Positive matrix factorization and trajectory modelling for source identification: A new look at Indian Ocean Experiment ship observations

S.G. Bhanuprasad, Chandra Venkataraman*, Mani Bhushan

Department of Chemical Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400076, India

Abstract

The sources of aerosols on a regional scale over India have only recently received attention in studies using back trajectory analysis and chemical transport modelling. Receptor modelling approaches such as positive matrix factorization (PMF) and the potential source contribution function (PSCF) are effective tools in source identification of urban and regional-scale pollution. In this work, PMF and PSCF analysis is applied to identify categories and locations of sources that influenced surface concentrations of aerosols in the Indian Ocean Experiment (INDOEX) domain measured on-board the research vessel Ron Brown [Quinn, P.K., Coffman, D.J., Bates, T.S., Miller, T.L., Johnson, J.E., Welton, E.J., et al., 2002. Aerosol optical properties during INDOEX 1999: means, variability, and controlling factors. Journal of Geophysical Research 107, 8020, doi:10.1029/2000JD000037]. Emissions inventory information is used to identify sources co-located with probable source regions from PSCF. PMF analysis identified six factors influencing PM concentrations during the INDOEX cruise of the Ron Brown including a biomass combustion factor (35–40%), three industrial emissions factors (35–40%), primarily secondary sulphate–nitrate, balance trace elements and Zn, and two dust factors (20–30%) of Si- and Ca-dust. The identified factors effectively predict the measured submicron PM concentrations (slope of regression line = 0.90 ± 0.20; $R^2 = 0.76$). Probable source regions shifted based on changes in surface and elevated flows during different times in the ship cruise. They were in India in the early part of the cruise, but in west Asia, south-east Asia and Africa, during later parts of the cruise. Co-located sources include coal-fired electric utilities, cement, metals and petroleum production in India and west Asia, biofuel combustion for energy and crop residue burning in India, woodland/forest burning in north sub-Saharan Africa and forest burning in south-east Asia. Significant findings are equivalent contributions of biomass combustion and industrial emissions to the measured aerosol surface concentrations, the origin of carbonaceous aerosols largely from biomass combustion and the identification of probable source regions in Africa, west Asia, the Arabian peninsula and south-east Asia, in addition to India, which affected particulate matter concentrations over parts of the INDOEX domain covered by the Ron Brown cruise.

Keywords: Positive matrix factorization (PMF); Potential source contribution function (PSCF); Source identification; INDOEX; Atmospheric aerosols
1. Introduction

Observational studies during the Indian Ocean Experiment or INDOEX (e.g., Ramanathan et al., 2001; Ramachandran and Jayaraman, 2002) and more recent field campaigns (e.g., Ramachandran, 2005; Moorthy et al., 2005) show widespread existence of aerosols and trace gases of anthropogenic origin over the Indian continent and ocean regions adjoining India. Interior continental regions have a predominance of coarse-mode dust aerosols, significant particulate black carbon concentrations at the surface and in elevated layers (e.g., Tripathi et al., 2005) and high anthropogenic aerosol load especially in the Indo-Gangetic plain (e.g., Ramachandran et al., 2006). These studies point to large spatial and temporal variability of surface and elevated aerosols over the Indian subcontinent and adjoining oceans, the contribution of both local emissions and long-range transport, and regionally specific aerosol chemistry, potentially mediated by dust aerosols.

The possible sources of aerosols on a regional scale over India have only recently received attention. Using back trajectory information INDOEX investigators identified channels of flow from different geographical regions influencing their observations (e.g., Quinn et al., 2002) and qualitatively linked aerosol chemical and optical properties to broad source categories of anthropogenic, dust and sea-salt aerosols. Further, general circulation model simulations with newly available regional emissions information for atmospheric simulations over south Asia during the INDOEX period (e.g., Verma et al., 2007) indicated the possible contribution of regions outside India to INDOEX aerosols.

Receptor modelling for quantitative source apportionment of aerosols includes the chemical mass balance (Friedlander, 1973) and positive matrix factorization (PMF) (Paatero, 1997) approaches. Additionally, identification of source locations by combining back trajectory and composition information for a given aerosol dataset includes methods such as the potential source contribution function (PSCF) (Hopke et al., 1995). Both PMF and PSCF have found wide application on an urban scale (Pekney et al., 2006a,b) and are effective tools in source identification of long-range transported pollution as well (e.g., Kim and Hopke, 2004). In this work, we attempt quantitative source identification of INDOEX aerosols towards developing an understanding of the south Asian aerosol budget, i.e., the relative influence of emissions from within the region and from long-range transport from distant regions, on the south Asian aerosol load. We apply PMF and PSCF analysis to identify categories and locations of sources that influenced the surface concentrations of aerosols in the INDOEX domain measured on-board the research vessel Ron Brown (Quinn et al., 2002).

2. Modelling approach

2.1. Aerosol approach

2.1.1. Data description

Aerosol chemical data were collected aboard the Ron Brown during its cruise in the Indian Ocean as part of the INDOEX (Quinn et al., 2002) from 22 February to 30 March 1999. These included submicron \((D_{\text{aero}} < 1.1 \mu m)\) and super micron \((1.1 \mu m < D_{\text{aero}} < 10 \mu m)\) particle fractions, collected using a two-stage impactor (Quinn et al., 2002). This work is limited to source identification for the submicron fraction, more likely to be subject to regional-scale transport. Particle samples for different chemical analysis were collected for different sampling durations: inorganic ions on a 6-h duration (four samples per day), carbon on 12-h duration (two samples per day) and trace elements on 24-h duration (one sample per day). Chemical species included in the concentration matrix input to PMF analysis were \(K^+, Na^+, NH_4^+, Ca^{2+},\) methyl sulphonic acid (MSA), \(Cl^-, NO_3^-,\) non-sea-salt sulphate (nss-SO_4^{2-}), carbon fractions of organic carbon (OC) and elemental carbon (EC), and trace elements Al, Si, P, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, As and Pb (Table 1). Analysis methods used (Quinn et al., 2002) were ion chromatography (IC) for ions, a thermographic method for OC and EC (Neususs et al., 2002), and X-ray fluorescence (XRF) spectroscopy for trace elements. The thermographic method for carbon analysis did not use laser transmission to identify the OC/EC split or correct for pyrolytic OC, per standard TOT/TOR methods, which could lead to overestimation of EC and corresponding underestimation of OC. The primary data were averaged to give 24-h mean concentrations and uncertainty estimated by propagation of variance, to give a concentration data matrix of 27 daily mean values for 23 chemical species.
2.1.2. Below detection limit and missing values

The data screening methodology used here was adapted from earlier works involving PMF (Kim and Hopke, 2004; Pekney et al., 2006a; Reff et al., 2007). Below detection limit (BDL) values were as follows: 76 out of total 101 samples for inorganic ions, 3 out of 40 samples for carbon species and 150 out of 364 samples for trace elements. The concentrations of BDL values were replaced with half of the method detection limit (MDL) (e.g., Pekney et al., 2006a). Missing samples, which totalled 16, of which 8 were for inorganic ions and 8 for carbon species, were replaced with geometric mean of their measured concentrations. Time series plots (not shown here) were analyzed for identifying extreme values in the concentration dataset. Since no extreme values were found from the time series plot, all the samples were retained for further analysis.

2.1.3. Prescription of measurement uncertainties

Uncertainties in the measurement dataset were estimated using an uncertainty proportional parameter $p_j$ for each species $j$, operated on the measured concentration, $X_{ij}$, and adding 1/3 of the method detection limit (MDL) to it (e.g., Pekney et al., 2006a), as shown as follows:

$$s_{ij} = p_j X_{ij} + \text{MDL}_j / 3$$

(1)

Uncertainties assigned to measurements for input to PMF should estimate the variability from the measurement method as well as variability among individual sources in a given source category. Variability within source categories in the Indian region (i.e., among individual biomass burning sources or individual power plants) is unknown because of the absence of regional source profile measurements and of networks measuring aerosol

<table>
<thead>
<tr>
<th>Species</th>
<th>MDL (ng m$^{-3}$)</th>
<th>Percent BDL values$^d$ (%)</th>
<th>Percent missing values$^d$ (%)</th>
<th>Geometric mean$^\epsilon$ (ng m$^{-3}$)</th>
<th>Arithmetic mean$^\epsilon$ (ng m$^{-3}$)</th>
<th>Minimum (ng m$^{-3}$)</th>
<th>Maximum (ng m$^{-3}$)</th>
<th>$p_j$ (%)</th>
<th>Signal-to-noise (S/N) ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>K$^+$</td>
<td>3$^a$</td>
<td>5</td>
<td>0</td>
<td>77</td>
<td>185</td>
<td>2</td>
<td>637</td>
<td>8</td>
<td>4.45</td>
</tr>
<tr>
<td>Na$^+$</td>
<td>4$^a$</td>
<td>0</td>
<td>1</td>
<td>35</td>
<td>42</td>
<td>9</td>
<td>119</td>
<td>34</td>
<td>0.89</td>
</tr>
<tr>
<td>NH$_4^+$</td>
<td>6$^a$</td>
<td>0</td>
<td>2</td>
<td>504</td>
<td>772</td>
<td>32</td>
<td>2843</td>
<td>7</td>
<td>0.78</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td>0.9$^b$</td>
<td>9</td>
<td>0</td>
<td>21</td>
<td>41</td>
<td>1</td>
<td>175</td>
<td>9</td>
<td>3.89</td>
</tr>
<tr>
<td>MSA</td>
<td>2$^a$</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>23</td>
<td>5</td>
<td>44</td>
<td>8</td>
<td>0.45</td>
</tr>
<tr>
<td>Cl$^-$</td>
<td>10$^a$</td>
<td>23</td>
<td>0</td>
<td>17</td>
<td>24</td>
<td>10</td>
<td>121</td>
<td>8</td>
<td>1.93</td>
</tr>
<tr>
<td>NO$_3^-$</td>
<td>1.5$^a$</td>
<td>39</td>
<td>0</td>
<td>6</td>
<td>14</td>
<td>7</td>
<td>52</td>
<td>9</td>
<td>2.99</td>
</tr>
<tr>
<td>nss-SO$_4^{2-}$</td>
<td>4$^a$</td>
<td>0</td>
<td>2</td>
<td>2318</td>
<td>3488</td>
<td>290</td>
<td>16,010</td>
<td>8</td>
<td>0.73</td>
</tr>
<tr>
<td>OC</td>
<td>67.8$^c$</td>
<td>5</td>
<td>10</td>
<td>294</td>
<td>417</td>
<td>80</td>
<td>1760</td>
<td>15</td>
<td>0.24</td>
</tr>
<tr>
<td>EC</td>
<td>28.8$^c$</td>
<td>2.5</td>
<td>10</td>
<td>357</td>
<td>555</td>
<td>70</td>
<td>2010</td>
<td>38</td>
<td>0.24</td>
</tr>
<tr>
<td>Al</td>
<td>3$^b$</td>
<td>18</td>
<td>0</td>
<td>23</td>
<td>52</td>
<td>7</td>
<td>203</td>
<td>38</td>
<td>1.28</td>
</tr>
<tr>
<td>Si</td>
<td>2.2$^b$</td>
<td>14</td>
<td>0</td>
<td>66</td>
<td>189</td>
<td>6</td>
<td>848</td>
<td>8</td>
<td>6.12</td>
</tr>
<tr>
<td>P</td>
<td>1.9$^b$</td>
<td>46</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>87</td>
<td>30</td>
<td>1.58</td>
</tr>
<tr>
<td>Ti</td>
<td>0.81$^b$</td>
<td>29</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>14</td>
<td>17</td>
<td>2.39</td>
</tr>
<tr>
<td>V</td>
<td>0.69$^b$</td>
<td>36</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>50</td>
<td>0.78</td>
</tr>
<tr>
<td>Cr</td>
<td>0.57$^b$</td>
<td>86</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>50</td>
<td>0.45</td>
</tr>
<tr>
<td>Mn</td>
<td>0.52$^b$</td>
<td>36</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>33</td>
<td>1.18</td>
</tr>
<tr>
<td>Fe</td>
<td>0.11$^b$</td>
<td>18</td>
<td>0</td>
<td>10</td>
<td>44</td>
<td>0</td>
<td>122</td>
<td>6</td>
<td>8.27</td>
</tr>
<tr>
<td>Ni</td>
<td>0.05$^b$</td>
<td>68</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>100</td>
<td>0.50</td>
</tr>
<tr>
<td>Cu</td>
<td>0.05$^b$</td>
<td>54</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>14</td>
<td>25</td>
<td>1.98</td>
</tr>
<tr>
<td>Zn</td>
<td>0.05$^b$</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>7</td>
<td>6.93</td>
</tr>
<tr>
<td>As</td>
<td>0.05$^b$</td>
<td>64</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>50</td>
<td>0.98</td>
</tr>
<tr>
<td>Pb</td>
<td>0.06$^b$</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>63</td>
<td>50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$^a$Adapted from Xie et al. (1999).

$^b$Adapted from IMPROVE data guide (1995).

$^c$Calculated as one standard deviation of blanks.

$^d$Calculated from total number of samples: inorganic ions—101; OC/EC—40; trace elements—28.

$^\epsilon$Below detection limit values were replaced by half of the reported detection limit values for the geometric mean and arithmetic mean calculations.
chemical constituents. Therefore, we match the measurement method for each species in the present observations with that available in literature from operating networks to select representative \( p_j \) and MDL, for each species. Uncertainty parameters from the Environmental Protection Agency’s Speciation Trends Network (EPA STN) (Flanagan and Rickman, 2004) and the IMPROVE network (IMPROVE Data Guide, 1995) were reviewed. While IMPROVE uncertainties may better reflect the uncertainties in this dataset (both being from regional-scale transport), EPA STN uncertainties were more recently measured. Based on availability of uncertainty parameters for our complete list of species, we used uncertainty proportional parameters from the EPA STN compilation and MDLs from the IMPROVE network (IMPROVE Data Guide, 1995) and Xie et al. (1999). Larger expected uncertainties in the thermographic method used in this work (see Section 2.1.1) led to adoption of MDL’s for OC and EC as one standard deviation of blanks for OC and EC, reported in this dataset. Uncertainties for BDL values were set to 5/6 MDL and four times the estimated geometric mean concentration for missing samples. The signal to noise ratio (S/N) (Table 1) was concentration for missing samples. The signal to and four times the estimated geometric mean concentration for missing samples. The signal to noise ratio (S/N) (Table 1) was >2 for \( K^+ \), \( Ca^{2+} \), \( NO_3^- \), \( Si \), Ti, Fe and Zn, identifying them as strong species and between 0.2 and 2 for the rest, identifying them as weak species (Table S1, supplementary online data), thereby increasing their input uncertainty by a factor of 3 (Paatero and Hopke, 2003). No species had an S/N ratio <0.2, and hence none were excluded. The estimated uncertainty in most species was dominated by the \( p_jX_j \) term with the second term contributing <15%. Strong species, e.g., Zn, Fe, Si, \( Ca^{2+} \), \( NO_3^- \), \( K^+ \) typically had lower values of the proportionality parameter, but high measured concentrations. Weak species had higher values of the proportionality parameter, but lower measured concentrations. Among the strong species, Zn had a larger fraction of measurements below detection limit (Table 1).

2.1.4. Mass balance closure on measurements

When the measured sum of chemical species is not close to measured PM, the PM time-series would provide additional information to the PMF model. Hence, in such cases, PM is included as an explicit species albeit with a large uncertainty of four times the concentration (e.g., Pekney et al., 2006a). Mass balance closure was tested by a straight line fit between measurements of gravimetric mass and sum of chemical species. In the sum of chemical constituents, non-sea-salt sulphate was included in preference to total sulphate to distinguish the anthropogenic contribution. Lower measurement uncertainties were associated with ion chromatography than XRF (Flanagan and Rickman, 2004; IMPROVE Data Guide, 1995). Therefore, ionic species nss-SO_4^{2-}, \( Ca^{2+} \), \( K^+ \), and Cl~ measured by IC were included but elemental S, Ca, K, Cl measured by XRF were excluded, to avoid double counting (Reff et al., 2007). Also, since Indian regional aerosol is expected to contain soluble \( Ca^{2+} \) and \( K^+ \) (Rastogi and Sarin, 2005), it was preferable to retain the ionic forms. Organic mass to OC ratios (OM/OC), used for mass balance closure, range 1.4–1.8 for fresh to aged aerosols in north American studies (e.g., Pekney et al., 2006a). Limited regional measurements indicate higher average OM/OC of 2–2.8 in primary emissions from sources like biofuel combustion (Habib, 2005), but as this ratio is uncertain we retained OC as such in the mass balance closure analysis. Hence, (Fig. S1, supplementary online data), gravimetric PM mass was statistically larger (slope 1.92±0.41, 95% CI) than the sum of measured species, leading to inclusion of PM in the PMF concentration matrix. The resulting PMF generated source contributions and source profiles are rescaled by the PM mass concentration apportioned to the corresponding factor in order to get their values in meaningful units, namely factor profiles in dimensionless units (\( \mu g \cdot m^{-3} \)) and factor contributions in concentration units (\( \mu g \cdot m^{-3} \)). PMF was also run on the chemical species concentrations matrix, after excluding PM, but gave invalid solutions as explained in Section 3.1.

2.2. Positive matrix factorization

PMF models, e.g., the two dimensional model (PMF-2, Version 4.2) use non-negativity constraints on the predicted source compositions and contributions (Paatero, 1997) with individual weights for specific data points. PMF-2 uses a weighted least squares approach to estimate the solution. The mathematical expression of the model is given by the equation as follows:

\[
X = GF + E
\]

where \( x_{ij} \) is the concentration of \( j \)th species (\( m \) species) in \( i \)th sample (\( n \) samples), \( g_{ik} \) is the contribution of \( k \)th factor to the receptor site for \( i \)th
sample, $f_{kj}$ is the fraction of $j$th species in $k$th factor, $e_{ij}$ is the residual associated with the concentration of $j$th species measured in the $i$th sample and $p$ is the total number of factors. The goal in PMF is to estimate the factor contributions ($G$) and factor profiles ($F$) matrices that best explain $X$. Further, the entries of $G$ and $F$ are constrained to be non-negative. Towards this end, a weighted least squares approach is used. It involves minimization of an objective function $Q$, given as

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

where

$$h_{ij} = \begin{cases} 1 & \text{if } |e_{ij}/s_{ij}| \leq \alpha \\ |e_{ij}/s_{ij}|/\alpha & \text{otherwise} \end{cases}$$

where parameter $\alpha$ is the threshold outlier distance. Appropriate down weighting of outliers in PMF datasets has been done in literature (Paatero, 1997; Pekney et al., 2006a) by assigning a value typically between 2 and 4 to this parameter $\alpha$. Since well defined guidelines for appropriate specification of $\alpha$ are not available in literature (Reff et al., 2007), we have used a principal component analysis (PCA) based heuristic for specifying $\alpha$ (Bhanuprasad, 2007; for details please see Supplementary online information, pp. 7–8). Based on this analysis, four significant outliers were identified corresponding to 4, 6, 20, and 26 of March with the respective squared distances being 9.39, 9.44, 8.63, and 8.08. The $p$-value (area to the right) for the smallest of these outlier-squared distances was then calculated from $\chi^2$ distribution. This $p$-value was then used to find the threshold $\alpha$ for outlier distance from a $N(0,1)$ normal distribution. For the data we had, this $\alpha$ was found to be about 2 and hence a threshold of 2 for outlier distance was used in Eq. (3) during $Q$ minimization (Bhanuprasad, 2007).

### 2.3. Potential source contribution function

The PSCF analyses air trajectories in conjunction with atmospheric pollutant concentrations or factor/source contributions identified by PMF (Hopke et al., 1995; Pekney et al., 2006b) to identify the probable source location or preferred atmospheric transport pathways from source to receptor. PSCF estimates the conditional probability function which describes the spatial distribution of probable source locations that contributed pollutants measured during a given time period. The source domain is divided into a number of grid cells of a specific resolution, typically $1^\circ \times 1^\circ$ (Kim and Hopke, 2004). PSCF is defined as follows:

$$P_y = \frac{P[B_y]}{P[A_y]} = \frac{m_{ij}}{n_{ij}}$$

where $A_{ij}$ is the event that a trajectory endpoint falls in the $ij$th cell (there are $n_{ij}$ such trajectories). $B_{ij}$ is the event that a trajectory endpoint falling in the $ij$th cell arrived at the receptor site with pollutant concentration exceeding a threshold value (number of such trajectories is $m_{ij}$), and $P_y$ is the conditional probability that an air parcel which has passed through the $ij$th cell has a high concentration upon arrival at the receptor site. These conditional probabilities are computed for different pressure levels (corresponding to different heights). Hence the total probability of material transfer from various heights is calculated from the conditional probabilities associated with each height (Hopke et al., 1995). The total potential source contribution function (TPSCF) is given as

$$\text{TPSCF}_{ij} = \frac{P[B_y]}{\sum_k P[A_{yk}]} = \frac{\sum_k P[B_{yk}]P[A_{yk}]}{\sum_k P[A_{yk}]}$$

where index $k$ denotes the height, six levels in this study. To correct for anomalously high PSCF values because of low $n_{ij}$ in a given cell, a multiplicative weighting function (e.g., Hopke et al., 1995) is operated on the computed TPSCF for such cells. The weighting function used in this work (Eq. (6)) is described in the following section.

### 2.4. Back trajectory calculations

Back trajectories were computed using HYSPLIT4 model and the archived meteorological data from National Oceanic Atmospheric Administration’s (NOAA) Air Resources Laboratory (ARL) web site (HYSPLIT4, 1997). Ten-day back trajectories at 2-h intervals (12 per day) were computed at arrival heights of 10, 50, 500, 1000, 2500, and 5000 m, for each of the 27 daily mean samples. This resulted in 324 back trajectories per arrival height. Plots of the trajectory heights against time for trajectories arriving at 10 m height at the receptor revealed a high frequency (240 out of 324) of downward mixing of trajectories originating at higher heights of 50–1000 m, leading to adoption of TPSCF. We calculated separate TPSCF for
surface transport (10–1000 m) and free-tropospheric transport (2500 and 5000 m), as trajectories tended to originate from different geographic regions. The region covered by the trajectories was divided into grid cells of 1° × 1° latitude and longitude typical of large-scale transport in the mid latitudes (Kim and Hopke, 2004), compared to higher latitudes (Hopke et al., 1995) where larger grids are used. PSCF values in cells are usually weighted down when the number of endpoints in a cell is less than three times the average number of trajectory endpoints over all the cells. This is done using a weighting function (see Eq. (6)) that multiplies the computed TPSCF, in such cells. A few extremely distant trajectories led to a too large domain and too low n_{ij,avg} value in our TPSCF calculation thereby making the weighting function ineffective. High TPSCF values were computed in some cells without real sources, e.g., Atlantic Ocean (47–20°W) and the southern Indian Ocean (20–53°S). Therefore, the source domain was restricted to 10°S–40°N and 30–120°E, for the TPSCF calculation, excluding 61 trajectories which crossed domain boundaries. The weighting function then became more effective with a doubling of n_{ij,avg} from 6 to 12, and doubling of cells in which it was active, from 13% to 25% of the domain. The weighting function was applied with n_{ij,avg} = 12, as follows:

\[
W(n_{ij}) = \begin{cases} 
1.0 & 36 < n_{ij} \\
0.7 & 18 < n_{ij} \leq 36 \\
0.4 & 9 < n_{ij} \leq 18 \\
0.2 & n_{ij} \leq 9 
\end{cases} 
\]  

(6)

We expect source regions identified by elevated transport TPSCF analysis (trajectory arrival heights of 2500–5000 m) to not significantly impact factors that arise from surface emissions (such as residential or transport fuel use) but to have a more significant impact on factors that arise from elevated emissions (say large industrial stacks or large fires in forests and croplands or oil-well/refinery flares). Source identification is done by matching spatial and temporal distribution of source emissions from emissions inventory with the probable source locations from TPSCF analysis. Within India, a greater level of detail is available from Indian emissions inventory information from our group (Reddy and Venkataraman, 2002; Venkataraman et al., 2006), and outside details to the extent available in published work are used (Streets et al., 2003a,b; Tansey et al., 2004).

3. Source identification using positive matrix factorization

3.1. Base-case solution selection

To identify the likely number of factors, 20 random runs (corresponding to different starting points) were used and the run with the minimum estimated Q value was retained (e.g., Lee et al., 1999) for 2–10 factors. Reduction in Q with increase in number of factors and agreement of estimated Q with its theoretical value Q_{theo} were used to identify probable solutions. Estimated Q decreased with increasing number of factors (Fig. S2a, supplementary online data) and its slope levelled off at six factors, with a small decrease on moving to a 7-factor solution. Solutions of 5, 6 and 7 factors were carefully examined. The maximum individual column mean, IM, and the maximum individual column standard deviation, IS, of the scaled residual matrix (e.g., Lee et al., 1999) were used to identify the minimum number of factors needed for a well-constrained solution. Values of IM and IS dropped sharply for solutions with six or more factors (Fig. S2b, supplementary online data). The physical interpretation of factors in terms of likely source categories was the main criterion used in retaining solutions for further analysis. The 5-factor solution had relatively poor factor resolution with several mixed factors such as Si+trace elements, Ca + sea-salt + Ti, NaCl+trace elements. In the 7-factor solution, the trace elements factor split, putting Fe in a separate factor, with little physical basis. Additional PMF runs excluding PM as a species had drawbacks like highly mixed factors, physically unrealistic factors (e.g., Si and NaCl factors contaminated by trace elements) and negative coefficients on regression of factor contributions against PM mass leading to invalid solutions (Kim and Hopke, 2004). In comparison, solutions including PM as an explicit species gave superior factor resolution and more realistic factor composition. This analysis led to the selection of the 6-factor solution, including PM as a species in the PMF analysis, for further analysis.

3.2. Constraining rotational freedom

PMF can lead to multiple valid solutions of the form

\[
X = GTT^{-1} F + E 
\]  

(7)
For different choices of $T$ and $T^{-1}$ some solutions may result in source compositions, $T^{-1}F$, with factor loadings closer to 100% or factor composition that more clearly corresponds to the emissions composition from a given source than the base-case solution, $F$. Such rotations are prescribed in PMF-2 through a function $F_{\text{peak}}$ (Paatero et al., 2002). Twenty pseudorandom runs of the 6-factor solution were made with PMF-2 operated in robust mode and $F_{\text{peak}} = 0$ to identify the global minimum solution (e.g., Lee et al., 1999), and further rotated using $F_{\text{peak}}$ to identify optimal factor resolution and factor composition.

The base-case 6-factor solution had a single mixed-combustion factor containing biomass-burning tracers $K^+$, OC and EC along with tracers for fossil fuel combustion, $NH_4^+$ and nss-$SO_4^{2-}$. Two dust factors, one Si-dust (from desert) and the other Ca-dust (clay or limestone based) were resolved. One secondary sulphate–nitrate factor, one Zn factor and one trace elements factor explaining most of the Fe, Ti, Al mass were identified. Correlation between measured species and model predicted species concentration was determined to examine model goodness-of-fit. Species with $R^2 > 0.7$ were found to be: $K^+$, $NH_4^+$, $Ca^{2+}$, $NO_3^-$, nss-$SO_4^{2-}$, Al, Si, Ti, Mn, Fe, and Zn (Table S2, supplementary online data). Regression of the measured species concentrations against PMF-2 estimated factor contributions identified tracer species (Table S3, supplementary online data) which correlated well with each factor. The original solution was rotated using different values of $F_{\text{peak}}$ (Paatero et al., 2002) and the variation in $Q$ with $F_{\text{peak}}$ showed a stable $Q$-regime between $F_{\text{peak}}$ of −0.1 to 0.8 (Fig. S3, supplementary online data). The $F_{\text{peak}} = 0.8$ solution gave a higher loading of tracer species in many factors (increase in the number of zeros in the $F$-matrix to 72 from 50 in the base case—see Table S2, supplementary online data) and was retained for further analysis.

The $F_{\text{key}}$ function was used to pull down species in factors where their presence was not physically realistic. Several factors contained significant amounts of nss-$SO_4^{2-}$ and $NH_4^+$, requiring pulling down in factors contaminated by them (e.g., Lee et al., 1999). We wished to resolve the mixed-combustion factor into separate biomass combustion and fossil fuel combustion factors, as the extent of their contribution to aerosol loading in the Indian region remains unresolved (Novakov et al., 2000). The measurements dataset used here had a mean Si/Al ratio of over 3 and high correlation of Al with Si ($R^2 = 0.89$), suggesting Al origin from soil, also expected in desert dust with a Si/Al ratio typically 1.5 in east Asia to 2–4 in regions like the Sahara and China/Tibet (Arimoto et al., 2006). However, we caution that the XRF signal of Al in PM samples tends to be much smaller than for other elements, leading to large uncertainties in its measurement (Smiley, 2005).

$F_{\text{key}}$ values of 2–10 were operated on $NH_4^+$ and nss-$SO_4^{2-}$ along with an $F_{\text{key}}$ of 10 on Al. Mild pulling with $F_{\text{key}}$ values of 2–8 on any combination of factors gave invalid solutions where factor profiles had negative loadings of $NH_4^+$ and nss-$SO_4^{2-}$. Variation of $Q$, IM and IS with $F_{\text{key}}$ (Fig. S4a–c, supplementary online data) following Lee et al. (1999) showed a drop in these values for solutions where nss-$SO_4^{2-}$ and $NH_4^+$ were pulled down with $F_{\text{key}}$ of 10 each in either 4 or 5 factors and $F_{\text{key}}$ of 10 on Al on all other than the Si-dust factor. It was found that these solutions were identical, as evidenced by their same $Q$ value (Fig. S4a, supplementary online data), and identical factor composition and were accepted as the final solution.

4. Interpretation of factors

4.1. Identification of source categories

The accepted 6-factor solution had $F_{\text{peak}} = 0.8$ and $F_{\text{key}}$ matrix with all elements set to zero, except a value of 10 for nss-$SO_4^{2-}$ and $NH_4^+$ in five factors, other than the secondary sulphate–nitrate factor, and for Al in all factors other than the Si-dust factor. The resolved factors were secondary sulphate–nitrate factor derived from fossil fuel combustion explaining 100% of $NH_4^+$ and nss-$SO_4^{2-}$, 44% of EC and 92% of $NO_3^-$, a biomass combustion factor explaining 100% of $K^+$, 90% of OC and 53% of EC, a Si-dust factor now containing all the Si and Al, with other factors unchanged from the base solution, being the Ca-dust, trace elements and Zn factors. A comparison of the daily mean reconstructed submicron PM concentrations from all sources (Fig. 1), with the measured submicron PM concentrations shows that the identified factors effectively reproduce the measured mass and account reasonably for the variation in the submicron PM concentrations (slope = 0.90 ± 0.20; $R^2 = 0.76$).

The biomass combustion factor (Fig. 2a) is composed of OC (44%), $K^+$ (34%) and EC...
Biomass combustion in India includes residential biomass fuel combustion for cooking and open burning of forests and agricultural residue (Venkataraman et al., 2005, 2006; Streets et al., 2003a). The PMF analysis does not distinguish between these two source categories. The resolved factor explained 100% of K⁺, a tracer for biomass burning (Kuykendal and Masser, 1989) and 91% and 53% of the OC and EC, respectively. Source measurements show that OC predominates in aerosol emissions from large fires in open biomass burning (Andreae and Merlet, 2001) while the BC constitutes up to 50% of particle mass in emissions.

Fig. 1. Goodness of fit of PMF predictions versus measurements of particle mass concentrations.

Fig. 2. Factor profiles or relative abundance of chemical species on PMF resolved factors. (a) Biomass combustion, (b) secondary sulphate-nitrate, (c) trace elements, (d) zinc, (e) Si-dust and (f) Ca-dust.
from small, shielded fires in cooking stoves (Venkataraman et al., 2005).

The secondary sulphate–nitrate, Zn and trace elements factors are all associated with energy production and industrial processes, e.g., coal-fired electric power, iron and steel, cement, refinery and captive power plants in many industrial sectors (e.g., Kuykendal and Masser, 1989) and are industrial emissions factors. The secondary sulphate–nitrate factor is composed of 78% nss-SO$_4^{2-}$ and 18% NH$_4^+$. About 4% of the factor mass comes from EC (explaining 44% of its loading) and around 0.5% from NO$_3^-$ (explaining 93% of its loading) (Fig. 2b). Some of the trace element masses are also explained well in this factor such as V (40%), Ni (57%) and Cr (29%), among which the model predicts V reasonably, but Ni and Cr poorly (Table S2 supplementary online data). These tracers are present in emissions from coal-fly ash, open hearth steel furnaces and steel sintering plants and refinery process heaters. The trace elements factor is composed of 89% Si, and 9% Al, explaining all the Si and Al mass (Fig. 2e). Among a long list of sources, Fe, Al and Ti are all tracers of bituminous coal-fired power plants, coal-fly ash and soil dust, with Fe additionally emitted from open-hearth steel production. The Zn factor has 42% Zn and 12% EC and 25% MSA (Fig. 2d). The factor explained 100% of Zn, 19% of MSA. Zn is a tracer for iron and steel and non-ferrous metal production and residual oil combustion.

Dust was resolved into two factors, Si-dust and Ca$^{2+}$-dust. This is physically realistic in the Indian region since Ca$^{2+}$ is a predominant soil component in western India (Rastogi and Sarin, 2005), while Si-dust arises from desert regions (NW India, west Asia, north Africa). The Si-dust factor is composed of 89% Si, and 9% Al, explaining all the Si and Al mass (Fig. 2c). The factor explains almost all the Pb mass and about a third of the Cu, Cr and P mass, but these species are all poorly resolved by the model (Table S2, supplementary online data). The Ca-dust factor is composed of 40% Ca$^{2+}$, 18% OC, 28% Na$^+$, and 14% Cl$^-$ (Fig. 2f). The factor explains all the Ca$^{2+}$ mass (slope = 0.99; $R^2$ = 1), and a large part of the Na$^+$ and Cl$^-$ mass (however, predicted very poorly by the model with $R^2$ of 0.23 and 0, respectively, Table S2, supplementary online data). It may be noted that Na$^+$ has large measurement uncertainties and NaCl associated with dust aerosols is believed to be of marine origin (Rastogi and Sarin, 2005).

4.2. Identification of probable source regions and collocated sources

Examination of the factor contributions (Fig. 3) showed trends of high and low concentrations (corresponding to periods of high or low contributions) corresponding to four different time periods, during 4–8 March (ship traverse off the west coast of India), 10–16 March (north–south traverse in the Arabian Sea 0–20°N), 17–22 March (south–north traverse in the Indian ocean 0–10°S) and 26–29 March (west–east traverse towards the Bay of Bengal). In each time period, we estimate mean factor contributions (Table 2). Further, we identify probable source regions using TPSCF analysis, for each factor, using the 50th-percentile factor-contributed PM concentration as the criterion value.

4.2.1. Sources influencing coastal India and the northern Arabian Sea (4–8 March)

During 4–8 March, mean PM concentrations arose from biomass combustion (35%), the three industrial emissions factors (35%) and Si-dust (20%) (Table 2). TPSCF analysis showed similar transport pathways for the biomass combustion and the three industrial emissions factors. Probable source areas from surface transport (10–1000 m arrival heights) were in the eastern Indo-Gangetic plain, with preferred transport over south India or the Bay of Bengal (Fig. 4a). It may be noted that in the PSCF literature related to long-range transport (e.g., Hopke et al., 1995) high PSCF grids are found over ocean regions and represent preferred transport pathways, if there are no expected sources (like dimethyl sulphide from ocean plankton) in those grid cells.

The Si- and Ca-dust factors had probable source regions in north-west Indian desert regions, transported down the west coast or through the east coast and south India (Fig. 4b). Elevated transport (2500–5000 m arrival heights) TPSCF showed preferred pathways from south-east Asia between 510°N and 10°N (Fig. 4c). Therefore, industrial, biomass combustion aerosols and dust aerosols were transported in mixed surface plumes over India and in elevated plumes from south-east Asia over south India. This is consistent with INDOEX observational studies reporting correlation among
multiple tracers in ship (e.g., Quinn et al., 2002) and aircraft measurements.

Emissions inventory information shows a high density of biomass fuel used for cooking in the Indo-Gangetic plain (Venkataraman et al., 2005), an area source leading to surface emissions (high emissions flux locations based on the inventory for different source categories are shown as symbols in Figs. 4a, 5a, c, 7a; details in Figs. S5 and S6, supplementary online data). In March, open burning of forest occurs in central and north-east India (Figs. 4a, S5b, supplementary online data) and of crop residues (Figs. 4a, S5c, supplementary online data) in the western Indo-Gangetic plain and southwest India (Venkataraman et al., 2006). Therefore, all three biomass sources in India could have contributed to the biomass combustion factor during 4–8 March in surface transport. Probable source

![Factor contributions or mass concentrations of particle matter from each factor in each 24-h sample.](image)
regions in south-east Asia have both biofuel combustion and open biomass burning during March (Streets et al., 2003a). Satellite detection of burned area (Tansey et al., 2004) shows an active fire cycle between February and April, in woodland and shrubland in most of south-east Asia, and in forests in Indonesia. Interannual variability in the detection of fire-frequency by satellite sensors in south and south-east Asia showed larger fire frequencies in south and south-east Asia in March 1999, compared to other months and other years. Surface TPSCF identified source regions for industrial emissions, have coal-fired electric utilities located in West Bengal and Bihar (Chandrapura and Durgapur) and in Andhra Pradesh and Orissa (Ramagundam and Talcher—Figs. 4a, S6a, supplementary online data). Additionally, iron and steel plants are co-located with the high TPSCF values in east India between 20°N and 25°N in West Bengal and Bihar (Bokaro and Jamshedpur), and in Andhra Pradesh and Orissa (Vishakapatnam and Rourkela) co-located with the high PSCF cells at lower latitudes between 13°N and 16°N (Figs. 4a, S6b, supplementary online data). Sources of Zn located in this belt include zinc production, iron and steel production and lead smelters. Elevated TPSCF source regions for industrial emissions lay in southeast Asia (Fig. 4c), where petroleum extraction and refineries and coal burning in electric utilities, cement, iron and steel production and non-ferrous metal smelters account for 30% of SO2 and 7% of NOx emissions in Asia (EDGAR, 1995). In summary, biomass combustion biofuel use for energy and forest/crop-residue burning in India, forest burning in south-east Asia, industrial emissions from India and south-east Asia and dust from north-west India are probable sources of aerosols during 4–8 March.

4.2.2. Sources influencing the Arabian Sea (10–16 March)

During 10–16 March, mean PM concentrations were predominantly from the biomass combustion (31%), secondary sulphate–nitrate (24%) and Ca-dust (15%) factors (Table 2). Probable source regions from surface TPSCF for the three industrial emissions factors and two dust factors were located in west Asia, the northern Arabian peninsula and regions to the north and west of India with southern transport into the Arabian Sea (Fig. 5a), with some pixels in central and east coast India. TPSCF using elevated trajectories (2500–5000 m) showed probable source regions for the industrial emissions and dust factors largely over the Indo-Gangetic plain, east coast India and south India (Fig. 5b). A few pixels in south-east Asia and the Arabian peninsula were identified. Probable source regions from surface TPSCF affecting the biomass combustion factor were spatially distinct from those for industrial and dust emissions during this period. They lay to the north and west of India (Fig. 5c). Two high TPSCF pixels in India were on the west coast in Maharashtra and in central India in Madhya Pradesh. Elevated transport pathways for biomass combustion were in ocean regions off the coast of north sub-Saharan Africa and from south-east Asia along with some high TPSCF pixels in east India (Fig. 5d).

Industrial emissions sources in high TPSCF source regions in west Asia include petroleum extraction and refineries, coal burning in electric utilities, cement and iron and steel production accounting for 30% of SO2 and 10% NOx emissions from Asia (EDGAR, 1995). Elevated TPSCF source regions in the Indo-Gangetic plain, have large point sources with emissions injected at stack heights of 200–250 m that may be lofted out of the boundary.

Table 2
Average source contributions from PMF during four identified time periods of ship cruise

<table>
<thead>
<tr>
<th>Source name</th>
<th>Average contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4–8 March</td>
</tr>
<tr>
<td></td>
<td>µg m⁻³</td>
</tr>
<tr>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Biomass burning</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>35</td>
</tr>
<tr>
<td>Secondary sulphate–nitrate</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>21</td>
</tr>
<tr>
<td>Trace elements</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Zinc</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Si-dust</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Ca-dust</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
layer and transported via the elevated transport pathway. High TPSCF values for the Ca and Si-dust factors were co-located with desert regions of west Asia and the Arabian peninsula in this period in contrast to the earlier period, when they arose from north-west India.

We are unable to identify significant biomass combustion sources in west Asia and large fires in this region appear to arise from oil-well flares associated with petroleum drilling (Streets et al., 2003a). The high TPSCF in this region associated with the biomass combustion factor (Fig. 5c) seems to be spurious and may result from the EC loading on this factor, which could also arise from oil-well flares. Significant burning of woodlands, shrubland and some broadleaf forest are assigned to both northern sub-Saharan Africa during December to March (Tansey et al., 2004). Emissions from biofuel combustion for energy and open biomass burning are assigned to south-east Asia (Streets et al., 2003a) during March. The two high TPSCF pixels in India are likely forest burning (Figs. 5b, 5b, supplementary online data—Madhya Pradesh, central India) and crop waste open burning (Figs. 5b, 5c, supplementary online data—Maharashtra, west coast). In summary, biomass combustion from open burning of forests or woodland in north sub-Saharan Africa (elevated transport), industrial emissions from west Asia (surface transport) and the Indo-Gangetic plain (elevated transport) and dust from west Asia and the Arabian peninsula are likely sources of aerosols measured during 10–16 March 1999. This is consistent with recent general circulation modelling with region-tagged emissions (Verma et al., 2007), which identified 8–12 March 1999 as a period when emissions from Africa and west Asia dominated surface aerosols and especially aerosol optical depth (from elevated aerosol layers) in the northern Arabian Sea.

4.2.3. Sources influencing the tropical Indian Ocean (17–22 March)

During 17–22 March, when the Ron Brown sailed in the Indian Ocean between 0°S and 10°S, predicted aerosol concentrations were dominated by biomass combustion (35%), secondary sulphate-nitrate (33%) and Ca-dust (13%) (Table 2). For two industrial emissions factors (except Zn), biomass combustion and dust factors, TPSCF analysis shows two surface transport pathways largely over ocean regions, arising west of India with a few pixels in north Arabian peninsula and in south-east Asia.
Source regions from elevated transport TPSCF were on the east coast of Africa, north Africa and south-east Asia (Fig. 6b). Contribution from the Zn factor was low compared to other periods (Table 2). The Zn factor had source regions in a few scattered grid cells off the coast of Africa, India and China and was transported along pathways distinct from the other factors. Biomass combustion sources active in these locations are open burning of woodland and forests in Africa and of forests in south-east Asia (Tansey et al., 2004; Streets et al., 2003a), along with some biofuel combustion for energy in both regions. Industrial emissions in west Asia include oil refineries, iron and steel plants, cement plants and oil-well flares from petroleum processing, as discussed earlier. High TPSCF for the dust factors is co-located with desert in west Asia (surface transport) and north and east Africa (elevated transport). In summary, biomass combustion contributions from forest burning in Africa and south-east Asia, industrial emissions from west Asia and south-east Asia, and dust from west Asia and north Africa were likely sources of aerosols during this period.

4.2.4. Sources influencing the south Bay of Bengal (26–29 March)

During 26–29 March 1999, when the Ron Brown made a west to east traverse into the Bay of Bengal, dominant contributions were from biomass combustion (40%), secondary sulphate–nitrate (25%) and Si-dust (15%). Again, TPSCF analysis showed
different source regions for Zn from other factors. Probable source regions for the industrial emissions, biomass combustion and dust factors were all coincident in this period and lay in west Asia, with surface transport down the Indian west coast and a pathway over the Indo-Gangetic plain, central India transported down the east coast of India (Fig. 7a). Probable source regions contributing through elevated transport were in south-east Asia and a couple of pixels on the east coast of India (Fig. 7b). The Zn factor had probable source regions only in east India in this period (elevated TPSCF branch down the east coast in Fig. 7a), possibly associated with the iron and steel and non-ferrous metals production in this region.

Sources of biomass combustion located in these regions are biofuel combustion and forest burning in central India (Figs. 7a, S5a, b, supplementary online data) and forest burning in south-east Asia, as discussed for an earlier period. Industrial emissions sources located in these regions are energy transformation sectors such as oil refineries and electric utilities, and industries such as iron and steel and cement plants in west Asia, coal-fired utilities (Talcher in Orissa) and iron and steel production (Vishakapatnam in Andhra Pradesh and Rourkela in Orissa) in central India (Figs. 7b, S6a, b, supplementary online data). High TPSCF regions for dust are coincident with desert in west Asia. In summary, biomass combustion contributions from biofuels for energy in India, forest burning in India and south-east Asia, industrial emissions from west Asia and India, and dust from west Asia are probable sources of aerosols in this period.

5. Conclusions

PMF analysis identified six factors influencing PM concentrations during the INDOEX cruise of the Ron Brown including a biomass combustion factor (35–40%), three industrial emission factors (35–40%), primarily secondary sulphate–nitrate, balance trace elements and Zn, and two dust factors (20–30%) of Si- and Ca-dust. The biomass combustion factor explains all the OC (45% of the factor loading) and most of the BC (20% of factor loading) concentrations. The three industrial emissions factors explain all the secondary sulphate, nitrate and ammonium and a small fraction of the BC concentrations. Probable source regions for the biomass combustion factor shifted from India (4–8 March) to Africa and south-east Asia (10–22 March) and back to India along with south-east Asia (26–29 March). Sources located in these regions are biofuels for energy and crop-residue/forest burning in India, woodland/forest burning in Africa and forest burning in south-east Asia. Probable source regions for industrial emissions were in India (4–8 March), west Asia and India (10–16 and 26–29 March) and in west Asia and south-east Asia (17–22 March). Collocated sources include coal-fired electric utilities, cement, metals and petroleum production in both India and west Asia. Dust aerosols arose from north-west India.
(4–8 March) and from west Asia during the remaining period, with some contribution through elevated transport from Africa (17–22 March). Significant findings from the PMF–PSCF source-apportionment modelling in this study, are equivalent contributions of biomass combustion and industrial emissions to measured aerosol surface concentrations, the origin of carbonaceous aerosols largely from biomass combustion and the identification of probable source regions in Africa, west Asia,
the Arabian peninsula and south-east Asia, in addition to India, which affected particulate matter concentrations over parts of the INDOEX domain covered by the Ron Brown cruise.

Acknowledgements

This paper is dedicated to the late Sheldon Friedlander, pioneer in receptor modelling, in fond reminiscence of his suggestion, in 1999, that source identification of INDOEX aerosols is a worthwhile goal for aerosol scientists. We acknowledge Dr. Particia Quinn, NOAA-PMEL, for sharing Ron Brown data and Prof. Greg Carmichael and Sarika Kulkarni of the University of Iowa for helpful discussions.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.atmosenv.2008.02.041.

References


