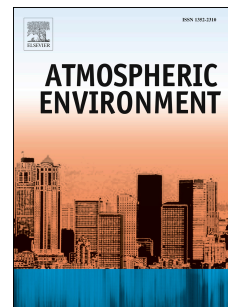


# Accepted Manuscript



A New Methodology to Assess the Performance and Uncertainty of Source Apportionment Models II: the Results of Two European Intercomparison Exercises

C.A. Belis, F. Karagulian, F. Amato, M. Almeida, P. Artaxo, D.C.S. Beddows, V. Bernardoni, M.C. Bove, S. Carbone, D. Cesari, D. Contini, E. Cuccia, E. Diapouli, K. Eleftheriadis, O. Favez, I. El Haddad, R.M. Harrison, S. Hellebust, J. Hovorka, E. Jang, H. Jorquera, T. Kammermeier, M. Karl, F. Lucarelli, D. Mooibroek, S. Nava, J.K. Nøjgaard, P. Paatero, M. Pandolfi, M.G. Perrone, J.E. Petit, A. Pietrodangelo, P. Pokorná, P. Prati, A.S.H. Prevot, U. Quass, X. Querol, D. Saraga, J. Sciare, A. Sfetsos, G. Valli, R. Vecchi, M. Vestenius, E. Yubero, P.K. Hopke

PII: S1352-2310(15)30485-4

DOI: [10.1016/j.atmosenv.2015.10.068](https://doi.org/10.1016/j.atmosenv.2015.10.068)

Reference: AEA 14226

To appear in: *Atmospheric Environment*

Received Date: 24 February 2015

Revised Date: 14 September 2015

Accepted Date: 24 October 2015

Please cite this article as: Belis, C.A., Karagulian, F., Amato, F., Almeida, M., Artaxo, P., Beddows, D.C.S., Bernardoni, V., Bove, M.C., Carbone, S., Cesari, D., Contini, D., Cuccia, E., Diapouli, E., Eleftheriadis, K., Favez, O., El Haddad, I., Harrison, R.M., Hellebust, S., Hovorka, J., Jang, E., Jorquera, H., Kammermeier, T., Karl, M., Lucarelli, F., Mooibroek, D., Nava, S., Nøjgaard, J.K., Paatero, P., Pandolfi, M., Perrone, M.G., Petit, J.E., Pietrodangelo, A., Pokorná, P., Prati, P., Prevot, A.S.H., Quass, U., Querol, X., Saraga, D., Sciare, J., Sfetsos, A., Valli, G., Vecchi, R., Vestenius, M., Yubero, E., Hopke, P.K., A New Methodology to Assess the Performance and Uncertainty of Source Apportionment Models II: the Results of Two European Intercomparison Exercises, *Atmospheric Environment* (2015), doi: 10.1016/j.atmosenv.2015.10.068.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please

note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## A New Methodology to Assess the Performance and Uncertainty of Source Apportionment Models II: the Results of Two European Intercomparison Exercises

Belis, C.A.<sup>a</sup>, Karagulian F.<sup>a</sup>, Amato F.<sup>b</sup>, Almeida M.<sup>c</sup>, Artaxo P.<sup>d</sup>, Beddows D.C.S.<sup>e</sup>, Bernardoni V.<sup>f</sup>, Bove M.C.<sup>g</sup>, Carbone S.<sup>h</sup>, Cesari D.<sup>i</sup>, Contini D.<sup>i</sup>, Cuccia E.<sup>g</sup>, Diapouli E.<sup>j</sup>, Eleftheriadis K.<sup>j</sup>, Favez O.<sup>k</sup>, El Haddad I.<sup>l</sup>, Harrison R.M.<sup>e</sup>, Hellebust S.<sup>m</sup>, Hovorka J.<sup>n</sup>, Jang E.<sup>e</sup>, Jorquera H.<sup>o</sup>, Kammermeier T.<sup>p</sup>, Karl M.<sup>q</sup>, Lucarelli F.<sup>r</sup>, Mooibroek D.<sup>s</sup>, Nava S.<sup>r</sup>, Nøjgaard J. K.<sup>t</sup>, Paatero P.<sup>u</sup>, Pandolfi M.<sup>b</sup>, Perrone M.G.<sup>v</sup>, Petit J.E.<sup>k,y</sup>, Pietrodangelo A.<sup>w</sup>, Pokorná P.<sup>n</sup>, Prati P.<sup>g</sup>, Prevot A.S.H.<sup>l</sup>, Quass U.<sup>p</sup>, Querol X.<sup>b</sup>, Saraga D.<sup>x</sup>, Sciare J.<sup>y</sup>, Sfetsos A.<sup>x</sup>, Valli G.<sup>f</sup>, Vecchi R.<sup>f</sup>, Vestenius M.<sup>h</sup>, Yubero E.<sup>z</sup>, Hopke P.K.<sup>aa</sup>

<sup>a</sup>European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via Enrico Fermi 2749, Ispra (VA) 21027, Italy

<sup>b</sup>Institute of Environmental Assessment and Water Research, Spanish Research Council (IDAEA-CSIC), c/Jordi Girona 18-26, 08034 Barcelona, Spain

<sup>c</sup>C2TN, Instituto Superior Técnico, Universidade de Lisboa, Estrada Nacional 10 km 139.7 2695-066 Bobadela LRS, Portugal

<sup>d</sup>Instituto de Física, Universidade de Sao Paulo, Rua do Matao, Traversa R, 187 05508-900 Sao Paulo, Brazil

<sup>e</sup>Division of Environmental Health and Risk Management, School of Geography, Earth and Environmental Sciences University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom; also at: Department of Environmental Sciences / Center of Excellence in Environmental Studies, King Abdulaziz University, PO Box 80203, Jeddah, 21589, Saudi Arabia

<sup>f</sup>Dept. of Physics, Università degli Studi di Milano & INFN-Milan, via Celoria 16, Milan 20133, Italy

<sup>g</sup>University of Genoa- Dept. of Physics and INFN, via Dodecaneso 33, 14146, Genova, Italy

<sup>h</sup>Finnish Meteorological Institute, Atmospheric Composition Research, P.O.Box 503, FI-00101 Helsinki, Finland

<sup>i</sup>Istituto di Scienze dell'Atmosfera e del Clima, ISAC-CNR Str. Prv. Lecce-Monteroni km 1.2 73100 Lecce, Italy

<sup>j</sup>Institute of Nuclear and Radiological Science & Technology, Energy & Safety, N.C.S.R. "Demokritos", 15341 Athens, Greece

<sup>k</sup>Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, France

<sup>l</sup>Laboratory of Atmospheric Chemistry (LAC), Paul Scherrer Institut, Villigen, Switzerland

<sup>m</sup>Centre for Research into Atmospheric Chemistry, Dept. Chemistry University College Cork, Ireland

<sup>n</sup> Institute for Environmental Studies, Charles University in Prague, Albertov 6, 128 43 Prague 2, Czech Republic

<sup>o</sup> Departamento de Ingeniería Química y Bioprocesos, Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860 Santiago 6904411, Chile

<sup>p</sup> IUTA e.V., Bereich Luftreinhaltung & Nachhaltige Nanotechnologie, Institut für Energie- und Umwelttechnik e.V. Bliersheimer Strasse 60, D-47229 Duisburg, Germany

<sup>q</sup> Urban Environment and Industry, Norwegian Institute for Air Research (NILU), P.O. Box 100, NO-2027 Kjeller, Norway

<sup>r</sup> Department of Physics and Astronomy and INFN, Firenze, Italy

<sup>s</sup> National Institute of Public Health and the Environment, Centre for Environmental Quality (MIL), Department for Air and Noise Analysis (ILG), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

<sup>t</sup> Department for Environmental Science, Aarhus University, Frederiksborgvej 399, PO Box 358, DK-4000 Roskilde, Denmark

<sup>u</sup> Department of Physics, University of Helsinki, Rikalan tie 6, FI-00970 Helsinki, Finland

<sup>v</sup> Department of Earth and Environmental Sciences, University of Milano-Bicocca, P.zza della Scienza 1, 20126 Milan, Italy

<sup>w</sup> C.N.R., Institute of Atmospheric Pollution Research, Area della Ricerca di Roma1, Via Salaria Km 29,300 Monterotondo (RM) 00015, Italy

<sup>x</sup> IN.R.A.S.T.E.S., NCSR Demokritos, P. Grigoriou and Neapoleos Str, 153 10, Agia Paraskevi, Greece

<sup>y</sup> CNRS LSCE, France

<sup>z</sup> Laboratory of Atmospheric Pollution (LCA), Miguel Hernández University, Av. de la Universidad s/n, Edif. Alcudia, 03202 Elche, Spain

<sup>aa</sup> Center for Air Resources Engineering and Science Clarkson University Box 5708 Potsdam, NY 13699-5708, United States of America

corresponding author: Claudio A. Belis, tel. +39 0332 786644, fax +39 0332 785837, e-mail: claudio.belis@jrc.ec.europa.eu

## Abstract

The performance and the uncertainty of receptor models (RMs) were assessed in intercomparison exercises employing real-world and synthetic input datasets. To that end, the results obtained by different practitioners using ten different RMs were compared with a reference. In order to explain the differences in the performances and uncertainties of the different approaches, the apportioned mass, the number of sources, the chemical profiles, the contribution-to-species and the time trends of the sources were all evaluated using the methodology described in Belis et al. (2015).

In this study, 87% of the 344 source contribution estimates (SCEs) reported by participants in 47 different source apportionment model results met the 50% standard uncertainty quality objective established for the performance test. In addition, 68% of the SCE uncertainties reported in the results were coherent with the analytical uncertainties in the input data.

The most used models, EPA-PMF v.3, PMF2 and EPA-CMB 8.2, presented quite satisfactory performances in the estimation of SCEs while unconstrained models, that do not account for the uncertainty in the input data (e.g. APCS and FA-MLRA), showed below average performance. Sources with well-defined chemical profiles and seasonal time trends, that make appreciable contributions (>10%), were those better quantified by the models while those with contributions to the PM mass close to 1% represented a challenge.

The results of the assessment indicate that RMs are capable of estimating the contribution of the major pollution source categories over a given time window with a level of accuracy that is in line with the needs of air quality management.

Keywords: source apportionment, receptor models, intercomparison exercise, model performance indicators, model uncertainty, particulate matter

Highlights:

Intercomparisons were used to test the performance and uncertainty of receptor models.

More than 85% of the reported sources met the model quality objectives.

Two thirds of the output uncertainties were coherent with those in the input data.

PMF v2, 3 and CMB 8.2 estimated the source contributions satisfactorily.

The accuracy of receptor models is in line with the needs of air quality management.

## 1 1. INTRODUCTION

2 Source Apportionment (SA) is the practice of deriving information about the pollution  
3 sources and the amount they contribute to measured concentrations. Receptor models  
4 (RMs) apportion the measured mass of pollutants to its emission sources by using  
5 multivariate analysis to solve a mass balance equation (Friedlander, 1973; Schauer et al.,  
6 1996; Thurston and Spengler, 1985). RMs derive information from measurements  
7 including estimations of their uncertainty and have been extensively used in Europe to  
8 estimate the contribution of emission sources to atmospheric pollution at a given site or  
9 area (Belis et al., 2013; Viana et al., 2008a). In the Chemical Mass Balance (CMB)  
10 approach, both chemical concentrations of pollutants, including their uncertainties, and  
11 chemical fingerprints of the sources (source profiles) are used as input. In the multivariate  
12 factor analytical approach (MFA), only environmental concentrations and uncertainties of  
13 pollutants are used as input data and the model computes the factor profiles and the mass  
14 contributed by the factors. The CMB approach is sensitive to the selection of sources, their  
15 stability and the collinearity among them. Differences between the methods used to  
16 analyze the source and ambient samples may also impact the results. On the other hand,  
17 MFA models identify factors that have to be attributed to emission sources. For a more  
18 thorough discussion about the pros and cons of the two approaches see Hopke (2010),  
19 Watson et al. (2008) and Belis et al.(2013).

20 Previous studies provided first estimates of the output variability by comparing the results  
21 of different RMs on the same dataset (Hopke et al., 2006; Larsen et al., 2008; Favez et al.,  
22 2010; Viana et al., 2008b; Pandolfi et al., 2008). In the present work, intercomparison  
23 exercises aimed at quantitatively assessing the performance and the uncertainty of RMs  
24 by comparing the results reported from different practitioners on the same dataset using  
25 different RM techniques.

## 27 2. METHODOLOGY

28 The methodology adopted in this research to assess the model results evaluates all the  
29 aspects of a source apportionment study, including the variability due to the influence of  
30 different practitioners using the same model on the same data (Belis et al., 2015). The  
31 procedure includes: complementary, preliminary and performance tests.

32 The “complementary tests” aim at providing ancillary information about the performance of  
33 the solutions in terms of apportioned mass and number of source categories. The  
34 “preliminary tests” are targeted at establishing whether the entities identified in the results,

35 either a factor or a source (hereon, factor/source), are attributable to a given source  
36 category. In addition to the correlation coefficient (hereafter, Pearson), the standardized  
37 identity distance (SID), that prevents the distortions caused by source profiles with  
38 dominant species, is used (more details in Belis et al., 2015). The “ff tests” are the  
39 comparison among factor/sources attributed by participants to the same source category  
40 in all the solutions while “fr tests” refer to the comparison between reported factor/sources  
41 and a reference value. The objective of the “performance tests” is to evaluate whether the  
42 source contribution estimates (SCEs) are coherent with a 50% standard uncertainty target  
43 value using the z-score performance indicator complemented by the z'-score and zeta-  
44 score indicators (Thomson et al., 2006; ISO 13528, 2005). In this study, SCE denotes the  
45 mass attributed to a source or factor in the results obtained with either CMB or MFA  
46 approaches. The methodology is fully described in the companion paper by Belis et al.  
47 (2015) and was implemented using the open source software R (and R-studio). Source  
48 categories with less than five factors/sources were not evaluated and profiles attributed by  
49 participants to more than one category were tested in each of the proposed categories.  
50 Considering that source apportionment studies are mostly targeted at identifying and  
51 quantifying the typical sources in the studied area, the performance tests were conducted  
52 on the average SCE over the whole time window represented in every dataset. Moreover,  
53 the SCE time series were evaluated using the root mean square error normalised by the  
54 standard deviation/uncertainty of the reference value ( $RMSE_u$ ), as discussed in Belis et al.,  
55 (2015).

56 The intercomparison exercises were structured in two rounds involving 16 and 21  
57 organizations respectively. In the first round, 22 results were reported and 25 were  
58 provided in the second one. A real-world  $PM_{2.5}$  dataset collected in Saint Louis (USA) was  
59 used in Round 1 (Table 1). The dataset used for the intercomparison was developed by  
60 merging two datasets: one of inorganic species collected every day (Lee et al., 2006) and  
61 one of organic species collected every sixth day over the same time window (Jaeckels et  
62 al., 2007). In the final dataset, the structure of the uncertainties of the different species was  
63 heterogeneous with differences between species deriving from the data treatment in the  
64 original datasets and variability within single species due to the different analytical batches  
65 that were necessary to cover the whole monitoring campaign. In addition, the uncertainty  
66 of organic tracers was complex to quantify due to the possible influence of atmospheric  
67 chemistry and radiation on the degradation of these compounds (Galarneau et al., 2008;  
68 Hennigan et al., 2010).



69 The site and time window in which the real-world dataset was collected was not revealed  
70 to the intercomparison participants. The dataset containing the concentrations of 44  
71 species in 180 samples with their analytical uncertainties was distributed to participants  
72 together with the analytical parameters (uncertainty of the method and minimum detection  
73 limits) and the emission inventory of the study area.

74 In Round 1, the following preliminary tests were performed: Pearson and SID between  
75 factor/source profiles, Pearson between log-transformed factor/source profiles, and  
76 Pearson between factor/source time trends. Only ff tests were accomplished in this round  
77 because of the absence of independent unbiased reference values.

78 In the performance tests of Round 1, the SCE reference value for each source category  
79 was the average of the results reported by the participants. The reference values were  
80 obtained by calculating the robust average (Analytical Methods Committee, 1989) using  
81 only the SCEs of source/factors that passed the preliminary tests (Table 2).

82 In the second round, a synthetic dataset with known reference values that were unbiased  
83 and independent from the results reported by participants was used (Supplementary  
84 Material S1). The chemical species included in the synthetic dataset (Round 2) are  
85 reported in Table 1 and the procedure followed to generate it is given in Belis et al. (2015).  
86 Since the site was not disclosed to participants, the emission inventory of the study area  
87 and a set of 23 local source profiles (more than one for every source category) were  
88 distributed to them in order to: a) provide the necessary information to create the input files  
89 for CMB models, and b) support the interpretation of the models' output.

90 In addition to the preliminary tests performed in the previous round, the Pearson between  
91 the factor/source contribution-to-species of the Round 2 results was also computed. All of  
92 the preliminary tests were performed by comparing factor/sources reported by participants  
93 with the reference source for the considered source category (fr tests).

94 The model abbreviations used in this document are: CMB8.2, Chemical Mass Balance v.  
95 8.2 by U.S. EPA ; ME, Multilinear Engine; PCA, Principal Component Analysis; APCS,  
96 Absolute Principal Component Score; FA-MLRA, Factor Analysis-Multilinear Regression;  
97 COPREM, constrained physical receptor model and PMF, Positive Matrix Factorization.

98 The code "PMF2" denotes the program PMF2 described by Paatero (1997). The codes  
99 "EPAPMF3, EPAPMF4, and EPAPMF5" denote the respective releases of the U.S. EPA  
100 program "EPA PMF".

101



## 102 3. RESULTS AND DISCUSSION

### 103 3.1. Complementary tests

#### 104 3.1.1. Mass apportionment

105 The sample-wise comparison between the sum of the SCEs in every solution and the  
106 gravimetric mass are summarised using normalised target diagrams (Fig. 1). More than  
107 70% of the solutions in Round 1 rank in the area of acceptance (outer circle). Most scores  
108 rank in the lower quadrants indicating a tendency to underestimate the observed mass  
109 (the distance to the horizontal axis is proportional to the  $PM_{2.5}$  mass that was not  
110 apportioned). On the contrary, the evident overestimation of the mass observed in two  
111 solutions is likely due to problems in the conversion of normalised data to concentration  
112 values rather than to errors in the apportionment of the mass. In Round 2, the majority of  
113 solutions (ca. 90%) rank in the area of acceptance and show little bias indicating that many  
114 solutions achieved a quite satisfactory apportionment of the gravimetric mass to its  
115 sources. In these tests, no clear relation between the type of model used and the  
116 performance is observed.

#### 118 3.1.2. Number of factor/sources

119 There are different techniques to determine the number of sources (e.g. Henry et al.,  
120 1984). The procedures followed by participants to determine the number of sources were  
121 based on multi-criteria, the most common of which were: a) the impact of the number of  
122 factors on the model diagnostics, b) the stability of factor profiles across different models  
123 set up, and c) the physical meaning of the factor profiles and their comparability with  
124 source profiles from the literature.

125 In Round 1, nine factor/sources per solution are reported on the average (Table 3). One  
126 half of the solutions identifies between six and ten factor/sources while six solutions report  
127 more than 10. An approximation of the expected number of factor/sources for this round is  
128 derived from the original solution of the inorganic dataset obtained using PMF (Lee et al.,  
129 2006), which identified 10 different source categories. In this round, the estimations of  
130 PMF and CMB are relatively close. In Round 2, more than half of the solutions report the  
131 exact number of factor/sources used to design the dataset (8) and all the solutions, except  
132 one, report between six and nine factor/sources.

133 The tests suggest that the reliability of the performance diagnostics influence the ability of  
134 the tools to establish the most suitable number of factor/sources. Often, unconstrained

135 MFA tools ranked far from the average. The higher number of factor/sources in COPREM  
136 is likely due to the attempt to apportion the secondary organic aerosols (not present in the  
137 synthetic dataset) and the split of ammonium sulphate into  $(\text{NH}_4)_2\text{SO}_4$  and  $(\text{NH}_4)\text{HSO}_4$ .  
138 No relevant differences in the number of factor/sources are observed between CMB8.2  
139 and the different versions of PMF.

## 140 **3.2. Identity and uncertainty of the factor/sources**

### 141 **3.2.1. Factor/source identity**

#### 142 **3.2.1.1. Chemical Profiles**

143 Fig. 2 shows the distribution of the Pearson and SID values used for comparing the  
144 chemical profile of each factor/source to all of the others attributed by practitioners to the  
145 same source category (ff tests) in Round 1. More than 75 % of the Pearson values are  
146 above the limit of acceptance (broken line), indicating that the majority of the source  
147 categories present relatively comparable chemical compositions. The most heterogeneous  
148 categories (SHIP, BRA, DUST, SEC, STEEL and ZINC) show between 25% and 75% of  
149 factor/sources in the rejection area.

150 In this step, the number of factor/sources passing the SID test is, in the majority of cases,  
151 lower than those passing the Pearson. Therefore, there are more categories with profiles  
152 in the rejection area (e.g. DIE and LEAD).

153 Considering the two indicators, SHIP and BRA are amongst the most heterogeneous  
154 categories. The dissimilarities observed within the SHIP source category are likely due to  
155 the variety of chemical profiles allocated to it in the reported solutions. Due to similar fuel  
156 and combustion conditions, SHIP source profiles may be difficult to distinguish from  
157 stationary sources such as energy plants, oil refineries and other industrial processes  
158 (Viana et al., 2014). Only six profiles were attributed to the heterogeneous category BRA.  
159 Some of them, obtained with unconstrained factor analysis (APCS), are of difficult  
160 interpretation due to the extremely high concentration of Ca or the absence of Ba.

161 In Round 2, Pearson and SID tests point out SALT and TRA as categories where a  
162 discrete number of chemical profiles diverge from the reference (Fig. 3; see discussion in  
163 sections 3.2.1.2 and 3.2.1.3). In addition, Pearson test highlights also factor/sources in  
164 INDU as poorly comparable to their reference source chemical profile. This source  
165 category is, by definition, quite heterogeneous considering that it includes factor/sources  
166 attributed to different types of industries, combustion processes, without excluding regional  
167 (secondary) aerosol. Because of their simple chemical composition,  $\text{SO}_4$  and  $\text{NO}_3$  are the  
168 source categories in which factor/source profiles resemble more the reference profile in

169 the Pearson tests. Nevertheless, these source categories are much less homogeneous  
170 when tested using SID, which gives more weight to minor components in the factor/source  
171 profiles. This may indicate there are different sources of precursors associated to these  
172 secondary compounds.

173 The very limited changes observed in the Pearson values with log-transformed data in the  
174 two steps suggest that this kind of transformation is not solving efficiently the problem of  
175 dominant species in the profiles. For a more detailed discussion about the indicators of  
176 similarity see the companion paper by Belis et al. (2015).

177 The correlation (Pearson) between factor/sources identified in Round 1, on the basis of  
178 their time series, is summarized in Fig. 4 (left). The time series of BioB, COPPER, LEAD,  
179 NO<sub>3</sub> and ZINC are quite comparable among the different reported results. For the  
180 industrial sources, the time correlation is attributed to the effect of the intermittent pattern  
181 determined by the changes in wind direction and the time windows in which the emitting  
182 facility was in operation. Other sources, such as BioB and NO<sub>3</sub>, are synchronous due to  
183 common seasonal patterns determined by the trends in the emission rates and in  
184 atmospheric variables (e.g. air temperature, thermal inversion).

185 Factor/sources in the categories BRA, DIE, INDU, SEC, SHIP, and TRA display different  
186 temporal patterns. Most of these sources show also medium to poor correlation among the  
187 different chemical profiles (Fig. 2). The poor time correlations in factor/sources of the  
188 categories TRA, DIE and GAS may, at least in part, be connected with the time resolution  
189 of the data used for Round 1. One sample every sixth day may not be optimal to capture a  
190 sufficient number of weekends to show the week day/weekend patterns.

191 In Round 2, the time trends of the factor/sources are quite comparable with the reference  
192 for the majority of the source categories.

193 Despite the good correlations among the reported chemical profiles, likely determined by  
194 the presence of a combination of organic carbon and characteristic trace elements (e.g.  
195 Cu, Sn and Cr), ROAD is the source category with the lowest correlation between the  
196 reported time trends and the reference. This has been interpreted as the influence, to  
197 varying extents in each solution, of elements like Si, Al, and Mg that are also typical of  
198 DUST profiles and that may blur the boundary between these two categories. Also INDU  
199 shows quite variable results in this test and the considerations made for Round 1 are valid  
200 also in this case.

201 Source categories with inhomogeneous chemical profiles, such as INDU, often present  
202 poorly correlated time trends suggesting that an imperfect separation and identification of  
203 the sources leads to a poor fit in both the chemical composition and the temporal pattern.

204 Nevertheless, this general rule is not always valid. For instance, the time trends of SALT in  
205 Round 2 are quite comparable (Fig. 4) even though the chemical profiles of the  
206 factor/sources attributed to it are not homogeneous (Fig. 3). This apparent contradiction is  
207 explained by the high variance between the SALT time trends in the different reported  
208 results that is not detected by the Pearson test because the oscillations are synchronous.  
209

### 210 **3.2.1.2. Contribution-to-species**

211 The contributions of sources to the mass of every single species in the dataset expressed  
212 as percentage (contribution-to-species) were reported only in Round 2 (Fig. 5). The results  
213 reported in the different solutions are quite comparable among each other and with the  
214 reference source. As already observed in the tests for chemical profiles, INDU and ROAD  
215 show a number of records in the action area. Also the factor/sources in NO<sub>3</sub>, that are  
216 comparable with the reference in terms of time trend, show a non-negligible share of  
217 scores in the action area. In this category, the lower scores observed in the contribution-to-  
218 species may be attributed to the lower influence of dominating species, like ammonium  
219 nitrate, and higher influence of minor species such as Ca, As, Mo, Rb, Cl and PAHs.  
220 On the other hand, factor/sources in the SALT category, which show poor correlation with  
221 the concentrations in the reference profile, are well correlated with the reference in terms  
222 of contribution-to-species. In the SALT chemical profiles, Cl and Na represent on average  
223 81% and 49% of the source mass, respectively, and their relationship is close to the  
224 stoichiometric ratio in sodium chloride. As for the contribution-to-species, the ratio between  
225 the two elements (39% and 58% of the SALT mass, respectively) indicates that the share  
226 of Cl in SALT is lower than the one it would have been if the only source consisted of  
227 NaCl. This mismatch indicates the contribution of additional sources to this element other  
228 than sea and road salt (e.g. INDU).

### 229 **3.2.2. Chemical Profile Uncertainty**

230 In order to assess the uncertainty of the factor/source profiles, the weighted differences  
231 (WD, Karagulian and Belis, 2012) between the source profiles reported by participants and  
232 the corresponding reference profiles were computed.

233 The interpretation of WD scores depends on the relevance of the reference value for the  
234 factor/sources being tested. If a factor/source has been attributed to the wrong source  
235 category, the reference is not appropriate to evaluate that factor/source. For that reason,  
236 WD are interpreted by taking into account the results of the chemical profile tests (see  
237 section 3.2.1.1).

238 In Round 1, the fr tests were carried out using external reference profiles available in the  
239 literature and are, therefore, used only for informative purposes (not reported).  
240 The WD test shows that, in Round 2, SALT is the category with the highest proportion of  
241 scores outside the area of acceptance (above the broken line) followed by NO<sub>3</sub>, INDU,  
242 SO<sub>4</sub> and ROAD (Fig. 6). The analysis of the chemical profile's uncertainty using the WD  
243 indicator shows that, in this round, 65 % of factor/sources present acceptable WD scores.  
244 In addition, the joint evaluation with the chemical profile test suggests that only 18% of the  
245 factor/source profiles, which allocation to source categories was confirmed,  
246 underestimated their uncertainty.

247

### 248 3.3. Performance tests

249 In this section the results of the tests aiming at evaluating the SCEs, the most important  
250 output of a source apportionment study, are presented. The assessment of the SCE time  
251 trends is discussed in the companion paper by Belis et al. (2015).

#### 252 3.3.1. Reported Source Contribution Estimates

253 The distributions of the SCEs reported by participants in Round 1 and 2 are shown in  
254 Supplementary Material S2. The coefficients of variation (CVs) of the SCE reported by  
255 participants for every source category are, on average, 0.77 and 0.45 in the first and  
256 second round, respectively. NO<sub>3</sub> and SO<sub>4</sub> are the source categories with the lowest CV  
257 (between 0.26 and 0.48). In Round 1, CVs higher than the unity are observed in DUST,  
258 SHIP, INDU and ZINC while GAS, DIESEL and BRA show values in the range 0.80 -1.00.  
259 In Round 2, the SCEs are higher, because of the higher PM levels, and their relative  
260 variability within source categories is lower than in Round 1. The highest CV is the one of  
261 SALT (0.70) followed by DUST and INDU (0.60 and 0.55, respectively). As in Round 1, the  
262 lowest CVs are those in SO<sub>4</sub> and NO<sub>3</sub> (0.28 and 0.31, respectively).

#### 263 3.3.2. Z-scores

264 Fig. 7 summarises the z-scores assigned to each factor/source reported by participants in  
265 Round 1. The z-scores are in the acceptance area 85% of the time, 3% in the warning  
266 area, and 12% in the action area. The majority of solutions, 19 out of 22, present at least  
267 75% of the scores in the acceptance area. Only solution G2 presents the majority of  
268 scores in the action area. Such performance is likely due to the problems in mass  
269 quantification highlighted in the complementary tests (section 3.1.1).

270 DUST is the source category with the highest variability and the highest number of scores  
271 in the action area due to overestimation (6 scores) while SHIP and BRA are the ones with  
272 the highest number of scores in the action area due to underestimation (4 and 2 scores,  
273 respectively). Source categories DIE, GAS, BIOB, INDU and ZINC present three or less  
274 profiles with scores in the upper action area each. Inaccuracy in the SCE estimation of  
275 DUST, SHIP and BRA have been associated with the lack of homogeneity in the chemical  
276 profiles of the source factors attributed to them, as pointed out in the preliminary tests.  
277 Alternatively, those factors/sources with poor scores in DIE and, GAS are likely connected  
278 to results affected by the limited number of weekend days included in the dataset, as  
279 indicated by the preliminary test on time trends. The few z-scores of INDU ranking in the  
280 action area may be associated with divergences in both time trends and chemical profiles.  
281 In Round 1, about 80% of the reported factor/sources were obtained either with  
282 EPAPMF3, PMF2 or CMB8.2. In each of these models, more than 80% of the z-scores are  
283 placed in the area of acceptance. Interpretation of the results of the other models should  
284 be made with caution due to the limited number of reported solutions obtained with them.  
285 An 89% of the z-scores assigned to factor/sources reported by participants in Round 2 are  
286 in the acceptance area, while 2% and 9% are in the warning and action areas, respectively  
287 (Fig. 8). The majority of solutions, 21 out of 25, had more than 75% of the scores in the  
288 acceptance area.

289 SALT is the only source category with more than half of the scores in the action area. The  
290 overestimation of the SALT SCEs in the majority of solutions is likely due to the small  
291 contribution of this source category, which represents only 1% of the total PM mass.  
292 These low-contributing factors are likely to be severely affected by the remaining ambiguity  
293 derived from scaling indeterminacy. Their contributions and composition could be  
294 underestimated/overestimated by a large unknown coefficient (Amato et al., 2009). The  
295 negative SCE reported in a result obtained with FA-MLRA also contributed to the poor  
296 performance in this source category and further highlights the limitations of fully  
297 unconstrained factor analytical methods. A common drawback of tools without non-  
298 negativity constraints is the attribution of negative SCEs to minor sources to compensate  
299 the excess of mass attributed to others.

300 As in Round 1, INDU shows some z-scores ranking either in the warning or in the action  
301 areas. The performance of this source category in the two rounds is likely caused by the  
302 poor match in the chemical composition and time trends between the factor/sources  
303 reported in the solutions and the reference values. A limited degree of overestimation is  
304 also observed in ROAD, as shown by one of the scores in the action area. As discussed in



305 section 3.2.1.2, this can be attributed to the interference of DUST, especially during windy  
306 days, that may also lead to inaccuracies in the time trends. A propensity to underestimate  
307 source categories with high SCEs such as NO<sub>3</sub> and to a lesser extent SO<sub>4</sub> (29% and 17%  
308 of the PM mass, respectively) is present in many solutions. Nevertheless, the bias is too  
309 small to give rise to poor scores.

310 In Round 2, about 75% of the reported SCEs derive from solutions obtained with  
311 EPAPMF3, PMF2 and CMB8.2 and their performances are comparable to those observed  
312 in Round 1. Although a limited number of solutions are available for the other models, it is  
313 worth mentioning the good performances of COPREM, EPAPMF5 and ME-2. FA-MLRA is  
314 the only model with 50% of the scores either in the warning or action areas.

315 The z'-score indicator was used in Round 2 to assess the difference between solutions  
316 and the reference value taking into account the reference's uncertainty. No substantial  
317 differences were observed between z-scores and z'-scores indicating that the uncertainty  
318 of the reference had no impact on the evaluation of participant's performance.

### 319 **3.3.3. The uncertainty of the source contribution estimates**

320 In source apportionment modelling, there are different sources of error: random error,  
321 modelling error (bias), and rotational ambiguity (Paatero et al., 2013). One important  
322 source of random error is the one present in the input data and is commonly approximated  
323 from their analytical uncertainty. Modelling error arises in situations in which the RM  
324 assumptions (Belis et al., 2013) are seriously infringed. It may derive from wrong number  
325 of sources or variation of sources in time and is mostly contributing to the bias kind of  
326 error. Also atmospheric composition and meteorology acting selectively on the degradation  
327 of organic tracers (Galarneau, 2008) are a component of the bias error.

328 Many RM tools supply the output uncertainty. In EPA PMF versions, the uncertainty of the  
329 output profiles is estimated using re-sampling and more recently also with displacement  
330 methods while the CMB EPA 8.2 model performs a propagation of the input analytical  
331 uncertainty. Many practitioners using non-US EPA tools compute the output uncertainty  
332 with resampling and error propagation techniques in post-processing. The rotational  
333 ambiguity is not discussed in this section because only one of the used of tools (EPA PMF  
334 v5) was designed to estimate this kind of uncertainty. More discussion about the  
335 uncertainty test can be found in the companion paper by Belis et al. (2015).

336 The tests described in the previous sections were mostly oriented to assess: a) the bias by  
337 comparison with a reference value and b) the reproducibility intended as the range of  
338 results that can be obtained from a single dataset (with a given degree of noise) by  
339 different practitioners using the same or different tools. In the following, the analysis will



340 focus on the assessment of the SCEs uncertainty estimation accomplished by RMs by  
341 comparing them with the one of the reference. Considering that unbiased reference values  
342 are available only for the synthetic dataset, in this section are discussed only the results of  
343 Round 2.

344 The mean of the reported relative standard uncertainties for the SCE of the whole time  
345 window in Round 2 is 13%. The lowest values are those in NO<sub>3</sub> source categories and the  
346 highest are those in INDU. As for the models, the lowest uncertainties are those reported  
347 in ME-2 and CMB8.2 solutions and the highest are those of COPREM solutions. No  
348 uncertainty was reported for the SCEs obtained with FA-MLRA. The uncertainty attributed  
349 to the reference was equivalent to the noise introduced in the synthetic dataset (20%  
350 standard deviation) that was derived from the analytical uncertainty in the input dataset  
351 (Belis et al., 2015). The zeta-score test indicates that a 68% of the declared factor/source  
352 SCE uncertainties are coherent with the one of the reference while a 19%, ranking in the  
353 action area, are likely underestimated (Fig. 9).

354 SALT is the only source category with the majority of the zeta-scores in the action area  
355 (75%). Likely, models do not allow for the higher relative uncertainty due the very low  
356 SCEs in this source category compared to the others. Uncertainty underestimation is  
357 observed also in ROAD, which shows a 60% of the scores either in the warning or in the  
358 action areas.

359 A considerable proportion of factor/sources obtained with EPAPMF4 and EPAPMF3 show  
360 underestimated uncertainties (29% and 24% of scores in the action area, respectively).  
361 COPREM showed uncertainties higher than the reference in a 31% of the factor/sources.  
362 The satisfactory performance of CMB8.2 (more than 90% successful scores) suggests that  
363 propagating the uncertainty of the source profiles can provide a satisfactory estimation of  
364 the SCEs uncertainty.

#### 365 **3.3.4. The impact of the operator**

366 The variability between solutions obtained by different practitioners using the same tool  
367 and the same input data are an indicator of the maximum impact of the operator  
368 subjectivity on the reproducibility. The tools with the highest number of reported solutions:  
369 EPAPMF3, PMF2, and CMB8.2 present a high consistency among solutions obtained by  
370 different practitioners using the same tool. The standard deviations of the SCE mean in  
371 each of these models ranges between 0.2 - 0.3  $\mu\text{g}/\text{m}^3$  and 1.4 - 1.7  $\mu\text{g}/\text{m}^3$ , in the first and  
372 second rounds, respectively. These values are, in addition, close to the standard deviation  
373 of the overall mean (0.2  $\mu\text{g}/\text{m}^3$  and 1.7  $\mu\text{g}/\text{m}^3$ , in the first and second rounds, respectively).  
374 These results suggest a limited impact of the practitioners' subjectivity, on average.

375 However, “outliers” were often associated with less experienced practitioners in terms of  
376 both years of use of the tool and number of studies performed.  
377

#### 378 4. KEY FINDINGS OF THE INTERCOMPARISON

379 The tests on chemical profiles confirmed, in the majority of cases (83%), the attribution of  
380 factors/profiles to source categories in the reported results and the majority of the SCEs  
381 (87%) reported by participants met the 50% standard uncertainty quality objective  
382 established for the performance test. A high share of the tested solutions (70% - 80%)  
383 apportioned a considerable amount of the PM<sub>2.5</sub> mass to its pollution sources and many  
384 solutions estimated a number of sources close to the expected value.

385 In this study, the estimation of source contribution was most critical for SALT, DUST, SHIP  
386 and categories associated with mobile sources. The majority of the solutions  
387 overestimated the SCE of SALT, a source category with a contribution of about 1% of the  
388 PM mass. Such relative contribution may be considered a first approximation of the lower  
389 limit that the tested methodologies are able to quantify. Poor scores attributed to some  
390 DUST and ROAD SCEs were ascribed to the similarities in the chemical composition  
391 between road dust and crustal material that may have interfered with the allocation of  
392 mass between these sources. The uncorrelated time trends and, in some cases, the  
393 heterogeneous chemical profiles observed in INDU and SHIP were attributed to the lack of  
394 a common definition of these categories. Sources with appreciable contributions and  
395 chemical profiles dominated by few species, such as NO<sub>3</sub> and SO<sub>4</sub>, were more efficiently  
396 recognised by the models even though there was a tendency to slightly underestimate  
397 their SCEs.

398 The most commonly used models, EPAPMF3, PMF2, and CMB8.2 showed quite  
399 satisfactory performance with successful z-scores ranging between 80% and 100%. The  
400 good agreement between CMB and PMF may be partially due to the main RM  
401 assumptions being substantially respected in the used datasets: limited alteration of the  
402 species between source and receptor and relatively stable source profiles. In addition,  
403 both types of tools account for the uncertainties in the input data, have built-in  
404 performance indicators and have been available long enough to allow a wide number of  
405 practitioners be familiar with them. For those models used in a limited number of solutions,  
406 only preliminary conclusions can be drawn at this stage. In general, fully unconstrained  
407 models which do not account for the input data uncertainty (e.g. FA-MLRA and APCFA)  
408 showed performances below the average. This result is likely because in these tools, the

409 noise deriving from the uncertainty structure of the datasets is incorporated into the  
410 factor/sources (Paatero and Hopke, 2003).

411 The tests used to assess the SCE uncertainty reported in the solutions confirmed that the  
412 RMs output uncertainty estimation is coherent with the analytical/random uncertainty of the  
413 input data. Other components of the uncertainty could be evaluated in specially designed  
414 intercomparisons where RMs are either compared with other types of models or synthetic  
415 datasets with known perturbing factors are used. Processes altering the factor/source  
416 chemical profiles could be detected in the preliminary tests by comparison with the  
417 reference source profiles. In addition, diagnostic ratios could be used to detect long-range  
418 transport or photochemical age of aerosols (Hien et al., 2004; DeCarlo et al., 2010).

419 The slightly better performance observed in Round 2 compared to Round 1 is likely  
420 connected to the differences between simulated and measured data. Round 1 was more  
421 challenging for the participants due to the inconsistencies in the uncertainties they had to  
422 deal with in a blind test with limited information about a non-European study area. On the  
423 contrary, the synthetic dataset contained internally consistent data with a lower level of  
424 noise and fewer source categories.

425 In the real-world, the variability of profiles in time and the chemical reactivity of organic  
426 species may affect the source/receptor relationships. Datasets from areas with complex  
427 atmospheric transport and chemistry are likely more challenging for models to quantify the  
428 sources (especially secondary and/or distant ones) than areas influenced mainly by local  
429 sources. In this study, there are no indications that the variability of profiles and  
430 degradation of markers affected the comparability of results among participants working on  
431 the same dataset. On the other hand, it was observed that the time resolution of the  
432 datasets influenced the ability of RMs to capture the time patterns of mobile sources.

## 433 5. CONCLUSIONS

434 The results of this study indicate that RMs are capable of estimating the contributions of  
435 the main pollution source categories within a given time window with a level of accuracy  
436 that is in line with the needs of air quality management.

437 Further intercomparisons evaluated with the same or comparable methodologies are  
438 needed to create a weight-of-evidence about the characteristics and capabilities of the  
439 models and tools.

440 Future work to improve the capacity of these models should focus on: a) the development  
441 and availability of source profiles relevant for the study area, b) better definition of the  
442 source categories, c) experimental design to improve the uncertainty estimation, d)

443 development of speciated PM data series with appropriate time resolution and extended  
444 set of markers.

445 Moreover, the implementation of common guidelines (Belis et al., 2014) would lead to  
446 more comparable results with recognised quality standards in line with those reported in  
447 the present work.

## 448 6. ACKNOWLEDGMENTS

449 The authors are grateful to J. Schauer, and R. Turner for making available the organic  
450 dataset collected in St. Louis for the first round of the intercomparison. We thank C.  
451 Samara and G. Argyropoulos for sharing their data elaborations and for the fruitful  
452 discussions during this work.

## 453 7. REFERENCES

- 454  
455 Amato, F., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Pey, J., Perez, N., and Hopke,  
456 P. K., 2009. Quantifying road dust resuspension in urban environment by Multilinear  
457 Engine: A comparison with PMF2. *Atmospheric Environment* 43, 2770-2780  
458 Analytical Methods Committee, 1989. Robust statistics – How not to reject outliers. Part 1:  
459 Basic Concepts. *The Analyst* 114, 1697-1698  
460 Belis, C. A., Karagulian, F., Larsen, B. R., and Hopke, P. K., 2013. Critical review and  
461 meta-analysis of ambient particulate matter source apportionment using receptor  
462 models in Europe. *Atmospheric Environment* 69, 94-108  
463 Belis, C. A., Pernigotti D., Karagulian F., Pirovano G., Larsen B. R., Gerboles M., and P.K,  
464 H., 2015. A new methodology to Assess the Performance and Uncertainty of  
465 Source Apportionment Models. *Atmospheric Environment* 119, 35-44  
466 Belis, C. A., Larsen, B., Amato, F., El Haddad, I., Favez, O., Harrison, R. M., Hopke, P. K.,  
467 Nava, S., Paatero, P., Prévot, A., Quass, U., Vecchi, R., and Viana, M. 2014,  
468 *European Guide on Air Pollution Source Apportionment with Receptor Models*, p.  
469 88. Publication Office of the European Union, Italy.  
470 Decarlo, P. F., Ulbrich, I. M., Crouse, J., De Foy, B., Dunlea, E. J., Aiken, A. C., Knapp,  
471 D., Weinheimer, A. J., Campos, T., Wennberg, P. O., and Jimenez, J. L., 2010.  
472 Investigation of the sources and processing of organic aerosol over the Central  
473 Mexican Plateau from aircraft measurements during MILAGRO. *Atmospheric*  
474 *Chemistry and Physics* 10, 5257-5280  
475 Favez, O., El Haddad, I., Piot, C., Boréave, A., Abidi, E., Marchand, N., Jaffrezo, J. L.,  
476 Besombes, J. L., Personnaz, M. B., Sciare, J., Wortham, H., George, C., and  
477 D'anna, B., 2010. Inter-comparison of source apportionment models for the  
478 estimation of wood burning aerosols during wintertime in an Alpine city (Grenoble,  
479 France). *Atmospheric Chemistry and Physics* 10, 5295-5314  
480 Friedlander, S. K., 1973. Chemical element balances and identification of air pollution  
481 sources. *Environmental Science and Technology* 7, 235-240  
482 Galarneau, E., 2008. Source specificity and atmospheric processing of airborne  
483 PAHs: implications for source apportionment. *Atmospheric Environment* 42, 8139-  
484 8149.  
485 Hennigan, C.J., Sullivan, A.P., Collett Jr., J.L., Robinson, A.L., 2010. Levoglucosan  
486 stability in biomass burning particles exposed to hydroxyl radicals. *Geophysical*  
487 *Research Letters* 37 (L09806), 4.

- 488 Henry R. C., C. W. Lewis, P. K. Hopke, and H. J. Williamson 1984. Review of Receptor  
489 Model Fundamentals. *Atmospheric Environment* 18:1507-1517.
- 490 Hien, P. D., Bac, V. T., and Thinh, N. T. H., 2004. PMF receptor modelling of fine and  
491 coarse PM<sub>10</sub> in air masses governing monsoon conditions in Hanoi, northern  
492 Vietnam. *Atmospheric Environment* 38, 189-201
- 493 Hopke, P. K., 2010. The application of receptor modeling to air quality data. *Pollution  
494 Atmospherique*, 91-109
- 495 Hopke, P. K., Ito, K., Mar, T., Christensen, W. F., Eatough, D. J., Henry, R. C., Kim, E.,  
496 Laden, F., Lall, R., Larson, T. V., Liu, H., Neas, L., Pinto, J., Stölzel, M., Suh, H.,  
497 Paatero, P., and Thurston, G. D., 2006. PM source apportionment and health  
498 effects: 1. Intercomparison of source apportionment results. *Journal of Exposure  
499 Science and Environmental Epidemiology* 16, 275-286
- 500 ISO 13528 (2005), Statistical methods for use in proficiency testing by interlaboratory  
501 comparisons, edited, (ISO) International Organization for Standardization.
- 502 Jaekels, J. M., Bae, M. S., and Schauer, J. J., 2007. Positive matrix factorization (PMF)  
503 analysis of molecular marker measurements to quantify the sources of organic  
504 aerosols. *Environmental Science and Technology* 41, 5763-5769
- 505 Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M. a. M., Helber, R., Arnone,  
506 R.A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill  
507 assessment. *J. Mar. Syst.* 76, 64-82.
- 508 Karagulian, F., and Belis, C. A., 2012. Enhancing source apportionment with receptor  
509 models to foster the air quality directive implementation. *International Journal of  
510 Environment and Pollution* 50, 190-199
- 511 Larsen, B. R., Gilardoni, S., Stenström, K., Niedzialek, J., Jimenez, J., and Belis, C. A.,  
512 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty  
513 characterization and sensitivity analysis of secondary and primary sources.  
514 *Atmospheric Environment* 50, 203-213
- 515 Larsen, B. R., Junninen, H., Monster, J., Viana, M., Tsakovski, P., Duvall, R. M., Norris, G.  
516 A., and Querol, X. 2008, The Krakow receptor modelling intercomparison exercise  
517 *Rep.*, JRC Scientific and Technical Reports, EUR 23621 EN 2008, Ispra.
- 518 Lee, J. H., Hopke, P. K., and Turner, J. R., 2006. Source identification of airborne PM<sub>2.5</sub>  
519 at the St. Louis-Midwest Supersite. *Journal of Geophysical Research D:  
520 Atmospheres* 111, D10S10, 1-12.
- 521 Paatero, P., 1997. Least squares formulation of robust non-negative factor analysis.  
522 *Chemometrics and Intelligent Laboratory Systems* 37, 23-35
- 523 Paatero, P., and Hopke, P. K., 2003. Discarding or downweighting high-noise variables in  
524 factor analytic models. *Analytica Chimica Acta* 490, 277-289
- 525 Paatero, P., Eberly, S., Brown, S. G., and Norris, G. A., 2013. Methods for estimating  
526 uncertainty in factor analytic solutions. *Atmos. Meas. Tech. Discuss.* 6, 7593-7631
- 527 Pandolfi, M., Viana, M., Minguillón, M. C., Querol, X., Alastuey, A., Amato, F., Celades, I.,  
528 Escrig, A., and Monfort, E., 2008. Receptor models application to multi-year  
529 ambient PM<sub>10</sub> measurements in an industrialized ceramic area: Comparison of  
530 source apportionment results. *Atmospheric Environment* 42, 9007-9017
- 531 Schauer, J. J., Rogge, W. F., Hildemann, L. M., Mazurek, M. A., Cass, G. R., and  
532 Simoneit, B. R. T., 1996. Source apportionment of airborne particulate matter using  
533 organic compounds as tracers. *Atmospheric Environment* 30, 3837-3855
- 534 Thomson, M., Ellison, S. L. R., and Wood, R., 2006. The international harmonized protocol  
535 for the proficiency testing of analytical chemistry laboratories. *Pure and Applied  
536 Chemistry* 78, 145-196
- 537 Thurston, G. D., and Spengler, J. D., 1985. A quantitative assessment of source  
538 contributions to inhalable particulate matter pollution in metropolitan Boston.  
539 *Atmospheric Environment - Part A General Topics* 19, 9-25



- 540 Viana, M., Pandolfi, M., Minguillón, M. C., Querol, X., Alastuey, A., Monfort, E., and  
541 Celades, I., 2008b. Inter-comparison of receptor models for PM source  
542 apportionment: Case study in an industrial area. *Atmospheric Environment* 42,  
543 3820-3832
- 544 Viana, M., Hammingh, P., Colette, A., Querol, X., Degraeuwe, B., Vlieger, I. D., and Van  
545 Aardenne, J., 2014. Impact of maritime transport emissions on coastal air quality in  
546 Europe. *Atmospheric Environment* 90, 96-105
- 547 Viana, M., Kuhlbusch, T. a. J., Querol, X., Alastuey, A., Harrison, R. M., Hopke, P. K.,  
548 Winiwarter, W., Vallius, M., Szidat, S., Prévôt, A. S. H., Hueglin, C., Bloemen, H.,  
549 Wählin, P., Vecchi, R., Miranda, A. I., Kasper-Giebl, A., Maenhaut, W., and  
550 Hitzenberger, R., 2008a. Source apportionment of particulate matter in Europe: A  
551 review of methods and results. *Journal of Aerosol Science* 39, 827-849
- 552 Watson, J. G., Chen, L. W. A., Chow, J. C., Doraiswamy, P., and Lowenthal, D. H., 2008.  
553 Source apportionment: Findings from the U.S. supersites program. *Journal of the*  
554 *Air and Waste Management Association* 58, 265-288  
555  
556

Table 1. Outline of the datasets used in every round of the intercomparison exercises.

Table 2 Source categories, codes and reference values used in every round of the intercomparison.

Table 3. Average number of reported factor/sources by model

ACCEPTED MANUSCRIPT



	Round 1	Round 2
Type of data	Real-world dataset	Synthetic dataset
Site	Saint Louis (USA)	Milan (Italy)
Time window	June 2001 – May 2003	January – December 2005
Pollutant	PM <sub>2.5</sub>	PM <sub>2.5</sub>
Number of samples	178, 24 h samples	364, 24 h samples
Number of chemical species	44	38
Carbonaceous species	OC/EC (steps)	OC/EC (total)
Ionic species	sulphate, nitrate, ammonium	sulphate, nitrate, ammonium, chloride
Elements	Al, As, Ca, Cr, Cu, Fe, K, Mn, Ni, Pb, Rb, Si, Sr, Ti, V, Zn	
	Ba, Co, Hg, P, Se, Zr	Sb, Sn, Na, Mo, Cd, Mg
Organic species	indeno(cd)pyrene, benzo(ghi)perylene, benzo(a)pyrene, coronene, benzo(e)pyrene, dibenz[a,h]anthracene, levoglucosan	
	benz(a)anthracene, fluoranthene, pyrene, benzo(b,k)fluoranthene, benzo(j)fluoranthene	chrysene, benzo(b)fluoranthene, benzo(k)fluoranthene

ROUND 1			ROUND 2		
Code	Source category	Reference SCE ( $\mu\text{g}/\text{m}^3$ )	Code	Source category	Reference SCE ( $\mu\text{g}/\text{m}^3$ )
BioB	Biomass burning / wood burning	1.59	BioB	Biomass burning / wood burning	4.33
BRA	Road dust / brake abrasion	0.83	SO4	Ammonium sulphate	7.12
COPPER	Copper production	0.57	NO3	Ammonium nitrate	12.69
DIE	Diesel vehicles	0.42	DUST	Soil dust/ crustal	4.01
DUST	Soil dust/ crustal	0.74	ROAD	Road dust	2.68
GAS	Gasoline vehicles	0.59	SALT	Sea salt / road salting	0.52
INDU	Industrial emissions/combustion	1.07	TRA	Exhaust emission from vehicles	6.63
LEAD	Lead smelter	0.42	INDU	Industrial emissions/point sources	5.11
NO3	Ammonium nitrate	2.98			
SEC	Secondary aerosol	6.36			
SHIP	Ship emissions	1.63			
SO4	Ammonium sulphate	5.99			
STEEL	Steel processing	1.57			
TRA	Traffic exhaust	2.44			
ZINC	Zinc smelter	0.58			

model	ROUND AVERAGE	CMB8.2	PMF2	EPA PMF3	EPA PMF4	EPA PMF5	ME-2	COPREM	PCA	APCS	FA- MLRA	REFERENCE
ROUND1	9	8	9	9	-	-	6	13	7	11	-	10*
ROUND2	9	8	8	7	7	8	8	13	-	-	6	8

\* indicative reference.

Fig. 1. Target diagrams summarizing the mass apportionment in the first (left) and second (right) rounds. The outer circle delimits the acceptance area and the inner circle represents the boundary of scores with Pearson equal to 0.7. Only scores outside the inner circle are labelled with the model abbreviation and solution code. RMSD': unbiased root mean square difference (Jolliff et al., 2009).

Fig. 2. Similarity of factor/source chemical profiles in each source category (ff tests) in Round 1 calculated using Pearson (left) and SID (right). Pearson: values above the broken line rank in the area of acceptance. SID: accepted values are those below the broken line. The number of tested factor/sources is reported on top of each bar.

Fig. 3. Comparison of factor/source chemical profiles with the reference profile for every source category (fr tests) in Round 2 calculated using Pearson (left) and SID (right). Pearson: values above the broken line rank in the area of acceptance. SID: accepted values are those below the broken line. The number of tested factor/sources is reported on top of each bar.

Fig. 4. Comparison of factor/source time series in Round1 (ff tests, left) and in Round 2 (fr tests, right) using Pearson. Values above the broken line rank in the area of acceptance. The number of tested factor/sources is reported on top of each bar.

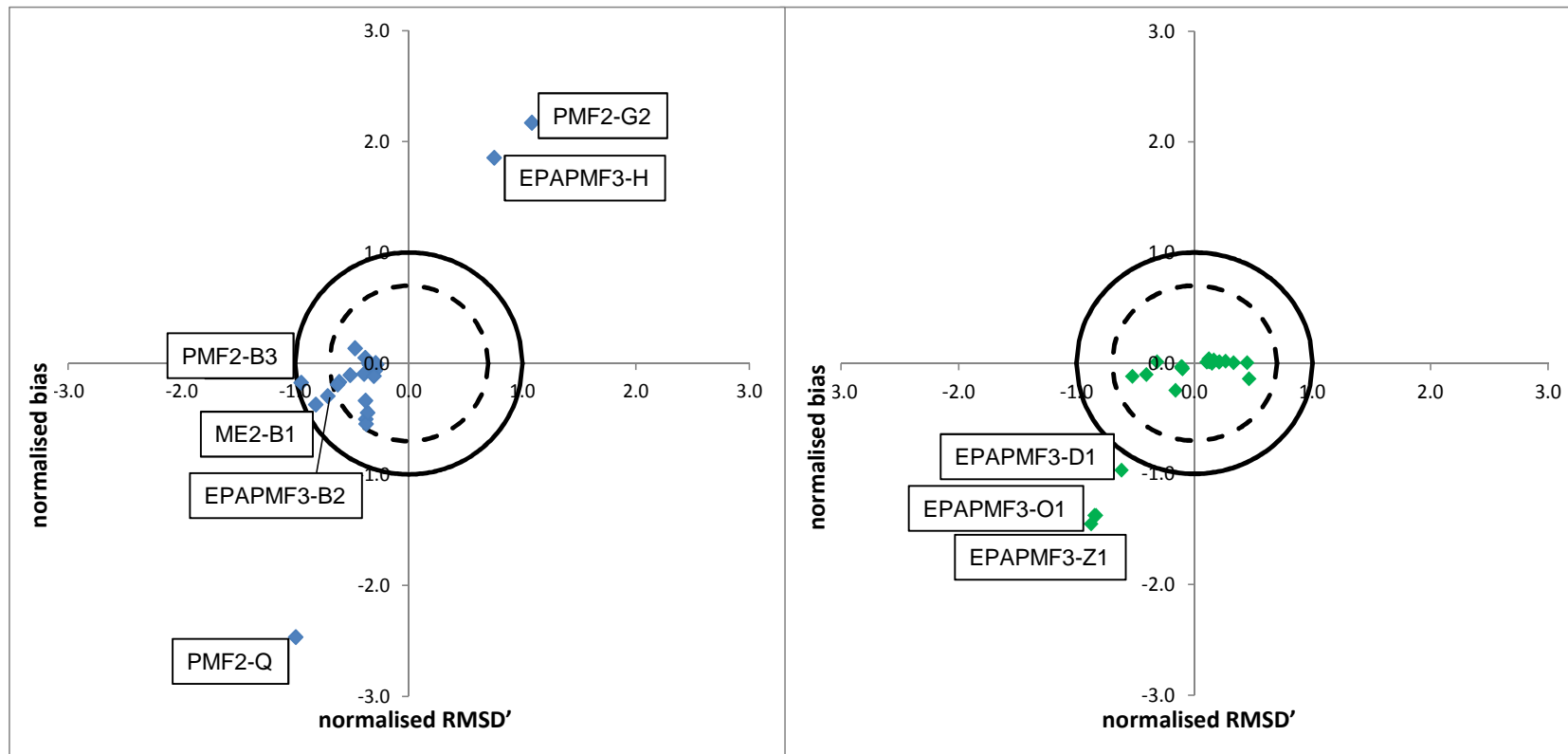
Fig. 5. Comparison of factor/source contribution-to-species with the reference profile for every source category (fr tests) in Round 2. Values above the broken line rank in the area of acceptance. The number of tested factor/sources is reported on top of each bar.

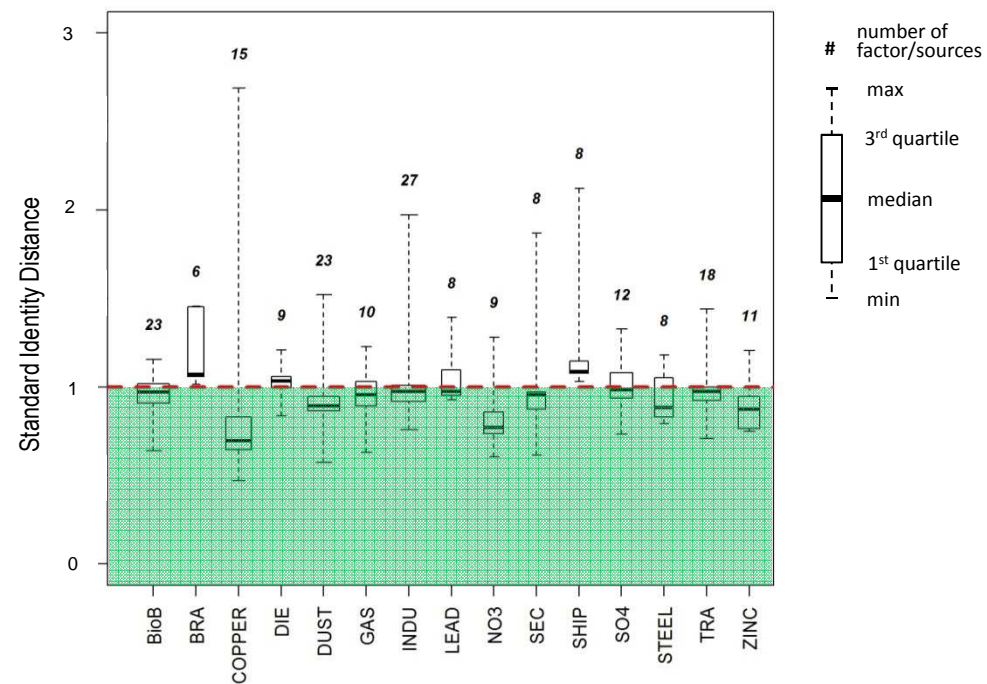
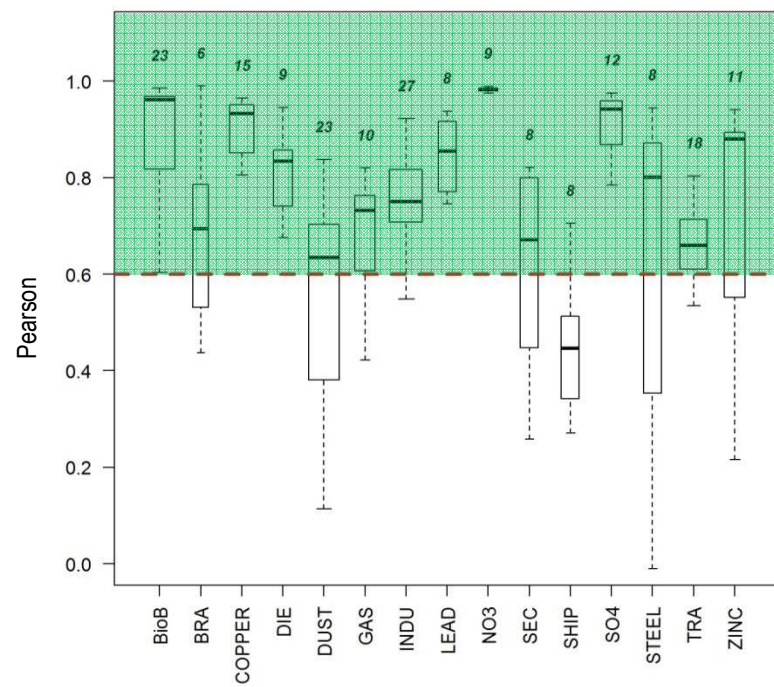
Fig. 6. Evaluation of chemical profiles uncertainties, using the weighted difference (WD) indicator in Round 2 (fr tests). Values below the broken line rank in the area of acceptance. The number of tested factor/sources is reported on top of each bar.

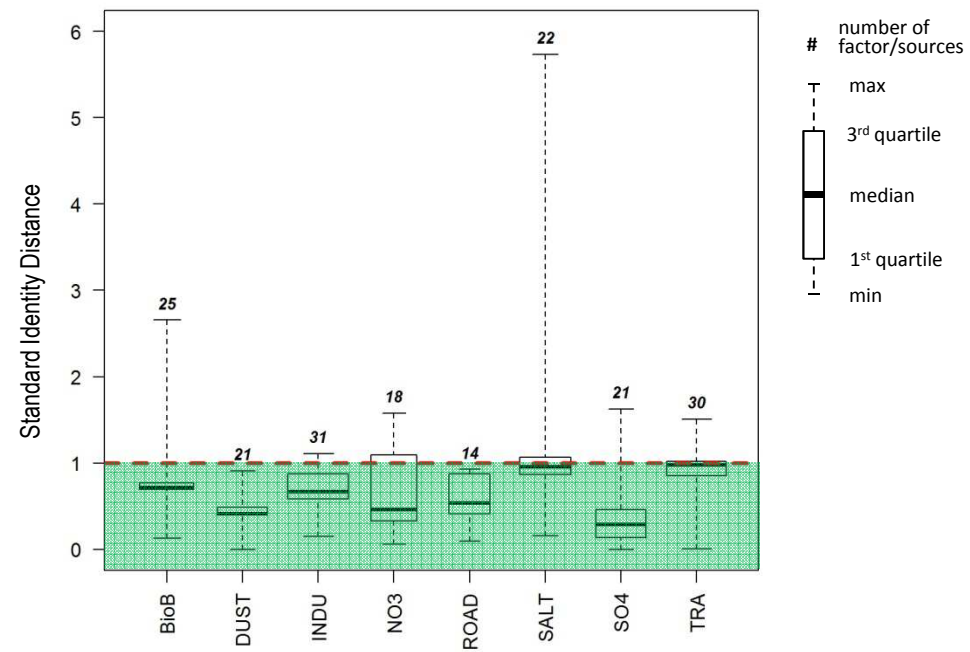
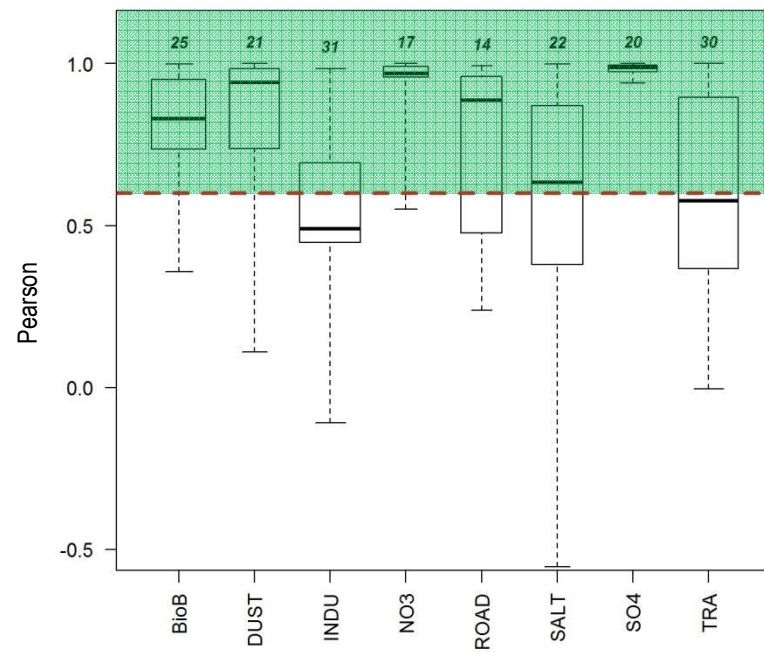
Fig. 7. Z-scores attributed to the factor/profiles in Round 1 arranged by source category (left) and by model (right). Scores outside the zone between continuous lines rank in the action area, those in the space between the continuous and the broken lines rank in the warning area and those in the zone within the broken lines rank in the acceptance area. The number of tested factor/sources is reported on top of each bar.

Fig. 8. Z-scores attributed to the factor/sources in Round 2 arranged by source category (left) and by model (right). Scores outside the zone between continuous lines rank in the action area, those in the space between the continuous and the broken lines rank in the warning area and those in the zone within the broken lines rank in the acceptance area. The number of tested factor/sources is reported on top of each bar.

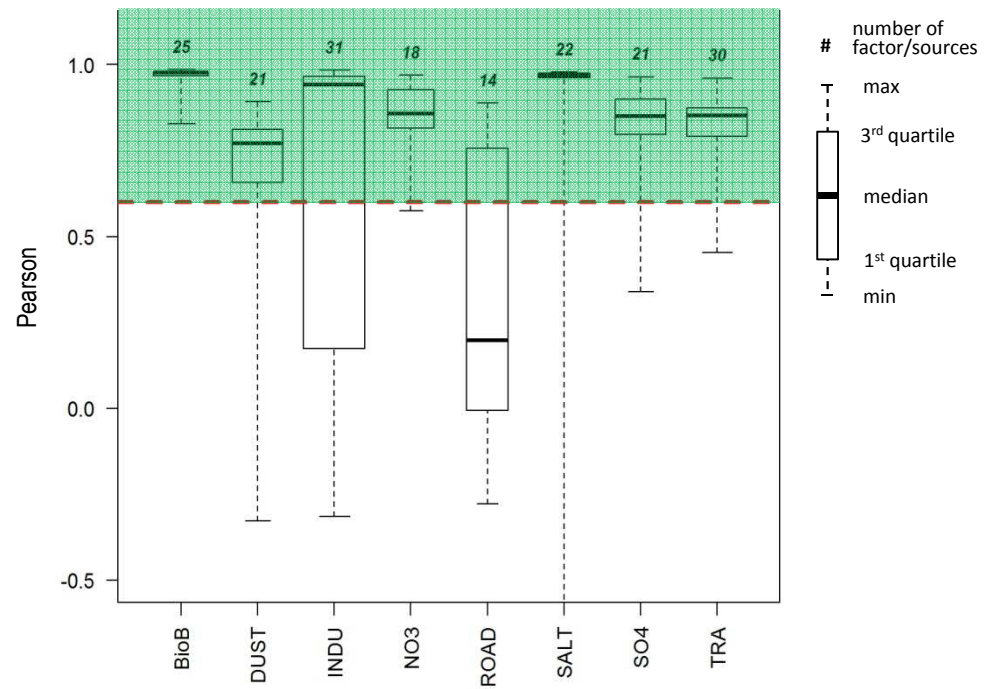
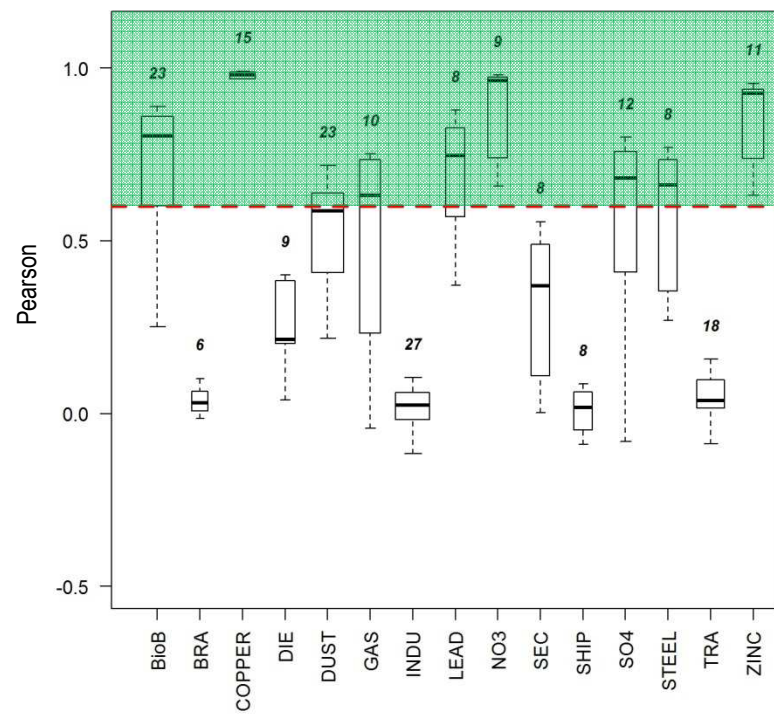
Fig. 9. Zeta-scores attributed to the factor/sources in Round 2 arranged by source category (left) and by model (right). Scores ranking above or below the continuous lines are in the action area, those in the space between the continuous and the broken lines are in the warning area and those in the zone within the broken lines are in the acceptance area. The number of tested factor/sources is reported on top of each bar.



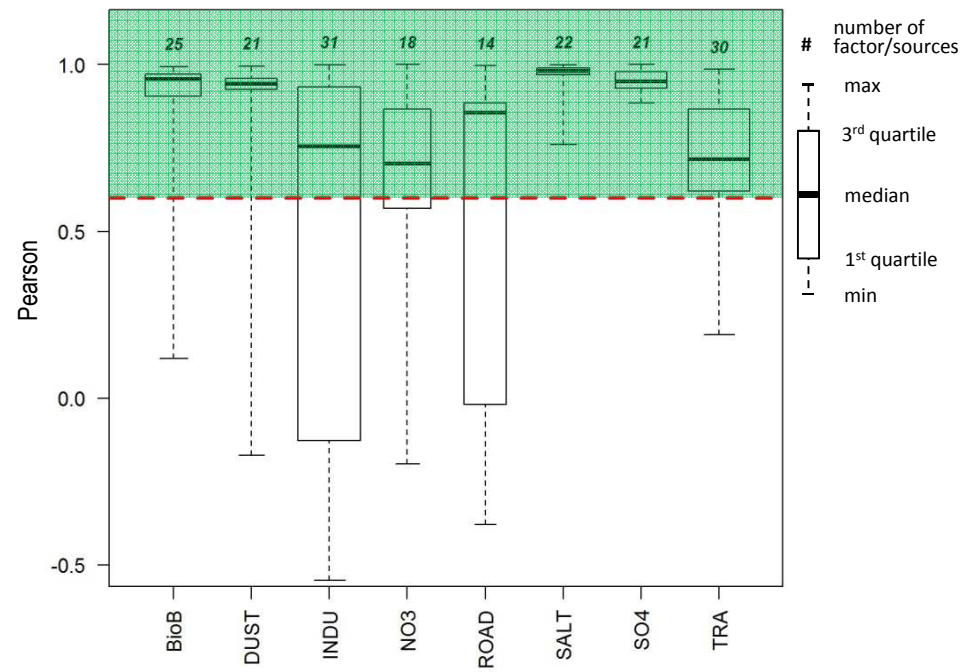


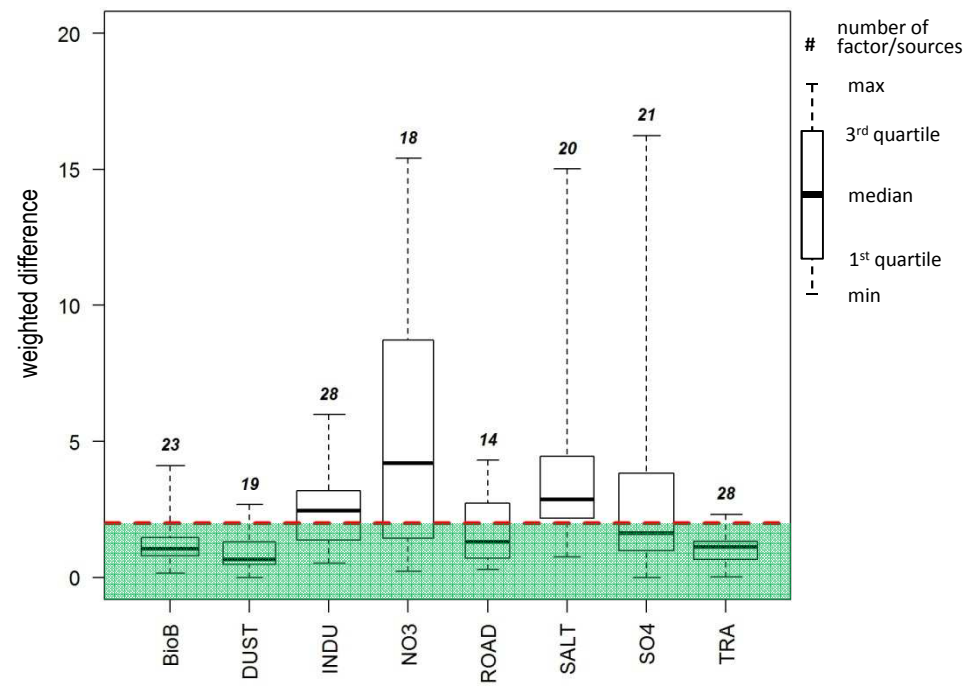


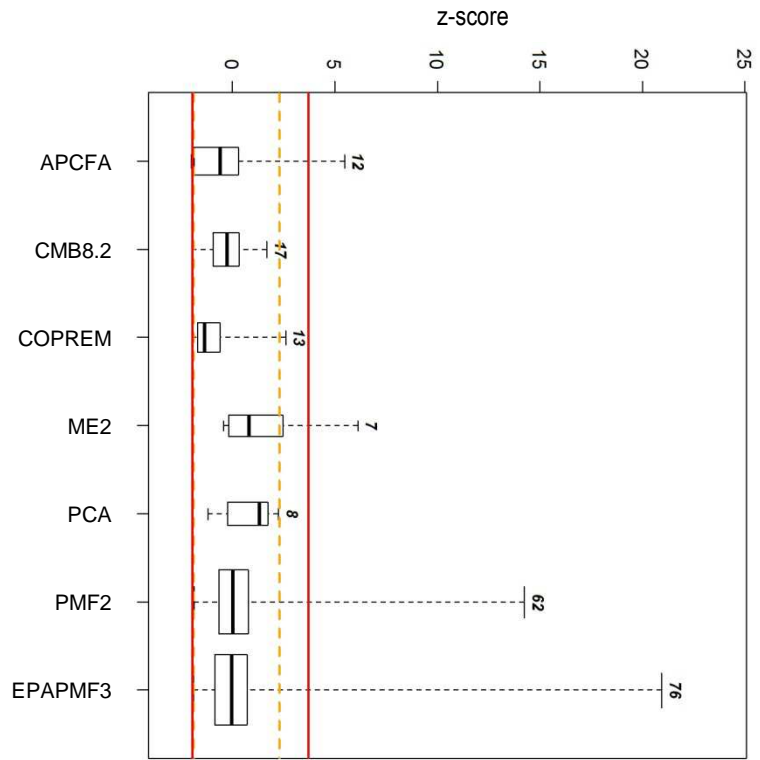
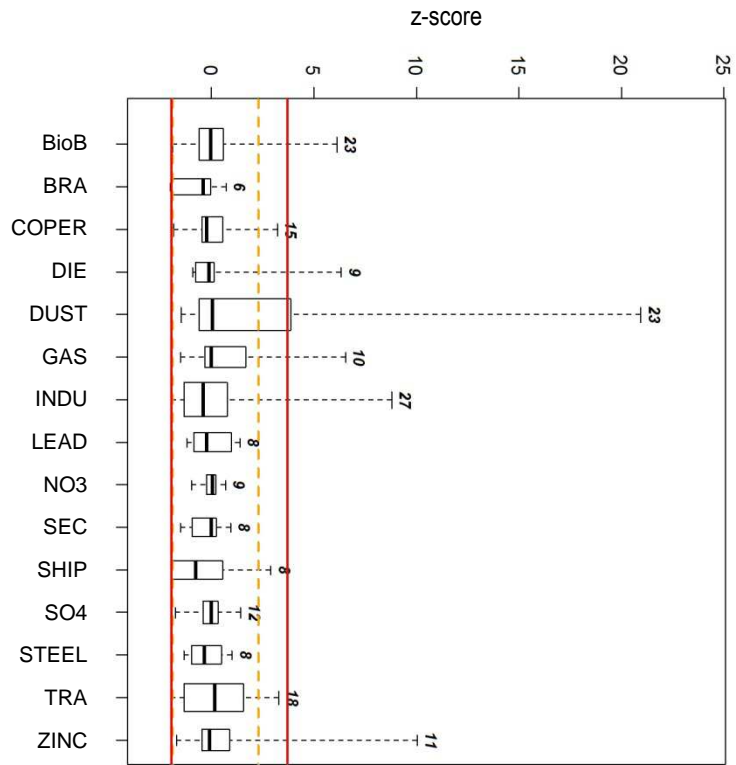




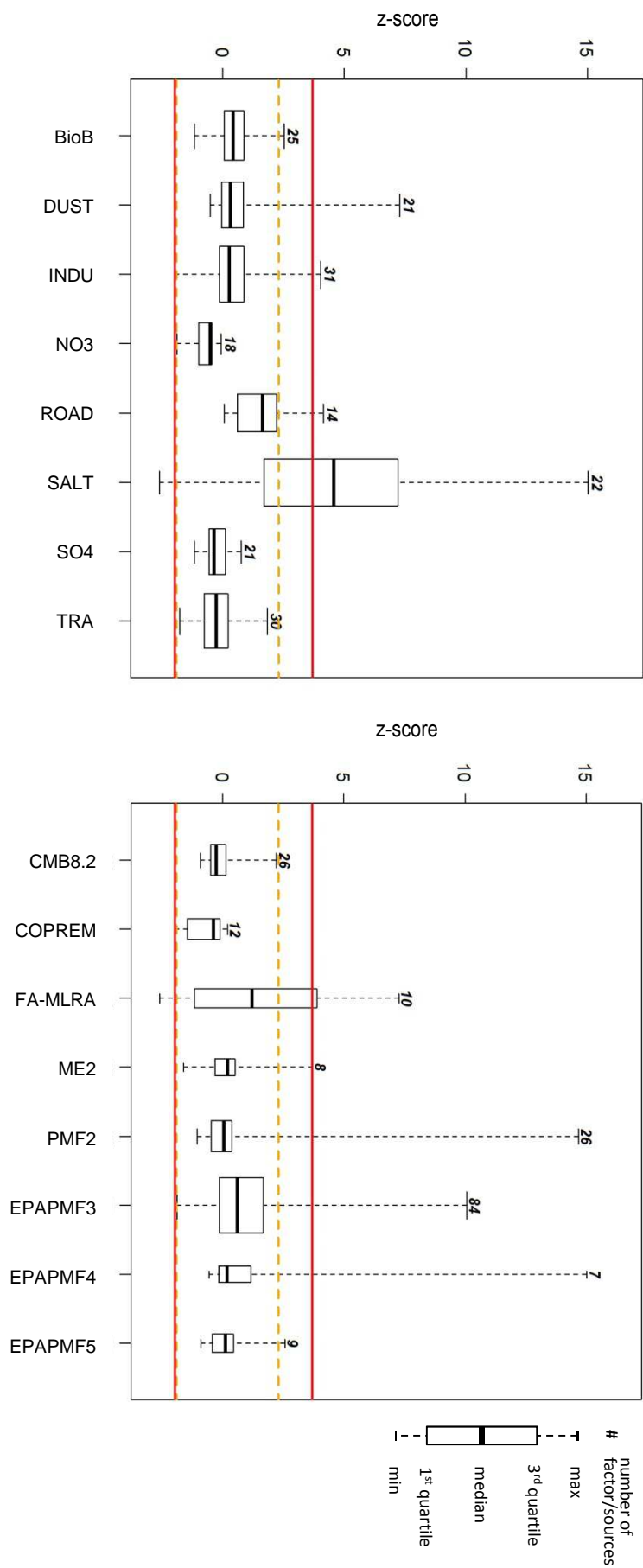
# number of factor/sources  
 T max  
 3<sup>rd</sup> quartile  
 median  
 1<sup>st</sup> quartile  
 min

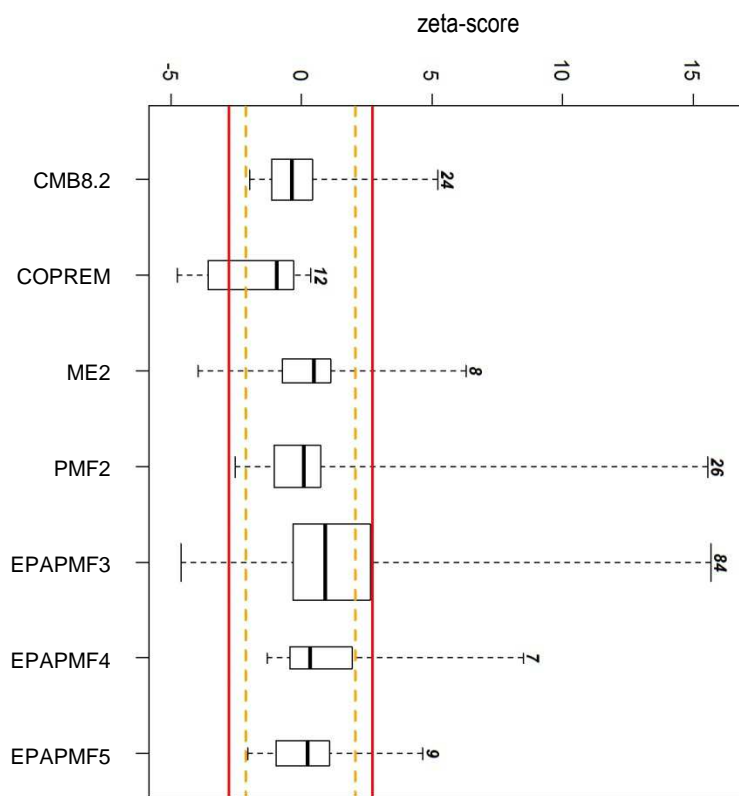
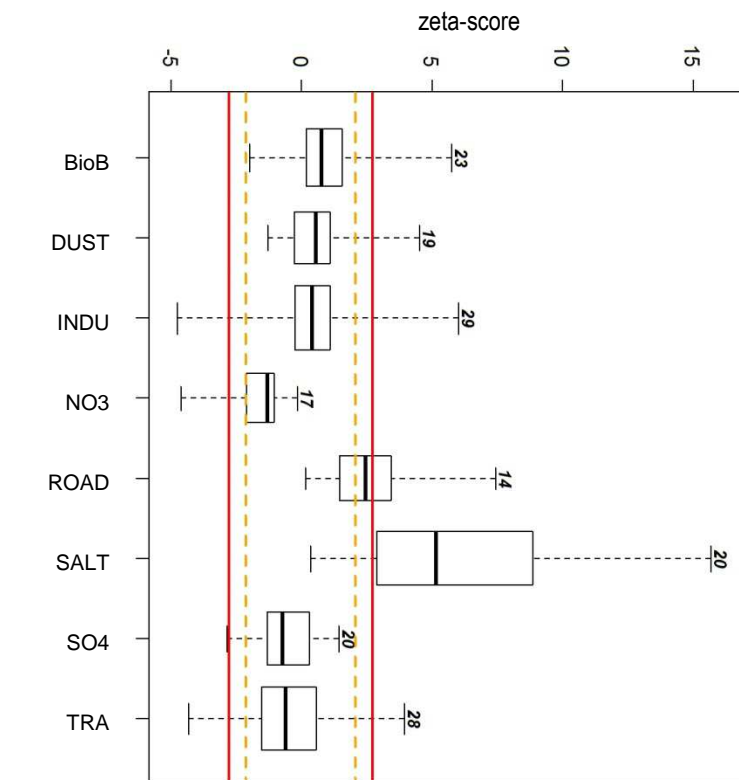






# number of factor/sources  
 T max  
 3<sup>rd</sup> quartile  
 median  
 1<sup>st</sup> quartile  
 min





# number of factor/sources  
 T max  
 3<sup>rd</sup> quartile  
 median  
 1<sup>st</sup> quartile  
 min

- 1 Intercomparisons were used to test the performance and uncertainty of receptor models.
- 2 More than 85% of the reported sources met the model quality objectives.
- 3 Two thirds of the output uncertainties were coherent with those in the input data.
- 4 PMF v2, 3 and CMB 8.2 estimated the source contributions satisfactorily.
- 5 The accuracy of receptor models is in line with the needs of air quality management.
- 6

ACCEPTED MANUSCRIPT