Article (refereed) - postprint

COMPARISON OF THREE TECHNIQUES FOR ANALYSIS OF DATA FROM AN AEROSOL TIME-OF-FLIGHT MASS SPECTROMETER

Chiara Giorio[1][*], Andrea Tapparo[1], Manuel Dall'Osto[2], Roy M. Harrison[3][,5], David C.S. Beddows[3], Chiara Di Marco[4] and Eiko Nemitz[4]

[1]Dipartimento di Scienze Chimiche
Università di Padova, Via Marzolo 1, 35131 Padova
Italy

[2]Institute of Environmental Assessment and Water Research (IDÆA)
Consejo Superior de Investigaciones Científicas (CSIC)
C/ LLuis Solé i Sabarís S/N, 08028 Barcelona
Spain

[3]National Centre for Atmospheric Science
School of Geography, Earth and Environmental Sciences
Division of Environmental Health and Risk Management
University of Birmingham
Edgbaston, Birmingham B15 2TT
United Kingdom

[4]Centre for Ecology & Hydrology
Bush Estate, Penicuik, Midlothian, EH26 0QB
United Kingdom

[5]Department of Environmental Sciences, Center of Excellence in Environmental Studies, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

*Corresponding author. Tel: +39 049 8275180; fax: +39 049 8275271. E-mail address: chiara.giorio@unipd.it (C. Giorio).

ABSTRACT

The Aerosol Time-of-Flight Mass Spectrometer (ATOFMS) is one of few instruments able to measure the size and mass spectra of individual airborne particles with high temporal resolution. Data analysis is challenging and in the present study, we apply three different techniques (PMF, ART-2a and K-means) to a regional ATOFMS dataset collected at Harwell, UK. For the first time, Positive Matrix Factorization (PMF) was directly applied to single particle mass spectra as opposed to clusters already generated by the other methods. The analysis was performed on a total of 56898 single particle mass spectra allowing the extraction of 10 factors, their temporal trends and size distributions, named CNO-COOH (cyanide, oxidised organic nitrogen and carboxylic acids), SUL (sulphate), NH4-OOA (ammonium and oxidized organic aerosol), NaCl, EC+ (elemental carbon positive fragments), OC-Arom (aromatic organic carbon), EC- (elemental carbon negative fragments), K (potassium), NIT (nitrate) and OC-CHNO (organic nitrogen). The 10 factor solution from single particle PMF analysis explained 45% of variance of the total dataset, but the factors are well defined from a chemical point of view. Different EC and OC components were separated: fresh EC (factor EC-) from aged EC (factor EC+) and different organic families (factors NH4-OOA, OC-Arom, OC-CHNO and CNO-COOH). A comparison was conducted between PMF, K-means cluster analysis and the ART-2a artificial neural network. K-means and ART-2a give broadly overlapping results (with 9 clusters, each describing the full composition of a particle type), while PMF, by effecting spectral deconvolution, was able to extract and separate the different chemical species contributing to particles, but loses some information on internal mixing. Relationships were also examined between the estimated volumes of ATOFMS PMF factors and species concentrations measured independently by GRAEGOR and AMS instruments, showing generally moderate to strong correlations.

KEYWORDS: Aerosol, ATOFMS, PMF, single particle analysis, K-means, ART-2a

1. Introduction

In the last decade numerous epidemiological studies have revealed a significant correlation between environmental particulate matter concentrations and adverse health effects. However, since most studies have used PM_{10} or $PM_{2.5}$ mass concentrations to investigate correlations with human health outcomes it is likely that the health impacts of PM have been in most cases underestimated (Harrison et al., 2010). Atmospheric aerosol is especially problematic to characterize because of its complex and variable composition, wide size range and a broad spectrum of both natural and anthropogenic sources. In this connection, on-line measurements deploying Mass Spectrometric techniques are very promising in order to characterize both aerosol size and chemical composition for a wide range of substances (Pratt and Prather, 2011). Aerosol Time-of-Flight Mass Spectrometry (ATOFMS) is particularly attractive as it allows size and chemical characterisation by measuring the aerodynamic diameter and positive and negative ion mass spectra of individual particles in real time within the diameter range of 0.1 to 3 μm (Rebotier and Prather, 2007; Gard et al., 1997; Dall'Osto et al., 2004; Drewnick et al., 2008). The ATOFMS can measure in a single campaign hundreds of thousands of single particle mass spectra which present a considerable data analysis challenge.

Successful analysis of ATOFMS data requires fast and reliable processing and interpretation of the huge amount of data generated. In order to reduce the time of analysis and the pre-deterministic nature of the manual classification, statistical methods can be used. The general aim of classification is to find a structure, i.e, groups of similar or related objects in the available data set (Hinds, 1999). The main difference between a clustering method and manual classification is that the clustering method has the ability to perform analysis over the whole spectrum, rather than as individual peaks. By applying a statistical algorithm to the ATOFMS dataset, the user bias of determining which chemical information is more important in the spectra is minimised. Therefore single particle data are usually treated with a clustering algorithm, such as K-means or ART-2a, in

89   order to group particles of similar size range and chemical composition (Rebotier and Prather, 2007;

90   Gross et al., 2010; Healy et al., 2009; Pekney et al., 2006).

91   In environmental studies, factor analysis techniques (PCA, PCFA, PMF) are widely used to perform

92   source apportionment from data taken at receptor sites. PMF analysis has been successfully applied

93   to 24h averaged data from analysis of particles collected on filters (Stortini et al., 2009; Jia et al.,

94   2010; Dogan et al., 2008; Bari et al., 2009; Alleman et al., 2010) whose principal limitation is the

95   possibility of losing the point source contributions as the characteristic time of plumes from local

96   sources is short. Thus the results obtained are usually limited to the extraction of the 3 or 4 main

97   sources like crustal, marine, combustion sources and secondary particulate matter, while other

98   sources can be extracted only with a wide range of chemical analyses, size segregation and more

99   frequent measurements (Pekney et al., 2006; Wexler et al., 2008). On the other hand, PMF applied

100   to high-resolution data (only obtainable for long periods with an on-line technique) can be a useful

101   tool for this purpose. For example PMF analysis was successfully applied to 1h semi-continuous

102   characterization data of both particulate and gas phase composition leading to the extraction of 6

103   main sources, while by combining ATOFMS and AMS (aerosol mass spectrometry) data to the

104   original dataset the PMF was able to identify 16 factors during a field campaign in Riverside, CA

105   (Eatough et al., 2008).

106   PMF has previously been applied to ATOFMS data *after* clustering by another technique (e.g.

107   McGuire et al., 2011), but not to data *before* clustering. In the present study, for the first time, PMF

108   analysis is directly applied to single particle mass spectra in order to deconvolve the different

109   chemical species which contribute to ambient particulate matter in a rural background location in

110   Harwell (UK). A comparison among three different data treatment techniques (PMF, K-means,

111   ART-2a) is also conducted. Hourly temporal trends of the factors extracted from single particle

112   analysis are compared to each other in order to highlight possible correlations and to study the

113   mixing state of ambient particles. Moreover, temporal trends of factors and clusters are compared

114 with independent ion (and non refractory organic carbon) measurements to evaluate the

115 performance of the data analysis.

116

117 **2. Methodology**

118 *2.1. Measurement Site and Instrumentation*

119 The sampling campaign was conducted in Harwell (51°34'32"N, 1°18'49"W), a rural background

120 site in Oxfordshire (UK) from the 4th October to the 17th October 2008 deploying two on-line mass

121 spectrometric instruments, an Aerosol Time-of-Flight Mass Spectrometer (ATOFMS TSI Model

122 3800-100) and an Aerosol Mass Spectrometer (Aerodyne high-resolution-ToF-AMS) (Drewnick et

123 al., 2005; DeCarlo et al., 2006; Canagaratna et al., 2007; Jimenez et al., 2003), and a GRAEGOR

124 (Thomas et al., 2009), which performs semi-continuous measurements of water-soluble trace gas

125 species (NH_3, HNO_3, HONO, HCl and SO_2) collected by two wet-annular rotating denuders and

126 their related particulate compounds (NH_4^+, NO_3^-, Cl^-, SO_4^{2-}) collected in series by two steam-jet

127 aerosol collectors (SJAC). Sample solutions are analyzed on-line by ion chromatography for anions

128 and flow injection analysis for ammonia and ammonium (Thomas et al., 2009). During the

129 campaign, the two inlets of GRAEGOR were placed at the same height (roughly XXX above

130 ground) collecting PTS and $PM_{2.5}$ simultaneously.

131 Hourly data for gaseous pollutant concentrations measured as part of the UK national air quality

132 network and local weather were obtained from the UK national air quality archive

133 (www.airquality.co.uk). Five day air mass back-trajectories arriving at Harwell at three different

134 altitudes (100, 500 and 1000 metres) were obtained using HYSPLIT (Hybrid Single Particle

135 Lagrangian Integrated Trajectory Model) (Draxler and Rolph, 2003). Details of Harwell aerosol

136 characterization and air mass trajectories have been provided in supplementary material.

137

138 *2.2. ATOFMS Technique*

139   The ATOFMS (TSI 3800-100) collects, in real-time, bipolar mass spectra of individual aerosol

140   particles. The instrument is constituted by an aerosol inlet, a sizing region and a mass spectrometer

141   detector. In the aerosol inlet, particles are introduced into a vacuum system region through a

142   converging nozzle, then focused through aerodynamic lenses into a narrow particle beam, which

143   travels through the sizing region. The aerodynamic diameter of individual particles is determined

144   from the time of flight between two continuous-wave laser beams ($\lambda = 532$ nm). After that, particles

145   enter into the mass spectrometer region where a pulsed high power desorption/ionization laser ($\lambda =$

146   266 nm) is triggered on the basis of the transit time of the particle measured in the sizing region.

147   Mass analysis is then provided by a bipolar time of flight reflectron mass spectrometer (Gard et al.,

148   1997; Dall'Osto et al., 2004; Drewnick et al., 2008).

149   During the campaign, the ATOFMS sampled aerosol through a 3/4 inch diameter copper pipe

150   mounted vertically and in-line with the Aerodynamic Focussing Lens (AFL). The inlet of the

151   copper pipe (roughly 4m above the ground) was protected using a simple hockey stick rain cap. The

152   ATOFMS itself was fitted with a TSI 3800-100 AFL which admitted the aerosol at nominal

153   volumetric flow rate of 0.1 L/min operating at a pressure of 2 torr. The device has a quoted size

154   range of 100-3000 nm (Su et al., 2004) although in practice during the sampling campaign our

155   system was capable of hitting 56898 particles with a measured aerodynamic diameter up to 3019

156   nm.

157   Before data analysis, single particles mass spectra were exported using the TSI MS-Analyze

158   software. The peak-list were constructed using the following parameters: minimum peak height of

159   20 units above the baseline, minimum area of 20 units and representing at least the 0.005% of the

160   total area in the particle mass spectrum. The data obtained were analysed using positive matrix

161   factorization (PMF), K-means cluster analysis and artificial neural network (ART-2a) analysis.

162

163   *2.3. Positive Matrix Factorization (PMF) Analysis*

164   The PMF analysis was performed using the program PMF2 (Paatero and Tapper, 1994; Paatero,

165   1998). Briefly, the positive matrix factorisation model (whose principles are detailed elsewhere

166   (Paatero, 1994; Paatero, 1998)) solves the following equation $X = GF + E$ where $\mathbf{X}$ is the original

167    $n \times m$ data matrix, $\mathbf{G}$ is the $n \times p$ scores matrix (factors weight) and $\mathbf{F}$ is the $p \times m$ loadings matrix

168   (factors profile), $\mathbf{E}$ represents the $n \times m$ residuals matrix. In the present case n is the number of

169   particles, m is the number of m/z signals of the spectra and p is the number of factors. The exact

170   number of factors to use was determined by monitoring the parameters suggested by Lee et al.

171   (1999) and the chemical interpretation of the factors profile.

172   *Data matrices.* Before the PMF analysis the dataset was reduced to 106 major m/z values (-146, -

173   144, -124, -121, -119, -104, -101, -99, -98, -97, -96, -95, -89, -88, -85, -84, -81, -80, -79, -76, -73, -

174   72, -71, -64, -63, -62, -61, -60, -59, -49, -48, -46, -45, -44, -43, -42, -37, -36, -35, -27, -26, -25, -24,

175   -17, -16, -15, -14, -13, -12, 7, 12, 15, 18, 23, 24, 27, 36, 37, 39, 41, 43, 46, 48, 49, 50, 51, 52, 53,

176   54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 69, 70, 71, 72, 73, 74, 75, 77, 81, 83, 84, 85, 86, 87, 88,

177   91, 94, 96, 108, 115, 118, 120, 128, 132, 138, 139, 207) and 55357 particles by eliminating the bad

178   variables (the ones that have more than 55000 zero point values on a total number of particles of

179   56898) and the particles with a diameter below the calibration range. Absolute area of peaks was

180   considered for the analysis, which was directly applied to single particle mass spectra.

181

182   *2.3.1. Data uncertainties*

183   Positive Matrix Factorization relies on the accuracy of error estimates to produce reliable non-

184   negative results and uses the estimates of the error in the data to provide both variable and sample

185   weighting. This is particularly important when less robust datasets have to be used because of the

186   presence of many missing or below detection limit values, as in the case of mass spectra, that could

187   have the ability to define real sources or even be source markers (Owega et al., 2004; Paatero and

188   Taper, 1994; Paatero, 1998; Zhang et al., 2008). The original noise of the data ($\bar{x}_b = 4$, $\sigma_b = 4$),

189   evaluated in zones of particle mass spectra without peaks, was added to the input matrix by

190   simulating it with random numbers between 0 and 8, to avoid multiple zero entries. In fact, circa

191   70% of data in the input matrix are null values. The detection limit was evaluated as the blank value

192   plus three times its standard deviation by integrating the mass spectra signals in several regions

193   without peaks. The uncertainty of the data was evaluated in a laboratory experiment in which

194   equimolar solutions of various salts were nebulised and analyzed with the ATOFMS. The data

195   reproducibility was about 50% and 80% on the average signals for positive and negative ions

196   respectively. Moreover, the particle diameter does not influence the signal intensity. These high

197   uncertainties reflect the principal limits of the ATOFMS analyzer which reside in the size-

198   dependent transmission losses (Allen et al., 2000, Wenzel et al., 2003), laser intensity shot-to-shot

199   variations (Bhave et al., 2002), ionization matrix effects (Reilly et al., 2000), different sensitivities

200   among chemical species that make a semi-quantitative analysis possible to achieve only beside

201   independent sampling measurements (Bhave et al., 2002; Gross et al., 2000; McGuire et al., 2011).

202   The data uncertainties used for the PMF analysis were then calculated as follow $s_{ij} = t + v \cdot x_{ij}$,

203   where $t=4DL=64$ and $v=0.4$ in order to give the same weight to both low and high intensity signals

204   and to avoid the effect of background noise upon the analysis. The data uncertainty of 40% was

205   chosen because there were no further improvements by using a higher uncertainty or different

206   uncertainties for positive and negative ions in terms of quality of the fit and explained variations.

207   Although Q/Qexp = 0.43 could indicate a slight overestimation of real data uncertainty, the

208   optimized value seems to be a good compromise considering laboratory experimental data.

209

210   *2.3.2. PMF solution*

211   The robustness of factor solutions was inspected by comparing the temporal trends of factors

212   through the different PMF solutions. The global minimum of the factor solution was achieved by

213   starting from 50 seeds (pseudorandom starting points). The rotational ambiguity was also tested by

214   modifying the Fpeak parameter from -2.5 to 2.5. The effect of this variation was not significant with

215   values in the range -0.5 – 0.5 while PMF analysis did not converge with larger Fpeak values. Thus

216 the PMF solution obtained could be considered unique and Fpeak=0 was used for the final analysis.

217 After the PMF analysis, factor loadings (**F**) and scores (**G**) obtained were respectively normalized

218 and weighted as follows: each factor loading vector was normalized by dividing it by a scalar value

219 $b_h = \sum_{j=1}^{m} f_{hj}$ and the corresponding score vector was weighted by multiplying it by the same scalar

220 b_h.

221

222 *2.4. Cluster Analysis*

223 *2.4.1. K-means*

224 ATOFMS particle mass spectra were directly imported into ENCHILADA, an open source single

225 particle mass spectra software package (Gross et al., 2010), and 56898 single particle mass spectra

226 were clustered using the K-means/Euclidean square algorithm (McQueen, 1967). K-means, which is

227 a non hierarchical clustering technique, starts with the random subdivision of objects (in this case

228 single particles) into a number of clusters previously defined by the operator. The algorithm

229 computes the total heterogeneity of the system $E_T = \sum_{c=1}^{C} \sum_{i=1}^{I_c} \sum_{v=1}^{V_c} (x_{ivc} - \bar{x}_{vc})^2$, which is related

230 to the Euclidean distance of every object to the centroid of the cluster to which the object belongs

231 to, and moves objects from a cluster to another until it finds the minimum of system heterogeneity

232 (McQueen, 1967; Gross et al., 2010). In the current study, data analysis was repeated several times

233 with increasing numbers of clusters. The exact number of clusters to use was chosen by monitoring

234 E_T and the chemical interpretation of the cluster centroid mass spectra.

235

236 *2.4.2. ART-2a*

237 The ATOFMS dataset was imported into YAADA (Yet Another ATOFMS Data Analyzer) and

238 single particle mass spectra were grouped with Adaptive Resonance Theory neural network, ART-

239 2a (Song et al., 1999). The parameters used for ART-2a in this experiment were: learning rate 0.05,

240 vigilance factor 0.85 and iterations 20. These are standard setting used in the ART-2a procedure on

241 ATOFMS data and further details of the parameters can be found elsewhere (Song et al., 1999;

242 Dall'Osto and Harrison, 2006; Rebotier and Prather, 2007). An ART-2a area matrix (AM) of a

243 particle cluster represents the average intensity for each m/z for all particles within a group. An

244 ART-2a AM therefore reflects the typical mass spectrum of the particles within a group.

245

246 *2.5. Positive Matrix Factorization of AMS data*

247 Standard unit mass resolution PMF analysis was carried out on the organic matrix of the AMS

248 dataset (Ulbrich et al. 2009). Two general factors were found: LV-OOA (low-volatile oxidized

249 organic aerosol) and a SV-OOA (semi-volatile oxidized organic aerosol). Whilst the mass spectrum

250 of LV-OOA factor was found to be equivalent to previous standard factor (Ulbrich et al. 2009), the

251 factor SV-OOA contains the standard aliphatic series together with a high m/z 44 and m/z 60

252 signals, indicating a contribution from biomass burning (Lanz et al. 2007).

253

254

255 **3. Results and discussion**

256 *3.1. PMF Analysis on Individual Particle Mass Spectra*

257 Single particle mass spectra were subjected to Positive Matrix Factorization analysis with solutions

258 varying from 3 to 15 factors. According to both mathematical parameters and chemical

259 interpretation of factor profiles, the 10 factor solution was selected. The factors extracted are:

260 • F1 "CNO-COOH", explaining 2% of variance, presents peaks of (CN^-) (m/z -26) and oxidised

261   species (CNO^-) (m/z -42), $(CHOO^-)$ (m/z -45) and (CH_3COO^-) (m/z -59), i.e. carboxylic acids

262   and organic nitrogen species (Angelino et al. 2001, Dall'Osto and Harrison 2006, Moffett et al.,

263   2008);

264 • F2 "SUL" explaining 2 % of variance, is characterized by the main peak of sulphate (m/z -97);

265 • F3 "NH4-OOA" with an explained variation of 4%, is characterized by peaks of (NH_4^+) (m/z 18)

266   and secondary organic species $(C_2H_3^+)$ (m/z 27) and $(C_2H_3O^+)$ (m/z 43);

267 • F4 "NaCl" explaining 6% of variance, is characterized by peaks of (Na^+) (m/z 23), (Na_2^+) (m/z

268   46), (Na_2O^+) (m/z 62), (Na_2OH^+) (m/z 63) and (Na_2Cl^+) (m/z 81/83);

269 • F5 "EC+" explaining 7% of data variation, contains the elemental carbon positive ions $(C^+, C_2^+,$

270   C_3^+ at m/z=+12,+24,+36);

271 • F6 "OC-Arom" explaining 5% of variance, contains signals related to organic carbon and the

272   benzene fragment (m/z 27, 41, 43, 51, 53, 55, 57, 63, 69, 77, 87, 91, 115) (McLafferty, 1983);

273 • F7 "EC-" explaining 3%, is characterized by elemental carbon signals in the negative mass

274   spectrum $(C^-, C_2^-, C_3^-$ at m/z=-12,-24,-36);

275 • F8 "K" explaining 7%, contains the potassium signals (m/z 39/41);

276 • F9 "NIT" explaining 4%, is characterized by the nitrate peaks (m/z -46/-62);

277 • F10 "OC-CHNO" with an explained variation of 5%, is characterized by organic carbon and

278   organic carbon related to nitrogen signals (m/z -26, 27, 37, 49-52, 60-63, 84-87).

279 The 10 factors obtained can explain only 45% of the total data variance but they are characterized

280 by clear and well defined chemical patterns (Figure 1). Despite the low explained variance, the

281   main signals constituting the factors are well represented and they account for up to 89% of the

282   variance of potassium for example. Sulphate is explained at 84%, while the majority of the bad

283   variables (m/z values with low signal/noise ratio, i.e. m/z=-146, -144, -124, -121, -119, -104, -101)

284   are not explained at all.

285   From inspection of residuals (Figure S1) it appears that the PMF analysis failed to extract a few

286   components: this includes chloride signals (m/z = -35,-37), which are not present in the NaCl factor,

287   water signals and some other signals probably related to m/z miscalibration problems (Dall'Osto

288   and Harrison, 2006); however, these signals do not influence the interpretation of factors. It should

289   be noted that despite the limited explained variance, which could be a problem in relation to

290   quantification, the factors' chemical profiles obtained are clear and well-defined and thus of

291   qualitative value with the only exceptions of chloride and water signals.

292   The results obtained demonstrate that Positive Matrix Factorization analysis applied to individual

293   particle mass spectra allows the deconvolution of the mass spectra into the contributing specific

294   chemical species (factors K, NIT, SUL, NaCl) or their related classes (factors EC+, EC-, OC-Arom,

295   OC-CHNO, CNO-COOH, NH4-OOA) as well as the extraction of their temporal trends and size

296   distributions (Figure S2). Positive and negative m/z signals are split into different factors (EC+ and

297   EC-, K, NIT, SUL for example) due to different temporal trends either representing changing

298   source contributions or varying relative ionization efficiencies (Bhave et al., 2002; Dall'Osto et al.,

299   2006; Gross et al., 2000). Unlike K-means or ART 2a, PMF does not cluster whole spectra, but

300   disaggregates them into chemical constituents, or groups of constituents. The factors are used to

301   reconstitute actual particle mass spectra as shown in Figure S3. From Figure S3, it may be seen that

302   more than one factor is necessary to reconstruct each particle mass spectrum, demonstrating the

303   mass spectral deconvolution made by PMF analysis on single particles. For example, in Figure S3a,

304   the particle mass spectrum is reconstructed by 10% CNO-COOH, 3% SUL, 6% NH4-OOA, 26%

305   OC-Arom, 10% EC-, 8% K, 8% NIT, and 29% remains unexplained.

306   Factor time-series were calculated as hourly sum of factor scores (not shown) and in equivalent

307   numbers of particle (Figure 2a). Factor time series in equivalent number of particles are calculated

308   as sum over each hour of particle fractions attributable to each factor by first calculating the fraction

309   of particle i attributable to factor h as

310   $$fF_{ih} = \frac{\sum_{j=1}^{m} g_{ih} f_{hj}}{\sum_{j=1}^{m} (g_{ih} f_{hj} + e_{ij})}$$ (1)

311   and then summing over each hour of particle fractions attributable to factor h:

312   $$NfF_{h,hour} = \sum_{hour} fF_{ih}$$ (2)

313   The number size distributions were calculated by summing the factor scores of particles within the

314   same size bin (size bin width of 0.01 μm). The factor size distributions are very similar to each

315   other and all are dominated by the accumulation mode. The only exceptions are F4-NaCl, which

316   presents a coarse distribution because of its origin from sea spray, and F9-NIT which presents both

317   an accumulation and a coarse mode (Figure S2). Moreover, EC- and OC-CHNO factors clearly

318   show a distribution that is shifted towards smaller particles with a tail in the direction of the Aitken

319   mode particles. Despite not being corrected for size-dependent inlet efficiencies, these distributions

320   show predictable differences.

321   The analysis of the correlations between temporal trends of the factors, obtained by summing the

322   score values of each factor within an hour, may give deeper insight into particle components and

323   their sources. Correlations between factors were studied through the correlation coefficients (Figure

324   3 and Table S1) in the Pearson correlation test. Almost every correlation is statistically significant

325   (p-value < 0.05) but to different degrees. The NaCl seems to be an independent factor because it has

326   no strong correlations with the other factors, according to the Cohen classification (Cohen, 1988)

327   and it is not correlated to sulphate, EC-, potassium, nitrate and OC-CHNO (p-value>0.05). Sulphate

328   is strongly correlated with potassium (r=0.64), nitrate (r=0.76) and the organic carbon factors, OC-

329   Arom (0.82) in particular. Potassium and nitrate are strongly correlated with almost every factor and

330   are the dominant species, present in the majority of the particles collected. This reflects, at least in

331   part, the very high sensitivity of the ATOFMS to these species (Gross, 2000).

332   Hourly temporal trends of EC- and EC+ present only a correlation of medium strength (r=0.41).

333   EC- is strongly correlated with OC-CHNO (r=0.64) while EC+ is correlated more with secondary

334   species (r coefficients for NH4-OOA, 0.87 and NIT, 0.61). This result suggests that the splitting of

335   elemental carbon signals into two factors may not only reflect different ionization and detection

336   efficiencies between positive and negative ions. It seems that the ionization pattern is influenced by

337   the matrix composition (Reilly, 2000) distinguishing two different elemental carbon components:

338   one probably freshly emitted (EC-) and one more aged (EC+), modified by oxidation reactions, and

339   internally mixed with secondary species. In fact, as proposed by Reinard and Johnston (2008)

340   secondary species like nitrate and sulphate could limit the electron availability, leading to a

341   suppression of elemental carbon fragments negatively charged, while potassium and sodium, on the

342   contrary, could lead to an enhancement of them. Moreover, the temporal trend of the EC- is

343   characterized by a peak event on 16/10/2008 probably due to a combustion event near the sampling

344   site.

345

346   *3.2. Cluster Analyses*

347   *3.2.1. K-means*

348   The K-means analysis separated 13 clusters. Clusters obtained from miscalibrated mass spectra

349   were eliminated and clusters with similar profiles and temporal trends were recombined to generate

350   a total of 9 clusters (mass spectra are reported in Figure S4a and their temporal trends, expressed as

351   the number of particles are reported in Figure 2b). The clusters are:

352   • K (14140 particles, 25%), which presents high potassium signals and some signals of low

353   intensity due to Na^+, cyanide, nitrate and sulphate;

354 • K-EC (3252 particles, 6%), which presents negative ions signals related to elemental carbon, and

355   to a lesser extent nitrate and sulphate signals, while in the positive mass spectrum it presents

356   signals of a low intensity, related to oxidized organic carbon, potassium and sodium;

357 • NaCl (10872, 19%), which mainly presents signals of sodium, chloride, potassium and nitrate;

358 • EC (9436 particles, 17%), which presents both positive and negative signals related to elemental

359   carbon and signals of nitrate and sulphate;

360 • K-SUL-OC-NIT (1832 particles, 3%) presents CN^-, NO^-, NO_2^-, SO_3^-, HSO_3^-, HSO_4^- signals in

361   the negative mass spectrum and potassium and OC aromatic signals in the positive mass

362   spectrum;

363 • OC (4625 particles, 8%) presents both aromatic, amine and oxygenated carbon signals and traces

364   of ammonium, nitrate, sulphate and cyanide;

365 • K-NIT (6829 particles, 12%) is mainly characterized by potassium and nitrate signals along with

366   the presence of cyanide, sulphate, ammonium and oxidized organic aerosol fragments (m/z =

367   +27/+43);

368 • OOA (2006 particles, 4%), composed of signals corresponding to $C_2H_3^+$, $C_2H_3O^+$ and carboxylic

369   acids along with ammonium, potassium, nitrate and sulphate;

370 • Fe-V (840 particles, 1%), characterized by signals at m/z = +51/+56/+67 that could be assigned

371   respectively to V^+, Fe^+ and VO^+ and by signals at m/z = +58/+60 that could be attributed to

372   nickel and, to a lesser extent, by sodium, potassium, elemental carbon and nitrate.

373

374 *3.2.2. ART-2a*

375 The ART-2a algorithm generated 389 clusters used to describe the dataset (total particles 56898).

376 The 50 most populated clusters represent more than 63% of the mass spectra from the study and

377 thus were used for the results presented in this paper. The remaining clusters were mostly made up

378 of a majority of miscalibrated mass spectra. By manually merging similar clusters according to their

379 chemical and temporal profiles with the standard procedure elsewhere described (Dall'Osto and

380 Harrison, 2006), the total number of clusters describing the whole database was reduced to 9,

381 representing about 63% of the total number of particles sampled (Figure S4b). The rest of the

382 particles presented low signal to noise ratios and therefore were not classified. The 9 clusters are:

383 • K-NIT (9613 particles, 17%) composed by potassium, nitrate, cyanide and sulphate;

384 • NaCl (7852 particles, 14%) characterized by a mass spectrum identical to K-means NaCl;

385 • OC (3172 particles, 6%) composed mainly by oxidized organic aerosol and aromatic

386   compounds, and potassium, cyanide, nitrate and sulphate signals;

387 • K-SUL (2355 particles, 4%) with high potassium and sulphate signals, along with ammonium,

388   nitrate and aromatic organic compounds;

389 • EC (5416 particles, 10%) which present a mass spectrum identical to the K-means EC;

390 • K (1656 particles, 3%) with an high potassium signal and Na^+, C_3^+, nitrate and sulphate signals;

391 • EC-Fe-V (1337 particles, 2%) composed by high signals of elemental carbon, and V^+, Fe^+ and

392   VO^+ in the positive mass spectrum while it does not present significant signals in the negative

393   mass spectrum;

394 • SOA (1066 particles, 2%) composed by ammonium, C_2H_3+, C_2H_3O+, potassium, low elemental

395   and organic carbon signals, nitrate and sulphate;

396 • K-EC (3390 particles, 6%) elemental carbon signals, potassium, sodium, cyanide, nitrate and

397   sulphate.

398 The mass spectra of the 9 clusters are shown in Figure S4b and their time-series, expressed as

399 number of particles are reported in Figure 2c. Results obtained from ART-2a analysis are very

400 similar to the K-means results. The two NaCl and EC clusters present the same mass spectrum. The

401 two K-EC clusters are similar but the ART-2a cluster is characterized by higher cyanide, nitrate and

402 sulphate signals. The OC ART-2a cluster presents higher aromatic signals than OC K-means

403 cluster. The K-NIT K-means cluster presents aromatic organic carbon signals which are indeed not

404 present in the K-NIT ART-2a cluster. The OOA K-means cluster has a mass spectrum similar to the

405 SOA ART-2a cluster, with a high contribution of NH_4^+, $C_2H_3^+$ and $C_2H_3O^+$, but the former presents

15

406   carboxylic acids signals while the latter presents more aromatic organic compounds signals. The

407   main differences reside in the abundance of the K cluster, which is probably overestimated by K-

408   means (25% of particles in the K-means clustering and 3% of particles in the ART-2a clustering),

409   the EC-Fe-V ART-2a cluster which exhibits strong elemental carbon signals that are not present in

410   the Fe-V K-means cluster, and K-SUL ART-2a cluster which has a different positive mass

411   spectrum, dominated by the potassium signal, while the K-SUL-OC-NIT K-means cluster has OC

412   aromatic signals in the positive mass spectrum.

413   The differences between the two techniques could reside in the different approach to clustering the

414   data. In K-means cluster analysis, all particles are assigned to the clusters by dividing them into

415   groups of similarity. The number of clusters is chosen by the operator who proceeds with a trial-

416   and-error approach by incrementing the number of clusters until the division into more clusters is

417   chemically meaningless (13 clusters in this case). On the contrary, ART-2a (running with standard

418   parameters) usually produces a huge number of clusters (389 in this case). After that, clusters made

419   by only few particles are eliminated and only the main contributing clusters are considered, and

420   clusters of similar composition and size distribution are merged manually. Thus, ART-2a may give

421   more clear and well defined clusters than K-means which considers more particles than the former

422   in the final solution.

423

424   *3.3. Comparison between Results of PMF Analysis on Single Particles, K-means Cluster Analysis*

425   *and ART-2a Artificial Neural Network Analysis*

426   PMF and cluster analysis can be viewed as complementary techniques. While K-means and Art-2

427   give a rapid classification of whole particles by dividing them into classes of similarity the PMF

428   analysis on single particle mass spectra permits the extraction of the chemical species constituting

429   the particles. Much of the information on internal mixing is lost.

430   The results of the correlation analysis (Pearson correlation test) among cluster and factor temporal

431   trends (in equivalent number of particles) which have a similar chemical profile show a good

432   agreement (Table 1). For instance, taking into account K-means results, PMF F8-K is correlated to

433   the K cluster with $r^2 = 0.99$ and p-value = <0.001 (Pearson correlation test); F4-NaCl is correlated to

434   the NaCl cluster ($r^2 = 0.91$, p-value = <0.001) and F3 NH4-OOA is correlated to the OOA cluster (r^2

435   = 0.83, p-value = <0.001). The cross correlations between factors and cluster temporal trends

436   confirm the conclusions obtained from the cross correlations between PMF factors (Figure 3 and

437   Table S1). The NaCl cluster presents a strong correlation only with the NaCl factor. In fact, NaCl is

438   an independent particle type which is directly associated with the sea spray source. The EC+ factor

439   is strongly correlated with clusters characterized by secondary aerosol (OOA r^2=0.53, K r^2=0.59, K-

440   NIT r^2=0.36) while EC- is not strongly correlated with any cluster, confirming the two different

441   elemental carbon contributions to aged (EC+) and fresh (EC-) particles. The cluster Fe-V is strongly

442   correlated with the EC+ factor (r^2=0.49) probably because of a common origin from oil based fuel

443   combustion (Korn et al., 2007) or transported from coal-fired power plants in Central Europe. In

444   fact, EC+ abundance increased during long-range transport of air masses from Central Europe (see

445   SI). Moreover, the K-EC cluster which is moderately correlated to organic factors as well as EC-,

446   could represent a biomass burning signature (Bi et al., 2011; Healy et al., 2012). CNO-COOH, SUL

447   and NH4-OOA PMF factors, as expected, are present in multiple clusters as they are highly

448   oxidized aerosol components produced during aging processes.

449

450   *3.4. Comparison of PMF Analysis Results with Independent Measurements*

451   Alongside the ATOFMS, inorganic water soluble components in the TSP and $PM_{2.5}$ size fractions

452   were measured by GRAEGOR, and in non-refractory PM_1 (NR-PM_1) by the AMS, defined as those

453   components within PM_1 that volatilise rapidly at the vaporiser temperature of 600°C. In order to

454   validate the PMF factor temporal trends, a correlation analysis (r-Pearson test) was made between

455   them and these independent measurements. In Figure 5 the sulphate, nitrate, chloride, ammonium

456   and organic concentrations are reported compared to the corresponding PMF ATOFMS factors.

457   For this purpose, factor temporal trends were calculated under the simplifying assumptions that all

458   particles are homogenous, spherical and a constant mass of material is ionized from each particle,

459   irrespective of their size (Dall'Osto et al., 2006). Particle volume was multiplied by the percentage

460   contribution of each factor to it. The hourly time-series (in volume) of the factors were then

461   calculated by summing the partial volume of each particle attributable to each factor (Figure 4). For

462   comparison with AMS PM_1 concentrations, PMF factor partial volumes were integrated for

463   particles of < 1 μm diameter. It is important to note that ATOFMS time-series were not corrected

464   for size-dependent inlet efficiencies (Dall'Osto et al., 2006).

465   The SUL factor (expressed in volume of particles) is significantly correlated with sulphate

466   concentrations in $PM_{2.5}$ ($r^2 = 0.34$, p-value = <0.001) and in AMS PM_1 ($r^2 = 0.41$, p-value =

467   <0.001). In the case of nitrate, the NIT PMF factor temporal trend is weakly correlated with nitrate

468   concentration in $PM_{2.5}$ ($r^2 = 0.07$, p-value = <0.001), but is strongly correlated with nitrate in NR-

469   PM_1 ($r^2 = 0.54$, p-value = <0.001). The difference in the correlations may reflect different

470   instrumental inlet characteristics leading to different large particle contributions to the temporal

471   patterns. In fact, while NR-PM_1 is fairly specific to NH_4NO_3, $PM_{2.5}$ can contain also significant

472   amount of $NaNO_3$, produced by sea salt processing through HNO_3. NIT PMF factor presents both

473   an accumulation and a coarse mode, and the latter could be measured with higher efficiency than

474   the former, and would also contain contributions that are not included in the NR-PM_1. The high

475   correlations seen for the NR-PM_1 fraction are however reassuring.

476   The NaCl factor is weakly but significantly correlated to the chloride measurements in $PM_{2.5}$ ($r^2 =$

477   0.11, p-value = <0.001). On the contrary AMS chloride is not significantly correlated with the

478   GRAEGOR chloride measurements (p-value = 0.32 for TSP and 0.46 for $PM_{2.5}$), which shows

479   much larger concentrations, because the AMS only detects the non-refractory fraction which is

480   thought to be dominated by NH_4Cl. The NH4-OOA factor, which contains both OOA and

481   ammonium signals, is correlated with the ammonium concentration in $PM_{2.5}$ ($r^2 = 0.61$, p-value =

482   <0.001) and in NR-PM_1 ($r^2 = 0.59$, p-value = <0.001) and to the organic component measured by

483 the AMS ($r^2 = 0.60$, p-value = <0.001). The non-refractory organic concentration measured by AMS

484 is strongly correlated with ammonium concentration and presents the highest correlation with the

485 NH4-OOA PMF factor rather than with other organic factors. ATOFMS factors were also compared

486 to TSP ion measurements, but because of different inlet characteristics, the correlations are in

487 general weak or not significant and the results are not reported.

488 The analysis shows clearly that PMF factors are highly significantly correlated with the

489 corresponding chemical species mass concentrations, with a better agreement with NR-PM$_1$ (if

490 ATOFMS PMF factors are integrated for particles <1 μm). On the contrary, clustering analytical

491 techniques such as K-means and ART-2a cannot disaggregate the contribution of the different

492 chemical species present in the particles. For this reason, a direct comparison between the time-

493 series of a cluster and the mass concentration of one of its components is not appropriate. In fact,

494 such correlation would be highly dependent on particle mixing-state. Thus, the disaggregation of

495 species made by the PMF analysis (on single particles) proves very useful for quantification

496 purposes of the principal substances or classes of substances constituting the particles. The

497 determination coefficient, slope and intercept of the linear regressions between ATOFMS factors

498 and the species concentrations measured by AMS in NR-PM1 are reported in Supplementary

499 Material. (Table S2). Moreover, the correlation between PMF factors and the corresponding species

500 concentrations may be even stronger if ATOFMS data are corrected for size-dependent transmission

501 losses (Jeong et al., 2011).

502

503 *3.5. Comparison between ATOFMS-PMF factors and AMS-PMF factors for secondary organic*

504 *aerosol*

505 In order to further validate the PMF analysis on single particle ATOFMS spectra, the factors

506 obtained were compared with standard factors (Ulbrich et al., 2009) extracted by PMF analysis on

507 the organic matrix of the AMS measurements (Table S3, Supplementary Materials). The

508 comparison was conducted considering ATOFMS-PMF factor time-series in volume (integrated

509   over particles of < 1 µm diameter) because AMS-PMF factors are expressed in mass concentration

510   ($\mu g/m^3$). The results show that ATOFMS-PMF factors associated with aged aerosol (NH4-OOA and

511   EC+) are better correlated with the most aged LV-OOA AMS-PMF factor (r^2=0.66 and 0.67 for

512   NH4-OOA and EC+ respectively) rather than with SV-OOA (r^2=0.55 and 0.43 for NH4-OOA and

513   EC+ respectively). On the contrary, fresh or less aged components (ATOFMS-PMF factors OC-

514   Arom and OC-CHNO) are better correlated with the less aged SV-OOA AMS-PMF factor (r^2=0.54

515   and 0.37 for OC-Arom and OC-CHNO respectively) rather than with LV-OOA (r^2=0.36 and 0.07

516   for OC-Arom and OC-CHNO respectively).

517   Unexpectedly, EC- presented a correlation of medium intensity with both AMS-PMF factors (r^2 =

518   0.45 for LV-OOA and r^2 = 0.43 for SV-OOA). However, the correlations are stronger, especially

519   with respect to the less aged SV-OOA if the EC- time-series is expressed as the equivalent number

520   of particles (r^2 = 0.62). This may be due to the fact that using time-series calculated in volume we

521   may further underestimate the contribution of small particles because of size-dependent

522   transmission losses (Gross et al., 2000; Dall'Osto et al., 2006).

523

524   *3.6. Harwell Aerosol Characterization*

525   From the study of the back-trajectories of air masses arriving in Harwell during the sampling

526   campaign (detailed in S.M.), it was clear that the NaCl factor was dominant during the sampling of

527   marine-polar air masses, while during periods of sampling continental air masses (from Central

528   Europe) elemental carbon, potassium, nitrate and sulphate concentrations increased. This is as

529   expected as Harwell is a rural background site and it should not be influenced substantially by local

530   primary sources. More interesting were two marine-continental periods. The first was characterized

531   by air masses coming from the ocean, crossing Scotland and England before arriving at the Harwell

532   site. It was characterized by high concentrations of NO_x and other primary gaseous pollutants, and a

533   high abundance of the OC-CHNO and EC- factors. The second was characterized by air masses

534   coming from the west coast of France, low concentrations of primary gaseous pollutants and a high

535 amount of CNO-COOH and EC+ PMF factors. Thus, the first period was characterized by freshly

536 emitted aerosol while the second period is characterized by aged and chemically oxidized particles.

537

538 **4. Conclusions**

539 PMF analysis has been applied to single particle ATOFMS mass spectra and allows the extraction

540 and separation of significant contributing chemical components. In general, PMF factor profiles

541 identify well defined chemical species or classes of substances from inorganic (NaCl, K, NIT, SUL)

542 to organic families (EC+, EC-, OC-Arom, OC-CHNO, CNO-COOH, NH4-OOA). There is a partial

543 loss of information on internal mixing of particles.

544 From the cross correlation analysis among temporal trends of PMF factors it was possible to

545 identify two elemental carbon components: the EC- factor, correlated to OC-CHNO, probably

546 related to anthropogenic primary emissions and the EC+ factor present in aged particles internally

547 mixed with secondary species. Furthermore, this is the first time in which different families of

548 organic carbon have been extracted from ATOFMS data, including aromatic, oxidized organic

549 compound and two different organic nitrogen components: primary (OC-CHNO) and oxidized

550 (CNO-COOH). Oxidized carbon in the form of oxidised organic nitrogen and carboxylic acids is

551 found only in aged aerosol while nitrate and sulphate are found in different proportions: the former

552 in less aged aerosol such as in urban plumes while sulphate arose predominantly from long-range

553 transport from continental sources.

554 From the comparison of different data treatment techniques it emerges that K-means cluster

555 analysis and ART-2a artificial neural network analysis give similar results, with particles grouped in

556 clusters of similar composition, reflective of aerosol sources, chemical processes and a combination

557 of both, while PMF analysis of single particle mass spectra allows the deconvolution of the mass

558 spectra and the extraction of some constituent components. Moreover, when expressed in volume,

559 the temporal trends of PMF factors are highly significantly correlated to the corresponding chemical

560 species concentration measured by independent instruments, even in the case of highly internally

561 mixed particles, while the correlation between cluster temporal trends and corresponding chemical

562 species concentration is highly dependent upon particle mixing state. Thus PMF analysis may prove

563 useful for the quantification of the main components of PM data collected with the ATOFMS

564 instrument. However, better repeatability of the ionization process and higher efficiency of particle

565 detection would improve its quantification capability.

566

573

574 **Appendix. Supplementary material**

575 Supplementary material related to this article can be found at ...

576

577

References

Alleman, L.Y., Lamaison, L., Perdrix, E., Robache A. and Galloo J.C., 2010. PM_{10} metal concentrations and source identification using positive matrix factorization and wind sectoring in a French industrial zone. Atmospheric Research 96, 612-625.

Allen, J.O., Fergenson, D.P., Gard, E.E., Hughes, L.S., Morrical, B.D., Kleeman, M.J., Gross, D.S., Gälli, M.E., Prather, K.A. and Cass, G.R., 2000. Particle detection efficiencies of aerosol time of flight mass spectrometers under ambient sampling conditions. Environmental Science & Technology 34, 211-217.

Angelino, S., Suess, D.T. and Prather K.A., 2001. Formation of aerosol particles from reactions of secondary and tertiary alkylamines: Characterization by aerosol time-of-flight mass spectrometry. Environmental Science & Technology 35, 3130-3138.

Bari, M.A., Baumbach, G., Kuch, B. and Scheffknecht G., 2009. Wood smoke as a source of particle-phase organic compounds in residential areas. Atmospheric Environment 43, 4722-4732.

Bhave, P.V., Allen, J.O., Morrical, B.D., Fergenson, D.P., Cass, G.R. and Prather K.A., 2002. A field-based approach for determining ATOFMS instrument sensitivities to ammonium and nitrate. Environmental Science & Technology 36, 4868-4879.

Bi, X., Zhang, G., Li, L., Wang, X., Li, M., Sheng, G., Fu, J. and Zhou Z., 2011. Mixing state of biomass burning particles by single particle aerosol mass spectrometer in the urban area of PRD, China. Atmospheric Environment 45, 3447-3453.

Canagaratna, M.R., Jayne, J.T., Jimenez, J.L., Allan, J.D. Alfarra, M.R., Zhang, Q., Onasch, T.B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L.R., Trimborn, A.M., Northway, M.J., DeCarlo, P.F., Kolb, C.E., Davidovits, P. and Worsnop, D.R., 2007. Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer. Mass Spectrometry Reviews 26, 185-222.

Dall'Osto, M., Beddows, D.C.S., Kinnersley, R.P. and Harrison, R.M.,2004. Characterization of individual airborne particles by using aerosol time-of-flight mass spectrometry at Mace Head, Ireland. Journal of Geophysical Research 109, D21302.

Dall'Osto, M., Harrison, R.M. and Beddows, D.C.S., 2006. Single-particle efficiencies of Aerosol Time-of Flight Mass Spectrometry during the North Atlantic Marine Boundary Layer Experiment. Environmental Science & Technology 40, 5029-5035.

Dall'Osto, M. and Harrison R.M., 2006. Chemical characterisation of single airborne particles in Athens (Greece) by ATOFMS. Atmospheric Environment 40, 7614-7631.

DeCarlo, P.F., Kimmel, J.R., Trimborn, A., Northway, M.J., Jayne, J.T., Aiken, A.C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K.S., Worsnop, D.R. and Jimenez, J.L., 2006. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. Analytical Chemistry 78, 8281-8289.

Doğan, G., Güllü, G. and Tuncel, G., 2008. Sources and source regions effecting the aerosol composition of the Eastern Mediterranean. Microchemical Journal 88, 142-149.

Draxler, R.R. and Rolph, G.D., 2003. HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) model v 4.9, NOAA Air Resource Laboratory, Silver Spring MD., http://ready.arl.noaa.gov/HYSPLIT.php

Drewnick, F., Hings, S.S., DeCarlo, P., Jayne, J.T., Gonin, M., Fuhrer, K., Weimer, S., Jimenez, J.L., Demerjian, K.L., Borrmann, S. and Worsnop, D.R., 2005. A new time-of-flight aerosol mass spectrometer (TOF-AMS)-instrument description and first field deployment. Aerosol Science & Technology 39, 637-658.

Drewnick, F., Dall'Osto, M. and Harrison, R.M., 2008. Characterization of aerosol particles from grass mowing by joint deployment of ToF-AMS and ATOFMS instruments. Atmospheric Environment 42, 3006-3017.

Eatough, D.J., Grover, B.D., Woolwine, W.R., Eatough, N.L., Long, R. and Farber, R., 2008. Source apportionment of 1h semi-continuous data during the 2005 Study of Organic Aerosols in Riverside (SOAR) using positive matrix factorization. Atmospheric Environment 42, 2706-2719.

Gard, E., Mayer, J.E., Morrical, B.D., Dienes, T., Fergenson, D.P. and Prather, K.A., 1997. Real-time analysis of individual atmospheric aerosol particles: design and performance of a portable ATOFMS. Analytical Chemistry 69, 4083-4091.

Gross, D.S., Atlas, R., Rzeszotarski, J., Turetsky, E., Christensen, J., Benzaid, S., Olson, J., Smith, T., Steinberg, L., Sulman, J., Ritz, A., Anderson, B., Nelson, C., Musicant, D.R., Chen, L., Snyder, D.C. and Schauer, J.J., 2010. Environmental chemistry through intelligent atmospheric data analysis. Environmental Modelling & Software 25, 760-769.

Gross, D.S., Galli, M.E., Silva, P.J. and Prather, K.A., 2000. Relative sensitivity factors for alkali metal and ammonium cations in single-particle Aerosol Time-of-Flight Mass Spectra. Analytical Chemistry 72, 416-422.

Harrison, R.M., Giorio, C., Beddows, D.C.S. and Dall'Osto, M., 2010. Size distribution of airborne particles controls outcome of epidemiological studies. Science of the Total Environment 409, 289-293.

Healy, R.M., O'Connor, I.P., Hellebust, S., Allanic, A., Sodeau, J.R. and Wenger, J.C., 2009. Characterisation of single particles from in-port ship emissions. Atmospheric Environment 43, 6408-6414.

Healy, R.M., Sciare, J., Poulain, L., Kamili, K., Merkel, M., Müller, T., Wiedensohler, A., Eckhardt, S., Stohl, A., Sarda-Estève R., McGillicuddy, E., O'Connor, I.P., Sodeau, J.R., and Wenger, J.C., 2012. Sources and mixing state of size-resolved elemental carbon particles in a European megacity: Paris. Atmospheric Chemistry and Physics 12, 1681-1700.

Hinds, W.C., 1999. Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles. John Wiley & Sons, New York.

Jeong, C.H., McGuire, M.L., Godri, K.J., Slowik, J.G., Rehbein, P.J.G., and Evans, G.J., 2011. Quantification of aerosol chemical composition using continuous single particle measurements. Atmospheric Chemistry and Physics 11, 7027-7044.

Jia, Y., Clements, A.L. and Fraser, M.P., 2010. Saccharide composition in atmospheric particulate matter in the southwest US and estimates of source contributions. Aerosol Science 41, 62-73.

Jimenez, J.L., Jayne, J.T., Shi, Q., Kolb, E., Worsnop, D.R., Yourshaw, I., Seinfeld, J.H., Flagan, R.C., Zhang, X., Smith, K.A., Morris, J.W. and Davidovits, P., 2003. Ambient aerosol sampling using the Aerodyne Aerosol Mass Spectrometer. Journal of Geophysical Research 108, (D7), 8425.

659 Korn, M.G.A., Santos, D.S.S., Welz, B., Rodrigues, M.G., Teixeira, A.P., Lima, D.C. and Ferreira,
660 L.C., 2007. Atomic spectrometric methods for the determination of metals and metalloids in
661 automotive fuels – A review. Talanta 73, 1-11.

662 Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot, A.S.H., 2007.
663 Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling
664 of aerosol mass spectra. Atmospheric Chemistry and Physics 7, 1503–1522.

665 Lee, E., Chan, C.K. and Paatero, P., 1999. Application of positive matrix factorization in source
666 apportionment of particulate pollutants in Hong Kong. Atmospheric Environment 33, 3201-3212.

667 McGuire, M.L., Jeong, C.H., Slowik, J.G., Chang, R.Y.W., Corbin, J.C., Lu, G., Mlhele, C.,
668 Rehbein, P.J.G., Sills, D.M.L., Abbatt, J.P.D., Brook, J.R. and Evans, G.J., 2011. Elucidating
669 determinants of aerosol composition through particle-type-based receptor modelling. Atmospheric
670 Chemistry & Physics 11, 8133-8155.

671 McLafferty, F.W., 1983. Interpretation of mass spectra. Third Edition, Mill Valley, CA University
672 Co. Books.

673 Moffet, R.C., de Foy, B., Molina, L.T., Molina, M.J. and Prather, K.A., 2008. Measurement of
674 ambient aerosols in northern Mexico City by single particle mass spectrometry. Atmospheric
675 Chemistry & Physics 8, 4499-4516.

676 Owega, S., Khan, B.U.Z., D'Souza, R., Evans, G.J., Fila, M. and Jervis, R.E., 2004. Receptor
677 modeling of Toronto $PM_{2.5}$ characterized by aerosol laser ablation mass spectrometry.
678 Environmental Science & Technology 38, 5712-5720.

679 Paatero, P. and Tapper, U., 1994. Positive Matrix Factorization: a non-negative factor model with
680 optimal utilization of error estimates of data values. Environmetrics 5, 111-126.

681 Paatero, P., 1998. User's Guide for Positive Matrix Factorization programs PMF2 and PMF3.

682 Pekney, N.J., Davidson, C.I., Bein, K.J., Wexler, A.S. and Johnston, M.V., 2006. Identification of
683 sources of atmospheric PM at the Pittsburgh Supersite, Part I: Single particle analysis and filter-
684 based positive matrix factorization. Atmospheric Environment 40, S411-S423.

685 Pratt, K.A. and Prather, K.A., 2011. Mass Spectrometry of atmospheric aerosols - recent
686 developments and applications. Part II: On-line mass spectrometry techniques. Mass Spectrometry
687 Reviews, DOI: 10.1002/mas.20330.

688 Rebotier, T.P. and Prather K.A., 2007. Aerosol Time-of-Flight mass spectrometry data analysis: A
689 benchmark of clustering algorithms. Analytica Chimica Acta 585, 38-54.

690 Reilly, P.T.A., Lazar, A.C., Gieray, R.A., Whitten, W.B. and Ramsey, J.M., 2000. The elucidation
691 of charge-transfer-induced matrix effects in environmental aerosols via real-time aerosol mass
692 spectral analysis of individual airborne particles. Aerosol Science & Technology 33, 135-152.

693 Reinard, M.S. and Johnston, M.V., 2008. Ion formation mechanism in laser desorption ionization of
694 individual particles. Journal of the American Society for Mass Spectrometry 19, 389-399.

695 Song, X.H., Hopke, P.K., Fergenson, D.P. and Prather, K.A., 1999. Classification of single particles
696 analyzed by ATOFMS using an artificial neural network, ART-2A. Analytical Chemistry 71, 860-
697 865.

698   Stortini, A.M., Freda, A., Cesari, D., Cairns, W.R.L., Contini, D., Barbante, C., Prodi, F., Cescon,
699   P. and Gambaro, A., 2009. An evaluation of the PM2.5 trace elemental composition in the Venice
700   Lagoon area and an analysis of the possible sources. Atmospheric Environment 43, 6296-6304.

701   Su, Y., Sipin, M.F., Furutani, H. and Prather, K.A., 2004. Development and characterization of an
702   Aerosol Time-of-Flight mass spectrometer with increased detection efficiency. Analytical
703   Chemistry 76, 712-719.

704   Thomas, R.M., Trebs, I., Otjes, R., Jongejan, P.A.C., Brink, H.T., Phillips, G., Kortner, M.,
705   Meixner, F.X. and Nemitz, E., 2009. An automated analyzer to measure surface-atmosphere
706   exchange fluxes of water soluble inorganic aerosol compounds and reactive trace gases.
707   Environmental Science & Technology 43, 1412-1418.

708   Ulbrich, I.M., Canagaratna, M.R., Zhang, Q., Worsnop, D.R., and Jimenez, J.L., 2009.
709   Interpretation of organic components from positive matrix factorization of aerosol mass
710   spectrometric data. Atmospheric Chemistry and Physics 9, 2891-2918.

711   Wenzel, R.J., Liu, D.Y., Edgerton, E.S. and Prather, K.A., 2003. Aerosol Time-of-Flight mass
712   spectrometry during the Atlanta Supersite Experiment: 2. Scaling procedures. Journal of
713   Geophysical Research-Atmosphere 108(D7), 8427.

714   Wexler, A.S. and Johnston, M.V., 2008. What have we learned from highly time-resolved
715   measurements during EPA's supersites program and related studies?. Journal of the Air & Waste
716   Management Association 58, 303–319.

717   Zhang,T., Claeys, M., Cachier, H., Dong, S., Wang, W., Maenhaut, N. and Liu, X., 2008.
718   Identification and estimation of the biomass burning contribution to Beijing aerosol using
719   levoglucosan as a molecular marker. Atmospheric Environment 42, 7013-7021.

720

Table 1. Coefficient of determination (r^2) values of the linear regressions between hourly temporal trends of PMF factors (equivalent number of particles) and K-means clusters or ART-2a clusters*.

PMF factors	r^2 (PMF factors vs K-means clusters)									r^2 (PMF factors vs ART-2a clusters)								
	K	K-EC	NaCl	EC	K-SUL-OC-NIT	OC	K-NIT	Fe-V	OOA	K-NIT	NaCl	OC	K-SUL	EC	K	EC-Fe-V	SOA	K-EC
CNO-COOH	0.43	0.34	0.00	0.09	0.28	0.09	0.25	0.02	0.12	0.29	0.00	0.19	0.27	0.10	0.09	0.01	0.04	0.16
SUL	**0.53**	0.20	0.00	0.28	**0.51**	0.12	0.45	0.10	0.37	0.47	0.00	0.33	0.42	0.31	0.30	0.08	0.22	0.07
NH4-OOA	**0.76**	0.01	0.11	**0.68**	0.01	0.17	**0.64**	0.27	**0.83**	**0.78**	0.12	0.19	0.46	**0.76**	**0.73**	0.23	0.48	0.03
NaCl	0.06	0.06	**0.91**	0.12	0.06	0.00	0.03	0.06	0.08	0.07	**0.82**	0.00	0.05	0.12	0.12	0.05	0.07	0.01
EC+	**0.59**	0.00	0.14	**0.99**	0.00	0.02	0.36	0.49	**0.53**	**0.53**	0.16	0.04	0.38	**0.87**	**0.60**	**0.66**	0.24	0.01
OC-Arom	**0.54**	0.31	0.00	0.25	0.36	0.29	**0.59**	0.08	0.42	**0.53**	0.00	0.45	0.34	0.30	0.29	0.05	0.26	0.11
EC-	**0.53**	0.35	0.02	0.42	0.06	0.17	0.42	0.16	0.39	0.43	0.02	0.17	0.26	0.42	0.32	0.18	0.23	0.08
K	**0.99**	0.03	0.04	**0.56**	0.03	0.04	0.47	0.17	**0.61**	**0.78**	0.05	0.11	**0.64**	**0.58**	**0.60**	0.21	0.25	0.07
NIT	**0.60**	0.12	0.00	0.37	0.12	0.19	**0.92**	0.12	**0.64**	**0.77**	0.00	**0.50**	0.21	**0.50**	**0.54**	0.08	0.42	0.06
OC-CHNO	0.14	0.40	0.00	0.08	0.11	**0.83**	0.49	0.03	0.30	0.24	0.00	**0.51**	0.03	0.14	0.12	0.01	0.36	0.11

*strong correlated results ($r^2>0.5$) are presented in bold
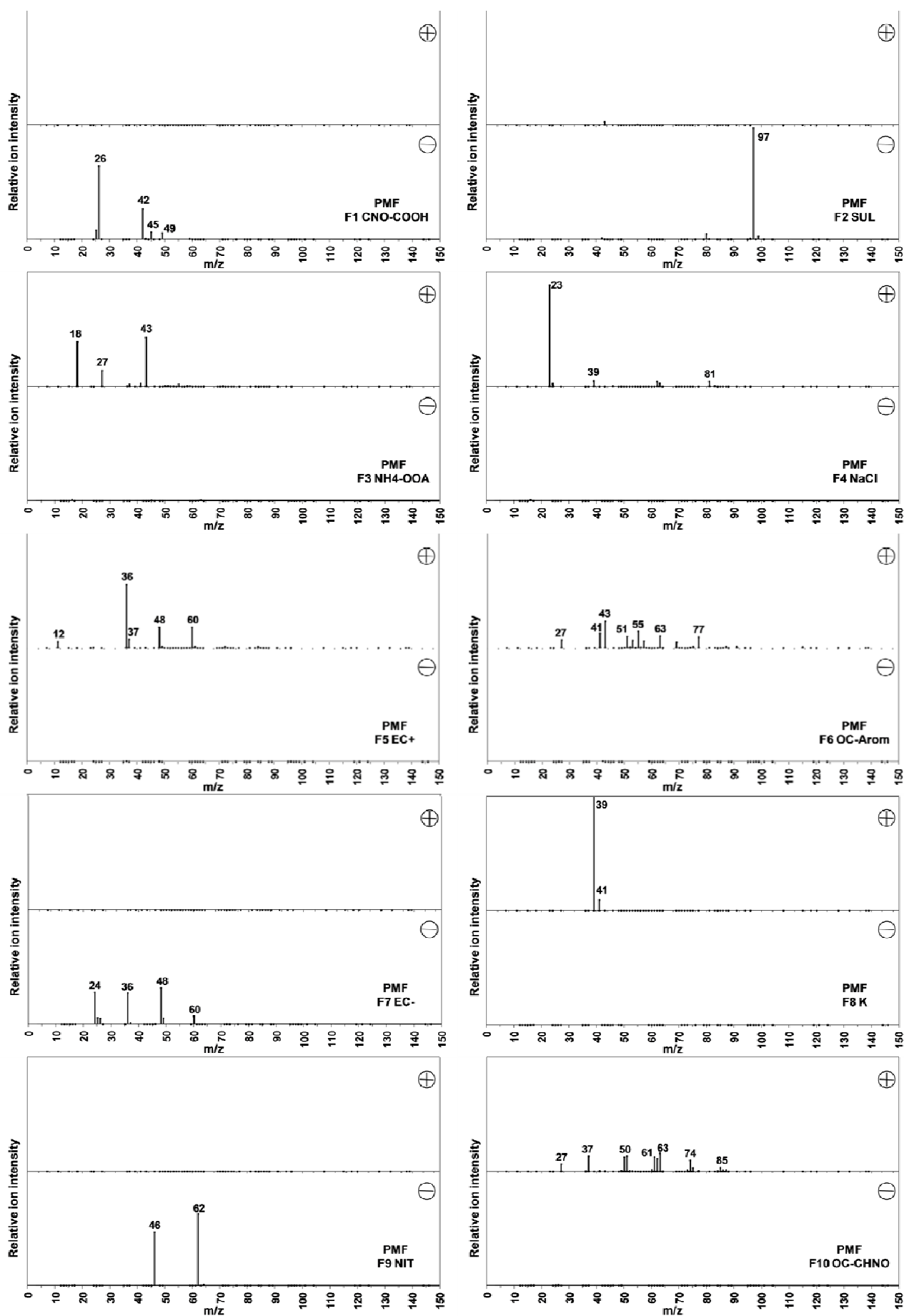
Figure 1. Mass Spectra of the 10 PMF factors.
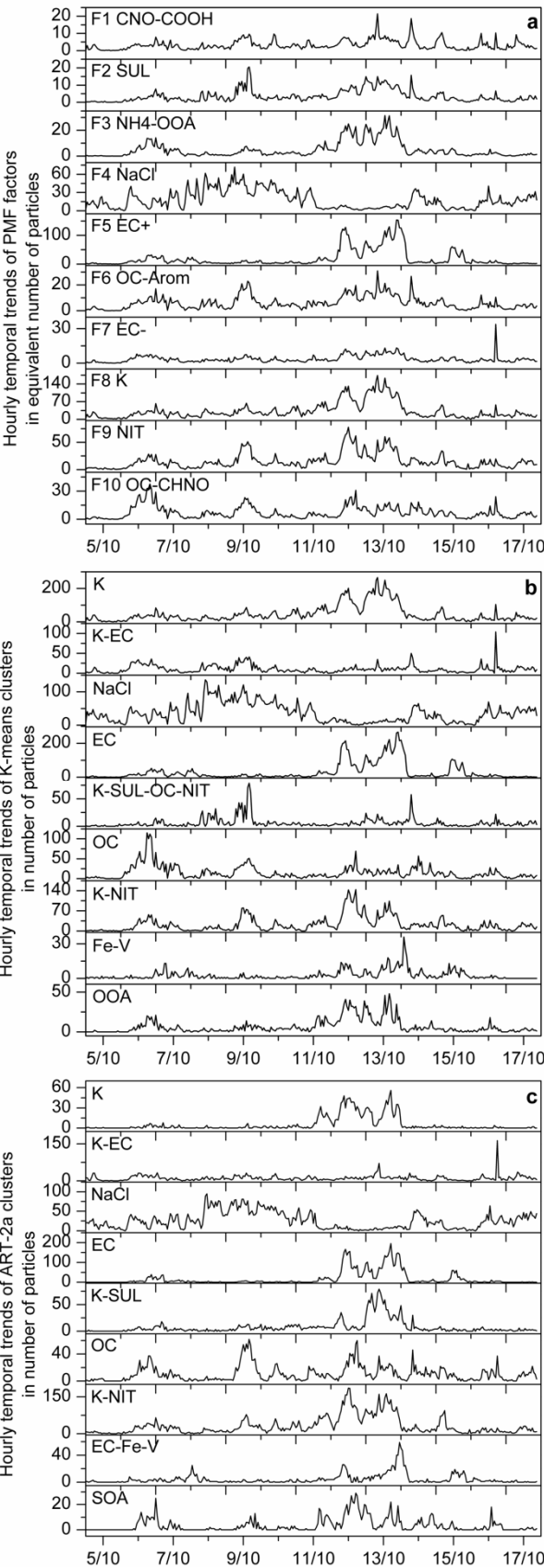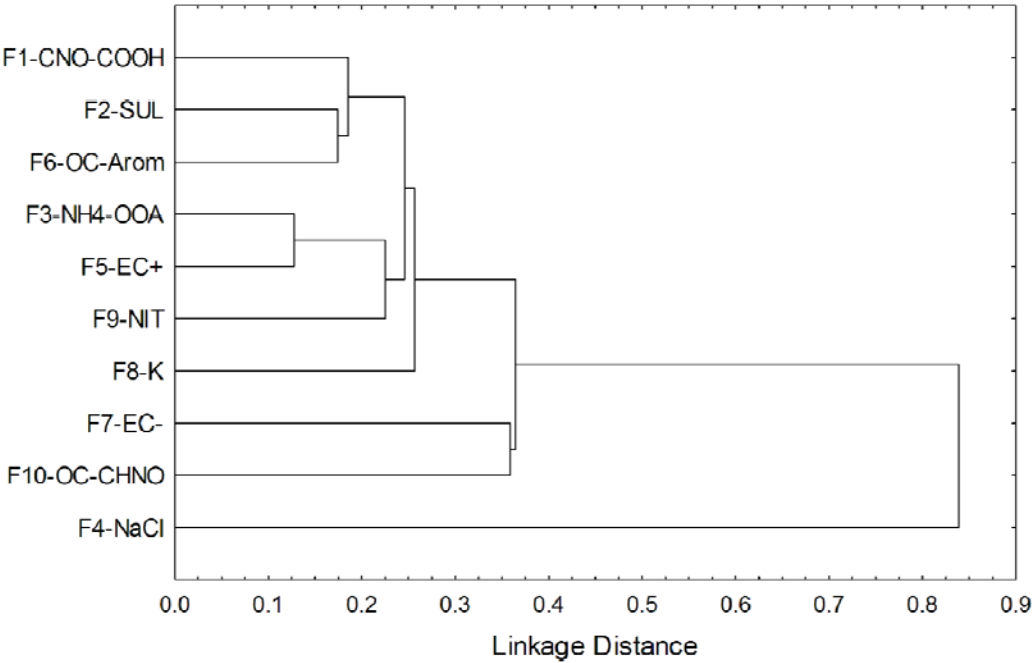
726

727

728

729

730   Figure 2. Temporal trends of (a) PMF factors expressed in equivalent number of particles, (b) K-
731   means clusters and (c) ART-2a clusters in number of particles.

732

733

734
735
736 Figure 3. Dendogram obtained from the hierarchical cluster analysis of the temporal trends of PMF
737 factors (single linkage method, r-Pearson correlation coefficient distance measure).
738
739
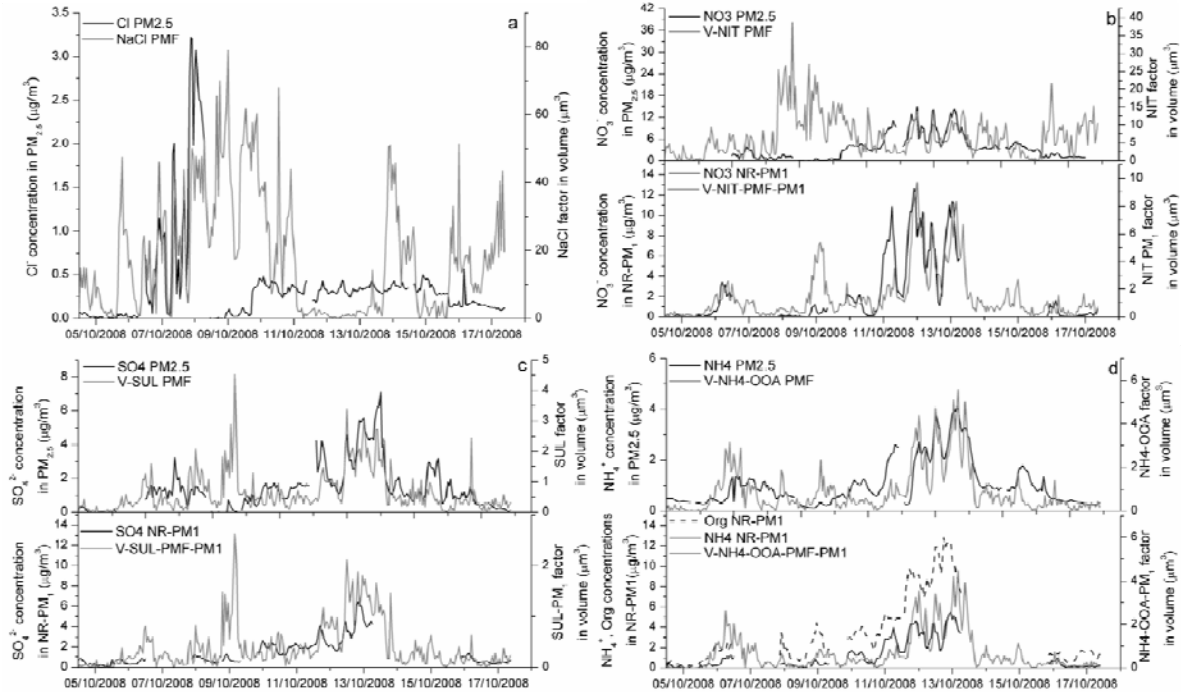
740
741 Figure 4. Hourly time-series of (a) chloride concentrations and NaCl factor, (b) nitrate
742 concentrations and NIT factor, (c) sulphate concentrations and SUL factor, (d) ammonium, organic
743 concentrations and NH4-OOA factor.
744
745
746

30