Sn), (Cd, Pb, Sb, Zn), (SO_4^{-2}, NO_3^-, NH_4^+), (V, As, Sr)	Pakistan
Yu Song et.al. 2006	Road Dust, Motor Vehicle Exhaust, Coal combustion & Biomass exhaust, Sec.SO_4^{-2} & Sec NO_3^-, Industrial Process	(Al, Si, Ca, Ti, Mg, Fe), (Pb, Zn), (OC, Cl, K), (SO_4^{-2}, NO_3^-), (Ni)	China

2.3.TRACERS AND MOLECULAR MARKERS FOR VARIOUS SOURCES

Organic molecular markers are individually quantifiable organic compounds emitted by a specific source or class of sources. Some common molecular marker compound classes include alkanes, cycloalkanes, polycyclic aromatic hydrocarbons (PAHs), steranes, fatty acids, sterols and methoxyphenols. Ideally, each molecular marker would originate from one unique source type. In reality, however, they can originate from multiple source types each with their own time-dependent source contribution. Much work has been done to characterize the unique molecular marker emissions from a wide range of anthropogenic and natural sources (e.g., motor vehicles, biomass combustion, food cooking).

Here this report present some typical molecular marker originates from unique source type.

Wood combustion: Use of retene as a molecular marker for wood smoke was first proposed by Ramdahl 1983. Many organic compounds found in wood smoke can be traced to the structure of the original wood or other biomass fuel burned (Standley et.al. 1990, Simoneit et.al. 1993, Rogge et.al. 1998).

These compounds survive the fire in altered or unaltered form. The useful tracer compounds include phytosterols, lignans, phenolic products from lignins, and diterpenoids from resins in the wood.

Meat charbroiling: One of the largest particle sources in urban areas that generally remains uncontrolled is the release of particles from food cooking operations. When cooking meat over a natural gas-fired charbroiler, meat fat drips from the meat, falls onto hot synthetic coals, and is volatilized or dispersed to form a fine aerosol mist which then escapes through the exhaust hood over the broiler, often forming a visible plume downwind of the kitchen.

Vegetative detritus: As plant leaves brush together in the wind, small wax protrusions present on the leaf surfaces break off and become entrained into the atmosphere. In rural and remote areas these plant fragments can be a significant fraction of the fine particle mass that is present in the atmosphere, while in urban areas they constitute one of the minor sources that is difficult to observe when mixed with the much larger amount of material that is contributed by man's activities.

Tire dust: Particles shed as motor vehicle tires wear down are found in the urban atmosphere. Tire dust particles are of interest not only because they contribute a small

amount to airborne particle mass concentrations, but also because natural rubber latex proteins to which some people may be allergic can be identified in tire dust extracts [Miguel.et.al.1996]. Probably the best molecular tracer for tire dust is the styrene / butadiene copolymer that forms a major portion of most synthetic tire tread.

Motor vehicle exhaust: Hopanes and steranes are present in the lubricating oil used by both gasoline-powered and diesel-powered motor vehicles, and are also found in diesel fuel. Thus hopanes and steranes are emitted in the particle phase from both vehicle types. Diesel engine exhaust particles historically have contained a much higher fraction of black elemental carbon (i.e. soot) than is the case for gasoline-powered vehicle exhaust [Hildemann et.al. 1991].

Paved road dust: Paved road dust entrained into the atmosphere by passing traffic is one of the largest sources of particle emissions in cities. The organic compounds present in paved road dust represent a mixture of material from most of the sources that exist on or near the highway. Deposited motor vehicle exhaust particles, tire dust, and plant fragments ground up as traffic passes over leaves that fall onto the street all are found in paved road dust [Rogge et.al. 1993]. But paved road dust is accompanied by large amounts of soil dust particles that are not found in the other sources just named.

Table 2.2 Marker elements associated with various emission sources:

Source name	Marker	Reference
Soil	Al, S, Mg, Ti, Fe, K, Ca, Si	Kulshrestha et. al; 1999, A. Chakraborty and T. Gupta, 2009, S.S. Park, Y.J. Kim 2005, de Foy et al., 2005
Road dust	Ca, Mg, Fe, Pb, Zn, Cr, Fe, V	A. Chakraborty and T. Gupta, 2010
Sea-salt	Na, Cl, Na^+, Cl^-, Br^-, I^-, Mg, Mg^{++}	Kulshrestha et. al; 1999
Oil Burning	Vn, Ni, Mn, Cr, Fe, As, S,	S.S. Park, Y.J. Kim, 2005
Coal burning	Cd, Pb, Cl, Se	Saolapurkar and Sharma, 2006,
Nonferrous metal Industries	Cu, Cd, Ni, Cr	Mohanraj et al., 2004; Chandra Mouli et al., 2006; Gioda et al., 2006
Refuse Incineration	K, Zn, Pb, Sb	S.S. Park, Y.J. Kim, 2005
Biomass Burning	K, OC, EC, Br, nss-K, Cl, Sn,	Kulshrestha et. al; 1999, Miranda et al., 1996; Malm et al., 1994
Automobile gasoline	Be, La, Pt, SO_4^{-2}, NO_3^-, Zn, Si,benzo[ghi]perylene, coronene	Kang, 1998
Automobile diesel	S, SO_4^{-2}, NO_3^-, EC, OC, Fe,	Kang, 1998
Secondary aerosol	SO_4^{-2}, NO_3^-, NH_4^+	Kulshrestha et. al; 1999, Saolapurkar and Sharma, 2006,
Vehicular emission	Ni, Cu, Zn, Cr, Pb, SO_4^{-2}	A. Chakraborty and T. Gupta, 2009, 2010
lubricating oil	17α (H)- 21 β (H)-29-norhopane, 17 α (H)-21 β (H)-hopane,	Michel J. Kleeman et.al. 2009
wood burning	levoglucosan	Michel J. Kleeman et.al. 2009
meat cooking	cholesterol	Michel J. Kleeman et.al. 2009

2.4. SOURCE IDENTIFICATION AND SOURCE QUANTIFICATION

Receptor modeling techniques like factor analysis (FA) and PMF are convenient to apply on aerosol measurements obtained from locations with little or no information on the pollution sources since they do not require “a priori” source profiles as input (Hopke 2003). Factor Analysis (FA), an advanced statistical technique, is extremely useful in identifying the relationships among variables that are driven by common processes such as sources, transport, and chemistry. Factor Analysis (FA) is particularly useful for source apportionment studies when there is no prior information available about the nature of the major aerosol sources affecting a particular receptor station. FA has been widely used for analyzing

atmospheric data and is well documented in the atmospheric literature (Huang *et al.,* 1999; Kulkarni 2004; Ma *et al.,* 2004).

2.4.1. RECEPTOR MODELING APPROACH

During the past three decades, receptor models have been widely used for source apportionment. The fundamental principle of receptor modelling is that mass conservation is assumed and a mass balance analysis can be used to determine and apportion ambient pollutant concentrations to individual emitting sources. A mass balance equation is written to account for all chemical species identified in the filter samples analysed, as contributions from independent sources. Receptor models represent a statistical evaluation of ambient measurements at different times and locations, thus the selection of the appropriate method depends on prior knowledge on the sources and source profiles. If the sources are known and detailed information on source profiles is available, Chemical Mass Balance (CMB) models can be applied, whereas in case the sources are unknown and there is limited information on source profiles, Principal Component Analysis (PCA) and Positive Matrix Factorization (PMF) methods are preferred. Apart from the different Chemical Mass Balance modelling tools, receptor models can also be categorised as Aerosol Evolution and Equilibrium models, which estimate how reduction in one precursor will affect PM end-products, and as Back Trajectory models, which identify the origin of polluted air masses usually transported to the receptor from long distance sources.

2.4.1.1 SOURCE PROFILE BASED RECEPTOR MODEL- CMB

I)Chemical Mass Balance (CMB):

The CMB receptor model (Friedlander, 1973; Gordon, 1980, 1988; Watson, 1984; Watson *et al.*, 1990; Hidy and Venkataraman, 1996) consists of a solution to linear equations that express each receptor chemical concentration as a linear sum of products of source profile abundances and source contributions. The CMB modelling approach is one of oldest and most widely used techniques for modelling the ambient air particulates and involves identification of the sources that contribute to pollutant, the proper selection of chemical species estimation of the uncertainty in the ambient concentrations and source profiles, and finally the solution of the chemical mass balance equations.

CMB studies concluded that the model performed well in apportioning primary particulates that do not undergo reactions in the atmosphere. This modelling method is easy and simple but the difficulty lies in determining the number and profile of sources before modelling. CMB quantifies concentrations from chemically distinct source types rather than contributions from individual sources. CMB cannot distinguish between sources with similar chemical and physical properties.

CMB model uses the chemical and physical characteristics of gases and particulate at a given receptor point to identify the presence of and/or quantify source contributions. Equation (1) is the basic relation corresponding to the selected receptor point. This equation expresses the relation between the concentrations of the chemical species measured at the receptor point (Main health center of Fahaheel) and the chemicals emitted from the source.

$$\Delta Ci = \Sigma Fij - Si \quad (1)$$

Where: ΔCi = the difference in concentration of a chemical compound i at the receptor point.
Fij = the fraction of concentration of the species i starting from the source j.
Si = the concentration of pollutant i at the receptor point.
The main theoretical limitations of this model may be summarised as follows:

- Need to use locally determined experimental source profiles. Source profiles for a vast number of sources are currently available in the literature, but certain specific profiles may date back to the 1990s (US-EPA, 1999) even though more recent source profiles are also available in the SPECIATE database (v4.0). Furthermore, emission profiles for a single source may strongly depend on the industrial process, the raw materials used, etc., and they may thus vary widely from one location to another. As a consequence, it is highly recommendable to use experimental (rather than literature) source profiles.
- Inability to identify unknown sources. Given that source profiles are provided as input to CMB, the model is unable to identify other unknown sources. The identification of a large difference between modelled and measured results is the only way to observe that an input source may be missing.
- Inability to apportion secondary aerosols linked to the input emission sources. The input data for CMB consist of emission profiles measured at the emission sources, and thus made up of only primary particulates. Secondary species must thus be modelled separately as secondary sources with chemical composition determined by stoichiometry (e.g. an ammonium sulphate source, an ammonium nitrate source, etc.).

2.4.1.2 STATISTICAL RECEPTOR MODEL- PCA, PMF, UNMIX

I) Principle component analysis (PCA):

PCA model belongs to the category of factor analysis (FA) techniques, i.e. it is a multivariate method used to study the correlations among the measured element concentrations at the receptor (Pio et.al.,1989; Paatero and Tapper, 1993). With this method, the principal components explaining the variance of the speciated data are extracted and then interpreted as possible sources. The main problem with PCA is that it does not provide a unique solution mainly because of its simple approach to factor analysis, which is in principle an ill posed problem (Henry, 1987). Despite this drawback, known as rotational ambiguity, PCA as been applied as a tool for source apportionment in many air quality studies (for example Viana et al., 2007; Karar and Gupta, 2007) because the results it provides are generally comparable with receptor modelling techniques like PMF or CMB.

II) Positive matrix factorisation:

Positive Matrix Factorisation (PMF) is a statistical factor analysis method, based on the law of mass conservation that follows a different approach to factor analysis. The PMF model calculates site-specific source profiles together with time variations of these sources based on correlations embedded in ambient data. PMF uses a point-by-point least-squares minimization scheme for the optimum scaling of data. Therefore, the profiles produced can be directly compared to the input matrix without transformation. Various PMF factor analysis least-squares problems are solved using a multilinear engine. PMF is comparatively a harder tool than PCA but provides better apportionment resolution. PMF has been successfully applied to particulate matter data in several studies Hopke et al, 2006.

PMF models require data on measured concentrations (of species/elements) for a number of samples, together with information on the associated uncertainties. Where appropriate (e.g. when analysing ambient PM samples), information on meteorological parameters and concentrations of associated gaseous species may also be used.

PMF models are expressed as follows:

$$C_{ij} = \sum_{p=1}^{n} g_{ip} f_{pj} + e_{ij}$$

where*: p* is the number of sources; *j* is the number of species, with $j \geq p$; C_{ij} is the measured ambient concentration of species *j* in sample *i*; f_{pj} (source profiles) is the fractional concentration of species *j* in the emissions from source p; g_{ip} is the concentration contribution of source *p* to sample *i;* and e_{ij} is the portion of the measured concentration that cannot be explained by the model.

III) UNMIX model:

The UNMIX model has been developed for the US EPA by Dr. Henry and has several unique features. UNMIX has an advanced computationally intensive algorithm to estimate the number of sources that can be seen above the noise level in the data (Henry et al., 1999; Park et al., 2000). UNMIX solves the problem using a geometrical approach. If the data consist of n observations of m species, then the data can be plotted in an m-dimensional data space, where the coordinates of a data point are the observed concentrations of the species during a sampling period. If **N** sources exist, the data space can be reduced to a (**N**-1)-dimensional space. It is assumed that for each source, some data points exist for which the contribution of the source is not present or small compared to the other sources. These are called edge points and UNMIX works by finding these points and fitting a hyperplane through them; this hyperplane is called an edge. If **N** sources exist, then the intersection of **N**-1 of these edges defines a point that has only one contributing source. Thus, this point gives the source composition. In this way, we determine the compositions of the **N** sources from which the source contributions are calculated (Henry, 2003).

UNMIX works fundamentally by searching for edges. To produce good edges, it is necessary to explore the data to check for influential points. These points are usually those that lie too far below or along the edge, enabling them to affect the location of the edge considerably more than the other points. UNMIX uses PCA to reduce the dimensionality of the data space. Geometrical concepts of self-modeling curve resolution are used to ensure the results obey (to within error) non-negativity constraints on source compositions and contributions. One of the natural constraints imposed to the model is the non-negativity condition since both source compositions and contributions must be zero or positive. However, the effects of analytical

errors require that UNMIX allows small negative values (Henry, 1997, 2000). UNMIX basically assumes linear combinations of all source contributions.

UNMIX model equation:

Let the data be organized in an n*m matrix C such that each row represents the observed concentrations of several species in one sample or time period. Thus, there are n time periods and m species with m is usually being much smaller than n. UNMIX begins with the SVD of the data matrix C, any ambient concentrations matrix, *C*, can be decomposed uniquely as:

$$C = UDV^T$$

where *C* is an ambient concentration matrix of *m* measurements by *n* species, U is the *m***m* eigenvectors of CC^T, *D* is an *m***n* diagonal matrix made up of square roots of the corresponding eigenvalues, and V^T is the *n***n* eigenvectors of C^TC. This result is called the SVD theorem. Traditionally, the columns of U and V and the elements of D are sorted in the order of decreasing singular values. The purpose of the SVD is to reduce the dimensionality of the data space from m to N, the number of sources that produced the data. The number of sources N is estimated by the Numfact algorithm (Henery et.al.1999 ;). Numfact uses a bootstrap method to estimate the signal-to-noise ratio of the components. The number of factors is the number of components with at least twice as much signal as noise. Once N is found, keep the first N columns of U and V and the first N rows and columns of D associated with the N largest singular values.

Then, D_N is a N*N diagonal matrix, and U_N and V_N are n*N and m*N matrices, respectively, and

$$C = U_N D_N V^T + €$$

$$C_N = U_N D_N V^T$$

Here, € represents the error term consisting of all the variability in C not accounted for by the first N principal components. The SVD is closely related to principal component analysis (PCA). Simply writing the product UD in equation as a single matrix gives the principal components decomposition of C. Indeed, the terminology of PCA and factor analysis is often used in SVD and the columns of V are spoken of as components or factors.

Each model has its own advantages and disadvantages. CMB, on the one hand, is the most commonly used method among those requiring a very detailed knowledge of sources and

emission profiles. On the other, PCA and PMF are the most commonly used models requiring relatively little quantitative knowledge of sources and emission profiles, even though they do require initial qualitative knowledge of the sources present in the study area. The major difference between the two techniques is that PCA does not have constraints on the values of either the component loadings or scores, but requires that the resulting components be orthogonal, while PMF requires component loadings and scores to be non-negative, but has no orthogonality requirement. The lack of a non-negativity requirement in PCA has the potential of giving physically unreasonable results in the form of negative values for quantities that must be non-negative.

UNMIX has an advanced computationally intensive algorithm to estimate the number of sources. UNMIX uses PCA to reduce the dimensionality of the data space. Geometrical concepts of self-modeling curve resolution are used to ensure the results obey (to within error) nonnegativity constraints on source compositions and contributions. This, however, is not sufficient to uniquely determine the source compositions and contributions (Henry, 1987). Additional constraints determined from the data itself are needed. These are estimated by looking for edges in the data determined by points where one source is small compared to the other sources (Henry, 1997, 2003). Some special features of UNMIX are the capability to replace missing data and the ability to estimate large numbers of sources (the current limit is 15) using duality concepts applied to receptor modelling (Henry, 2005).

UNMIX was more limited and identified a lower number of sources, this limitation is related to the model because it does not take into account the uncertainty in the experimental data but it is very sensitive to this one, excluding some variables. In this case, UNMIX excluded species where most of the measurements were replaced by half the detection limit. The UNMIX solution is highly dependent on the species that are selected. On the other hand, this model showed problems to identify those sources contributing with a low percentage to the total mass corroborating the weakness of UNMIX related to identify ubiquitous sources, infrequent sources and relatively small sources (contributing less than about 10% to the total mass).

UNMIX was the only model that slightly overestimated the concentration. This could be interpreted in terms that UNMIX was the best model regarding the explained mass.

CHAPTER-3

3.0. EXPERIMENTAL METHOD AND MODEL RUNNING

3.1. Input file preparation:

UNMIX software require METLAB on system for proper installation, here using UNMIX ver-6 software for present study. This software free to download on USEPA site.

For preparing input file following steps was taken in this study. Primarily make XL file in which Date should be placed in first column and species name should be in first row, first species should be MF (mass Fraction). In XL file date should be in the pattern of-mm/dd/yyyy.

Now species which have negative value or zero value should be replaced with half of the method detection limit (MDL) of the species. If the species have some missing value due to any reason then these space filled with the -99 value as mention in table 3.1. One more modification of input file was done i.e. Reconstruction of the input file in crustal origin element. Aerosol reconstructed mass was calculated to determine the extent of aerosol mass closure for the chemical species detected. Mass reconstructions were used to screen the data as outlined by Malm et al. (1994), Tolocka et al. (2001), Chen et al. (2002), and Lewis et al. To estimate the oxygen content of the crustal material, calcium concentrations were multiplied by 1.4, and iron by 1.43 (Chow et al., 1996).

Now this final XL was changed into the notepad format (text MS-DOS) because UNMIX software accept input file in the notepad form only.

Table-3.1: Input data file parameter

Data File Types	**Delimited *.txt, *.dat, *.csv, or *.xls**
Data Location: Species Names Date Only Date & Time	 First row of data file First column of data file First and second columns of data file
Data Format: Date Time Missing Values	MM/DD/YYYY (e.g. January 1, 2007 = 01/01/2007) HH:MM in 24 hour cycle (e.g. 1:45 PM = 13:45) Characters (e.g. XX) or Numerical Values (e.g. -99)
Recommended Units: NO_2 or SO_2 ppb CO ppm PM Mass & Species	 ppb ppm $\mu g/m^3$
Negative or Zero Data	Replace with 1/2 the method detection limit for the species

3.2 BASIC OPERATION FOR THE UNMIX VER.6 MODEL RUNNING

3.2.1 STEP 1: PROCESS FOR INPUT VALUE:

Input data file is read by selecting the ⟶ NEW Input data command (in the upper left corner of the main window).

- The input data file window will open as shown in Figure:3.1(appendix iii).
- Click on the "concentration" in the input data file then location of input data file will be open. Choose the appropriate file type and click open.
- Again select the data & hours information from same input data file (fig: 3.1, appendix iii).
- Now click "OK".

3.2.2 STEP 2: SUGGEST EXCLUSION FUNCTIONS:

- Exclusion of high SNR (signal to noise ratio) species is done for better result in Receptor models. It is recommended that species with more than 50 percent

of the variance due to error, or specific variance (SV) be considered for exclusion from further UNMIX modeling.

- If there is a species that is known to be important and it has only a little more than 50 percent error, then it should not be excluded from the modeling process.
- Exclusion species can be further evaluated by using the View/Edit Points and Observations (in particular the View Time Series Plots, View/Edit Observations and Points, and View/Edit Influential Points commands).
- It is best to first exclude all species that do not pass even the minimum quality requirements.
- These excludable species may be those with a low number of observations or species with a low number of concentrations above the detection limit.
- VEIW/EDIT POINTS AND OBSERVATIONS command can be used for more refining of exclusion species.

3.2.2.1 Suggest exclusion operation process:

- Now select the "Suggest Exclusion button" located in the data processing window and consider all species.
- After few moments, a message will be displayed stating that "some of the species have over 75% of their values missing".
- Select "**OK**" and the species that are recommended for exclusion will appear highlighted. For example Species such as NA are recommended for exclusion because they have a SV greater than 0.50 or 50 percent fig 3.2 (appendix iii).
- Select the Exclude button to move the selected species from the Included to Excluded Species box.
- Then selecting the OK button in fig 3.3 (appendix iii) then new window will be open like fig 3.4 (appendix iii).

3.2.3 STEP 3: SELECT THE TOTAL, NORM & TRACER SPECIES:

Before Selecting the total norm & tracer species first highlight the one species which you wish to be selected as total, norm, & tracer species in the selected species box.

Generally the Total & Norm species are same (MF). Specific examples of using the total, tracer, and normalization options for VOCs and PM can be found in Mukerjee et al. (2004), and Lewis et al. (2003).

The excluded species in the suggest exclusion step are simply not included in the analysis. but unselected species are available. They are essentially on the sidelines and can be brought in anytime for analysis.

Usually the normalization and TS are the same, as this gives a source composition as a mass fraction. However, in some applications, one may wish to normalize to some standard species to be consistent with other reported normalized compositions (more detail in UNMIX user manual).

3.2.4 STEP 4: "SELECT INITIAL SPECIES" FUNCTION

Selected species can be determined by the user or by using the UNMIX Species Selection Tools. The Select Initial Species command uses the species with the largest loadings in the varimax factor analysis of the data to find a selection of species that gives a 4 or 5 source model that has very good signal-to-noise properties. If a 4 or 5 source solution is not found, a 3 source solution is attempted.

- Before using the UNMIX Species Selection Tools, move the species with high mean mass concentration over to the Species box. For example, select the species with a mean mass concentration greater than $1\mu g/m^{3.}$
- For seeing the species of mean mass concentration greater than $1\mu g/m^3$ "Select the Data Processing button" in the upper right corner of the main window to go back to the Data Processing window in order to review the data. Then select all species have mean mass concentration greater than $1\mu g/m^3$.
- The species with mass concentrations greater than $1\mu g/m^3$ in the wdcpmdata are MF, EC, EC1, NO_3, OC, and SO_4 (species means are shown in the Data Processing window). Other species such as SI could be added, since it is a marker for soil or crustal material and it is typically present in quantities above the analytical method.
- Then Select the "**OK**" button to return to the main window.

3.2.4.1 "Select initial species" operation process:

- Select the "initial species" command from the species selection tool box as shown in fig 3.5 (appendix iii). Then initial species source profile will appear as like in fig 3.6 (appendix iii).
- The maximum number of potential sources is the number of sources with a signal-to-noise ratio greater than 1.5 (Analysis Diagnostics box).

- Replace missing value command also use after the step 2 but it is conditional. (see user manual).

3.2.4.2 Explanation of initial species selection output:

- The preamble to the UNMIX results is pretty much self-explanatory. The line that gives the min r^2, etc. does require some explanation. Min r^2 is the smallest r^2 value for any species in the model (r^2 for any species is greater than this value).
- The Solution Summary box displays the summary information for the run. The highlighted lien in the Solution Summary box tells us that there are 10 species and that there are 1171 observations. In addition, the r^2 values and the signal-to-noise ratios are listed for the run. Here in the example, the data can be explained with a five source model with a minimum r^2 value of 0.98. This means that at least 98% of the variance of each species can be explained by five sources. Thus, the number in the UNMIX display is the minimum r^2 value over all the species, not the overall r^2 of the fit.

3.2.5 STEP 5: "SUGGEST ADDITIONAL SPECIES"

- The Suggest Additional Species command is used to create a list of species that can be added to an UNMIX solution.
- Click on the "Suggest Additional Species" command from Species Selection Tool a new window will be open, in it Select the "All" button to run both the SAFER and Influential Points (IP) algorithms (fig 3.7, appendix iii).
- If the SAFER algorithm cannot suggest any additional species to add, only the influential species information will be displayed.
- Select species from the SAFER Algorithm that have a Max Signal to Noise (Max Sig/N) greater than 2 and Minimum Signal to Noise (Min Sig/N) less than 1. If "TS" was selected, the SAFER Algorithm output also includes columns labelled Big Mass and Big Mass Percent (fig 3.8, appendix iii).
- It is recommended to select species from the Influential Point Algorithm that have a high edge resolution (90th percentile greater than 0.80) or a low number of influential points.
- The Suggest Additional Species command can be run multiple times by evaluating the species with the highest Weighted Score in the UNMIX solution.

- If a feasible solution is found after adding the species to the Selected Species, use the Analyze Run, Fit Diagnostics command to evaluate the solution.
- For this example the following species are recommended: Se, Ca, and Si. If the number of significant and strong species abruptly decreases after adding a species, remove the species from the selected species box, and either try the next recommended species or finish using the Suggest Additional Species command

Two additional UNMIX options are available for selecting additional species:

- Auto UNMIX (AU) and
- Batch Mode:

The Batch Mode command can be used to evaluate the addition of species recommended by the Suggest Additional Species command. Batch mode will add all combinations of the Suggest Additional Species command. Batch mode will add all combinations of the highlighted species in the Unselected Species window to the Selected Species and evaluate each set of species.

3.2.5.1 Batch mode process:

To run the Batch Mode command, select the Batch mode radio button and select the Run button. Then a statement will be appear "would you like to use influential points algorithm and obtained batch mode species selection suggestion" click "**yes**" like fig 3.9 (appendix iii).

- Then new window will be appear for batch mode preference like fig 3.10 (appendix iii).
- Choose the output file then click the "OK" button.
- Now the Solution Summary box will list the Batch Mode results (Run Type "B").
- The Analyze Run and Diagnostic Plots buttons are not available for Batch Mode model results.
- Choose Run # 121 from the Solution Summary and select the Highlight Run Output button to view the results. The run number may be different due to the number of individual runs prior to the batch mode run but will have 18 species and be before the "NO FEASIBLE OR PARTIAL SOLUTION" and before the # Species increases.
- Then select the "highlighted run output button", new window will be appear like fig 3.11(appendix iii).

- Select the “Load Species” button to add these species to the Selected Species box and select the “Run” button. The new solution can now be evaluated with the Analyze Run and Diagnostic Plots options.

3.2.6 STEP 6: PLOT DISTRIBUTION

Select the MF line of Run # 2 (as e.g.) from the Analysis Results window then analysis result window will be open and click the Tools on this window in this choose Plot Distribution command to create a pie chart showing the average source contributions (figure 3.12, appendix iii).

3.2.7 STEP 7: EVALUVATING THE RESULTS

An individual run can be evaluated by selecting one of the buttons below the Solution Summary box in the main UNMIX window:

- ❖ Analyze Run
- ❖ Highlight Run Output
- ❖ Diagnostic Plots

3.2.7.1 Analyse run option:

When select the “analyse run” a new window will be open like fig 3.13 (appendix iii) in it select the “analyse result window” then click on **“**export”. This option gives the result in the text file form like fig 3.14 (appendix iii).

3.2.7.2 Highlighted run output option:

Select this option in solution summary box a highlighted run window will be appear with output.

3.2.7.3 Diagnostic plot option:

In the main UNMIX window, select the “Diagnostic Plots button”. The Diagnostic Plots window contains many plotting options and an example from the Fit Diagnostics option and source variability option are displayed in Figure 3.15, 3.16 respectively (appendix iii). The predicted vs. Measured concentrations are shown in the top plot and the regression results are shown in the bottom plot. A figure is generated for each of the selected species. Use the << Previous (Prev) and Next >> buttons to view each of the figures.

- The majority of the options in the Analyze Run and Diagnostic Plots window are self-explanatory and a Help button is available for details.
- Some explanation is required for the Source Contribution plots:
  - ❖ Normalized scale plots have source contributions that are scaled to range between 0 and 100,
  - ❖ Uniform scale plots have source contributions that are scaled to range between 0 and 1,
  - ❖ Actual scale plots are not scaled and these plots will display the true source contribution values.

3.2.7.4 Run Profiles:

Run profiles or selections can be saved using the Save Profile command as shown in figure 3.15 (appendix iii). The profiles are saved as .umx files that can be opened and edited in Microsoft Excel. The default folder for these files is C:\Program Files\EPA Unmix 6.0\Unmix Profiles. Using the data and selected select the File and Save Profile commands and save the file as wdcpmdata profile.umx.

3.2.8 STEP 8: ADVANCED OPERATIONS

The basic strength of UNMIX, as with all receptor models, is that it relies on the data; the basic weakness of UNMIX is that it relies on the data. A number of problems may afflict a species and make it unsuitable for selection to be part of the model.

i) A common problem is that the species may have missing concentration data, however, the "Replace Missing Values" command in UNMIX can be used to replace missing data.

ii) Another common problem is the existence of outliers in the values of the species. These can often be detected using scatter plots as described in the section on "View/Edit Observations & Points", or by using the "Influential Points" command.

3.2.8.1 Influential observations function:

The tools for evaluating "influential observations and points" are listed in the lower left corner of the Data Processing window under View/Edit Points and Observations.

This step will come after exclusion process (step1) "Select the View/Edit Points and Observations list" on the Data Processing window, and "consider all species" as seen in figure 3.17 (appendix iii). Select MF as he base species.

All of the species concentrations are plotted against the selected base species and the edges are displayed with dashed red lines (figure 3.18, appendix iii). The size of the figures can be increased by reducing the number of species or by selecting a figure and the Undock button.

3.2.8.2 Influential Observation Process operating steps:

- Select the Edit Data→Delete Selected Observation(s) command from the top left of the View/Edit Observations & Points window. After all of the species related to the deleted observation are removed, new edges are drawn (as displayed in figure 3.19, appendix iii) and the observation number of the deleted point is recorded in the Data Processing Report.
- The data points can be restored in the Edit Data command using the Restore Most Recently Added or Restore All options.
- Select the Data Processing Report button in the upper right corner of the Data Processing window to view the report, or export the report from the Analyze and Export Run Results window, Data Processing Report option.
- The Datacursor Mode command is turned off by re-selecting the Edit Data, Datacursor Mode command. After the Datacursor mode is turned off, the figures are re-drawn resulting in a delay in the window commands being available.
- Save the modified data set by selecting Save Changes. It should be noted that UNMIX saves the deleted point information in the Run profile.

3.2.9 STEP-9: FIT UNSELECTED SPECIES COMMOND

The unselected species are evaluated by selecting the Species Selection Tools, Fit Unselected Species command (figure 3.20, appendix iii). Figure 3.21(appendix iii) shows the results for the wdcpmdata species.

Each row is the result of a non-negative least squares regression of the source contributions (normalized to a mean of 1) and the species. The amount of the species in each of the sources, the amount not explained by the sources or constant, and coefficient of determination are displayed. Results are also sorted by the species r^2.

These results can also help guide the selection of additional species that can be added to the UNMIX solution. For example, OC_2 has a high r^2 value and after adding it to the Selected Species Window and selecting the Run button, the solution in figure 3.22 (appendix iii) is

produced. This command can be used to create a consistent set of species for comparing EPA UNMIX and EPA PMF profiles.

3.2.10 STEP 10: FACTOR ANALYSIS COMMAND

The UNMIX results can be compared to the typical factor analysis approach (varimax rotated factor analysis) by selecting the Species Selection Tools, Factor Analyze Selections command. Use the Factor Analyze Selections command to evaluate the UNMIX results shown in figure 3.23 (appendix iii).

The factor, matrix scores, covariance matrix scores, explained covariance, and Screen plot are displayed. Select the Display Loadings button to show the loadings. The number of factors can be changed by selecting a value in the Number of Factors list. Selecting the Display Loadings button again will show the loadings associated with the new number of factors (figure 3.24, appendix iii).

CHAPTER-4

4.0. RESULT

4.0 RESULTS AND DISCUSSIONS

A number of run were performed to find most appropriate sources of ambient aerosol. By default the model excludes the week species with specific variance greater than or equal to 0.5. However, it is recommended (EPA Unmix 6.0 Fundamental and user guide) if species is important and it has specific variance little more than 0.5, then it can be manually included in model run. Therefore, in selected run the species Ca (SV=0.53), K (SV=0.51) were added manually. The mass of crustal elements (Ca and Fe) were reconstructed to include the oxide form of these elements and improve the mass closure. However, the reconstruction neither change the number of factors nor improve the r^2 values therefore, did not considered in final model run presented in this study. The selected model run presented here had 8 species and produced signal to noise ratio 2.88 (>2.00) and r^2 as 0.88 which indicated the results are robust.

4.1 Correlation of aerosol constituents with PM_1:

UNMIX software package (Version 6) was applied on 51 days dataset. All the 13 species (NO_3^-, SO_4^{-2}, NH_4^+ Mn, Mg, Na, K, Cl, Zn, Ni, Fe, Cd, Cr, Cd, Cu and Pb) included for final analysis, but species with larger variation (Cu, Cr, Fe, Ni, Zn, Cd, Mg, and Pb) were excluded by model. A stability analysis was performed by removing sample one at a time to confirm that the converged result was obtained.

Scatter plots have been used in air quality studies to visualise the correlation between various parameters. The scatter plots generated by UNMIX (Figure 4.1) showed good to poor relationship between various aerosol constituents and PM1. Some species like Na, K, Mn, Ca, and Cl showed somewhat scattered data point over the plane and these species has poor r^2 (0.01-0.12) values however, NO_3^-, SO_4^{-2} and NH_4^+ showed moderate correlation (r^2=0.49-0.63) with PM_1.

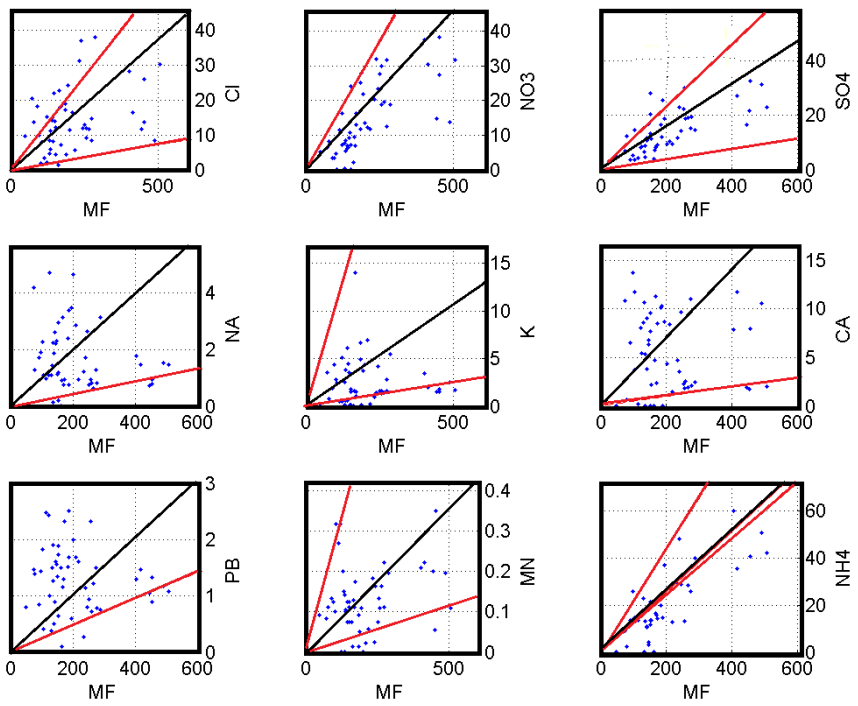

Figure.4.1. Scatter plots of various species concentrations (µg/m^3) versus PM_1 mass concentration (µg/m^3) for the IIT Delhi site.

Data points of K were closed to PM_1 axis, where certain days (12 March, 2010) showed relatively higher levels concentration. Henry (2000) defined a subjective term ‘‘good edge’’ to describe the species that show a clear relation with $PM_{2.5}$ concentration with identifiable, sharp lower and upper edges in the scatter plot. Species such as NO_3^-, SO_4^{-2}, and NH_4^+ can be qualitatively classified as those with ‘‘good edges’’ as demonstrated in fig 4.1.

4.2 Source identification using analyze run

UNMIX identified 3 factor associated with eigenvalue >1. UNMIX provided results in three different ways namely Analyze Run, Highlight Run Output, or Diagnostic Plots. The Highlight Run Output option simply highlights the solution in the Analysis Results window. The Analyze Run window, lists outputs (variability, variability distribution, Fit diagnostics and correlation). The diagnostic plot window present results in the graphical form (Signal to Noise, Confidence interval, Variability, Variability distribution, Edge plot and Fit diagnostics). Here Sigma based output (fit diagnostic option of analyze run), box plot based analysis (variability option of diagnostic plots) and factor loading based output from factor

analysis command option are compared. In sigma based output value '0' indicates source profile NOT within IQR, '1+' indicates source profile within IQR but not centred and not strongly influenced by outlier and '2+' indicates source profile within IQR and centred and little influence by outliers which means the species having 2+ value are significant contributor to the source.

Source 1 identified by model with Na, Ca, and Mn (1+: not strongly influenced by outliers). Source 2 cannot be clearly identified in this sigma based output however; further analysis is detailed in section 4.3, 4.4, and 4.5. Source 3 showed little influence by the outliers for SO_4^{-2} (2+), not strongly influenced by outliers for NH_4^+, NO_3^- (1+) (Table 4.1).

Table 4.1. Sigma based Output-Species fit diagnostics

Species	Source 1	Source 2	Source 3
MF	1+	1	0+
NO3	1	1	1+
SO4	0	1	2+
NA	1+	1	0
K	0	1	0
CA	1+	1	0
MN	1+	1	0
NH4	1	1	1+

4.3 Composition of sources and uncertainty

UNMIX also produced the source composition by using matrix of MxNx100 where M is the number of species, N is the number of sources and 100 is the feasible bootstrap runs. For each bootstrap run, the bootstrap matrix was normalized using the row sum (means sum of species concentrations), in addition, base run profile were also normalized by their row sum These normalized value from the base run was marked by "*" in the subplot. The resulted normalized data indicates the % contribution of various species to a particular source. The fraction of a species apportioned to a source was estimated by normalizing the MxNx100 matrix with sum of the column (sum of the concentration of a species in all the sources).

The box plots generated by above analysis showed the range of contribution of each species to a source (Figure 4.2) and fraction of species apportioned to the sources at 95% confidence interval (Figure 4.3). The lower whisker indicates 2.5 percentile and upper whisker indicates 97.5 percentile. The red lines within the boxes represent the median values and red plus signs represent the outliers.

For source profile 1 composition analysis (Figure 4.3a) showed high contribution from Mn, and Ca, but associated with large uncertainty as 95% CI as the data range was wide. Unlike sigma based output (detailed in section 4.1) Na was not identified as strong contributor to source profile 1 because it is associated with many outliers (Figure 4.2a). The contribution from NH_4^+ and SO_4^{-2} to source profile 1 was moderate to low with lower uncertainty range at 95% CI. Similarly, Figure 4.3a corroborated the inference from Figure 4.2a, the fraction apportionment of Ca, and Mn to source 1 did not show large variation in Figure 4.3a however, $SO4^{-2}$ and $NH4^+$ showed significantly large variation at 95% CI. Model also indicated that the $SO4^{-2}$ (enclosed in red box in Figure 4.3a) should be interpreted with caution. The above analysis revealed that source profile 1 was primarily influence by Ca and Mn and can be interpreted as crustal soil and was partially contaminated with $SO4^{-2}$ and $NH4^+$. One of the plausible explanation could be the crustal soil particle were enveloped by sulphate aerosol forming shell around the dust core.

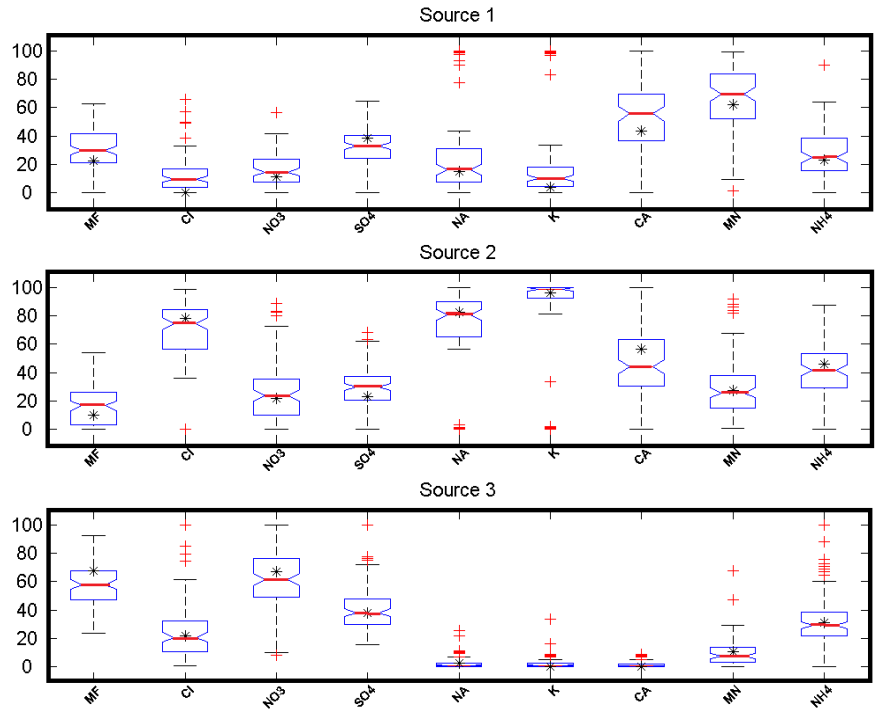

Figure4.2a,b,c BootstrapVariabilitybySpecies

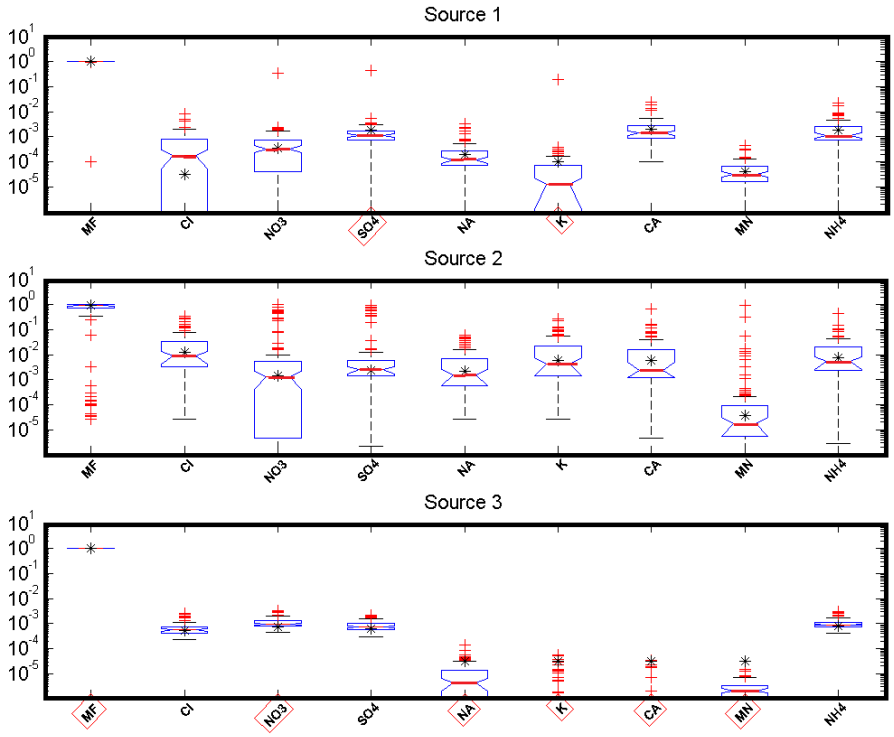

Figure 4.3 a,b,c Bootstrap Variability by Species

Cl, Na and K were the largest contributor to source profile 2 (Figure 4.2 b), however their contribution was associated with large uncertainty (Figure 4.2 b). The Figure 4.3 b showed low uncertainty in the fraction apportionment of Na and K to source profile 2. The presence of Na and K in the atmosphere can possibly be attributed to biomass burning however, further inference can be made from detailed analysis using other options in model such as factor analysis and also using the K/Na ratio for various sources which may contribute to the atmospheric aerosol loading of Delhi region. Figure 4.2c and 4.3c indicated sources profile 3 was significantly influenced by $SO4^{-2}$, $NO3^{-}$ and $NH4^{+}$ with large uncertainty, however, again the fraction apportionment of these species to source profile 3 (Figure 4.3c) indicated low variability. Also the $NO3^{-}$ was recommended by model to be interpreted with caution ($NO3^{-}$ enclosed in red box in Figure 4.3c). The formation of $SO4^{-2}$ and $NO3^{-}$ in the atmosphere takes place by photochemical reaction (Chakroburty and Gupta, 2010, Saolapurkar and Sharma, 2006), and are termed as secondary aerosol.

4.4 Comparison of Varimax Rotated Factor Analysis with UNMIX Result

The UNMIX results can be compared to the varimax rotated factor analysis by selecting the species selection tools, factor analyze selection command. Here the results from varimax rotated factor analysis have been discussed. The analysis showed the loading of species Ca (0.63), Mn (0.85) were high in factor 1/source 1 (Table 4.2) again corroborating the interpretation in previous section that source 1 can be interpreted as crustal soil.

Again the factor 2/source 2 highly loaded with elements K (0.84) and Na (0.90) can be interpreted as biomass burning. The evidence of open agricultural field burning and biomass fuel burning for space heating during winter have been reported in literature (Parmer et al., 2001; Nair et al., 2006, Chowdhury et al., 2007) for peninsular and Delhi region.

Factor 3/source 3 was significantly loaded with $NO3^-$ (0.88), $SO4^{-2}$ (0.84) and $NH4^+$ (0.89) (Table 4.2) indicating the secondary aerosol formation mechanism as a source. The photochemical reactions of SO_2 and NOx in the presence of UV radiation result in ammonium sulphate and ammonium nitrate aerosol formation. Singhai (2010) reported very dense fog in some of the days in December, 2009 and January, 2010, which were associated with high concentration of ions $SO4^{-2}$. The possible reason could be aerosol particles mediate the formation of fog in the atmosphere (Pandis et al., 1990; Seinfeld and Pandis, 1998) through the preferred heterogeneous nucleation of water vapour on pollution particles, at high RH or low supersaturation. Fog droplets further aid aerosol formation through aqueous-phase reactions of soluble gaseous precursors (e.g. SO_2 and H_2O_2), leading to higher aerosol concentrations of species like sulphate on fog abatement, which then nucleate subsequent fog–smog–fog cycles (Pandis et al., 1990). Particulate inorganic matter would also contain soluble inorganic ions (Cl^-, $NO3^-$, $NH4^+$) present in combustion aerosols (e.g. Habib et al., 2008), which could also mediate water uptake and fog formation (Mehta et al., 2009).

Table 4.2 Varimax Rotated Loadings for the selected species

Species	Factor 1	Factor 2	Factor 3
MF	0.02	-0.03	0.88
Cl	-0.28	0.41	0.68
NO3	-0.06	-0.09	0.89
SO4	0.17	-0.06	0.84
NA	-0.11	0.90	0.03
K	0.05	0.84	0.02
CA	0.63	0.62	0.00
MN	0.85	-0.10	0.19
NH4	0.20	0.19	0.88
Eigen Value	1.25	2.07	3.68

4.5 Validation of Sources using Source Signature Reported in Literature

As discussed in previous section the source 1 identified as crustal soil was contaminated with contribution from secondary sulphate aerosol. In order to validate this source the enrichment factor (EF_{crust}) analysis was presented in Table 4.3 The enrichment factor for crustal soil (EF_{crust}) was estimated as the ratio of concentration of element 'E' to reference element 'R' in air and in crustal soil. In this analysis Fe was chosen as reference element. If EF_{Crust} approaches unity, crustal soils are most likely the predominant source of element E. Operationally, if EF_{Crust} value is over 10, the element would have a significant contribution from non-crustal sources (Wang *et al.*, 2006a). The EF_{Crust} values of individual elements in PM1 size fractions for the current study are presented in Table 4.3. The EF_{crust} varied 2-9 for Ca and 4-8 for Mn in different months indicating these elements are predominantly from crustal soil confirming source 1 as crustal soil.

Table 4.3 EF_{Crust} values for different elements.

Species	NOV.09	DEC.09	Jan.10	Feb.10	Mar.10
Na	2	1	2	5	5
K	4	3	4	7	10
Mg	1	1	2	6	3
Ca	2	2	5	9	5
Pb	1805	1588	2491	6793	5119
Zn	156	255	364	758	442
Fe	1	1	1	1	1
Mn	ND	4	6	8	6
Cu	16	100	52	108	133
Cd	ND	2805	3392	4167	1560
Cr	ND	12	28	40	20
Ni	ND	37	52	87	45

The source 2 was validated using the K/Na ratio for different biomass burning reported in literature (Table 4.4). The K/Na ratio varies 0.6 to 2.92 for various type of biomass burned. The ratio estimated in present study also varied from 0.98 to 1.86 lies in the range of values reported in literature indicting the strong influence of biomass burning during winter in Delhi region.

Table 4.4 K/Na ratio for sampling month

Month	No. of samples	K/NA
NOV	6	1.66±0.6
DEC	14	1.65 ±0.5
JAN	7	1.86 ±0.4
FEB	12	0.98 ± 1.2
MAR	12	1.57 ±1.3

Table 4.5 K/Na ratio for different type of sources

K/Na	Types of source	References
0.035	Ocean	Andrea et al, 1998
0.814	Soil	Andrea et al, 1998
2.92	Zambia Sawana	Andrea et al, 1998
1.72	Sugarcane	Andrea et al, 1998
0.60	Subtropical Forest	Andrea et al, 1998

4.6 Model performance analysis

The model predicted element concentrations were correlated with observed concentration (Figure 4.4). The predicted Vs measured PM_1 concentrations plot (r^2=0.77) showed that the UNMIX model worked well and the predicted and measured PM1 mass concentrations are in good agreement. The NO_3^-, K, NH_4^+ $SO4^{-2}$ and Ca showed high correlation with r^2 value greater than 0.7 (Figure 4.6), while Cl, Na, and Mn showed moderate correlation with r^2 varied between 0.52-0.67.

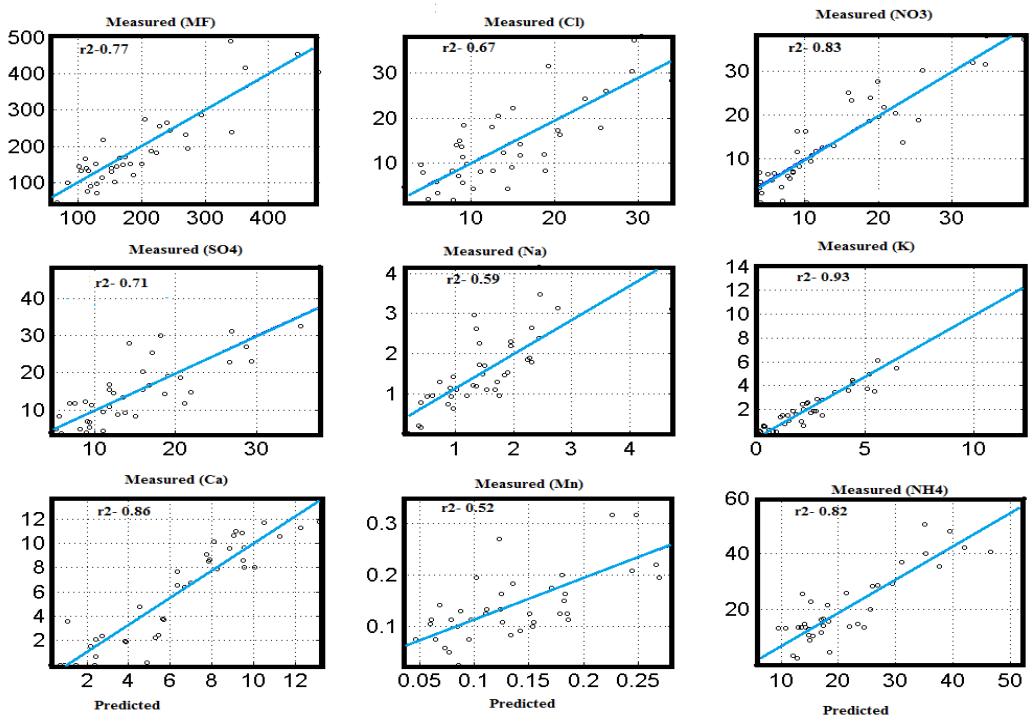

Figure 4.4 Diagnostic Plot under Fit Diagnostic Plot option (Predicted Vs Measured)

4.7 Source quantification:

Three source were identified by UNMIX model version 6.0 of EPA. The secondary aerosol formation is the major contributor to atmospheric aerosol loading over Delhi region followed by crustal soil and biomass burning (Figure 4.5). During winter the photochemical reactions of precursor gases (SO_2 and NO_x) in the presence of UV radiation result in aerosol formation. Singhai (2010) reported high concentration of $SO4^{-2}$ ions in densely foggy days of December, 2009 and January, 2010. Aerosol particles mediate the formation of fog in the atmosphere (Pandis et al., 1990; Seinfeld and Pandis, 1998) through the preferred heterogeneous nucleation of water vapour on pollution particles, at high RH or low super saturation. Fog droplets further aid aerosol formation through aqueous-phase reactions of soluble gaseous precursors (e.g. SO_2 and H_2O_2), leading to higher aerosol concentrations of species like sulfate on fog abatement, which then nucleate subsequent fog–smog–fog cycles (Pandis et al., 1990). Particulate inorganic matter would also contain soluble inorganic ions (Cl^-, $NO3^-$, $NH4^+$) present in combustion aerosols (e.g. Habib et al., 2008), which could also mediate water uptake and fog formation (Mehta et al., 2009) and hence again enhance the sulphate aerosol formation.

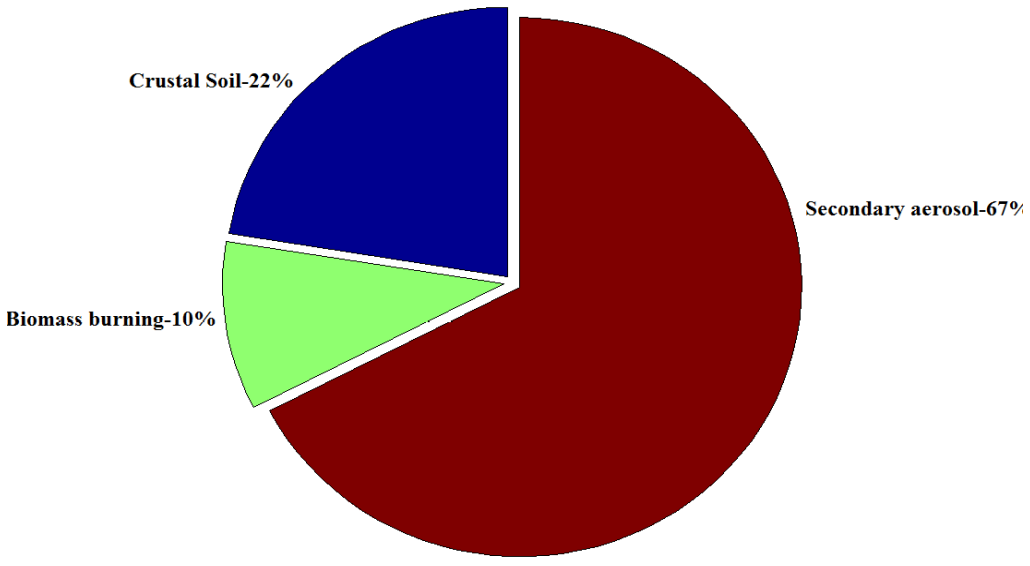

Figure 4.5 Distribution of source contribution for PM_1

The second largest source of aerosol during winter is crustal soil. Singhai (2010) reported dust events during winter attributed to construction activity for common wealth games. Biomass burning the third contributor to aerosol concentration. The evidence of open

agricultural field burning and biomass fuel burning for space heating during winter have been reported in literature (Parmer et al., 2001; Nair et al., 2006, Chowdhury et al., 2007) for peninsular and Delhi region.

CHAPTER-5

5.0. REFERENCE

- Abhishek Chakraborty, Tarun Gupta, 2010. Chemical Characterization and Source Apportionment of Submicron (PM1) Aerosol in Kanpur Region, India, *Aerosol and Air Quality Research*, 10: 433–445.
- A.G., Miguel, G.R., Cass, J., Weiss, M.M. Glovsky, 1996. Environ. Health Perspect. 104 1180-1186.
- A., Chakraborty and T., Gupta, 2009. Chemical Characterization of Submicron Aerosol in Kanpur Region: a Source Apportionment Study, 1:2.
- Arifa Lodhi, Badar Ghauri, M., Rafiq Khan, S., Rahman and Shoaib Shafique, 2009. Particulate Matter ($PM_{2.5}$) Concentration and Source Apportionment in Lahore; J. Braz. Chem. Soc., Vol. 20, No. 10, 1811-1820.
- B.R.T., Simoneit, W.F., Rogge, M.A., Mazurek, L.J., Standley, L.M., Hildemann, G.R., 1993. Cass, Environ. Sci. Technol. 27 2533-2541.
- Chandra Mouli P, Mohan SV, Balaram V, Praveen Kumar M, Reddy SJ. 2006. A study on trace elemental composition of atmospheric aerosols at a semi-arid urban site using ICPMS technique. Atmos Environ;40:136–46.
- Chen, L.-W.A., Doddridge, B.G., Dickerson, R.R., Chow, J.C., Henry, R.C., 2002. Origins of fine aerosol mass in the Baltimore–Washington corridor: implications from observation, factor analysis, and ensemble air parcel back trajectories. Atmospheric Environment 36, 4541–4554.
- Chow, J.C., Watson, J.G., Lu, Z., Lowenthal, D.H., Frazier, C.A., Solomon, P.A., Thuillier, R.H., Magliano, K., 1996. Descriptive analysis of PM2.5 and PM10 at regionally representative locations during SJVAQS/AUSPEX. Atmospheric Environment 30, 2079–5102111.
- de Foy, B., Caetano, E., Magana, V., Zit ´acuaro, A., C´ardenas, B., Mart´ınez, A. P., Retama, A., Ramos, R., Reyes, R., Sosa, G., Molina, L. T., and Molina, M. J., 2005. Mexico City basin wind circulation during the MCMA-2003 field campaign, Atmos. Chem. Phys., 5, 2267–2288.
- Friedlander, S.K., 1973. Chemical element balances and identification of air pollution sources. *Environ. Sci. Technol.* **7:** 235-40.
- Gioda A., Perez U., Rosa Z., Velez BJ., 2006. Concentration of trace elements in

airborne PM10 from Jobos Bay National Estuary, Puerto Rico. Water Air Soil Poll;174:141–59.

- Gordon, G.E., 1988. Receptor models. *Environ. Sci. Technol.* **22:** 1132-1142.
- Habib, G., Venkataraman, C., Bond, T.C., Schauer, J.J., 2008. Climate relevant properties of primary particles from the combustion of biofuels. Environ. Sci.
- Technol. 42 (23), 8829–8834.
- Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometrics and Intelligent Laboratory Systems 65, 179-189.
- Henry, R.C., 1997. History and fundamentals of multivariate air quality receptor models. Chemometrics and Intelligent Laboratory Systems 37, 37–42.
- Henry, R.C., 2000. UNMIX theory and applications. In: Willis, R.D. (Ed.), Final Report of Workshop on UNMIX and PMF as Applied to PM2.5. United States Environment Protection Agency, Washington, EPA/600/A-00/48.
- Henry R.C., et al, 1999.Comparinga new algorithm with the classic methods for estimating he number of factors. Chemom Intell Lab Systems 48: 91–97.
- Henry R.C., 1987 Current factor analysis receptor models are illposed,Atmos. Environ. 21 1815– 1820.
- Henry R.C., 2005 Duality in multivariate receptor models. Chemom Intell Lab Systems 77: 59–63.
- Henry R.C., Park E.S., Spiegelman C.H.,1999. Comparing a new algorithm with the classic methods for estimating the number of factors, Chemom. Intell. Lab. Syst. 48 91–97.
- Hidy, G.M., and C., Venkataraman, 1996. The chemical mass balance method for Estimating atmospheric particle sources in Southern California. *Chem. Eng. Comm.* **151:** 187-209.
- Hopke, P. K., 2003. "Recent developments in receptor modeling." Journal of Chemometrics **17**(5): 255-265.
- Hopke, P.K., Ito, K., Mar, T., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Liu, H., Neas, L., Pinto, J., Sto¨lzel, M., Suh, H., Paatero, P., & Thurston, G.D. (2006). PM source apportionment and health effects: 1. Intercomparison of source apportionment results, Journal of Exposure Science and Environmental Epidemiology, 16, 275–286.

- Huang, S. L., K. A. Rahn and R. Arimoto, 1999. "Testing and optimizing two factoranalysis techniques on aerosol at Narragansett, Rhode Island." Atmospheric Environment **33**(14): 2169-2185.
- J. Air Waste Manage. Assoc. 51 (2001) 489; Atmos. Environ. 36 (2002) 1911.
- Jim J., Lin*, Lien-Chih Lee, 2004. Characterization of the concentration and distribution of urban submicron (PM1) aerosol particles, 38 (2004) 469–475.
- Kang, B.W., 1998. Source Apportionment of PM2.5 using a CMB Receptor Model—in Chongju Area. Ph.D. dissertation, Kon-kuk University, Seoul, Korea, 1998.
- Kulkarni, S., 2004. Surface elemental composition of aerosols at Beijing (China), Gosan (Korea), and Tango (Japan) during ACE-Asia. M.S. Thesis, Chemical and Biochemical Engineering. Iowa City, The University of Iowa**:** xiv, 183 leaves, bound.
- L.J. Standley, B.R.T., Simoneit, 1990. Atmos. Environ. 24B 67-73.
- L.M., Hildemann, G.R., Markowski, G.R., Cass, 1991. Environ. Sci. Technol. 25 744-759.
- Malm, W.C., Sisler, J.F., Huffman, D., Eldred, R.A., Cahill, R.A., 1994. Spatial and seasonal trends in particle concentration and optical extinction in the United States. Journal of Geophysical Research 99 (D1), 1347–1370.
- Ma, C.-J., S., Tohno, M. Kasahara and S. Hayakawa, 2004. "Properties of individual Asian dust storm particles collected at Kosan, Korea during ACE-Asia." Atmospheric Environment **38**(8): 1133-1143.
- Michel J., Kleeman et.al.,2009. Source Apportionment of Fine (PM1.8) and Ultrafine (PM0.1) Airborne Particulate Matter during a Severe Winter Pollution Episode; *Environ. Sci. Technol., 43,* 272–279.
- Miranda, J., Andrade, E., L´opez-Su´arez, A., Ledesma, R., Cahill, T. A., andWakabayashi, P. H., 1996. A receptor model for atmospheric aerosols from a southwestern site in Mexico City, Atmos. Environ., 30, 3471–3479,.
- M. O.,Andrea, T., W., Andrea, H., Annega, J., Beer, H., Cachie, P le Canut, W., Elbert, W., Maenhaut, I., Salma, F., G., Wienhold, T., Zenke,1998. Air borne studies of aerosol emissions from savanna fires in southern Africa: 2. Aerosol chemical composition, journal of geophysical research, vol. 103, no. D24, pages 32,119-32,128, december 27.
- Mohanraj, R., Azeez PA., Priscilla, T., 2004. Heavy metals in airborne particulate matter of urban Coimbatore. Arch Environ Contam Toxicol; 47:162–7.

- M., Viana, T.A.J., Kuhlbusch, X., Querol, A., Alastuey, R.M., Harrison, P.K., Hopke, et.al., 2008. Source apportionment of particulate matter in Europe: A review of methods and results Aerosol Science 39, 827 –849.
- Nair, P.R., George, S.K., Sunilkumar, S.V., Parameswaran, K., Jocab, S., Abraham, 2006. A.: Chemical composition of aerosols over peninsular India during winter. Atmos. Environ. 40, 6477–6493.
- Pandis, S.N., Seinfeld, J., Pilinis, C., 1990. The smog–fog–smog cycle and acid deposition. J. Geophys. Res. 95 (D11), 18489–18500.
- Park, E.S., Henry, R.C. and Spiegelman, C.H. 1999. *Determining the Number of Major Pollution Sources in Multivariate air Quality Receptor Models*, NRCSE, TSR No.34.
- Park E.S., et al., 2000. Estimatingt he number of factors to include in a high-dimensional multivariate bilinear model. Commun Stat Simul Comput: 29: 723–746.
- Pitts, B.J., Pitts Jr., J.N., 2001. Atmospheric chemistry, fundamentals and experimental techniques. Wiley, New York (1986) Parmar, R.S., Satsangi, G.S., Kumari, M., Lakhani, A., Srivastav, S.S., Prakash, S.: Study of size distribution of atmospheric aerosol at Agra. Atmos. Environ. 35, 693–702.
- Pueschel, R.F., VanValin, C.C., Castillo, R.C., Kandlech, R.C., and Ganor, E., 1986, Aerosols in polluted versus non polluted air masses: long range transport and effect on clouds, Journal of Climate and Applied Meteorology, 25, pp 1908-1917.
- Riad Ladji, Noureddine Yassaa, Catia Balducci, Angelo Cecinato, Brahim Youcef Meklati, 2009. Distribution of the solvent-extractable organic compounds in fine (PM1) and coarse (PM1–10) particles in urban, industrial and forest atmospheres of Northern Algeria, 408, 415–424.
- Rosa Caggiano, Maria Macchiato, Serena Trippetta, Levels, 2010. chemical composition and sources of fine aerosol particles (PM1) in an area of the Mediterranean basin, 408, 884–895.
- Saolapurkar, T., and Sharma, M., 2006. *Characterization of PM10 and Formation of Secondary Particles: A Source Apportionment Study*, *M.Tech thesis*, IIT Kanpur.
- Suresh Tiwari1, Manoj K., Srivastava2, and Deewan S., Bisht1, 2008. Chemical Characteristics of Water Soluble Components of Fine Particulate Matter, PM2.5, at Delhi, India, Journal Earth Science India, Vol. I (III), pp. 72-86.

- Tolocka, M.P., Solomon, P.A., Mitchell, W., Norris, G.A.,Gemmill, D.B., Wiener, R.W., Vanderpool, R.W., Homolya, J.B., Rice, J., 2001. East versus west in the US: chemicalcharacteristics of PM2.5 during the winter of 1999. Aerosol Science Technology 34, 88–96.
- U. C., Kulshrestha, Monika Jain, R., Sekar, M., Vairamani, A. K., Sarkar, 1999. Chemical characteristics and source apportionment of aerosols over Indian Ocean during INDOEX-1999and D. C. Parashar.
- US-EPA, 1999. SPECIATE: EPA's repository of total organic compound and particulate matter speciated profiles for a variety of sources for use in source apportionment studies. U.S. Environmental Protection Agency OAQPS, Research Triangle Park, NC, 1999< /http://www.epa.gov/ttn/chief/software/speciate>.
- Yu Song, Shaodong xie, Yuanhang Zhang, Limin Zeng, Lynn G salmon, Mei Zheng, 2006.source apportionment of $PM_{2.5}$ in Beijing using principle component analysis/absolute principle component score and UNMIX; 372, 278-286.
- Wang, Y., Guoshun, Z., Xingying, Z., Kan, H., Chang, X., Aohan, Jianmin, C., and Zhisheng, A., 2006a. The Ion Chemistry, Seasonal Cycle, and Sources of PM2.5 and TSP Aerosol in Shanghai. *Atmos. Environ.* 40: 2935–2952.
- Watson, J.G., 1984. Overview of receptor model principles. *JAPCA* **34:** 619-23. Watson, J.G., Cooper, J.A. and J.J. Huntzicker, 1984. The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos. Environ* **18:** 1347-55.
- Watson, J.G., N.F. Robinson, J.C. Chow, R.C. Henry, B.M. Kim, T.G. Pace, E.L. Meyer And Q. Nguyen, 1990. The USEPA/DRI chemical mass balance receptor model, CMB 7.0. *Environ. Software* **5:** 38-49.
- W.F., Rogge, L.M., Hildemann, M.A., Mazurek, G.R., Cass, B.R.T., Simoneit, 1998. Environ. Sci. Technol. 32, 13-22.
- W.F., Rogge, L.M., Hildemann, M.A., Mazurek, G.R., Cass, B.R.T., Simoneit, 1993. Environ. Sci. Technol. 27, 1892-1904.
- Zhenxing Shen, Junji Cao, Zhi Tong, Suixin Liu, Lingala Siva Sankara Reddy, Yuemei Han, Ting Zhang, Juan Zhou, 2009. Chemical Characteristics of Submicron Particles in Winter in Xi'an, Vol. 9, No. 1, pp. 80-93.
- Zohir Chowdhury, Mei Zheng, James J., Schauer, Rebecca J., Sheesley, Lynn G., Salmon, Glen R., Cass, and Armistead G., Russell, 2007. Speciation of ambient fine

organic carbon particles and source apportionment of PM2.5 in Indian cities. Journal of geophysical research, vol. 112, D15303, doi:10.1029/2007JD008386.

APPENDIX I - Input data file

Date	MF	Cl	NO3	NA	K	MG	CA	PB	ZN	FE	MN	CU	CD	CR	NI	NH4
11-18-09	200.2	11.4	13.5	4.6	4.4	0.5	4.4	2.0	0.6	2.5	0.0	0.0	0.0	0.0	0.0	-99.0
11-19-09	135.4	11.2	8.6	2.2	3.8	0.7	5.3	1.9	0.5	5.8	0.0	0.0	0.0	0.0	0.0	-99.0
11-20-09	164.1	22.7	9.5	0.9	1.6	0.2	4.7	0.1	0.2	1.7	0.0	0.0	0.0	0.0	0.1	-99.0
11-25-09	185.7	19.1	11.8	3.4	6.7	1.1	10.3	2.5	1.1	7.0	0.0	0.1	0.0	0.0	0.0	-99.0
11-26-09	123.9	15.3	7.1	4.7	5.1	0.8	5.4	1.0	0.9	3.4	0.0	0.1	0.0	0.0	0.0	-99.0
11-27-09	208.7	16.6	12.2	2.8	6.9	1.0	8.4	1.7	1.2	12.1	0.0	0.1	0.0	0.0	0.0	-99.0
12-02-09	258.0	11.8	29.9	0.8	1.7	1.3	2.6	0.7	0.3	4.0	0.1	0.4	0.0	0.0	0.1	-99.0
12-03-09	275.3	14.7	31.6	0.8	1.6	1.6	2.1	0.8	1.4	2.1	0.2	0.4	0.0	0.0	0.2	-99.0
12-04-09	252.4	13.0	27.5	0.8	1.6	1.2	2.3	1.0	0.4	2.0	0.1	0.4	0.0	0.1	0.1	-99.0
12-09-09	443.6	11.8	14.8	0.7	1.5	1.2	1.9	1.1	1.1	3.9	0.2	0.4	0.0	0.1	0.1	-99.0
12-10-09	451.7	13.8	27.3	0.8	1.5	1.3	1.8	0.9	0.7	2.0	0.1	0.3	0.0	0.1	0.1	-99.0
12-11-09	265.9	11.7	19.6	0.6	1.5	1.2	1.9	1.2	0.8	3.0	0.2	0.3	0.0	0.0	0.1	28.3
12-14-09	244.7	11.9	23.4	2.6	4.2	1.5	9.6	0.8	0.8	4.7	0.2	0.2	0.0	0.1	0.1	28.8
12-15-09	193.9	24.3	21.9	3.5	3.7	2.1	8.7	0.6	1.0	3.3	0.1	0.2	0.0	0.0	0.1	36.9
12-16-09	288.0	38.1	18.8	3.1	5.5	1.7	2.4	0.8	0.9	3.9	0.2	0.2	0.0	0.1	0.2	40.2
12-23-09	169.3	18.1	12.9	0.7	1.1	1.1	2.4	1.3	0.6	2.9	0.1	0.2	0.0	0.0	0.2	15.7
12-24-09	233.1	31.4	20.3	1.2	1.0	1.2	4.7	0.3	1.0	4.0	0.2	0.0	0.0	0.1	0.1	29.2
12-25-09	239.5	37.0	31.9	1.7	1.5	1.0	3.6	1.1	0.6	2.5	0.0	0.0	0.0	0.1	0.1	48.2
12-30-09	504.8	30.3	31.6	1.5	1.7	1.2	2.1	1.1	0.9	1.9	0.1	0.3	0.0	0.0	0.2	42.3
12-31-09	453.8	26.0	38.2	0.9	1.8	0.6	8.0	1.3	0.8	4.1	0.4	0.1	0.0	0.0	0.1	40.7
01-01-10	403.6	28.2	37.3	1.8	3.5	0.3	7.9	1.3	0.8	4.6	0.2	0.1	0.0	0.0	0.2	60.1
01-06-10	488.9	8.4	13.8	1.5	2.5	0.7	10.5	1.2	0.6	3.4	0.2	0.1	0.0	0.1	0.2	50.6
01-07-10	415.4	16.4	30.3	1.5	2.0	1.5	11.7	1.5	0.9	1.7	0.2	0.1	0.0	0.2	0.1	35.5
01-08-10	274.7	8.3	12.5	1.3	2.8	1.7	11.0	1.5	1.0	0.9	0.1	0.1	0.0	0.1	0.1	25.8

1/27/2010	151.9	11.3	0.0	1.2	1.5	0.8	9.6	0.6	1.0	1.7	0.1	0.2	0.0	0.1	0.0	21.5
1/28/2010	130.7	8.5	0.4	1.1	2.6	1.6	8.6	0.4	1.1	2.3	0.2	0.1	0.0	0.0	0.1	22.8
1/29/2010	104.0	8.0	0.0	1.1	2.5	1.7	11.8	0.4	1.1	1.6	0.3	0.2	0.0	0.2	0.1	14.9
03-02-10	172.0	7.2	16.2	0.9	1.5	1.2	8.5	2.3	1.0	3.6	0.2	0.1	0.0	0.1	0.1	14.2
04-02-10	122.0	2.2	16.2	1.1	0.5	1.0	8.0	2.4	0.9	3.3	0.3	0.1	0.0	0.2	0.1	16.2
05-02-10	114.0	3.5	11.5	1.0	0.5	1.1	0.2	2.5	1.6	2.3	0.3	0.1	0.0	0.1	0.2	13.6
10-02-10	255.7	9.1	18.7	0.9	1.4	0.5	1.8	2.3	1.6	2.0	0.1	0.1	0.0	0.1	0.1	13.4
11-02-10	217.4	5.7	11.7	0.9	0.2	0.3	1.5	1.6	0.8	1.2	0.1	0.1	0.0	0.1	0.1	13.0
12-02-10	187.9	4.5	27.8	0.8	0.2	1.1	0.0	1.7	1.2	2.3	0.1	0.1	0.0	0.1	0.1	14.8
2/17/2010	152.0	12.3	24.0	0.2	0.1	1.2	0.0	1.7	0.6	1.3	0.1	0.1	0.0	0.1	0.1	13.8
2/18/2010	134.7	6.0	5.5	0.2	0.6	1.1	0.0	1.3	1.7	1.0	0.1	0.1	0.0	0.1	0.1	13.3
2/19/2010	133.0	4.8	7.7	1.4	0.1	6.6	10.6	1.5	0.7	1.0	0.2	0.1	0.0	0.1	0.1	0.0
2-24-2010	150.8	14.0	10.7	1.1	0.1	1.9	7.6	1.3	0.8	1.4	0.1	0.1	0.0	0.0	0.1	11.7
2-25-2010	184.1	17.2	25.2	2.4	4.9	2.5	10.1	1.5	0.7	1.0	0.1	0.1	0.0	0.1	0.2	20.1
2-26-2010	140.6	14.9	6.9	2.6	0.6	1.7	9.1	0.9	0.6	0.3	0.1	0.1	0.0	0.1	0.1	10.5
03-03-10	152.5	4.5	7.1	3.0	1.8	1.5	2.2	1.9	0.6	1.0	0.1	0.1	0.0	0.1	0.1	13.5
03-04-10	101.8	2.0	4.6	2.3	2.8	1.7	3.7	1.8	1.0	1.4	0.1	0.1	0.0	0.0	0.1	13.2
05-03-10	73.2	20.5	6.5	4.2	3.5	1.6	10.8	1.5	1.2	1.2	0.1	0.1	0.0	0.1	0.0	9.0
3-10-10	90.4	18.4	3.4	1.7	1.9	2.3	6.6	0.8	0.5	1.0	0.1	0.1	0.0	0.1	0.1	13.4
3-11-10	97.6	9.8	3.3	1.9	4.3	2.0	13.7	1.4	0.5	1.1	0.1	0.1	0.0	0.0	0.1	25.7
12-03-10	167.3	17.9	6.9	3.1	14.0	1.7	11.3	1.2	0.9	1.3	0.1	0.1	0.0	0.1	0.0	16.5
3-17-2010	162.1	1.5	2.0	0.0	0.0	0.0	0.0	1.5	0.5	5.5	0.1	0.1	0.0	0.1	0.0	0.0
3-18-2010	146.4	22.1	9.3	2.3	3.6	1.2	6.4	1.6	0.5	1.1	0.1	0.6	0.0	0.0	0.0	4.4
3-19-2010	140.1	14.2	6.1	1.8	6.1	2.8	3.8	0.5	0.5	1.2	0.1	0.1	0.0	0.1	0.1	10.8
3-24-2010	47.7	9.6	5.0	0.0	0.0	0.0	0.0	0.8	0.5	0.9	0.1	0.1	0.0	0.0	0.0	0.0
3-25-2010	78.2	13.7	8.3	1.3	0.7	0.8	0.7	1.4	0.3	1.0	0.1	0.1	0.0	0.0	0.0	2.3
3-26-2010	145.4	8.2	6.4	2.2	4.4	1.0	6.8	1.6	0.8	1.2	0.1	0.0	0.0	0.0	0.1	3.2

APPENDIX II -Result of Model Run

Variamax Rotated Principal Components(PC) with Total Variable loading > 0.01

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
MF	0.97	-0.13	0.12	0.10	0.08	-0.04	0.03	0.02	-0.01	-0.00
Cl	0.36	0.04	-0.03	0.84	0.07	0.21	-0.03	0.12	-0.29	0.10
NO3	0.70	0.03	0.04	0.33	0.17	-0.01	-0.08	-0.06	0.17	0.57
SO4	0.68	0.03	0.06	0.11	0.71	-0.04	0.07	-0.02	-0.03	0.09
NA	-0.01	0.02	-0.10	0.17	-0.02	0.89	0.24	0.33	-0.04	-0.00
K	0.02	0.01	-0.01	0.08	-0.01	0.28	0.19	0.93	-0.10	-0.02
CA	0.06	0.02	0.18	-0.03	0.03	0.21	0.93	0.19	-0.10	-0.02
PB	-0.03	-0.02	0.02	-0.17	-0.01	-0.04	-0.09	-0.09	0.97	0.04
MN	0.16	0.02	0.97	-0.01	0.03	-0.07	0.15	-0.01	0.02	0.01
NH4	0.82	0.42	0.15	0.25	0.13	0.10	0.11	0.07	-0.17	0.06

Species with Largest Loading in Each Component > 0.5

1 NH4 NO3

3 MN

4 Cl

5 SO4

6 NA

7 CA

8 K

9 PB

10 NO3

****** Run # 1 ******

05-Apr-2011 07:05:37

File: C:\Users\hp\Desktop\LUCKY.txt

Tracer: None

TOTAL: MF

Normalization: MF

9 Species, 40 Obs., 3 Sources,

Min r^2 = 0.88, Min Sig/Noise= 2.88

Unmix Source Composition

Species	**Source 1**	**Source 2**	**Source 3**
MF	44.3	20	133

Cl -0.00808 0.249 0.0683

NO3 0.0161 0.031 0.0956

SO4 0.0807 0.0491 0.0802

NA 0.0084 0.0463 0.00146

K 0.00466 0.121 -0.00194

CA 0.0905 0.119 -0.00178

MN 0.00181 0.00079 0.00031

NH4 0.0801 0.162 0.109

ELAPSED TIME = 0 MINUTES 0.1 SECONDS

Fitting remaining species

Species Constant Source 1 Source 2 Source 3 r^2 # Obs

8 PB 1.2825 0.0007 0.0000 0.0000 0.0024 40

Requested Output from run number: 1

1 NOTE: For run number 1, 140 runs were necessary to generate 100 runs with feasible results.

1 Block Length: 5

1 Base-Bootstrap Data Sets Correlation: LOW

1 Number of rejected bootstrap data sets: 0

1 **Source Composition Variability from Run # 1**

1 Species Source #1 Source #2 Source #3

1 MF 30.5 19.6 34.4

1 Cl 0.0728 0.904 0.0244

1 NO3 0.0311 0.376 0.0152

1 SO4 0.0368 0.245 0.0147

1 NA 0.0157 0.16 0.00212

1 K 0.0115 0.557 0.0033

1 CA 0.0834 0.3 0.00917

1 MN 0.00141 0.00419 0.000219

1 NH4 0.0771 0.444 0.0256

1 **Variability Details - Parametric Version**

1 Source #1

1 Species Composition 1 sigma Comp/2*sigma

1 MF 44.3 30.5 0.727

1 Cl -0.00808 0.0728 -0.0555

1 NO3 0.0161 0.0311 0.259

1	SO4	0.0807	0.0368	1.1
1	NA	0.0084	0.0157	0.267
1	K	0.00466	0.0115	0.203
1	CA	0.0905	0.0834	0.543
1	MN	0.00181	0.00141	0.645
1	NH4	0.0801	0.0771	0.519

1 Source #2

	Species	**Composition**	**1 sigma**	**Comp/2*sigma**
1	MF	20	19.6	0.51
1	Cl	0.249	0.904	0.138
1	NO3	0.031	0.376	0.0412
1	SO4	0.0491	0.245	0.1
1	NA	0.0463	0.16	0.145
1	K	0.121	0.557	0.109
1	CA	0.119	0.3	0.198
1	MN	0.00079	0.00419	0.0942
1	NH4	0.162	0.444	0.183

1 Source #3

1	Species	Composition	1 sigma	Comp/2*sigma
1	MF	133	34.4	1.93
1	Cl	0.0683	0.0244	1.4
1	NO3	0.0956	0.0152	3.14
1	SO4	0.0802	0.0147	2.73
1	NA	0.00146	0.00212	0.344
1	K	-0.00194	0.0033	-0.295
1	CA	-0.00178	0.00917	-0.097
1	MN	0.00031	0.000219	0.707
1	NH4	0.109	0.0256	2.13

1 **********

1 Variability Details - Percentile Version

1 **********

1 **Source # 1 Statistics**

		Bootstrap Percentiles						
1	**Species Profile**	**2.5th**	**5th**	**25th**	**50th**	**75th**	**95th**	**97.5th**
1	MF	44.3227	14.7401	20.2420	37.4742	58.3248	82.9797	119.8451	124.4309
1	Cl	-0.0081	-0.0589	-0.0558	-0.0121	0.0128	0.0371	0.0917	0.1051
1	NO3	0.0161	-0.0407	-0.0356	0.0023	0.0222	0.0427	0.0641	0.0682
1	SO4	0.0807	-0.0225	-0.0120	0.0567	0.0734	0.0888	0.1221	0.1362
1	NA	0.0084	0.0004	0.0019	0.0051	0.0079	0.0118	0.0261	0.0415
1	K	0.0047	-0.0288	-0.0192	-0.0045	0.0008	0.0054	0.0140	0.0170
1	CA	0.0905	0.0450	0.0484	0.0667	0.0857	0.1151	0.2045	0.3106
1	MN	0.0018	0.0006	0.0008	0.0012	0.0019	0.0024	0.0045	0.0056
1	NH4	0.0801	-0.0223	0.0074	0.0599	0.0804	0.1245	0.2060	0.2618

1 Source # 2 Statistics

		Bootstrap Percentiles						
1	**Species Profile**	**2.5th**	**5th**	**25th**	**50th**	**75th**	**95th**	**97.5th**
1	MF	19.9925	-18.3087	-13.7373	5.0757	13.9623	28.0009	50.1897	52.6868
1	Cl	0.2485	-2.2440	-1.1060	0.1346	0.2177	0.3840	1.3973	2.0949

1 NO3	0.0310	-0.7786	-0.3464	-0.0103	0.0338	0.0794	0.3357	0.6049
1 SO4	0.0491	-1.1546	-0.2329	0.0334	0.0669	0.0931	0.1933	0.2336
1 NA	0.0463	-0.3427	-0.2419	0.0221	0.0352	0.0774	0.2543	0.3511
1 K	0.1215	-0.9301	-0.6205	0.0550	0.1030	0.2515	0.7503	0.9775
1 CA	0.1188	-0.4894	-0.2164	0.0352	0.0665	0.1683	0.4675	0.6550
1 MN	0.0008	-0.0066	-0.0032	-0.0000	0.0006	0.0011	0.0076	0.0108
1 NH4	0.1621	-1.3554	-0.3508	0.0865	0.1359	0.2145	0.6125	0.7768

1 Source # 3 Statistics

Bootstrap Percentiles

1 Species Profile	**2.5th**	**5th**	**25th**	**50th**	**75th**	**95th**	**97.5th**
1 MF 133.0797	49.4678	54.3757	89.2775	109.6844	128.2690	176.8516	185.5010
1 Cl 0.0683	0.0228	0.0305	0.0514	0.0625	0.0799	0.1067	0.1158
1 NO3 0.0956	0.0838	0.0871	0.0984	0.1052	0.1166	0.1395	0.1511
1 SO4 0.0802	0.0585	0.0651	0.0748	0.0838	0.0961	0.1134	0.1161
1 NA 0.0015	-0.0029	-0.0028	-0.0012	0.0005	0.0013	0.0037	0.0041
1 K -0.0019	-0.0103	-0.0094	-0.0043	-0.0022	-0.0008	0.0022	0.0028
1 CA -0.0018	-0.0313	-0.0290	-0.0162	-0.0101	-0.0034	0.0011	0.0021

1	MN	0.0003	-0.0003	-0.0001	0.0001	0.0003	0.0004	0.0005	0.0005
1	NH4	0.1092	0.0471	0.0509	0.0902	0.1045	0.1195	0.1345	0.1456

1 **

1 Variability Details - Discrete Difference Percentile Version

1 **

1 **Source # 1 Statistics**

			Confidence Levels (%)					
1	**Species**	**Profile**	**90%(Tot.)**	**95%(Tot.)**	**Percent(Rel.)**	**90%(Rel.)**	**95%(Rel.)**	
1	MF	44.3227	140	170	22	31	38	
1	Cl	-0.0081	0	0	0	0	0	
1	NO3	0.0161	315	337	5	16	17	
1	SO4	0.0807	67	115	23	15	26	
1	NA	0.0084	154	292	25	37	70	
1	K	0.0047	0	0	9	32	50	
1	CA	0.0905	75	145	65	48	87	
1	MN	0.0018	109	174	58	63	99	
1	NH4	0.0801	118	177	17	18	27	

1 Source # 2 Statistics

			Confidence Levels (%)				
1	**Species**	**Profile**	**90%(Tot.)**	**95%(Tot.)**	**Percent(Rel.)**	**90%(Rel.)**	**95%(Rel.)**
1	MF	19.9925	156	187	10	15	18
1	Cl	0.2485	545	843	36	194	294
1	NO3	0.0310	0	0	4	52	87
1	SO4	0.0491	377	579	6	24	35
1	NA	0.0463	537	778	62	332	463
1	K	0.1215	543	833	102	547	818
1	CA	0.1188	294	482	39	100	174
1	MN	0.0008	0	0	11	73	128
1	NH4	0.1621	280	673	15	42	77

1 Source # 3 Statistics

			Confidence Levels (%)				
1	**Species**	**Profile**	**90%(Tot.)**	**95%(Tot.)**	**Percent(Rel.)**	**90%(Rel.)**	**95%(Rel.)**
1	MF	133.0797	51	59	67	34	39
1	Cl	0.0683	55	68	66	36	44
1	NO3	0.0956	33	46	91	30	39
1	SO4	0.0802	33	41	70	23	29
1	NA	0.0015	0	0	13	34	39
1	K	-0.0019	0	0	0	0	0
1	CA	-0.0018	0	0	0	0	0
1	MN	0.0003	114	150	30	34	44
1	NH4	0.1092	42	54	68	28	36

1 Zero entries in the confidence levels columns indicate that the 25th percentile

1 values of the bootstrap source profile are negative and hence likely not significant.

1 Tot. - Total Variation

1 Rel. - Relative Variation

Requested Output from run number: 1

1 **Source Composition Uncertainties and Stability Factors from Run # 1**

05-Apr-2011 07:07:34

1 **Source #1**

1 Species	**Comp.**	**sigma(#1)**	**sigma(#2)**	**sigma(#3)**	**sigma(#4)**	**sigma(#5)**	**Mean**	**SD**	**CV(%)**
1 Cl	-0.0081	0.0618	0.0502	0.0860	0.0875	0.0809	0.0733	0.0165	22.46
1 NO3	0.0161	0.0550	0.0455	0.0434	0.0505	0.0483	0.0486	0.0045	9.25
1 SO4	0.0807	0.0402	0.0355	0.0364	0.0369	0.0358	0.0370	0.0019	5.13
1 NA	0.0084	0.0124	0.0103	0.0171	0.0169	0.0154	0.0144	0.0030	20.61
1 K	0.0047	0.0139	0.0112	0.0179	0.0173	0.0162	0.0153	0.0027	17.80
1 CA	0.0905	0.0956	0.0754	0.0935	0.1037	0.0945	0.0925	0.0104	11.21
1 MN	0.0018	0.0020	0.0016	0.0017	0.0019	0.0017	0.0018	0.0002	9.20
1 NH4	0.0801	0.1054	0.0840	0.0887	0.1351	0.1236	0.1074	0.0220	20.46

1 Source #2

1 Species	**Comp.**	**sigma(#1)**	**sigma(#2)**	**sigma(#3)**	**sigma(#4)**	**sigma(#5)**	**Mean**	**SD**	**CV(%)**
1 Cl	0.2485	2.7742	2.1063	10.7413	9.3187	8.4370	6.6755	3.9597	59.31
1 NO3	0.0310	0.5220	0.5830	3.2798	2.8467	2.5670	1.9597	1.3096	66.82
1 SO4	0.0491	0.4063	0.3343	0.8151	0.7166	0.6699	0.5885	0.2075	35.26

1	NA	0.0463	0.5712	0.4284	1.3533	1.1765	1.0828	0.9224	0.4011	43.47
1	K	0.1215	2.3040	1.6567	3.7938	3.3017	3.0501	2.8213	0.8445	29.93
1	CA	0.1188	0.7621	0.6603	1.3903	1.2242	1.2855	1.0645	0.3299	30.98
1	MN	0.0008	0.0122	0.0102	0.0165	0.0145	0.0139	0.0135	0.0024	17.71
1	NH4	0.1621	1.0355	0.8741	3.4121	2.9627	2.6828	2.1935	1.1617	52.96

1 Source #3

1	**Species**	**Comp.**	**sigma(#1)**	**sigma(#2)**	**sigma(#3)**	**sigma(#4)**	**sigma(#5)**	**Mean**	**SD**	**CV(%)**
1	Cl	0.0683	0.0233	0.0230	0.0244	0.0232	0.0230	0.0234	0.0006	2.57
1	NO3	0.0956	0.0152	0.0190	0.0201	0.0193	0.0190	0.0185	0.0019	10.43
1	SO4	0.0802	0.0142	0.0150	0.0151	0.0148	0.0151	0.0149	0.0004	2.51
1	NA	0.0015	0.0017	0.0019	0.0021	0.0020	0.0020	0.0019	0.0001	6.56
1	K	-0.0019	0.0033	0.0031	0.0033	0.0034	0.0035	0.0033	0.0001	4.43
1	CA	-0.0018	0.0093	0.0096	0.0095	0.0093	0.0096	0.0095	0.0002	1.61
1	MN	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0000	4.68
1	NH4	0.1092	0.0240	0.0251	0.0245	0.0233	0.0233	0.0240	0.0008	3.31

Requested Output from run number: 1

1 Fitting remaining species

1 **Species Constant Source 1 Source 2 Source 3 r^2 # Obs**

1 8 PB 1.2825 0.0007 0.0000 0.0000 .0024 40

1 **********

1 Correlation, differences, and regression coefficients related to run number 1

1 predicted(model) and measured(input) values of selected species

1 Number of Sources: 3

1 **Species**	**r-Pearson**	**Mean diff**	**RMSE**	**Slope**	**Intercept**	**r^2**	**Outliers (Out of 40)**
1 Cl	0.82	0.00	5.58	0.94	0.87	0.67	0
1 NO3	0.91	-0.04	4.39	0.99	0.14	0.83	0
1 SO4	0.84	0.08	5.36	1.02	-0.35	0.71	0
1 NA	0.77	-0.08	0.62	0.85	0.30	0.59	0
1 K	0.96	0.31	0.67	1.03	-0.39	0.93	0
1 CA	0.93	0.44	1.58	1.12	-1.20	0.86	0
1 MN	0.72	-0.01	0.05	0.80	0.03	0.52	0
1 NH4	0.91	1.12	6.28	1.20	-5.38	0.82	0

1 Total Variable:

1 MF 0.88 -0.84 55.68 0.98 4.10 0.77 0

1 Average species r-Pearson value (excluding Total variable): 0.74

1 **********

1 **Source Composition Variability Details**

1 Block Length: 5 (Base-Bootstrap Data Sets Correlation: LOW)

1 Note: For this run, 145 157 153 153 156 runs were necessary to generate 100 runs with feasible results.

1 **********

1 Negative Bias in the chosen run

1 None found to be significant

1 **********

Significant/Strong Species in Sources (sigma-based)

Strong Species: Source Composition >= 1*sigma

Significant Species: Source Composition >= 2*sigma

Source 1

Strong - CA, MN, NH4

Significant - SO4

Source 2

Strong - None

Significant - None

Source 3

Strong - None

Significant - Cl, NO3, SO4, NH4

1 **********

1 Species Report (percentiles based)

1 **********

1 Legend:

0 = Source profile NOT within IQR

1 = Source profile within IQR but not centered

2 = Source profile within IQR and centered

+ = 2.5th percentile value of bootstrap source profiles > 0.00001

1 Species	**Source 1**	**Source 2**	**Source 3**
1 MF	1+	1	0+
1 Cl	1	1	1+
1 NO3	1	1	0+

1 SO4 0 1 2+

1 NA 1+ 1 0

1 K 0 1 0

1 CA 1+ 1 0

1 MN 1+ 1 0

1 NH4 1 1 1+

1 **********

Requested Output from run number: 1

1 Correlation matrix of sources for run number 1

1 Sources	**#1**	**#2**	**#3**
1 #1	1.000	-0.267	-0.193
1 #2	-0.267	1.000	-0.057
1 #3	-0.193	-0.057	1.000

Requested Output from run number: 1

Factor loading

Varimax Rotated Loadings for the selected species

Species	Factor 1	Factor 2	Factor 3
MF	0.88	-0.03	0.02
Cl	0.68	0.41	-0.28
NO3	0.89	-0.09	-0.06
SO4	0.84	-0.06	0.17
NA	0.03	0.90	-0.11
K	0.02	0.84	0.05
CA	0.00	0.62	0.63
MN	0.19	-0.10	0.85
NH4	0.88	0.19	0.20
Eigen Value	3.68	2.07	1.25

APPENDIX III -Figure of Model running methodology

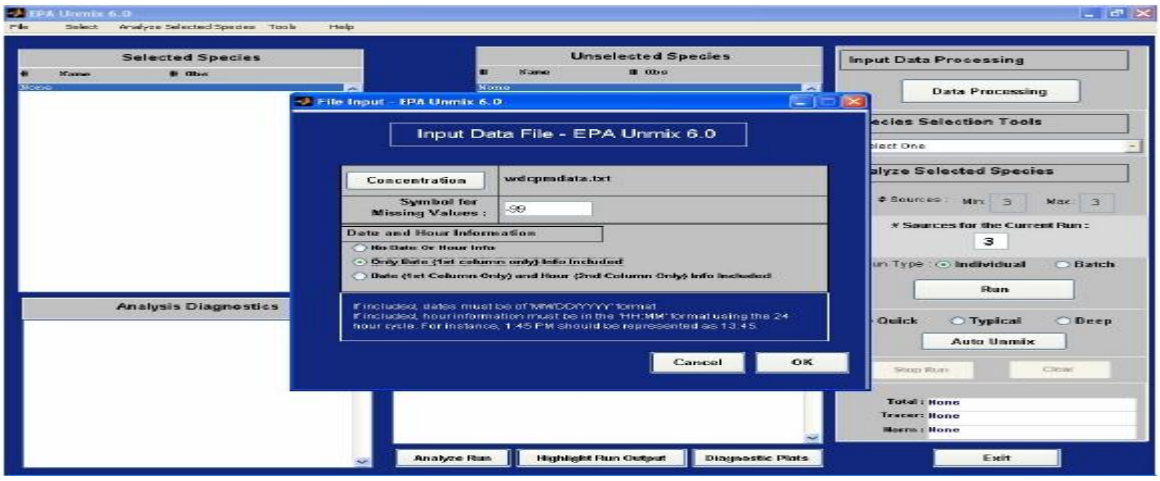

Figure: 3.1: Input Data File

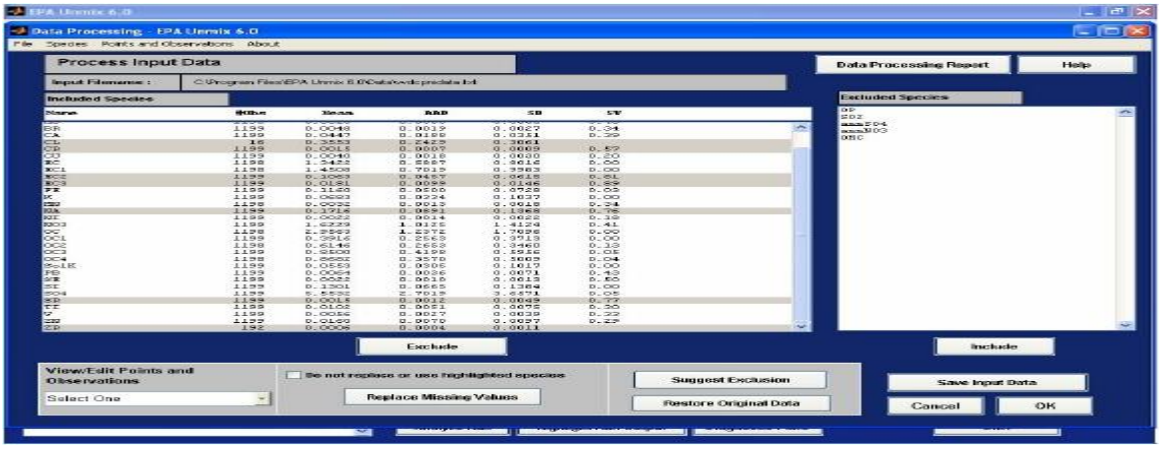

Figure 3.2: Suggest Exclusion

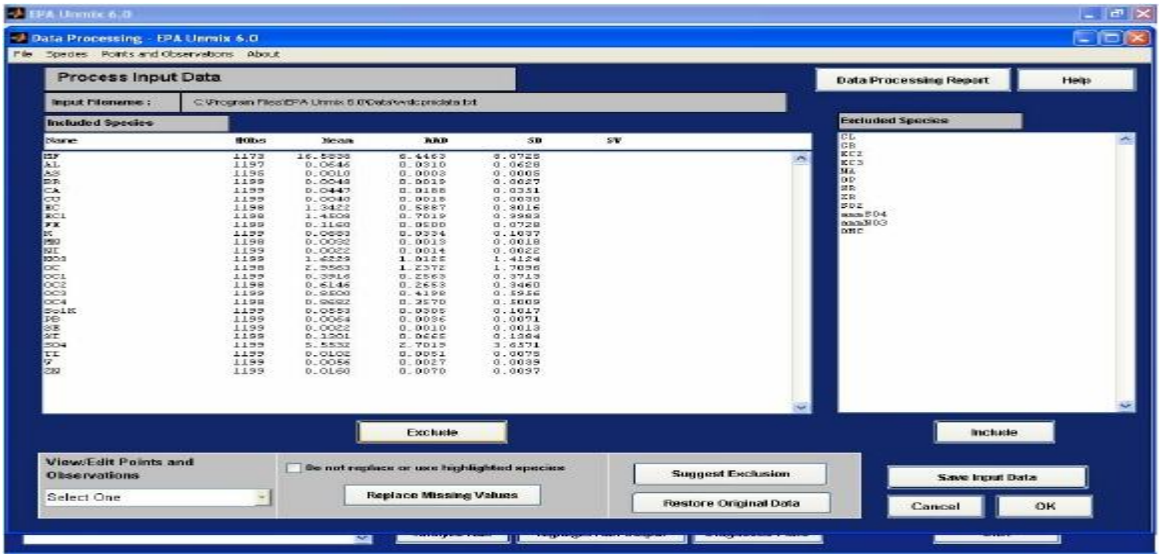

Figure3.3: Included and Excluded Specie

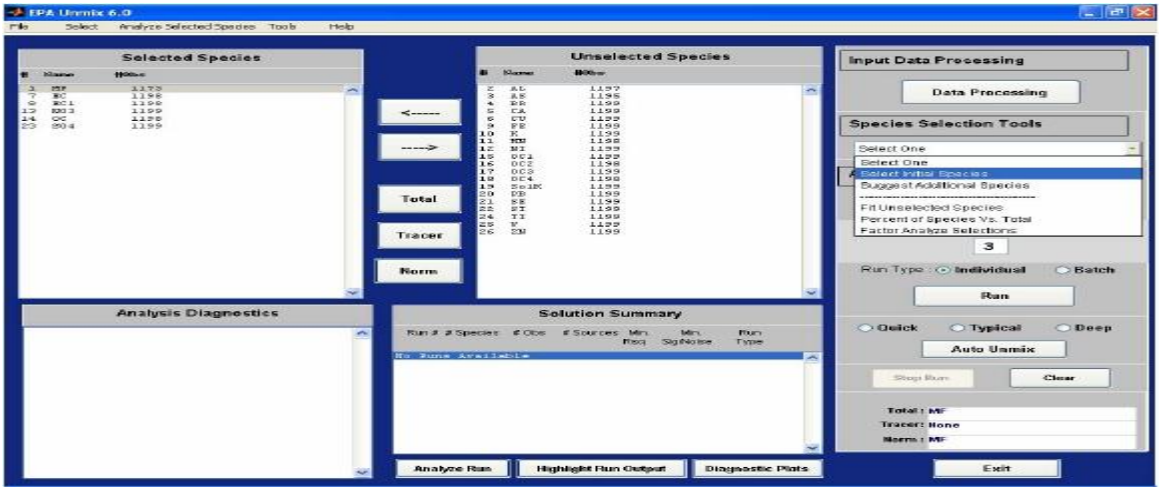

Figure 3.4: Select Initial Species

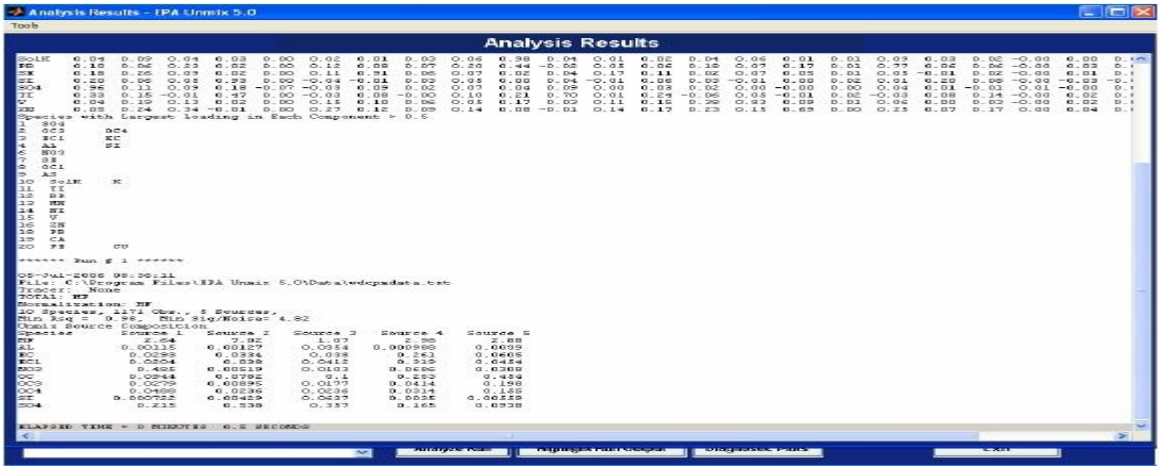

Fig 3.5: Initial Species Source Profiles

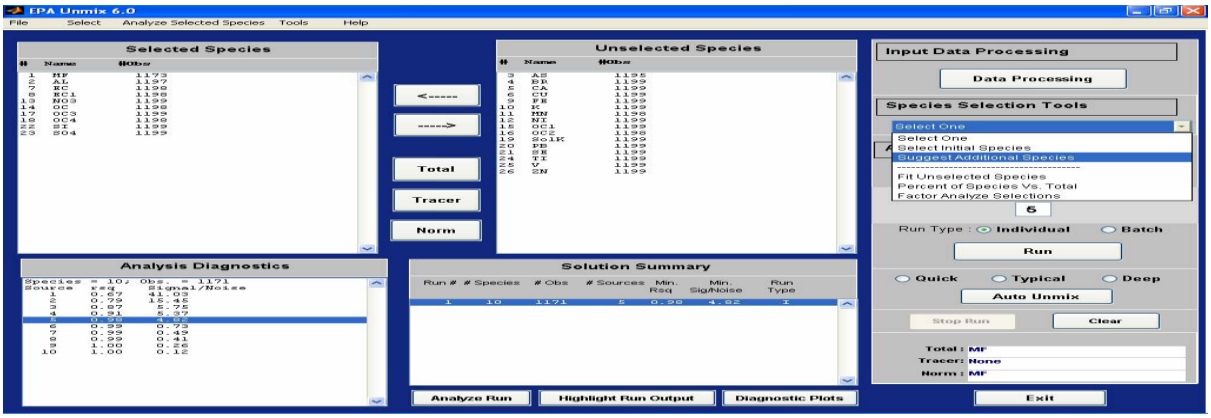

Figure3.6: Suggest Additional Species

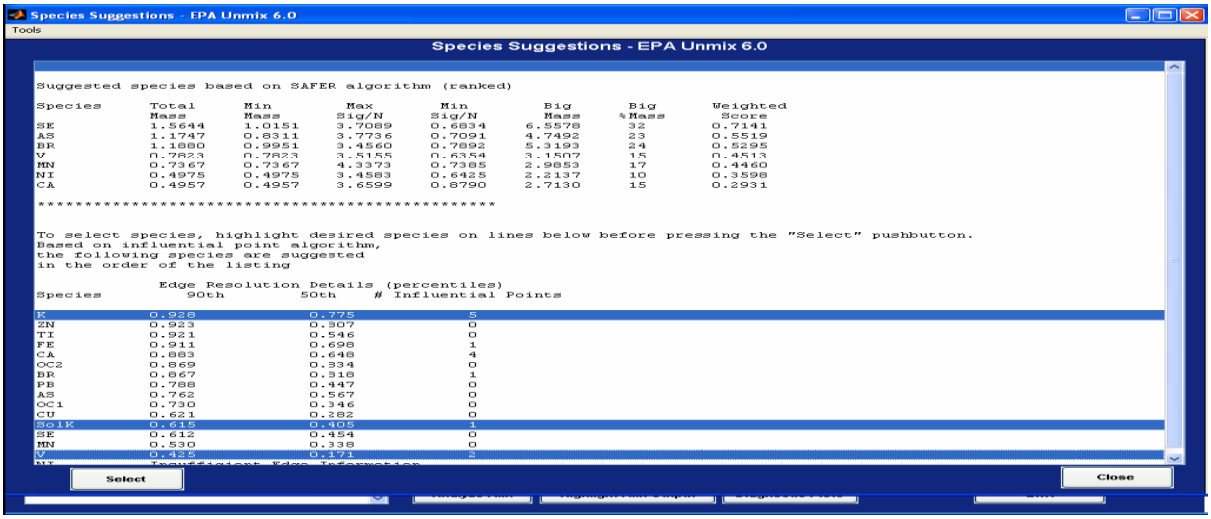

Figure 3.7: Suggestion Species

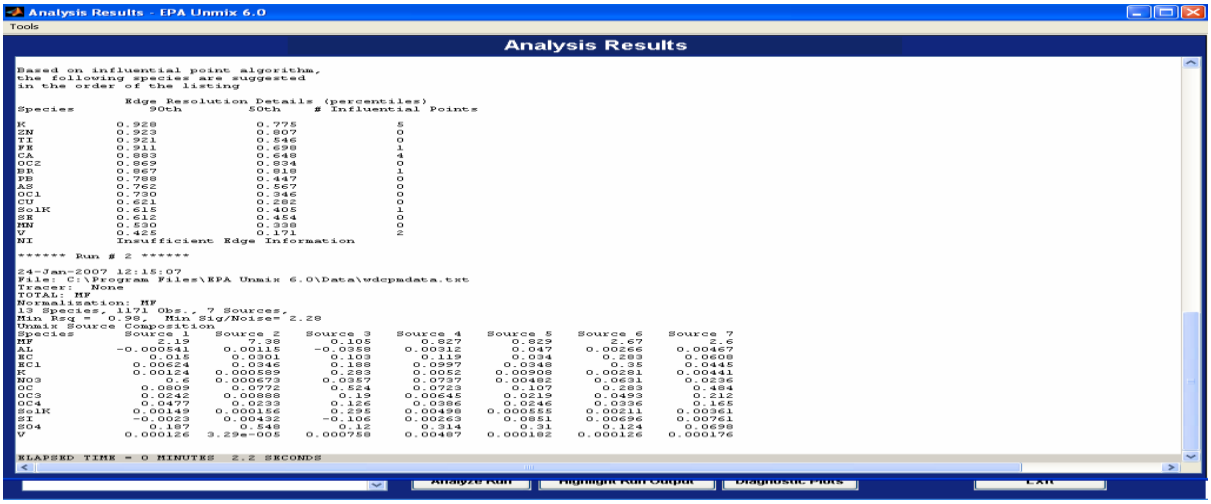

Figure 3.8: Suggest Additional Species Source Profiles

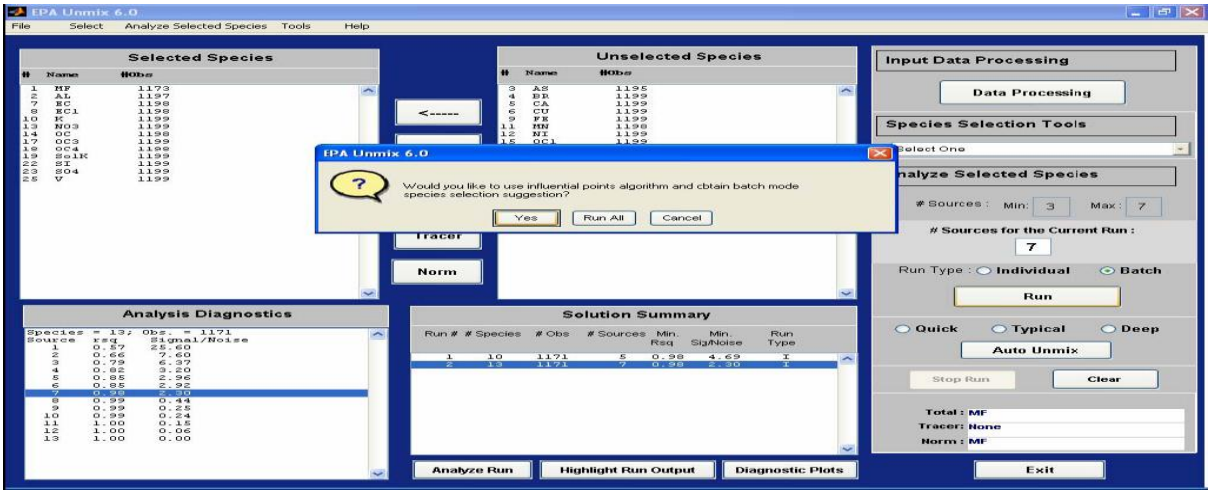

Figure 3.9: Batch mode influential point option

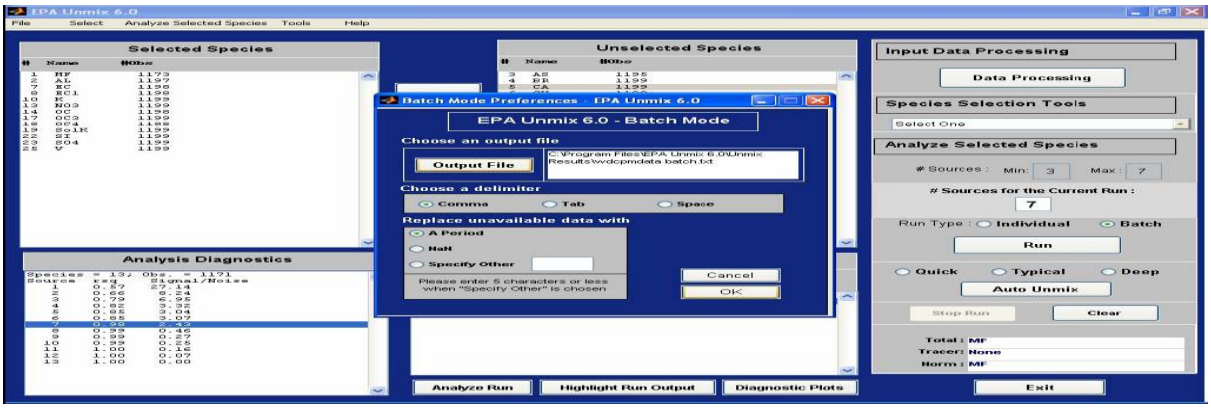

Figure 3.10: Batch Mode Preferences

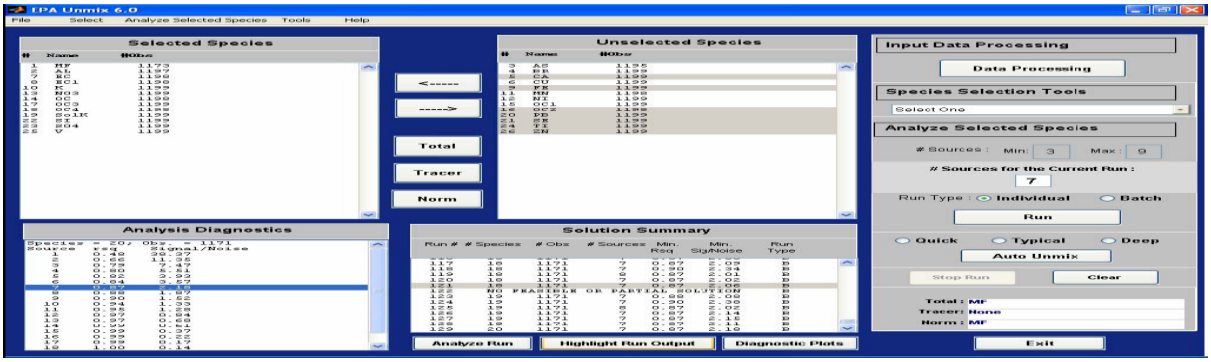

Figure 3.11: Batch Mode Solution Summary

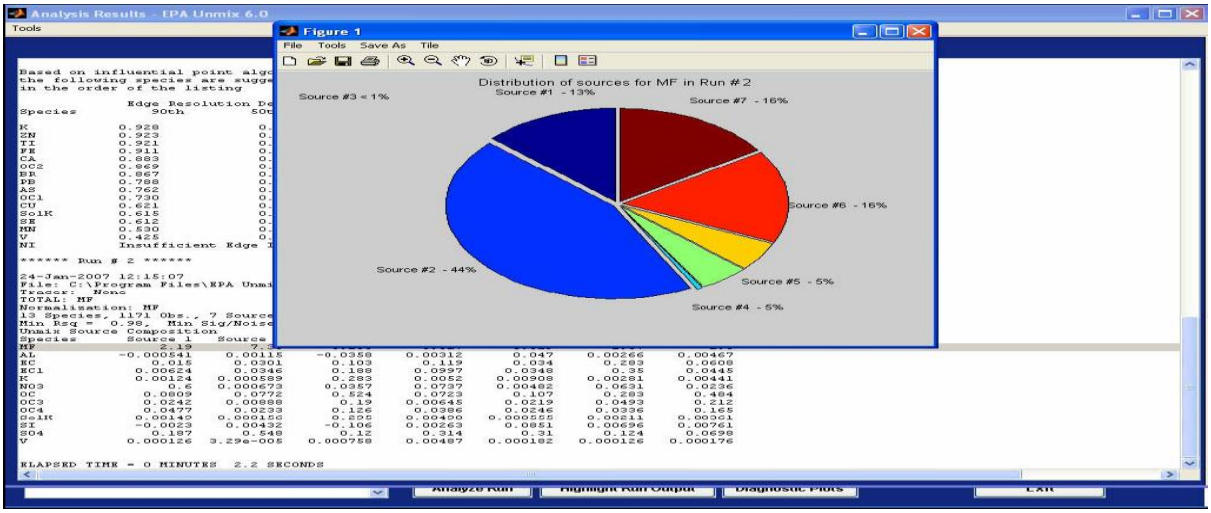

Figure 3.12: Analysis Results - Plot Distribution

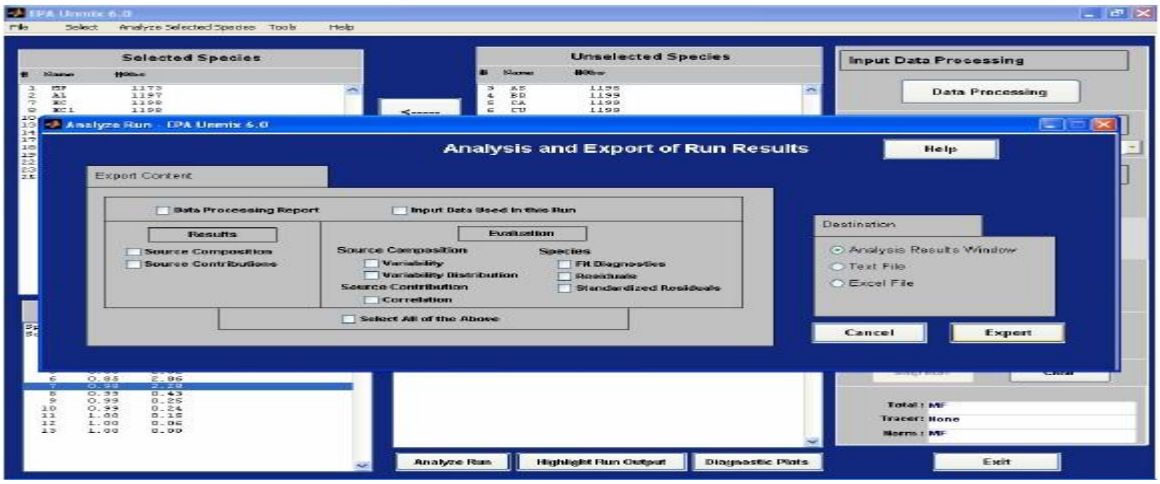

Figure 3.13: Analyze Output window

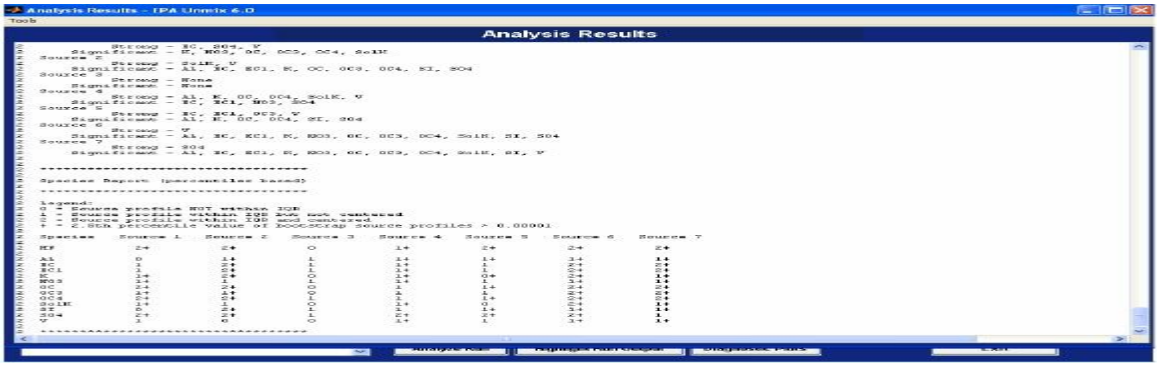

Figure 3.14: Species Report (percentile based)

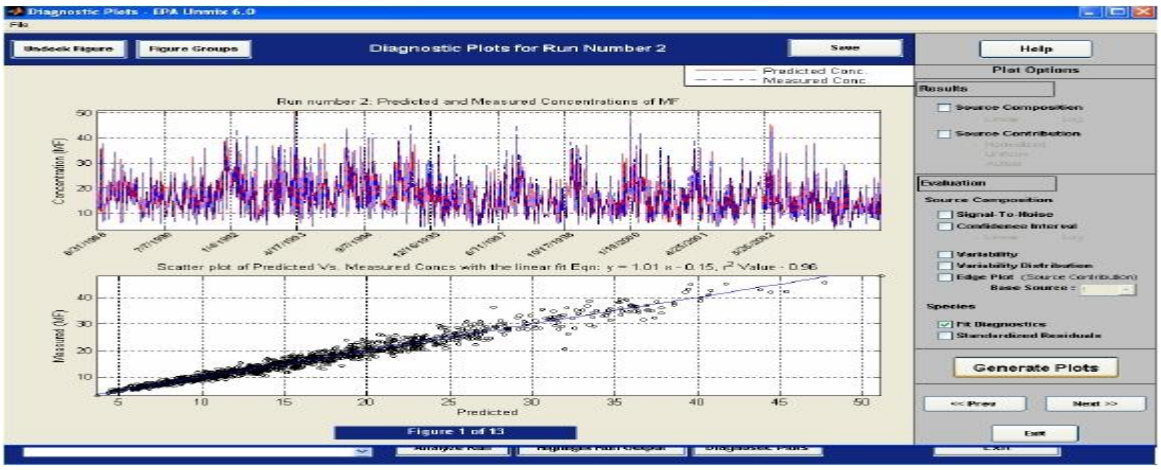

Figure 3.15: Diagnostic Plots – Fit Diagnostics Example

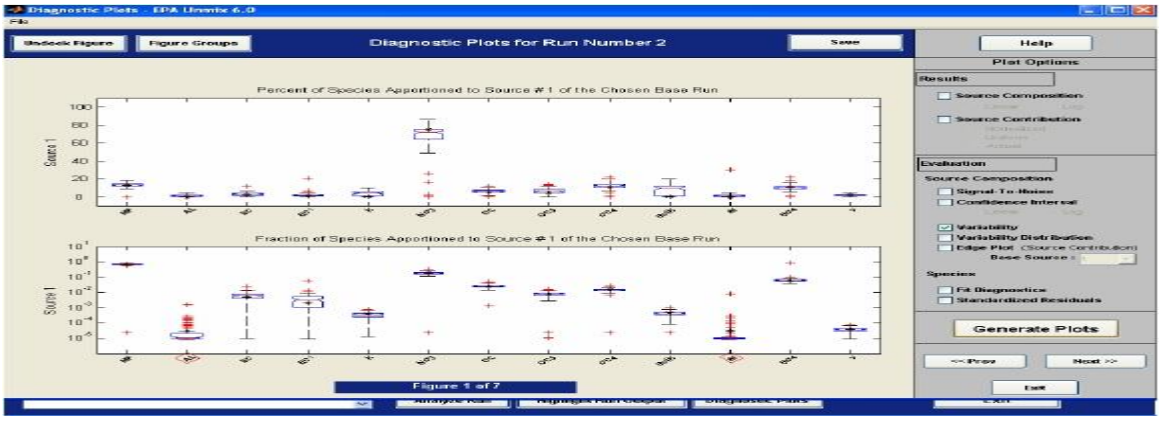

Figure 3.16: Source Profile Variability Plot

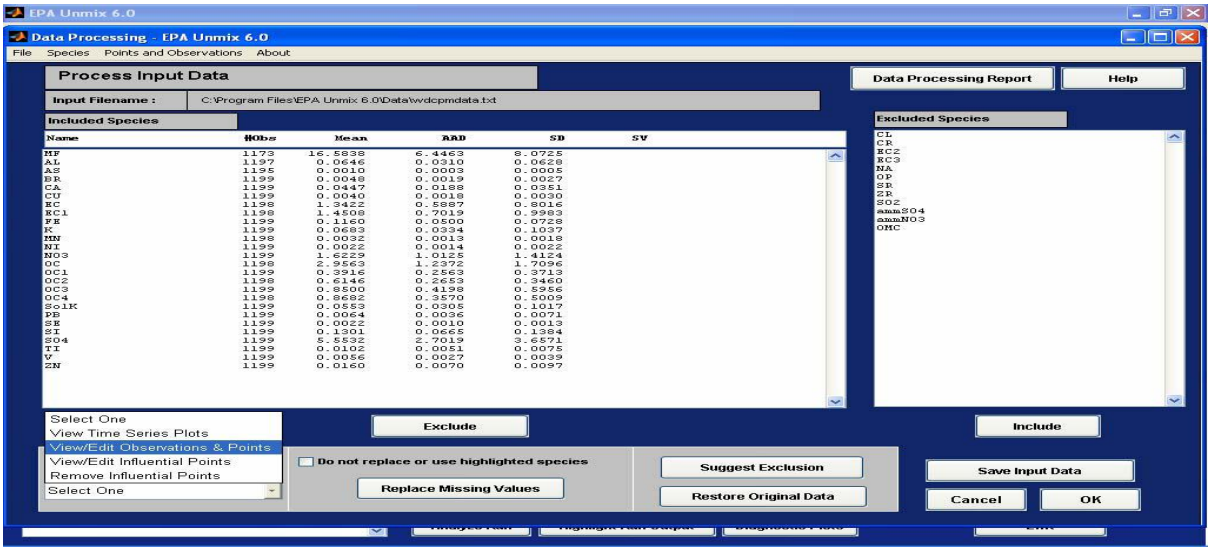

Figure : 3.17 View/Edit Observations & Points

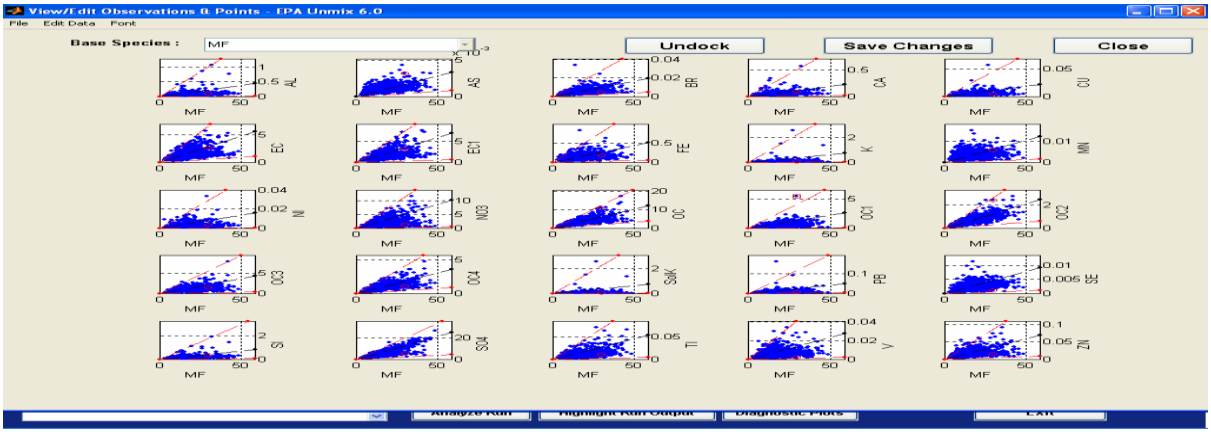

Figure 3.18: View/Edit Observations & Points plot high OC1 point

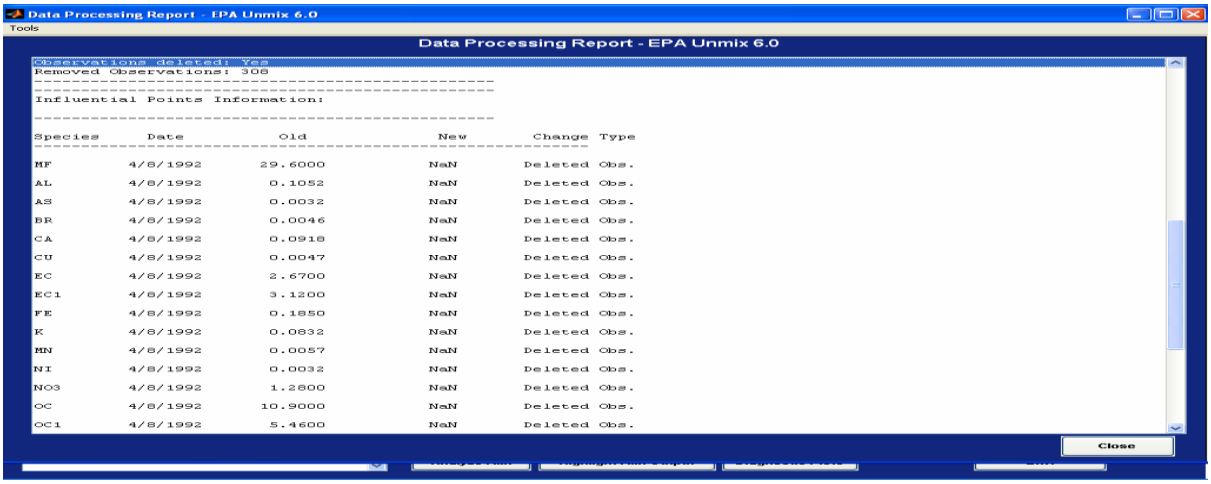

Figure 3.19: Datacusor Mode

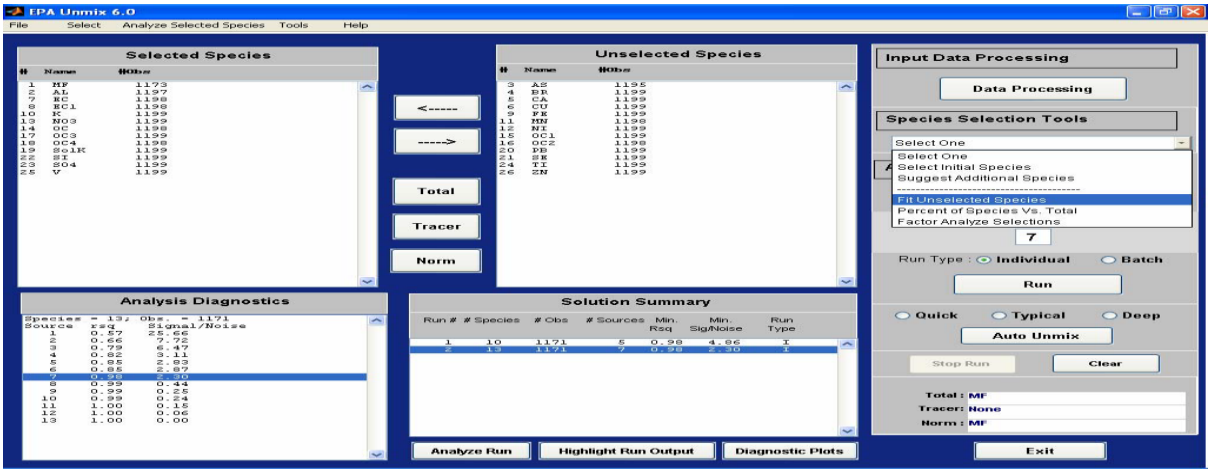

Figure 3.20: Fit Unselected Species command

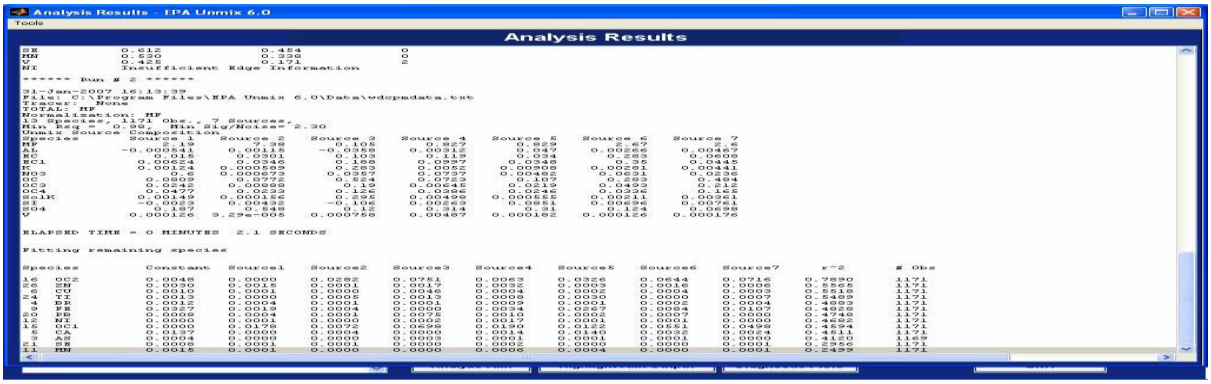

Figure 3.21: Fit Unselected Species Results

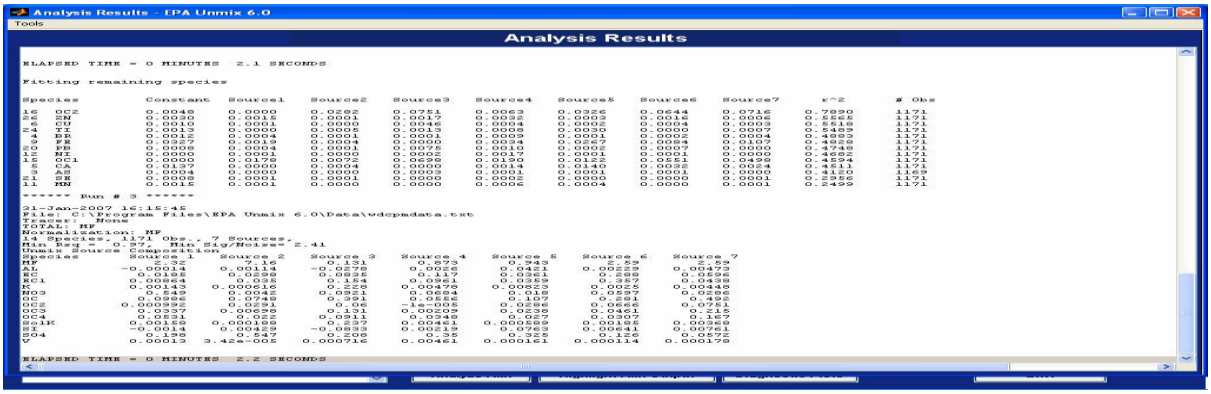

Figure 3.22: Adding species from Fit Unselected Species

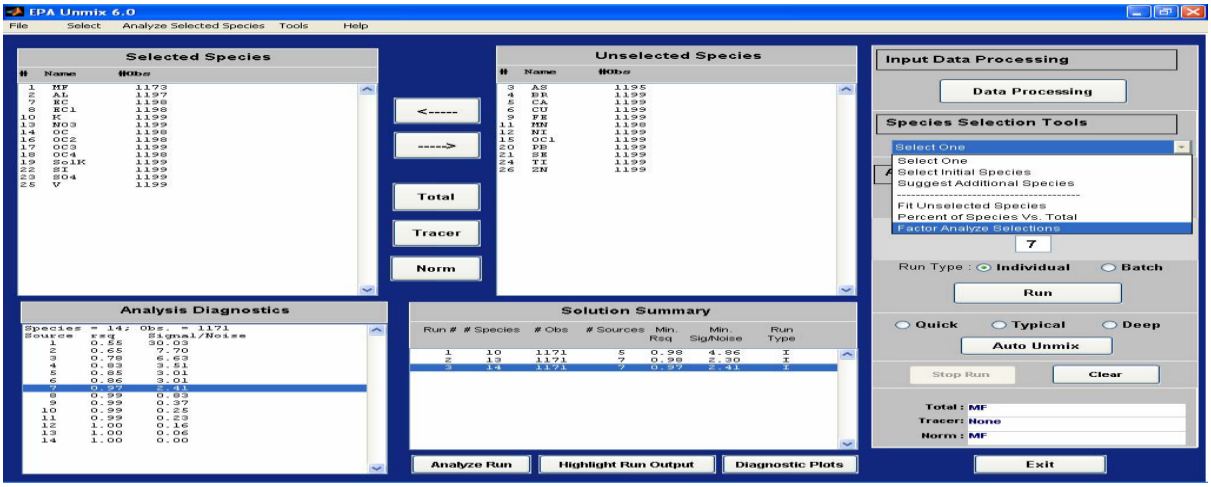

Figure 3.23: Factor Analyze Selections command

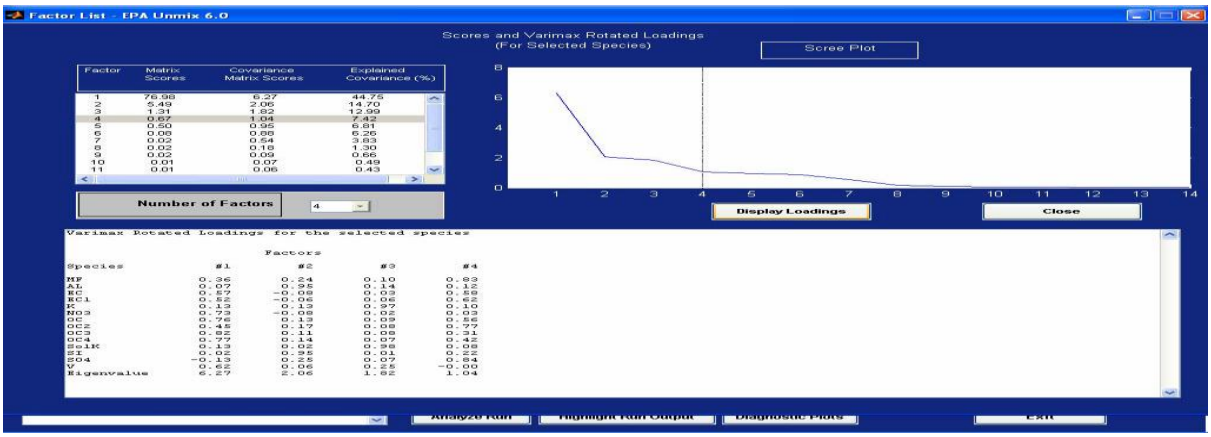

Figure 3.24: Factor Analysis Results