B-O-2

MATRIX FACTORIZATION MODELING FOR AIR POLLUTION SOURCE IDENTIFICATION APPLICATION TO FIELD DATA FROM KANPUR

Bhavin Mehta[1], ChandraVenkataraman[1*], Mani Bhushan[1] and \ Sachchida Nand Tripathi[2]

[1]Department of Chemical Engineering, IIT Bombay, Powai, Mumbai 400076.
[2]Department of Civil Engineering, IIT Kanpur, Kanpur 208016.

Introduction

Receptor modeling approaches such as positive matrix factorization (PMF) are effective tools in source identification of regional scale pollution over spatial scales of 100-2000 km and temporal scales of weeks to seasons (Kim and Hopke, 2004). The sources of aerosols on a regional scale over India have only recently received attention. Collocated measurements of physical, optical and chemical aerosol parameters were made at Kanpur (Tripathi et al., 2006; Tare et al., 2006) as part of a field campaign in December 2004 to understand regional fog and haze formation. In this work we use the time series aerosol chemical composition data from Kanpur to deduce probable sources using PMF.

Modeling Method

Pollutant source-receptor relationships can be expressed as X = GF + E, where X is the known pollutant elemental concentration matrix, E is the associated error matrix, F is the unknown source composition matrix or loading of different pollutants on the estimated factors and G is the unknown source contribution matrix or amount of material contributed by each factor/source (Paatero, 1997). PMF uses a weighted least squares approach to identify F and G. This model is solved using PMF2 software by minimizing the sum of squares of scaled residuals as given in equation (1), with the constraint that all elements of the estimated G and F matrices are non-negative (Paatero, 2004)

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{s_{ij}} \right)^2 \qquad (1)$$

* Corresponding Author, Phone: 022-25767224, Fax: 022-25726895, Email: chandra@iitb.ac.in

where n is the number of samples, m is the number of species, p is the number of estimated factors, and s_{ij} is the uncertainty associated with j^{th} species in i^{th} sample. PMF is used to identify factors, each with a distinct relative abundance of species, corresponding to emissions composition from different source categories, like biomass combustion, industrial emissions or dust, and estimate their contribution to particle concentrations.

Results and Discussion

The mass concentrations of particulate matter and constituent species including NO_3^-, SO_4^{2-}, Cl^-, NH_4^+, Na^+, Mg^{2+}, K^+, Ca, Ca^{2+}, Al, and Fe were measured in daily 8 h average samples during December 1-29, 2004 along with BC using an aethelometer (Tripathi et al., 2006). Ca was excluded from PMF model run to avoid double counting.

All species were above detection limits and 3 missing values of BC were replaced by the geometric mean of the remaining data (Table 1). Sample uncertainties, s_{ij}, estimated using an uncertainty proportional parameter (P_j) and method detection limit (MDL_j) are used to calculate signal to noise ratio, S/N (Table 1). Seven species had high S/N (greater than 2) and are strong species. Low S/N between 0.2 to 2 in Na^+, Mg^{2+}, Al and BC led to prescription of higher uncertainties for these species.

Table 1 Summary of data screening.

Species	Geometric mean (μg/m^3)	Arithmetic mean (μg/m^3)	Min (μg/m^3)	Max (μg/m^3)	Method detection limit (ng/m^3)	P_j (%)	Missing values (%)	S/N
NO_3^-	15.9	16.9	7.88	2.91	30.0	9	0	5.6
SO_4^{2-}	14.8	15.6	6.56	27.81	30.0	8	0	6.2
Cl^-	2.9	3.0	1.44	4.68	10.0	8	0	6.2
NH_4^+	8.4	3.1	1.62	18.85	6.0	7	0	7.1
Na^+	4.2	4.3	2.76	6.16	18.0	34	0	1.5
Mg^{2+}	0.2	0.2	0.05	0.36	25.0	171	0	0.3
K^+	5.1	5.2	2.89	6.95	20.0	8	0	6.2
Ca^{2+}	1.3	1.5	0.29	3.48	0.9	9	0	5.5
Al	0.8	1.0	0.32	2.66	3.0	38	0	1.3
Fe	1.9	2.0	1.17	3.27	0.1	6	0	8.3
BC	10.0	10.6	4.78	19.83	10.0	38	10	1.0

To identify the likely number of factors, ten random runs (corresponding to different initial guesses) were used and the run with the minimum estimated Q value was retained for two to eight factors. Discontinuity in the slope of Q with change in factor number (Zhao and Hopke, 2004) and agreement of the estimated Q with its theoretical value (Figure 1a) are used to identify probable solutions. These considerations (Q = 186; Q_{theo} = 184) suggested selection of a solution with four factors. Reduction in IM the maximum individual column mean, and IS the maximum individual column standard deviation of the residual matrix, with increasing number of factors (Lee et al., 1999) indicates three factors as the lower bound for solutions secure from lack of fit (Figure 1b). This analysis led to selection of the 4-factor solution for further interpretation.

The PMF resolved factors will be examined for the relative abundance of different species, which will be compared with known tracers for given sources, e.g. K for biomass combustion or SO_4^{2-} as secondary sulphate for fossil fuel SO_2 emissions. Further, emissions inventory information will be used to identify specific sources in the vicinity of Kanpur that are likely contributors to the measured aerosols.

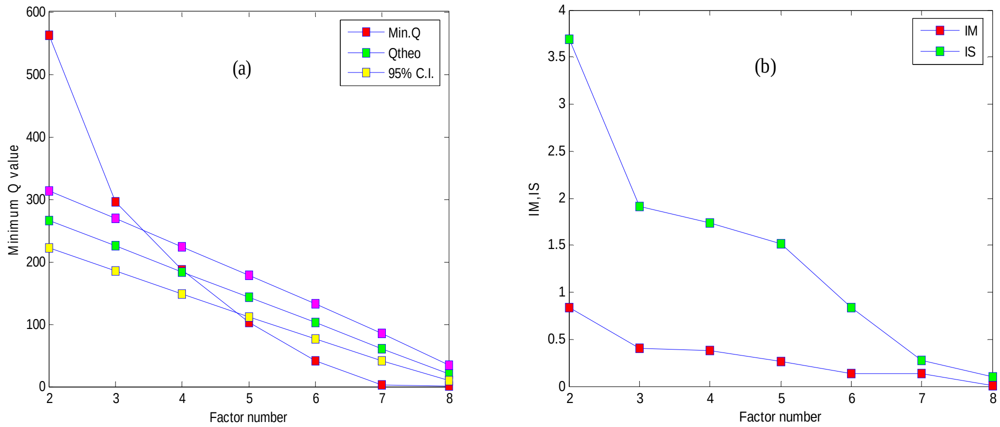

Figure 1 Variation in (a) Q value and (b) IM and IS with number of factors.

References

1. Kim, E., Hopke P.K., (2004). “Improving source identification of fine particles in a rural northeastern U.S. area utilizing temperature resolved carbon fraction”, *Journal of Geophysical Research*, 111, D09204, doi10.1029/2003JD004199.
2. Lee, E., Chan, C.K., Paatero, P., (1999). “Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong”, *Atmospheric Environment*, 33, 3201-3212.
3. Paatero, P., (1997), “Least squares formulation of robust, non negative factor analysis”, *Chemoetrics and Intelligent Laboratory Sys*tems, 37, 23-35.
4. Paatero, P., (2004), “User’s guide for Positive matrix factorization programs PMF2 and PMF3. part 1tutorial”.
5. Tare, V., Tripathi, S. N., Chinnam, N., Srivastava, A. K., Dey, S., Manar, M., Lal, R.B., Kanwade, Agarwal, A., Kishore, S., Lal, R.B., Sharma, M., (2006), “Measurements of atmospheric parameters during Indian Space Geosphere Biosphere Program Land Campaign II at a typical location in the Ganga Basin 2. Chemical properties”, *Journal of Geophysical Research*, 111, D23210, doi10.1029/2006JD007279.
6. Tripathi, S. N., Tare, V., S. N., Chinnam, N., Srivastava, A. K., Dey, S., Agarwal, A.,Kishore, S., Lal, R.B., Manar, M., Kanwade, V.P., Chauhan, S.S.S., Sharma, M., Reddy, R.R., Rama Gopal, K., Narasimhulu, K., Siva Sankara Reddy, L. Gupta, S. ,Lal S., (2006), “Measurements of atmospheric parameters during Indian Space Geosphere Biosphere Program Land Campaign II at a typical location in the Ganga Basin 1. Physical and optical properties”, *Journal of Geophysical Research*, 111, D23209, doi10.1029/2006JD007278.
7. Zhao, W., and Hopke, P.K., (2004), “Source apportionment for ambient particles in the San Gorgonio wilderness”, *Atmospheric Environment*, 38, 5901-5910.