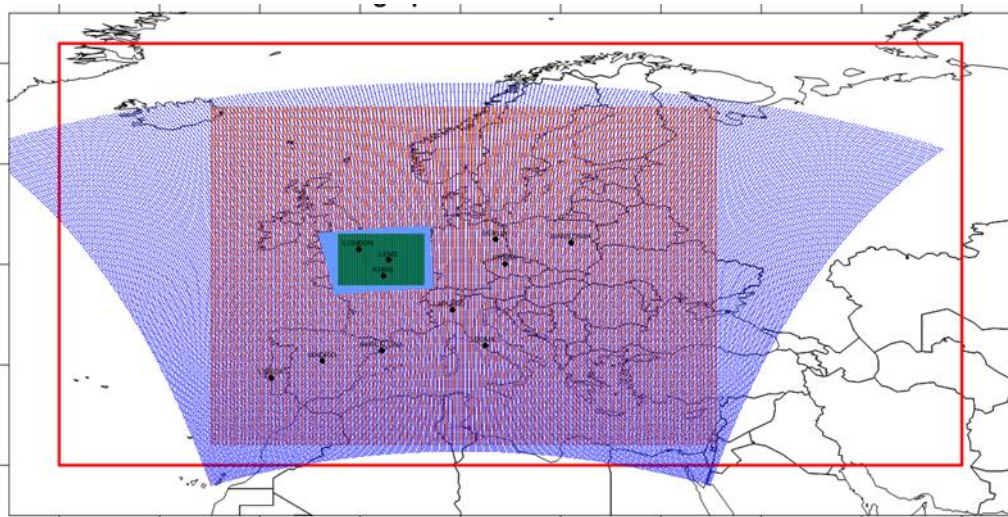# JRC TECHNICAL REPORTS

Results of the first European Source Apportionment intercomparison for Receptor and Chemical Transport Models

C. A. Belis, D. Pernigotti, G. Pirovano

In collaboration with: O. Favez, J.L. Jaffrezo, J. Kuenen, H. Denier van Der Gon, M. Reizer, M.T. Pay, M. Almeida, F. Amato, A. Aniko, G. Argyropoulos, S. Bande, I. Beslic, M. Bove, P. Brotto, G. Calori, D. Cesari, C. Colombi, D. Contini, G. De Gennaro, A. Di Gilio, E. Diapouli, I. El Haddad, H. Elbern, K. Eleftheriadis, J. Ferreira, G. Foret, M. Garcia Vivanco, S. Gilardoni, S. Hellebust, R. Hoogerbrugge, Y. Izadmanesh, H. Jorquera, A. Karppinen, Z. Kertesz, T. Kolesa, K. Krajsek, R. Kranenburg, P. Lazzeri, F. Lenartz, N. Liora, Y. Long, F. Lucarelli, K. Maciejewska, A. Manders, M. Manousakas, H. Martins, M. Mircea, D. Mooibroek, S. Nava, D. Oliveira, P. Paatero, M. Paciorek, M. Paglione, M. Perrone, E. Petralia, A. Pietrodangelo, S. Pillon, P. Pokorna, A. Poupkou, F. Pradelle, P. Prati, V. Riffault, D. Salameh, C. Samara, L. Samek, D. Saraga, S. Sauvage, F. Scotto, K. Sega, G. Siour, R. Tauler, G. Valli, R. Vecchi, E. Venturini, M. Vestenius, G. Yarwood, E. Yubero

2018

How to cite this report:
Belis C. A., Pernigotti D., Pirovano G., Favez O., Jaffrezo J.L., Kuenen J., Denier van Der Gon H., Reizer M., Pay M.T., Almeida M., Amato F., Aniko A., Argyropoulos G., Bande S., Beslic I., Bove M., Brotto P., Calori G., Cesari D., Colombi C., Contini D., De Gennaro G., Di Gilio A., Diapouli E., El Haddad I., Elbern H., Eleftheriadis K., Ferreira J., Foret G., Garcia Vivanco M., Gilardoni S., Hellebust S., Hoogerbrugge R., Izadmanesh Y., Jorquera H., Karppinen A., Kertesz Z., Kolesa T., Krajsek K., Kranenburg R., Lazzeri P., Lenartz F., Liora N., Long Y., Lucarelli F., Maciejewska K., Manders A., Manousakas M., Martins H., Mircea M., Mooibroek D., Nava S., Oliveira D., Paatero P., Paciorek M., Paglione M., Perrone M., Petralia E., Pietrodangelo A., Pillon S., Pokorna P., Poupkou A., Pradelle F., Prati P., Riffault V., Salameh D., Samara C., Samek L., Saraga D., Sauvage S., Scotto F., Sega K., Siour G., Tauler R., Valli G., Vecchi R., Venturini E., Vestenius M., Yarwood G., Yubero E., 2018 Results of the first European Source Apportionment intercomparison for Receptor and Chemical Transport Models, EUR 29254 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-86573-2, doi: 10.2760/41815, JRC 111887.

Contents

Authors

C.A. Belis, D. Pernigotti, Joint Reserch Centre, Directorate C Energy, Transport and Climate

G. Pirovano, RSE SpA

O. Favez, INERIS

J.L. Jaffrezo, D. Salameh, CNRS – LGGE

R. Kranenburg, J. Kuenen, A. Manders, H. Denier van Der Gon , TNO

M. Reizer, K. Maciejewska, M. Paciorek, Warsaw University of Technology

M.T. Pay, Barcellona Supercomputer Centre

M. Almeida, IST, University of Lisbon

F. Amato, Y. Izadmanesh, R. Tauler, IDAEA-CSIC

A. Aniko, Z. Kertesz, Institute for Nuclear Research, Atomiki, HAS

G. Argyropoulos, C. Samara, Aristotle University of Thessaloniki

S. Bande, ARPA Piemonte

I. Beslic, K. Sega, IMROH

M. Bove, P. Brotto, P. Prati, University of Genoa

G. Calori, ARIANET

D. Cesari, D. Contini, ISAC –CNR Lecce

C. Colombi, ARPA Lombardia

G. De Gennaro, A. Di Gilio, ARPA Puglia/ University of Bari

E. Diapouli, K. Eleftheriadis, M. Manousakas, D. Saraga, NCSR "Demokritos"

I. El Haddad, Paul Scherrer Institute

H. Elbern, K. Krajsek, Rhenish Institute for environmental Research, Univ. of Cologne

J. Ferreira, H. Martins, University of Aveiro

G. Foret, G. Siour, LISA-CNRS

M. Garcia Vivanco, CIEMAT

S. Gilardoni, M. Paglione, ISAC –CNR Bologna

S. Hellebust, University College Cork

R. Hoogerbrugge, D. Mooibroek, RIVM

H. Jorquera, Pontifical Catholic University of Chile

A. Karppinen, M. Vestenius, Finnish Meteorological Institute

T. Kolesa, Slovenian Environment Agency, ARSO

P. Lazzeri, APPA Trento

F. Lenartz, Institut Scientifique de Service Public - Wallonie

N. Liora, A. Poupkou, Aristotle University of Thessaloniki

Y. Long, F. Pradelle, G. Yarwood, Ramboll Environ, Inc.

F. Lucarelli, S. Nava, INFN

M. Mircea, E. Petralia, ENEA

D. Oliveira, V. Riffault, S. Sauvage, Ecole des Mines de Douai – SAGE

P. Paatero, University of Helsinki

M. Perrone, University of Milano-Bicocca

A. Pietrodangelo, IIA - CNR

S. Pillon, ARPA Veneto

P. Pokorna, Clarkson University, CARES

L. Samek, AGH-UST, University of Science and Technology

F. Scotto, ARPA Emilia-Romagna

G. Valli, R. Vecchi, University of Milan

E. Venturini, University of Bologna

E. Yubero, Miguel Hernandez University

Abstract

In this study, the performance of the source apportionment model applications were evaluated by comparing the model results provided by 44 participants adopting a methodology based on performance indicators: z-scores and $RMSE_u$, with pre-established acceptability criteria. Involving models based on completely different and independent input data, such as receptor models (RMs) and chemical transport models (CTMs), provided a unique opportunity to cross-validate them. In addition, comparing the modelled source chemical profiles, with those measured directly at the source contributed to corroborate the chemical profile of the tested model results. The most used RM was EPA- PMF5. RMs showed very good performance for the overall dataset (91% of z-scores accepted) and more difficulties are observed with SCE time series (72% of $RMSE_u$ accepted). Industry resulted the most problematic source for RMs due to the high variability among participants. Also the results obtained with CTMs were quite comparable to their ensemble reference using all models for the overall average (>92% of successful z-scores) while the comparability of the time series is more problematic (between 58% and 77% of the candidates' $RMSE_u$ are accepted). In the CTM models a gap was observed between the sum of source contributions and the gravimetric PM_{10} mass likely due to PM underestimation in the base case. Interestingly, when only the tagged species CTM results were used in the reference, the differences between the two CTM approaches (brute force and tagged species) were evident. In this case the percentage of candidates passing the z-score and $RMSE_u$ tests were only 50% and 86%, respectively. CTMs showed good comparability with RMs for the overall dataset (83% of the z-scores accepted), more differences were observed when dealing with the time series of the single source categories. In this case the share of successful $RMSE_u$ was in the range 25% - 34%.

1 Introduction

Assessing the performance of Source Apportionment (SA) model results is essential to guarantee reliable information on source contributions to be used in the development of pollution abatement strategies and for reporting to the Commission.

The performance of the source apportionment model application is evaluated by comparing the model results with reference values by means of performance indicators and assessing whether the difference falls within pre-established acceptability conditions.

Considering that it is not possible to measure the contribution of pollution sources at a given site or area, intercomparison exercises are the only option to generate SA reference values with the ensemble method. Such approach consists in averaging the output from different models/users' runs using the same input data. An alternative method, used only for RMs so far, is the development of synthetic datasets where the contributions from all sources are known. Intercomparison exercises give, in addition, the chance to measure the overall output uncertainty. In this intercomparison we have followed the method based on similarity and performance tests described in Belis et al (2015 a and b) which has been adopted in the previous FAIRMODE WG3 intercomparisons (see section 3).

One of the distinctive features of this inter-comparison is that both receptor models (RMs) and chemical transport models (CTMs) were applied on the same study area so called "reference site". This approach allowed the creation of an unprecedented dataset of both source oriented and receptor oriented evaluation of source contribution estimates (SCEs).

Involving models based on completely different and independent input data, such as RMs and CTMs provides a unique opportunity to cross-validate the obtained results from different types of models. In addition, comparing the results of the models with source chemical profiles measured directly at the emitting source (e.g. from SPECIEUROPE database, Pernigotti et al., 2016), contributes to corroborate the chemical profile of the tested sources.

The overall assessment of the SA results is also expected to provide insights to understand the models behaviour in terms of influence of specific factors (e.g. input data, type of site, type of pollutant, meteorological conditions, etc…).

Moreover, the intercomparison between RMs and CTMs is intended as a first step towards the integration of the two families of models in order to take advantage of the strength of every approach and control their limitations.

The improved definition of sources and characterisation of model performance and uncertainty is going to contribute to a better integration between source apportionment and planning activities in FAIRMODE and provide the basis for incorporating in integrated assessment tools like SHERPA.

Last but not least, the experience gained in the intercomparison will contribute to the standardisation process on SA model performance in progress under CEN WG44.

2 Intercomparison overview and methodology

This Intercomparison Exercise (IE) for Receptor and Source Oriented Models was organised within the framework of FAIRMODE Working Group 3 on Source Apportionment. The list of participants is given in Annex 1. The main objective of the IE was to assess the performance and the uncertainty of the SA methodologies and to compare different approaches.

This was the first intercomparison ever designed to test both receptor oriented models (RMs) and source oriented models (CTMs, in particular) using a comprehensive method based on model quality indicators and pre-established criteria. The target pollutant of this IE was PM_{10}.

The main input data for RMs was a real-world dataset of PM_{10} measurements with a high number of organic species while the input for CTMs included an emission inventory, meteorological fields and chemical boundary conditions.

The MACC emission inventory (Kuenen et al. 2014, 2015) used for this intercomparison contained enhanced details on fuels for the SNAP categories 2 (domestic and commercial) and 7 (road transport) that made it possible a better comparison of the results from the two families of models.

3 The methodology used for the intercomparison evaluation

The methodology adopted for the assessment of the SA model performance consists of several indicators and tests to evaluate different aspects of a source apportionment result (Figure 1). It encompasses three types of tests: a) complementary tests that provide information about the SA result as a whole, b) similarity tests that aim at comparing the candidate sources with a reference on the basis of their chemical composition and the time trends, and c) the performance tests that assess whether the mass of a pollutant attributed to a source category is coherent with the reference value on the basis of pre-established quality criteria. See list of abbreviations used in the text.
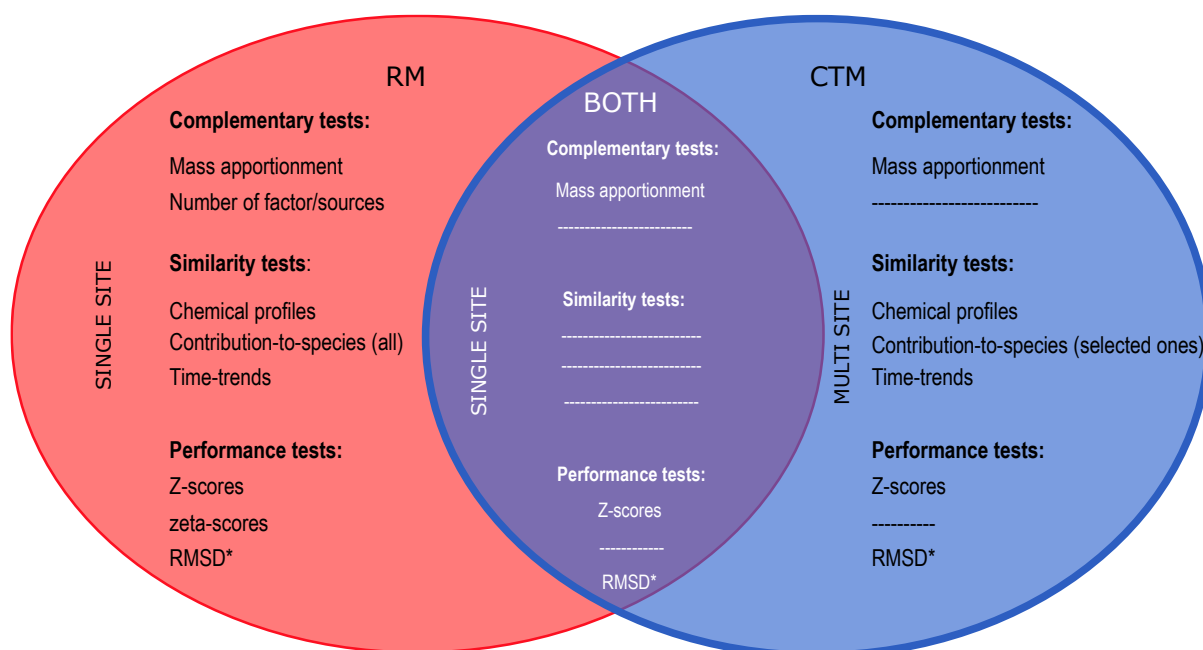

Figure 1 – Tests used to compare SA results from RMs and CTMs within each of these model approaches (left and right) and among both of them (centre). Single site: only Lens; multisite: Lens plus other nine sites.

The "complementary tests" provide an overall indication of the SA result in terms of apportioned mass. In order to test the agreement between the apportioned mass and the total PM mass, the sum of the mass of all the candidate sources in every SA result is compared with the gravimetric mass using the RMSD* according to Jolliff et al., (2009) and Thunis et al., (2012). Values ≤ 1 are considered indicators of good agreement.
In RMs also the number of sources (hereon candidates) reported in the different results is compared with their average. Deviations of more than 3 sources require further investigation.
The "similarity tests" (previously known as preliminary tests) are targeted at establishing whether the sources reported in the participants' results (candidates) are attributable to the source category indicated by the participant itself. In this test, the indicators are the Pearson Distance (1 - Pearson) and the Standardised Identity Distance (SID, Belis et al., 2015a) with acceptability thresholds: ≤0.4 and ≤1, respectively. Such indicators are

used to assess the chemical composition and the time trends of the candidates (SID only for chemical profiles) in two types of tests. The "f tests" are the comparison among factor/sources attributed by participants to the same source category in all the solutions while "r tests" refer to the comparison between reported candidates and a reference value.

The "performance tests" evaluate the mass attributed to every source: Source Contribution Estimates (SCEs). Considering that source apportionment studies are mostly targeted at identifying and quantifying the typical sources in the studied area, the performance tests were conducted on the average SCE over the whole time window represented in the dataset using the z score indicator with $\sigma_p=0.5$ of the reference SCE and the acceptability interval: -1.96 and 3.99. Moreover, the SCE time series were evaluated using the root mean square error ($RMSE_u$) normalised by the uncertainty of the reference value in every time step (standard deviation), as discussed in Belis et al. (2015a). As in RMSE*, $RMSE_u$ values ≤ 1 are considered indicators of good performance. Source categories with too few candidates were not evaluated and profiles attributed by participants to more than one category were tested in each of the proposed categories.

The same methodology was used for RMs and CTMs with different arrays of tests depending on the specific characteristics of these two modelling approaches, as shown in Figure 1. The reported results obtained with RMs and CTMs were first evaluated separately. For RMs only the reference site of Lens was available while a set of 10 different sites were used in the comparison between CTMs. In a second step, a cross comparison between the two approaches was accomplished. For methodological reasons, in this test the RMs were set as the reference. For a detailed description of the methodology refer to Belis et al. (2015a, 2015b).

4 The evaluation of receptor models (RMs)

The RM intercomparison was carried out using a real-world dataset of speciated PM_{10} collected at the reference site. Analytical detection limits and uncertainties were provided to participants for the estimation of the input uncertainty. Ancillary information consisting of gaseous pollutant concentrations, meteorological data and emission inventory relevant for the study area were also provided. Participants decided autonomously the method to perform the source apportionment.

4.1 The RM intercomparison dataset

The dataset contains 116 PM_{10} daily 24h concentrations collected every third day between March 2011 and March 2012 in the city of Lens (France). This study site is located in a background monitoring station in a large urbanised area (about 500,000 inhabitants within 100 km). Daily samples were collected on Quartz filter using a high volume (30 m^3/h) sampler.

The dataset has been produced within the framework of the CARA project designed and managed by INERIS, as part of the French reference laboratory for air-quality monitoring (LCSQA), involving also IGE, ATMO and LCME.

For every sample are reported 98 species: ions (anions and cations), OC/EC, trace elements, PAHs, anhydrosugars, hopanes, alkanes, POA-markers and the total PM_{10} mass.

The testing dataset including the concentrations, the analytical detection limits (ng/m3) and the relative uncertainties (%) was distributed to participants in July 2015.

4.2 RM complementary data

In addition to the testing dataset, supplementary information is provided.

1. <u>Meteorological data.</u>

A short set of most commonly used meteorological parameters: air temperature, atmospheric pressure, relative humidity, wind speed and direction from the nearest meteorological station (Lesquin) were provided. For this location, also a complete set of meteorological data with hourly time resolution was distributed. Only precipitation is provided for the city of Lens.

2. <u>Gaseous pollutants</u>

The concentrations of nitrogen oxides are provided for the monitoring station in Lens, while ozone and sulphur dioxide are provided for the Lens surroundings (Harnes).

3. <u>Emission Inventory (EI) of the study area</u>

The emissions in an area of few kilometres around the monitoring site were provided. Extensive information on emissions was available from the emission inventory described in section 5.3.

4. <u>Source profiles</u>

Reference source profiles for this exercise were those in the repositories SPECIEUROPE (http://source-apportionment.jrc.ec.europa.eu/Specieurope/index.aspx) and SPECIATE (https://www.epa.gov/air-emissions-modeling/speciate-version-45-through-40).

4.3 Receptor models' results

The Intercomparison involved 33 participants who delivered a total of 38 different RM results. The participants are listed in Annex 1. Each result is labelled with a letter from A to Z and then from *A to *L. Each of the results delivered by participants consisted of a set of candidate sources (hereafter referred to as '**result**', '**participant**' and '**candidates**', respectively). The candidates are encoded with an alphanumeric string consisting of a letter corresponding to the result and a number corresponding to the sequence of the candidate in the reported result matrix.

Table 1: Summary of the source categories identified in the results reported by participants using RM with the corresponding SPECIEUROPE code.

code	Sources categories	abbr.	hierarchy of categories	n. of reference profiles	n. of candidate sources
1	traffic	tra	1	286	34
2	exhaust	exh	2_1	130	20
3	diesel	die	3_2_1	49	2
4	gasoline	gas	4_2_1	10	3
5	road	roa	5_1	154	16
7	brake	bra	7_5_1	2	3
10	soil	soi	10	235	34
12	marine	ss	12		39
20	industry	ind	20	433	24
30	fuel oil	fue	30	88	31
31	coal	coa	31	47	8
37	ship	shi	37	14	9
40	biomass	bib	40	139	39
41	wood	woo	41_40	96	8
60	SIA	sia	60		20
61	ammonium nitrate	amn	61_60	1	20
62	ammonium sulfate	ams	62_60	1	21
69	metallurgy	met	69_20	43	2
70	POA	poa	70		33
71	aged sea	as	71	1	21
72	agriculture	agr	72	20	2

The majority of the results were obtained using EPA-PMF5 (31). The other tools were used in a few results: ME-2 scripts in two cases, and RCMB (robotic CMB), MLPCA (multilinear PCA), EPA-PMF4, EPA-PMF3 and PMF2 in one result each.

Table 2. Overview of reported RMs results

Result code.	Result letter	Model	n. of candidates	n. of species	n. of time steps
ATO	A	PMF5	6	47	108
AGH	B	PMF5	6	98	116
APP	C	PMF5	11	51	116
ARP E	D	PMF5	8	27	115
ARP L	E	PMF5	11	90	115
ARP P	F	PMF5	9	28	116
ARS	G	PMF5	8	89	88
AUT	H	RCMB	10	98	116
CAR	I	PMF5	9	27	116
ISC_B	J	PMF5	7	85	109
CNR	K	PMF5	9	38	116
ENE 12	L	PMF5	12	27	116
ENE 9	M	PMF5	9	27	116
INF	N	PMF5	10	30	116
FMI	O	PMF5	8	51	116
UNIG	P	PMF5	9	29	116
IDA_T	Q	MLPCA	7	92	116
IDA_A	R	PMF5	9	41	116
IMR	S	PMF5	8	98	116
ISS	T	PMF5	8	29	116
IST	U	PMF5	8	32	116
PSI	V	ME2	8	75	116
LGGE+	W	PMF5	10	36	116
NCS_5	X	PMF5	10	27	116
NCS_2	Y	PMF2	10	35	116
ISA_LE3	Z	PMF3	10	34	116
ISA_LE5	*A	PMF5	10	34	116
PUC	*B	PMF5	5	20	116
RIV1	*C	ME2	9	48	116
RIV2	*D	PMF5	9	48	116
RIV3	*E	PMF5	9	40	116
SAG	*F	PMF5	8	36	116
UCC	*G	PMF5	8	71	116
UMH	*H	PMF5	8	27	116
UNIB	*I	PMF4	12	49	116
UNIM	*J	PMF5	10	27	116
UNMB	*K	PMF5	8	34	116
WUT	*L	PMF5	10	36	116

The PM$_{10}$ mass uncertainty has been set to the maximum uncertainty accepted for PM$_{10}$ daily limit value (25%), so that the average over the period was 26.0 ± 9.0 μg/m^3.

The participants attributed each of their candidates to one or more source categories. In order to harmonise the nomenclature, the candidates were encoded in conformity with the SPECIEUROPE database source categories (Pernigotti et al., 2016). The source chemical profiles from this database are also referred to as 'reference' profiles (Table 1).

Worth of mention that SPECIEUROPE source categories are hierarchical (certain categories include others). The hierarchy of categories is given in the fourth column of Table 1. Only the source categories with at least two candidates are reported. No measured profiles for SIA (Secondary Inorganic Aerosol), POA (Primary Organic Aerosol) and aged sea are present in the source profile databases SPECIEUROPE (Pernigotti et al., 2016) and SPECIATE (Simon et al., 2010). An overview of the results is given in Table 2. The used model and the number of candidates, species and time steps are indicated therein.

4.3.1 Complementary tests

In 8 results (G, L, M, Q, V, Y, *I and *L) the sum of the SCE averages provided by participants do not fall within ±20% of the measured PM_{10} (gravimetric) average.

The test to assess the coherence between the sum of the average SCE of all the candidates (sources) in every reported result with the gravimetric PM_{10} mass (in $\mu m/m^3$) is summarised in Figure 2. RMSE* values in the green circle are those passing the test. The results **G, J** and ***E** do not meet the acceptability criterion for this test.
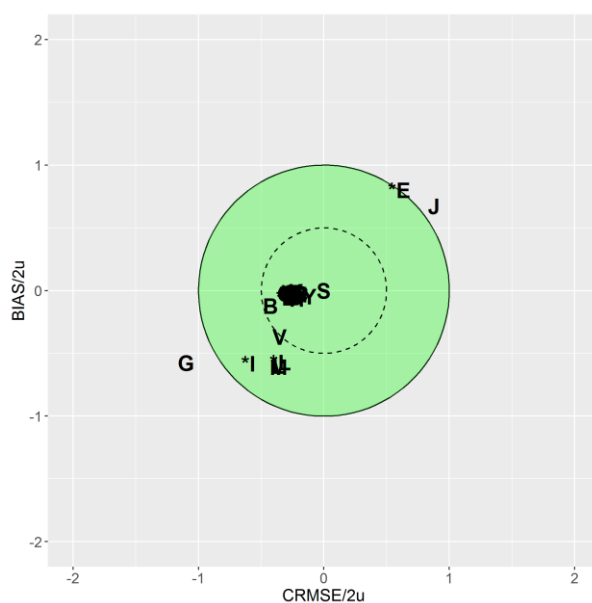

Figure 2. Target diagram representing RMSD* according to Jolliff et al., (2009) and Thunis et al., 2012. Values ≤ 1 (green circle) are considered indicators of good agreement between the sum of SCEs and the gravimetric PM_{10} mass.

The participants **O, *I** and **Q** were excluded from the list of candidates used in the calculations of the reference as these results presented a strong bias (one order of magnitude or more) in the mass of some chemical species. Candidates **G2** and **T7** were also excluded as their time series was always zero. Moreover, four candidates (**C2** and **C10**; **D3** and T7) were excluded from the analysis because attributed to 'undefined' sources.
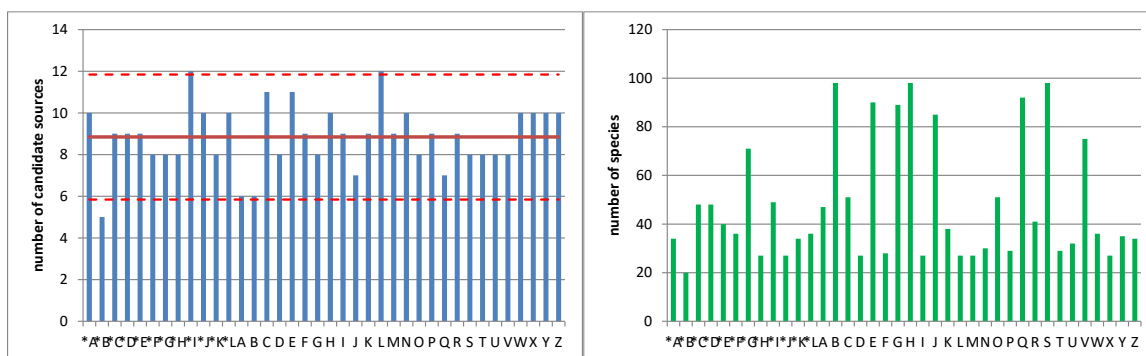
Figure 3. Number of candidate sources and number of chemical species used by participants in the reported results.

Result *B reports a number of candidates (5) which is beyond the range ±3 of the median number of candidates (9)(Figure 3).

In the 84% of results, the number of sources falls within the range average ±2 sources. In the 55% of results were used less than 40 species while only in 18 % of them were used more than 80 species. This indicates that many results were obtained with merged species or species were discarded.

4.3.2 Similarity tests

The similarity tests were performed to exclude the candidates whose chemical profiles differ significantly from the reference profiles (from SPECIATE/SPECIEUROPE) and their time trends differ from the other candidates of the same source category. For that purpose the SID and the PD were calculated and compared with their respective acceptability criteria (1 and 0.4).
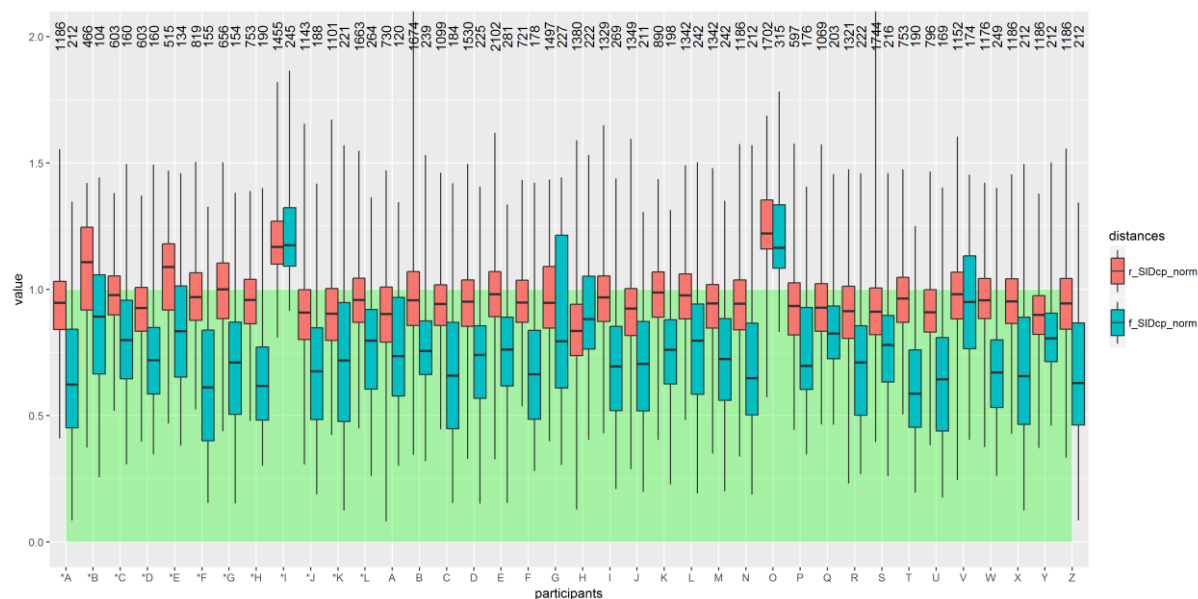
Figure 4: SIDcp with references (r) in red and among candidates (f) in blue. The green background indicates the acceptability area.
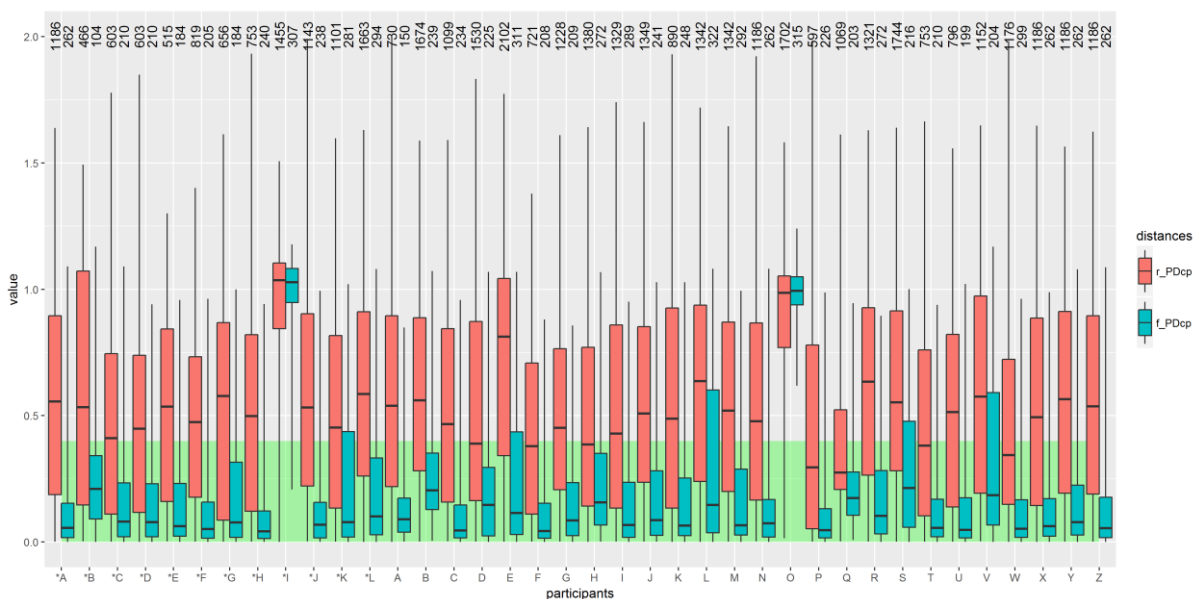
Figure 5: PDcp with references (r) in red and among candidates (f) in blue. The green background indicates the acceptability area.

Figures 4 and 5 summarise the distances between the candidates in every result and the reference profiles (r). These figures also depict the distance between the candidates in every result and all the candidates belonging to the same source category in the other results (f), excluding the ones mentioned in the complementary tests section. Two participants with bias in the total mass (O and *I) presented also very high distances from the reference and from the other candidates. On the other hand, the distances for candidates in result Q are quite comparable with the rest of the results indicating the inconsistency observed in SCE of this result (expressed as concentration) in the similarity tests is likely due to a wrong scaling factor and therefore is not affecting the chemical profiles (expressed in relative concentration).

In general, the distances to the reference profiles in every source category are greater than those among candidates. In particular, the PD between candidates and reference profiles comply with the acceptability criterion only in a limited number of cases.

In Figure 6 the distances are plotted arranged by source category. Unlike the rather uniform picture observed in the plots of distances arranged by participants (Figures 4 and 5), the distances between candidates vary considerably among sources. Distances are relatively small in sea salt (12), biomass burning (40) and wood burning (41), and greater in industry (20), fuel oil (30), coal combustion (31), ship (37) and gasoline (4). Distances in the "secondary sources" ammonium nitrate (61) and ammonium sulphate (62) are quite variable, with PD much smaller than SID due to the different sensitivity of the two indicators to dominating species (e.g. sulphate and nitrate). PD is extremely variable in the results reporting the two categories merged (60, SIA).
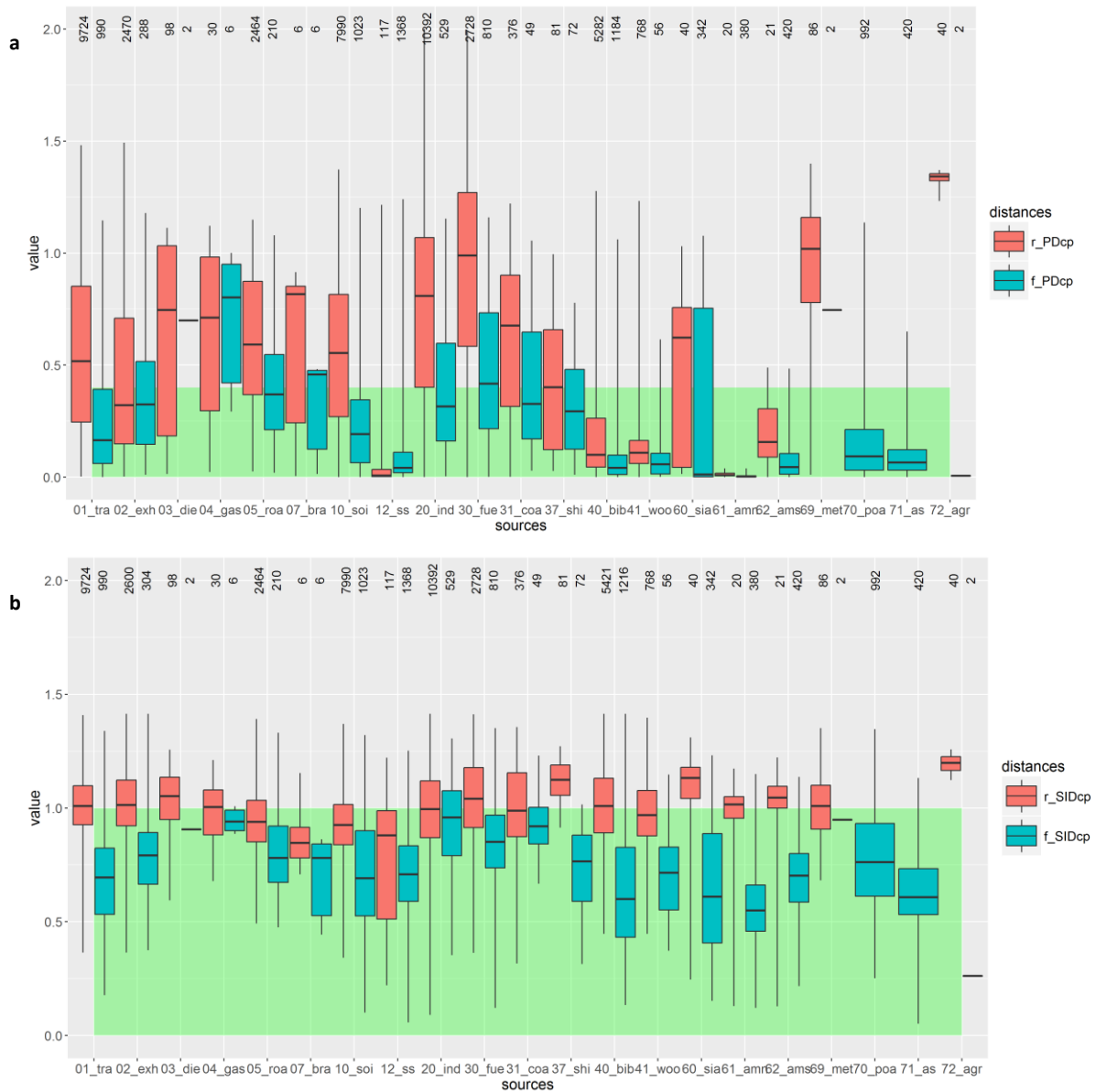
13

Figure 6: SID (panel a) and PD (panel b) distances arranged by source category (also the sources with only two candidates are plotted).

The summary of PD for the contribution-to-species (c2s) and the time series of source contributions estimates (sct) is depicted in Figure 7. A general good agreement between participants is observed. Results H, O, Q and S present atypical time trends. Results B, D, H, O, Q and S show contribution-to-species not comparable with the other results.
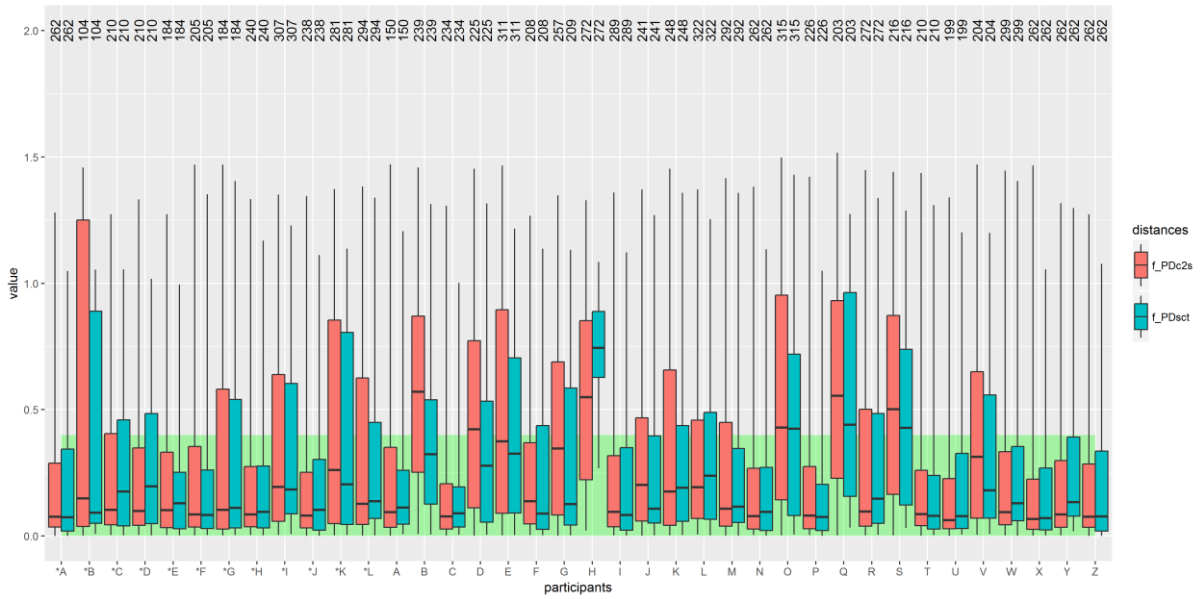
Figure 7: Pearson distance (PD) for the contribution-to-species (c2s) and the source contribution time series (sct) among candidates of the same source category

As shown in Figure 8, the variability between the time trends of the candidates classified in the same source category differs considerably. The source categories exhaust (2), gasoline (4), road dust (5), Industry (20), fuel oil (30), coal combustion (31) and ship (37) show on average values beyond the acceptability criterion. The picture observed in the contribution-to-sources test confirms the critical sources pointed out in the chemical profile test.
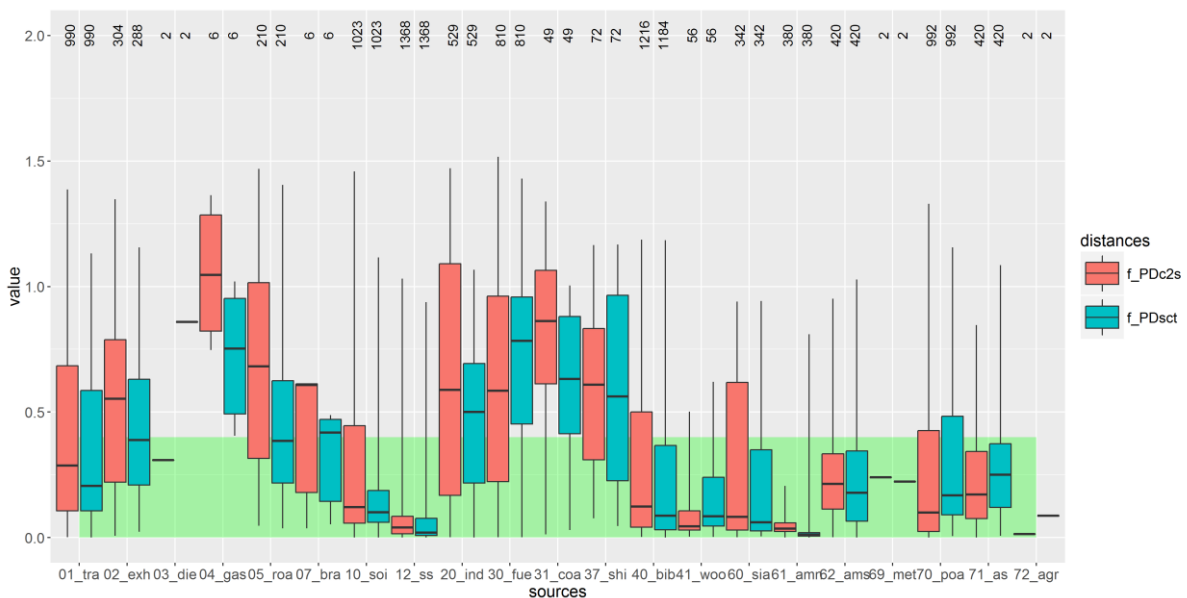
Figure 8: PD c2s and sct among candidates as a function of the source category.

According to the followed methodology, for each source category the candidates not passing 50% of the similarity tests: SIDcp r, PDcp r (or among candidates for the sources without reference profiles) and PDsct (among candidates) are marked as potential outliers and, therefore, not considered in the calculation of the reference for that source category. In this exercise it was decided to drop the PDc2s test because was not providing significant additional information with respect to the chemical profiles tests.

Using the above mentioned criteria 51 candidates (14%) were excluded from the calculation of the respective source category references.

4.3.3 Performance tests

The performance tests were executed following the methodology described in section 3. Not all the sources shown in the complementary and similarity test are represented in the performance tests. This is due to the lack of a minimum number of profiles (in general 3) for the reference to be calculated or the performance tests to be executed.
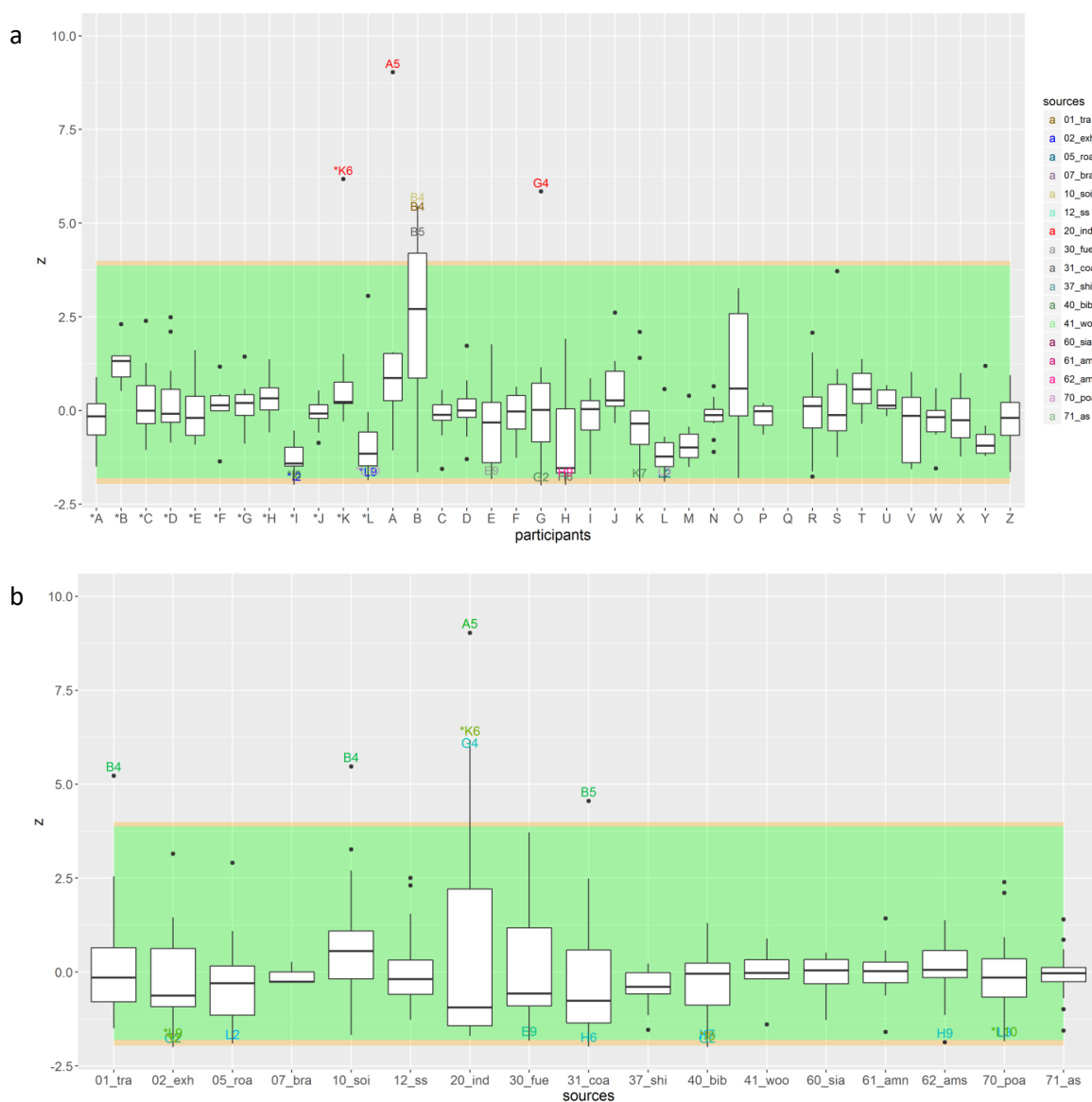

Figure 9. z-scores performance indicator values arranged by participant (a) and by source categories (b). Only candidates with warning or bad scores are indicated in the plot.

16

The z-score test results are plotted in Figure 9. The z-score indicates the performance of the SCE for the overall studied time window (all the time steps or samples). A 91% of the candidates fall within the area of acceptability indicating a general good agreement between the reported results and the reference values with a tolerance of 50% (Figure 9a). The results Q (scaling problems), B and G are those with the higher number of
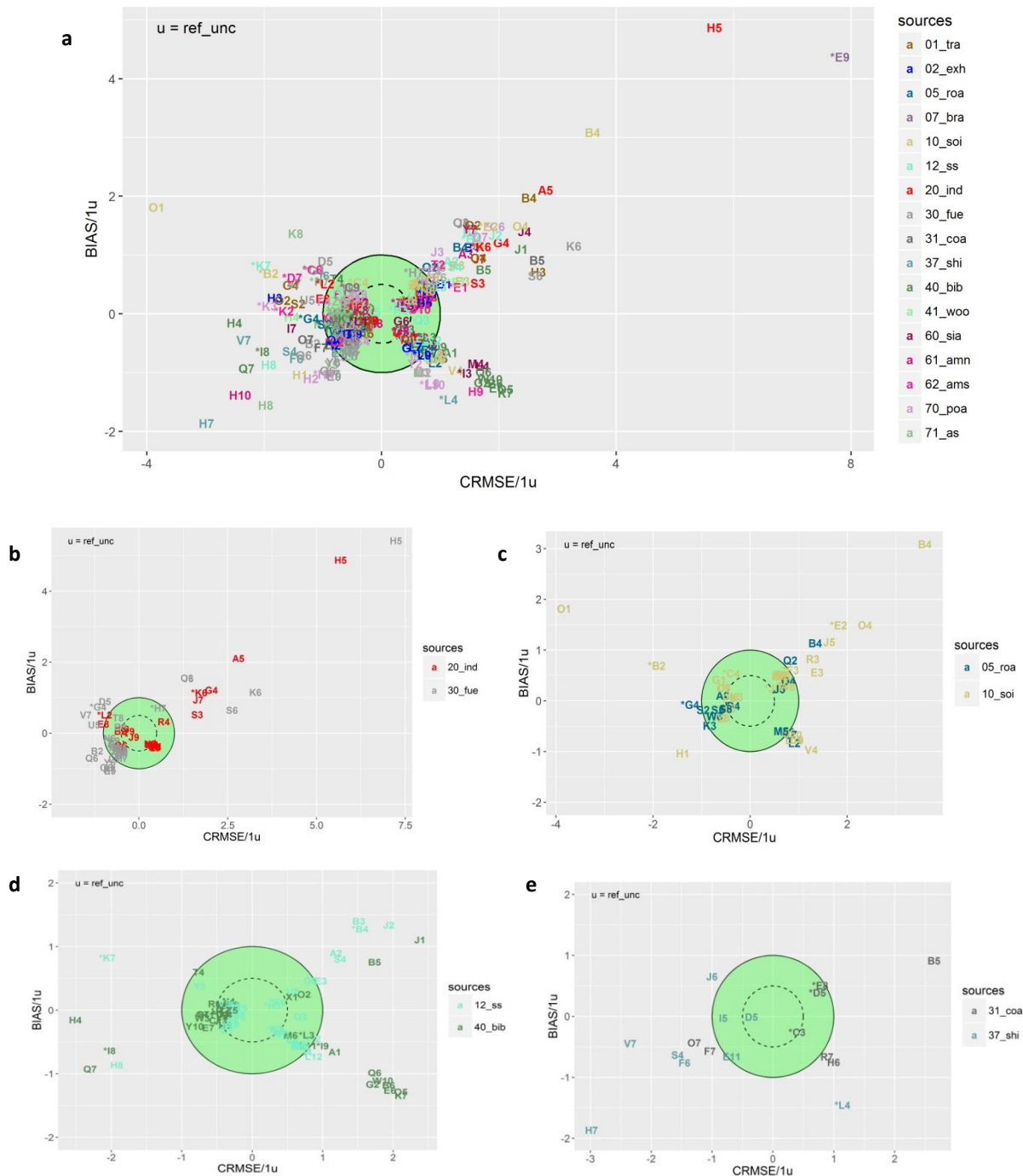

Figure 10: target plot performance indicator. Panel a: all the candidate sources; panel b: industry and fuel-oil sources; panel c: road and soil sources; panel d: biomass and sea salt sources; panel e: coal and ship sources.. Scores <1 (inside the green circle) are acceptable.

candidates out of the acceptability zone. Most of the overestimated SCE are in the industry source category (representing 15% of the reported candidates) with a few in the categories: traffic, soil and coal combustion (Figure 9b). On the other hand, the underestimated SCE are observed in exhaust, biomass burning and coal combustion categories.

The target plot showing the $RMSE_u$ is presented in Figure 10. In this intercomparison the $RMSE_u$ values used for assessing the SCE time trends are normalised by the uncertainty of the reference (u) which represents a more stringent criterion than the one used in previous intercomparisons (2u).

In the $RMSE_u$ test a 72% of the candidates fall in the acceptability area (Figure 10a). The share of results beyond the acceptability threshold is higher than in the z-scores test. The candidates falling in the rejection area represent a variety of source categories the most frequent of which are industry (Figure 10b), soil (Figure 10c), fuel oil (Figure 10b), biomass burning (Figure 10d), POA, and marine (Figure 10d). The source categories showing the highest percentage of candidates with poor scores in this test are ship (75%), coal burning (71%), and fuel oil (60%) (Figure 10 b and e). The 30% of the candidates' time series in the industry source category were rejected. The rejected industry candidates presented positive bias and amplitude problems with respect to the reference (i.e. fall in the upper right quadrant of the target plot, Figure 10b).

Positive bias is also observed in soil and traffic while biomass burning and primary organic aerosols are among the underestimated sct.

Results H, B, J, L, O and *E are those with the highest number of candidates in the rejection area. On the contrary, all the candidates in results C, N, P, X, Z, *A and *J ranked in the acceptability area.

In general, the lack of coherence with the reference are due to problems either in the variance or in the temporal correlation. Extreme cases of variance higher than the reference are candidates H5 (industry), *E9 (fuel oil) and B4 (soil).

5 The Evaluation of Chemical Transport Models

The obtained ensemble of SCEs were analysed and evaluated by means of the same methodology used for RMs, described in section 3, with minor adaptations to account for the differences between these models.

Unlike RMs, whose application is mostly based on the analysis of an observed dataset available at the receptor site, CTMs require the design of a whole model application including:

a)   the definition of a modelling domain;

b)   the reconstruction of the input data set;

c)   the CTM application and the evaluation of model performance;

d)   the source contribution evaluation.

5.1 Computational domains

The TNO inventory, specifically released for FAIRMODE encompasses a lat/lon regular grid of 720x672 cells with a DLON x DLAT step of 0.125 x 0.0625 [deg] (Figure 11). In order to reduce the emission pre-processing phase of most modelling teams, also the CTMs were asked to deliver their results over a geographical grid too.

The CTMs were run over two computational domains covering the whole Europe as well as the reference site area. The EU scale domain (Table 3) has been defined in order to provide suitable regional boundary conditions to the reference area simulations. Moreover, EU simulations could provide coarse resolution SCEs that could be used for a sensitivity analysis of source apportionment results with respect to grid resolution.

As already mentioned, the CTM domain was defined as a subset of the TNO grid, therefore anthropogenic emissions should not need any spatial interpolation.

The domain includes a portion of North Africa to account for dust emissions, while the northern boundary limited around latitude 65.0 to limit the spatial distortions at high latitudes when using a lat/lon grid. The grid step corresponds to a factor 3 of the TNO grid step, roughly corresponding to 18-20 km and being adequate to describe background concentrations.

The CTM computational domains are schematized in Figures 11 and 12 for the European and Lens domains, respectively.

Table 3 – Definition of the FAIRMODE-EU domain (geographical coordinates).

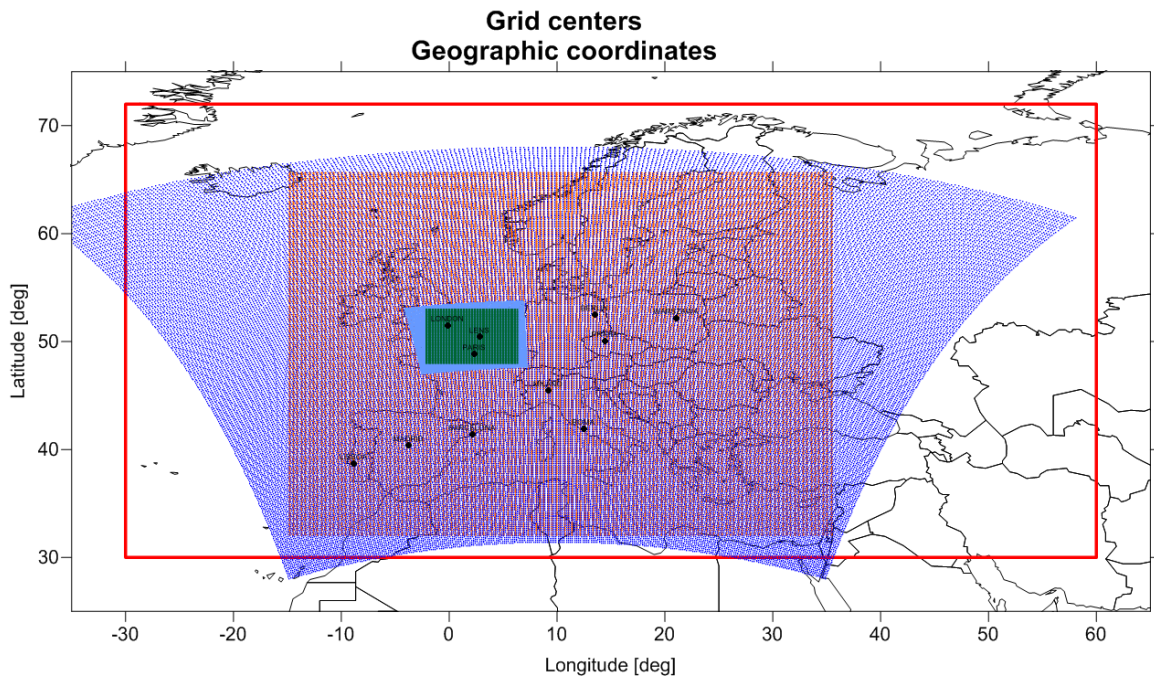| deg	X	Y
D	0.375	0.1875
N	135	180
SW_centre	-14.8125	32.09375
SW_corner	-15.0000	32.0000
NE_centre	35.4375	65.65625
NE_corner	35.625	65.750

Figure 11 – TNO emission domain (red square); FAIRMODE-EU domain (orange dots); FAIRMODE-LENS domain (green dots); WRF_FAIRMODE domains (blue and light blue dots).
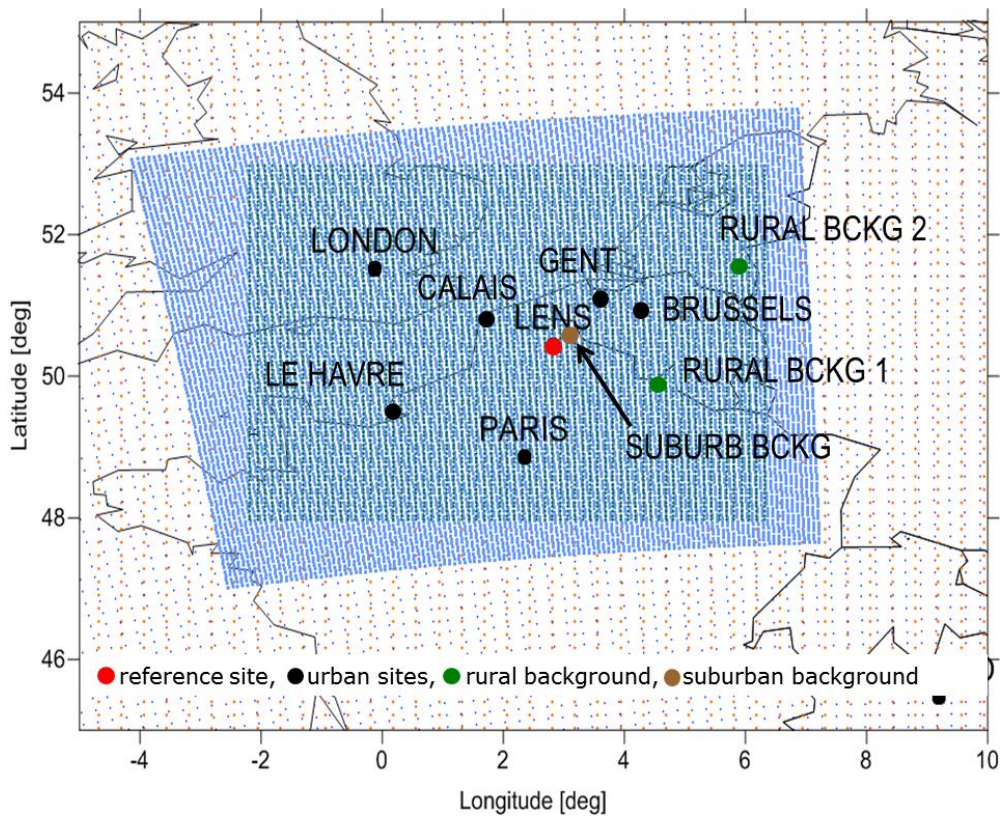

Figure 12 – Map zoomed on Lens domain with indication of the receptors selected for reporting results. FAIRMODE-LENS domain (green dots); WRF_FAIRMODE domains (blue and light blue dots).

The FAIRMODE-LENS domains (Figure 12, Table 4) is defined as a subset of the TNO grid as well as of EU grid, once again to avoid any interpolation of emission data. The domain is centred over Lens, but it is large enough to allow a reasonable description of the PM fate in atmosphere, limiting the influence of boundary conditions.

The domain includes both London and Paris, thus allowing the influence of both urban areas to be simulated at a proper resolution. Moreover it includes most part of the rural area of Belgium and The Netherlands, enough to take into account the influence of agriculture emissions coming from the eastern side.

To preserve as much as possible the original gradients in anthropogenic emissions the adopted horizontal grid step matches the TNO grid resolution. It roughly corresponds to 6-7 km, being adequate to describe urban background concentrations. Finally, it can easily allow further nesting levels, in case bottom up inventories should be available.

Table 4 – Definition of the FAIRMODE-LENS domain.

	X	Y
D	0.125	0.0625
N	69	81
SW_centre	-2.1875	47.96875
SW_corner	-2.2500	47.9375
NE_centre	6.3125	52.96875
NE_corner	6.3750	53.0000

5.2 Period of study

Observed data for Lens case study are available for the interval 9/3/2011 - 6/3/2012 as daily averages every third day. On the basis of observed data availability two periods for CTM modelling were defined:

Summer: from 1/6/2011 to 31/08/2011
Winter: from 15/11/2011 to 15/2/2012

These two were selected to be representative for both "hot" and "cold" seasons in France, being also long enough to include peak episodes and low concentration situations. Moreover, being observed data available every third day, three months of simulation were needed to pair at least 30 daily source apportionment results of RM and CTM.

5.3 Emissions

PM concentrations in ambient air are determined by both anthropogenic and natural emissions and by atmospheric advection and chemical processes. The dataset delivered by TNO was implemented to cover the anthropogenic sources. The main sources of natural emissions that can influence PM concentrations are: dust re-suspension, sea salt and biogenic VOCs, the latter representing precursors of Secondary Organic Aerosol.

The reconstruction of the different emission terms was left to the modelling teams, because in most cases such emission modules are embedded in their own modelling chain.

5.3.1 Anthropogenic emissions

Anthropogenic emissions have been provided by TNO. The TNO_SoAp_2011 emission dataset was prepared upon a request by the European Commission Joint Research Centre (JRC) with the specific aim to support and facilitate the CTM–RM inter-comparison exercise. The TNO_SoAp_2011 data set was derived from the TNO_MACC-III emission data, which is an update of the TNO_MACC-II emission inventory (Kuenen et al., 2014), in combination with fuel use information by country by sector from literature, IIASA GAINS model and TNO internal information.

In this dataset emissions are available on regular lon/lat grid at 0.125x0.0625 deg covering the domain shown in Figure 11. This emission inventory was developed specifically for the CTM-RM intercomparison thus including an enhanced source classification detailing fuels for macrosectors 2 and 7 (Table 5). Particularly, emissions due to combustion in the civil sector were split according to five fossil fuels plus solid biomass. Emissions from road transport sector were split according to three main fuels (gasoline, diesel and LPG/natural gas), while non-exhaust sources include evaporation and wear. Among non-road transport emissions, international shipping was specifically accounted for introducing two fuel categories.

Table 5 – SNAP source classification and description.

SNAP	SNAP_Name
1	Energy industry
21	Residential and commercial/institutional combustion, coal
22	Residential and commercial/institutional combustion, light liquid fuel
23	Residential and commercial/institutional combustion, medium liquid fuel
24	Residential and commercial/institutional combustion, heavy liquid fuel
25	Residential and commercial/institutional combustion, gas
26	Residential and commercial/institutional combustion, solid biomass (wood)
34	Industry (combustion & processes)
5	Fugitive emissions from fuels
6	Product use including solvents
71	Road transport, exhaust, gasoline
72	Road transport, exhaust, diesel
73	Road transport, exhaust, LPG/natural gas
74	Road transport, non-exhaust, gasoline evaporation
75	Road transport, non-exhaust, wear
8	Non-road transport
81	International shipping, marine diesel oil
82	International shipping, heavy fuel oil
9	Waste treatment
10	Agriculture

An example of emission data available in TNO inventory is shown in Figure 13. Differently, Figure 14 presents a comparison between TNO and EMEP national emissions. National estimates are in fairly good agreement in most countries, with the exception of a few eastern countries, where TNO emissions are generally higher than EMEP estimates. Some differences are shown also for Spain.
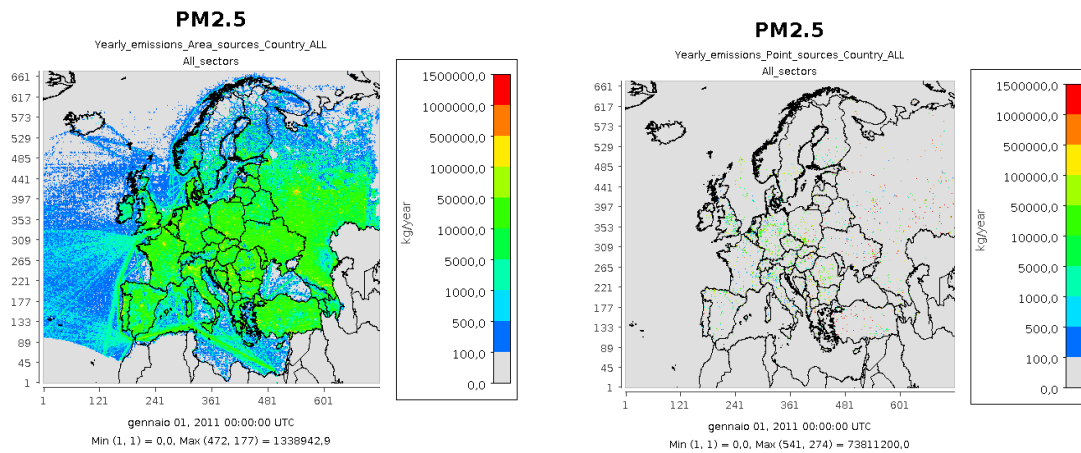
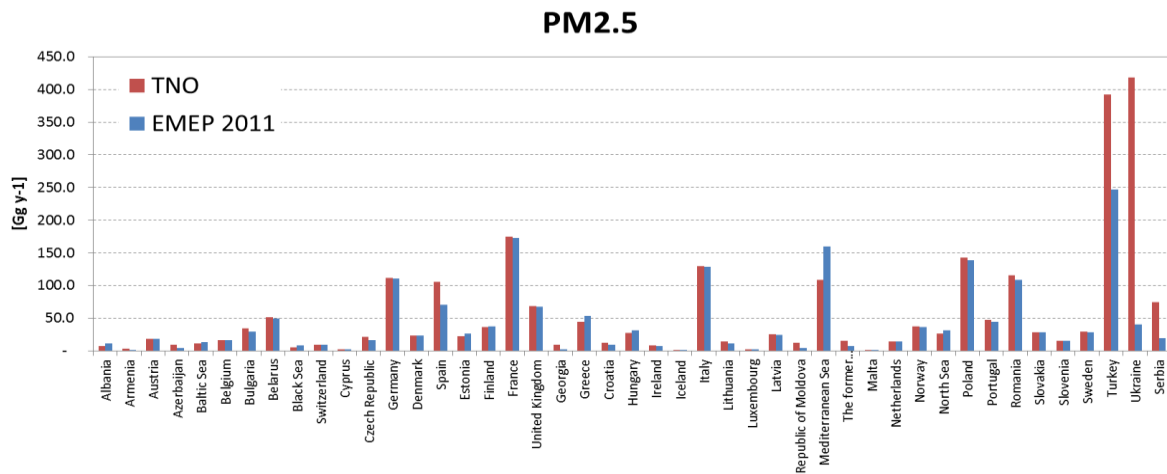Figure 13 – PM2.5 total emissions for 2011: area sources (left) and point sources (right).

Figure 14 – comparison of PM2.5 national emissions for 2011 estimated by TNO (red) and EMEP (blue) inventory.

A set of additional information, needed by modellers to perform temporal modulation and chemical speciation of emission data, was provided with the emission inventory.

5.3.2 Chemical speciation

- PM speciation factors allows to split $PM_{2.5}$ and PM coarse emissions into: EC, OC, Na, SO_4 and Other minerals; factors are available for each country and SNAP
- NO_X emissions are supposed to be split into NO (97%) and NO_2 (3%) for each country and sector
- SO_X emissions are supposed to be split into SO_2 (98%) and SO_4 (2%) for each country and sector
- No information is available for NMVOC speciation, therefore this aspect has been left in charge to each modelling team. In Europe a widely adopted approach is applying SNAP dependent NMVOCs speciation profiles which are based on Passant (2012). Those profiles consider SNAP from 1 to 9.

23

5.3.3 Height distribution

Point emissions were distributed along height according to Table 6.

Table 6 – Vertical distribution profiles [m agl] of total emissions according to SNAP category.

SNAP1	category	10	90	170	310	470	710	990
1	public power stations	0	0	0	0.08	0.46	0.29	0.17
2	Residential and commercial/institutional combustion	0.06	0.44	0.5	0	0	0	0
34	industry	0	0	0.04	0.19	0.41	0.3	0.06
5	extraction fossil fuel	0.1	0.8	0.1	0	0	0	0
6	solvents	1	0	0	0	0	0	0
7	road transport	1	0	0	0	0	0	0
8	other mobile, international shipping	1	0	0	0	0	0	0
9	waste	0.011	0.089	0.15	0.4	0.35	0	0
10	agriculture	1	0	0	0	0	0	0

5.3.4 Time profiles

Time profiles are available as Monthly, Daily and Hourly factor. Monthly and daily factors were defined for each country, SNAP and chemical species. Differently hourly factors were defined simply as a function of the SNAP emission category.

5.3.5 Natural emissions

Dust emissions

The term "dust", actually, implies several modelled contributions:

- Saharan dust taken into account just through boundary conditions

- Saharan dust emissions modelled inside the domain

- Natural dust other than Saharan

- Traffic re-suspension

RMs usually detect a "dust source", but it is not always possible to allocate it to the above mentioned subcategories. Modelling teams included any of these terms were asked to share information about the adopted approach.

Sea salt emissions and BVOC emissions

Also these two categories were included.

5.4 Meteorological fields

Meteorological fields were provided by WRF, run by the Technical University of Warsaw. WRF was run in a nested (one way) configuration in order to provide fields for both EU and LENS domain.

WRF simulations were carried out over a Lambert conformal domain, in addition to the lat/lon grid previously defined for CTMs. This choice implies a pre-processing phase of the meteorological fields in order to feed CTMs, however, it should be considered that:

1. most of the modelling teams may need to process meteorological fields before feeding their CTM, even if they in the same lat/lon CTM grid (e.g. to compute turbulence and other additional parameters);
2. as already mentioned, the use of a lat/lon grid over northern Europe is discouraged due to considerable distortion effects.

The WRF domains were defined as follows:

- both domains cover the corresponding CTM/output domains leaving also a border area;
- to limit the degradation of the meteorological information during the interpolation phase the WRF-EU domain adopted a grid step of 18 km, corresponding to the one of the CTM;
- likewise, WRF-LENS domain adopted a grid step of 6 km.

The adopted grid steps easily allow for a further nesting level at 2 km, to drive more CTM runs using more detailed emission inventories (e.g. for additional areas and/or runs)

WRF-EU run can be used to drive additional WRF simulations over other local areas (e.g. Po Valley, South France/Barcelona, Poland).

The definition of WRF domains is summarised in Table 7.

Initial and boundary conditions for WRF simulation were derived from GFS run at NCEP (http://www.ncep.noaa.gov/) and subsequently validated.

Table 7 – Definition of the WRF domains (LCP projection).

	EU	LENS
parent_grid_ratio	1	3
i_parent_start	1	85
j_parent_start	1	99
e_we	271	121
e_sn	225	112
dx	18000	6000
dy	18000	6000
map_proj	'lambert'	
ref_lat	50	
ref_lon	10	
truelat1	30	
truelat2	60	
stand_lon	10	
# of cells	60480	13320

The WRF configuration was defined as indicated in Table 8:

Table 8 – WRF physical configuration.

Physical process	*Option*
Microphysics	WRF Single-Moment 6-class scheme,
Longwave Radiation	Rapid Radiative Transfer Model scheme,
Shortwave Radiation	Dudhia scheme,
Surface Layer	Revised MM5 Monin-Obukhov scheme,
Land Surface Model	unified Noah land-surface model,
Planetary Boundary layer	Yonsei University scheme, minutes between boundary-layer physics calls every time step,
Cumulus Parameterization	New Grell scheme

5.5 Boundary conditions fields.

Boundary conditions were derived by MACC global model, the same approach that has been adopted in recent European initiatives like EURODELTA-III and AQMEII-3 (http://aqmeii.jrc.ec.europa.eu/).

MACC fields have the following features:

Projection: Geographical coordinates

Horizontal domain: Longitudes min/max -55.125/79.875 Latitudes min/max 16.875/76.5

Horizontal resolution: 1.125x1.125 deg.

Vertical coordinate system: hybrid sigma pressure coordinates (60 levels)

Temporal resolution: 3-hourly

Variables and file contents

4 files per day:

- GRG_yyyymmdd_EU_AQ.nc

O3,NO,NO2,HNO3,HO2NO2,OH,H2O2,CH4,CH2O,CO,C2H6,CH3CHO,PAN,BIGENE,BIGALK ,ISOP, TOLUENE,SO2

Units : mixing ratio (mol/mol)

Further details on MACC gas species can be found in Emmons et al. (2010) and Kinnison et al. (2007)

- AER_yyyymmdd_EU_AQ.nc
var1 Sea Salt (0.03-0.5 microns) ; units: microgram/m3
var2 Sea Salt (0.5-5 microns) ; units: microgram/m3
var3 Sea Salt (5-20 microns) ; units: microgram/m3
var4 Desert Dust (0.03-0.55 microns) ; units: microgram/m3
var5 Desert Dust (0.55-0.9 microns) ; units: microgram/m3
var6 Desert Dust (0.9-20 microns) ; units : microgram/m3
var8 Organic matter ; units: microgram/m3
var 10 Black carbon ; units: microgram/m3
var 11 Sulfate (SO4) ; units: microgram/m3

5.6 Definition of receptor and sources

For the source apportionment reporting were selected 10 receptors corresponding to the Airbase monitoring sites or to other monitoring sites of interest representing different types of locations (urban, suburban, rural) (Table 9). In particular the site of Gent was selected due to availability of data with chemical composition of PM$_{10}$ for the studied period from the Chemkar PM$_{10}$ study. (data kindly made available by Jordi Vercauteren, VMM).

Two sets of source categories (mandatory and optional) with different degree of detail of fuel for sectors 2 and 7 and of natural sources for sector 11 were defined to maximize the comparability between CTMs and RMs (Table 10). The mandatory set consists of 7 sources plus one corresponding to the unapportioned mass (99 OTH). The optional set encompasses 13 sources plus one for the unapportioned mass.

Table 9 Receptor sites for the SA results

Station Code	Station Name	Region	Station Type	Area Type	LON	LAT
LENS_SA	Lens-CARA	FRANCE	Background	urban	2.83	50.44
CALAIS_SA	Sangatte	FRANCE	Background	suburban	1.77	50.95
LE_HAVRE_SA	Le Havre Henri Fabre	FRANCE	Background	urban	0.11	49.52
PARIS_SA	PARIS 6eme	FRANCE	Background	urban	2.34	48.85
LONDON_SA	LONDON N. KENSINGTON	UNITED KINGDOM	Background	urban	-0.21	51.52
BRUXELLES_SA	41R012 - UCCLE	BELGIUM	Background	suburban	4.36	50.80
GENT_SA	Gent	BELGIUM	Background	urban	3.73	51.06
SUBU_BKGD_SA	40MN01 - MENEN	BELGIUM	Background	suburban	3.11	50.79
RUR_BKGD1_SA	REVIN	FRANCE	Background	rural	4.63	49.91
RUR_BKGD2_SA	Vredepeel-Vredeweg	NETHERLANDS	Background	rural	5.85	51.54

Table 10 Set of source categories (mandatory and optional)

SNAP	Description	Mandatory 8 sources	Optional 14 sources
1	Energy industry	01_ENI	01_ENI
21	R & C combustion, coal	99_OTH	02_OTH
22	R & C combustion, light liquid fuel	99_OTH	02_OTH
23	R & C combustion, medium liquid fuel	99_OTH	02_OTH
24	R & C combustion, heavy liquid fuel	99_OTH	02_OTH
25	R & C combustion, gas	99_OTH	02_OTH
26	R & C combustion, solid biomass (wood)	02_BIO	02_BIO
34	Industry (combustion & processes)	34_IND	34_IND
5	Fugitive emissions from fuels	99_OTH	99_OTH
6	Product use including solvents	99_OTH	99_OTH
71	Road transport, exhaust, gasoline	07_RTR	71_RTG
72	Road transport, exhaust, diesel	07_RTR	72_RTD
73	Road transport, exhaust, LPG/natural gas	07_RTR	07_RTR
74	Road transport, non-exhaust, evaporation	07_RTR	07_RTR
75	Road transport, non-exhaust, wear	07_RTR	75_RTW
8	Non-road transport	99_OTH	99_OTH
81	International shipping, marine diesel oil	08_SHP	08_SHP
82	International shipping, heavy fuel oil	08_SHP	08_SHP
9	Waste treatment	99_OTH	99_OTH
10	Agriculture	10_AGR	10_AGR
11P	Dust	11_DST	11_DST
11	Sea Salt	99_OTH	11_SLT
11	Biogenic SOA	99_OTH	11_BSO

5.7 Participants and SA results

CTM SA results were reported by 7 teams. Some of the results were obtained by single teams, other were produced thanks to the collaboration between more teams to obtain a joint output (Table 11). A third case were teams who coordinated their efforts to perform different runs: one best case and two sensitivity runs, to observe the impact of the spatial resolution and the vertical diffusion coefficient on the SA performance.

CTM results are encoded with c (low case) followed by uppercase letters from A to F. No suffix was added for mandatory set of sources while "o" denoted optional set of sources and "s" sensitivity run.

Table 11 CTM SA results reported by participants

kind of collaboration	participant code	result code	model	mandatory set	optional set
coordinated results	RSE[1]	cA	CAMx	selected for reference	
		cAo	CAMx		selected for reference
	ARPAV	cAs	CAMx	sensitivity test	sensitivity test
		cAso	CAMx	sensitivity test	sensitivity test
	UNIV AVEIRO	cAs2	CAMx	sensitivity test	sensitivity test
joint result	ENEA /ARIANET/ARPA PIEMONTE	cB	FARM	selected for reference (*)	
		cBo	FARM		selected for reference (*)
independent result	TNO	cD	LOTOS	selected for reference	
		cDo	LOTOS		selected for reference
independent result	RIER- UNI KOLN	cE	EURAD	selected for reference (*)	
joint result	CIEMAT/LISA - CNRS	cF	CHIMERE	NH4 and NO3 not reported	NH4 and NO3 not reported

[1] with the contribution of: Aristotle University of Thessaloniki, Ramboll-Environ, University of Genova.

(*) when both tagged species and brute force approaches are used for the reference

In table 11 is indicated whether results were reported for the mandatory or optional set of sources and which ones were selected for the computation of the ensemble reference.

In order to support the interpretation of the SA performance, an evaluation of the performance was accomplished on all the model base case results (Annex 2) and some highlights are given below.

The temporal evolution of the PM_{10} concentration during the winter period was mainly driven by regional scale processes. Therefore, the selected modelling approach adopting a 7 km horizontal resolution should be adequate to reproduce it as well to perform the source apportionment analysis.

CTM underestimations are more frequently at high concentrations, suggesting that the analysis of source apportionment results should be limited to mean concentrations and not extended to exceedance days that were poorly reproduced by the models. The observed strong underestimation of OA may have influence the reliability of source contribution estimates, particularly concerning domestic heating during winter season, biogenic sources during summer season and road transport for both periods.

Results cB, cD and cE underestimated sulphates and this could have had an influence on the reconstruction of the source contribution from sources where sulphur is an important component.

The overestimation of nitrate and sulphate observed in models cA, cE and partially cF gave rise to a corresponding overestimation of the ammonium concentration during the summer season that may have influenced the estimation of contributions from agriculture.

Different reference values were calculated for the mandatory and optional sets. The sensitivity runs were not used to calculate the reference values. Also the result cF was excluded for the calculation of the reference due to the lack of ammonium and nitrate in the estimation of the source contributions. All the teams reported the mandatory set while only four teams reported also the optional set of sources.

5.8 Complementary tests

The target plot for the mass closure compares the sum of the sources to the gravimetric mass (Figure 15).

In Lens, the sum of the source contributions is well below the gravimetric mass and in some cases falls beyond the acceptability area of the target plot. As expected, the difference is most evident in the result without apportionment of ammonium and nitrate (cF).

A similar pattern is observed in the majority of the other sites with the exception of London and Paris where the underestimation is much lower and all the results fall within the acceptability area. In London there is one case of overestimation (cB) while Paris is the only site where the result cF matches almost perfectly the gravimetric mass. This behaviour is explained by a lower underestimation (in some cases even overestimation) of the modelled PM_{10} in the base cases of these two cities.

29
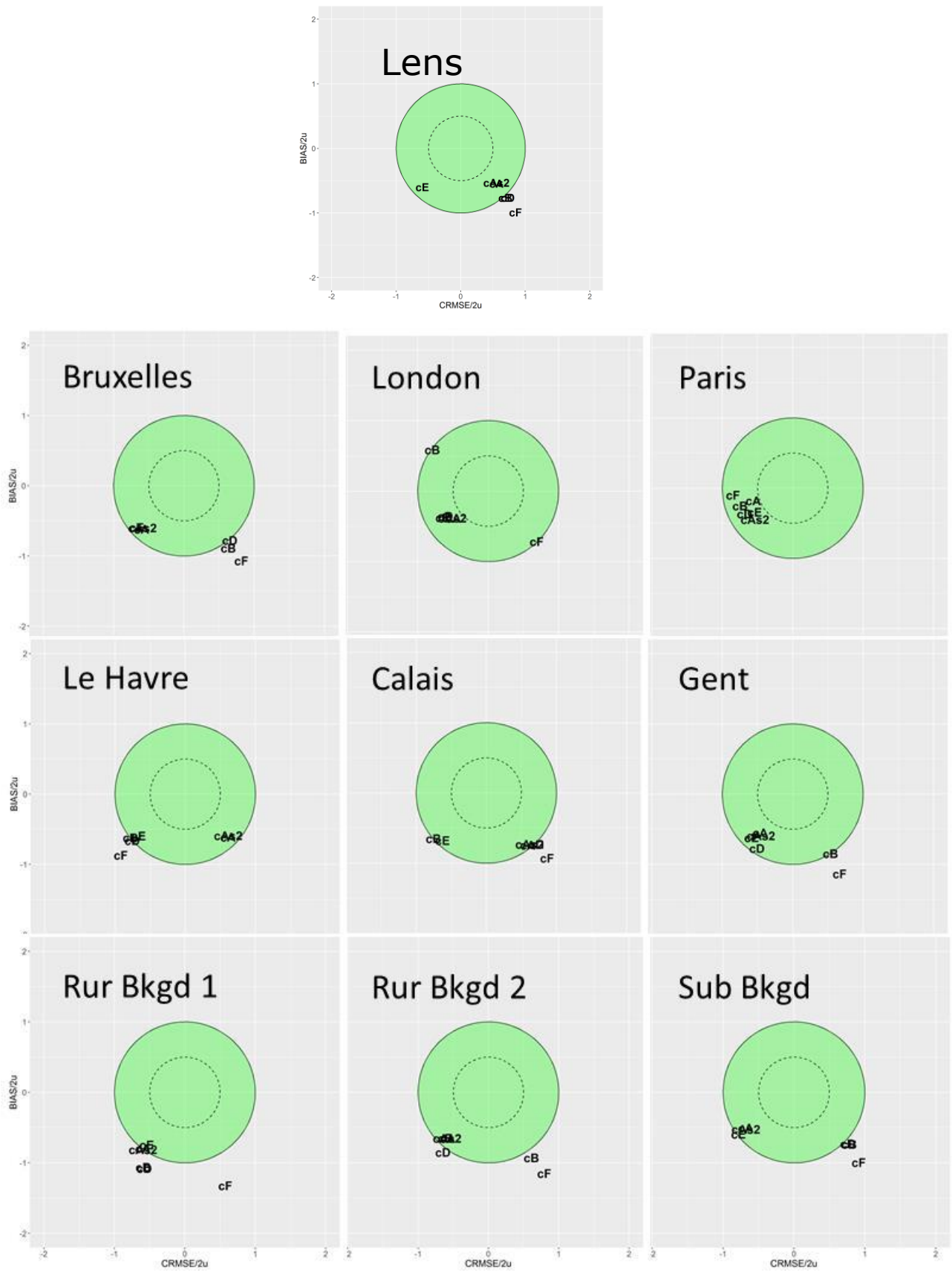
Figure 15. Target plot for the mass closure in Lens (top) and in the other receptor sites.

5.9 Similarity tests

In Figure 16 is presented the outcome of the similarity tests showing that candidates in the mandatory results are comparable among each other with the exception of cF. The candidates are less comparable with the reference cp but still within the tolerance of the test. There are no significant differences between results with mandatory and optional set of sources. Due to the missing contribution of ammonium and nitrate the test identifies cF as not similar to the others.
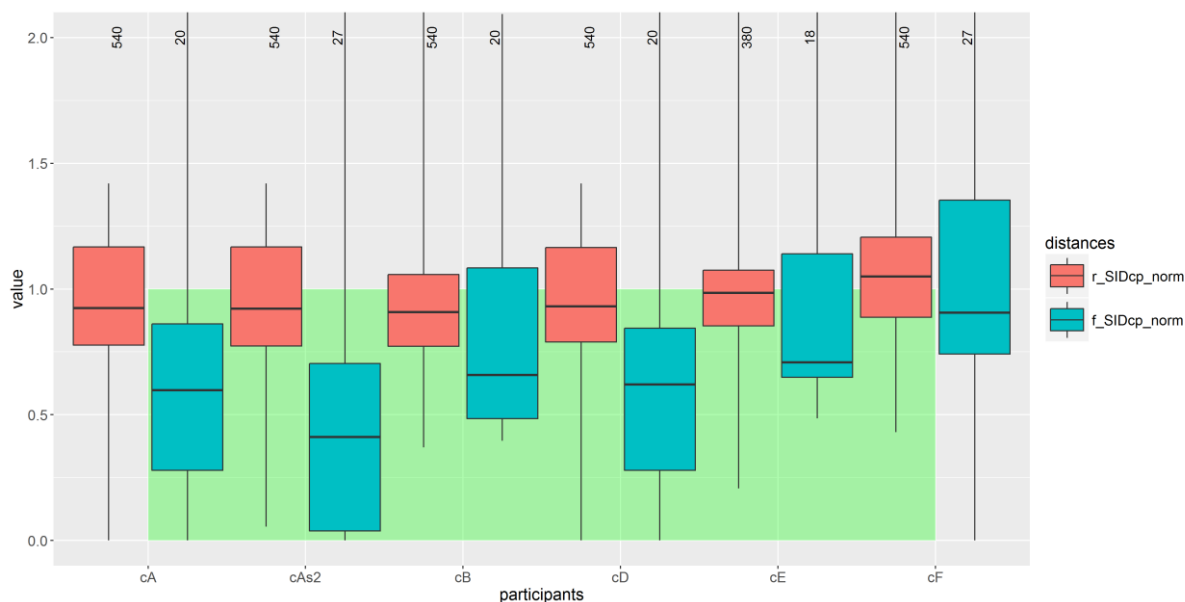
Figure 16. SID distance between candidate sources in mandatory results. r: distances to the reference chemical profiles (cp) in SPECIATE/SPECIEUROPE. f: distances among the candidate sources. Top: number of calculated distances, green background: acceptability threshold, normalised distances

The similarity tests arranged by source are shown in Figure 17 for the mandatory results and in Figure 18 for the optional results. The plots indicate that the chemical profiles of the sources in the different results are quite comparable. This holds for most of anthropogenic sources (such as road transport, industry and energy production), because they were reconstructed by all models on the basis of the same emission inventory and speciation profiles. The exceptions are dust, salt and to a lesser extend road dust. Indeed such sources were not included in the TNO inventory and they were simulated with different approaches by the participating models, sometimes not simulated at all. The comparability with reference source profiles is limited, however, in dust, road dust, industry and energy production. This is probably due to the limited number of species in the CTM profiles and the lack of specific markers. The similarity with the reference profiles is relatively good in biomass burning while in agriculture a high variability is observed.
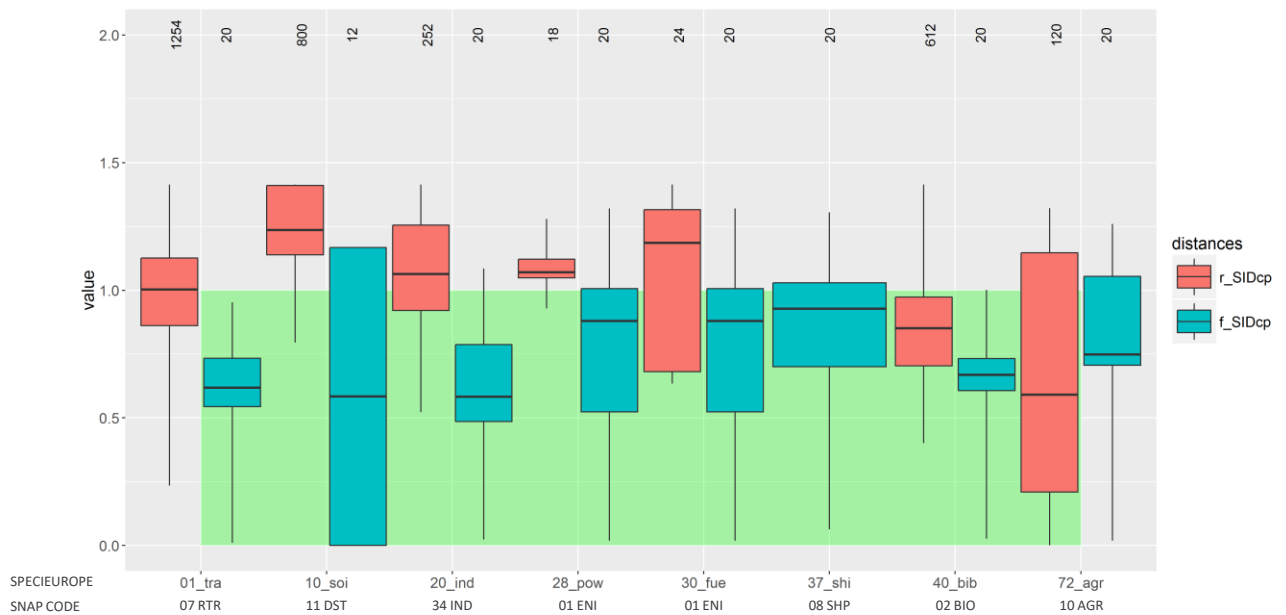
Figure 17. SID distance between candidate sources plotted for mandatory sources. r: distances to the reference chemical profiles (cp) in SPECIATE/SPECIEUROPE. f: distances among the candidate sources. top: number of candidate sources, green background: acceptability area, not normalised distances.
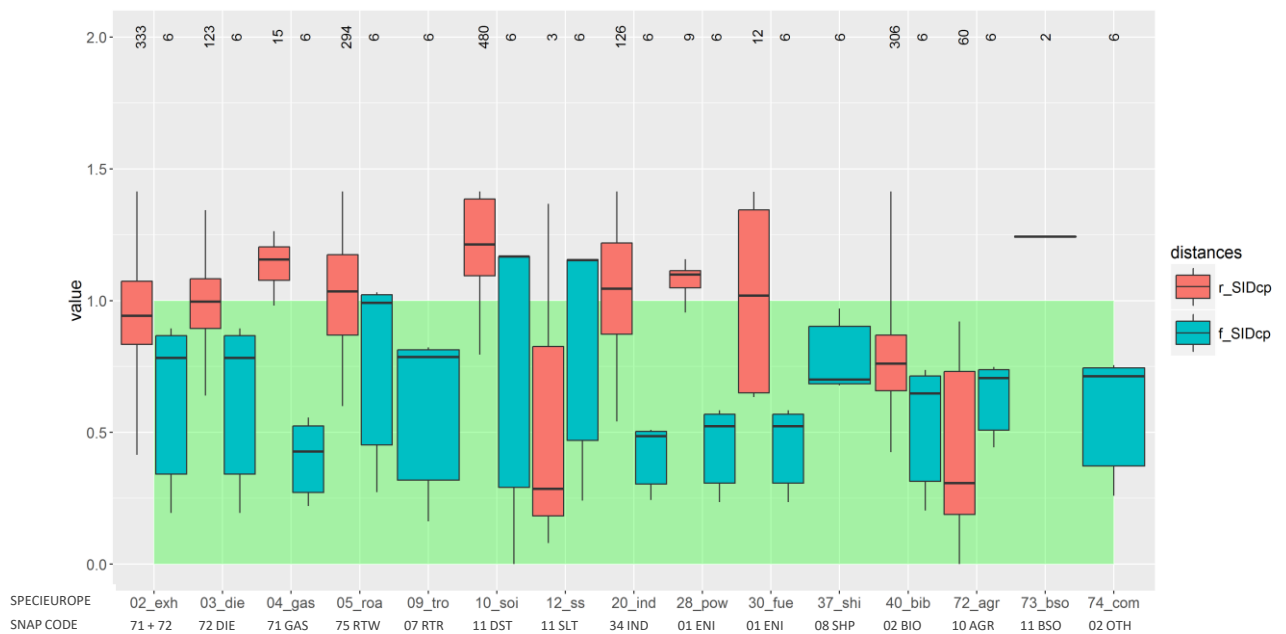
Figure 18. SID distance between candidate sources plotted for optional sources. r: distances to the reference chemical profiles cp) in SPECIATE/SPECIEUROPE. f: distances among the candidate sources. top: number of candidate sources, green background: acceptability area, not normalised distances.

Considering the high variability pointed out by the similarity tests of the agriculture source profiles, a more detailed investigation on the candidates of this source category was carried out. In Figure 19 the source profiles expressed as relative concentration of

the six considered chemical species are shown. The plot indicates there is a clear difference between the profiles reported by CTM using tagged species and those using brute force approach. In the first case the profile is dominated by ammonium with a variable contribution of OPA while in the second case the ammonium was present, however, the dominant component is nitrate. The differences in the profile are due to the way in which the two approaches estimate the contributions. The tagged species approach keep track of the source from which every chemical component derives. When ammonia from agriculture reacts with nitric acid deriving from NOx emitted by combustion process, the model attributes the mass of ammonium to agriculture and the one of nitrates to the respective combustion source. On the other hand, the brute force approach attributes the mass by estimating the difference between the base case and a simulation where the source is reduced or turned off. Since the abatement of agricultural emissions of ammonia leads to a drop in the concentration of ammonium nitrate with respect to the base case, using this method both compounds are attributed to agriculture.
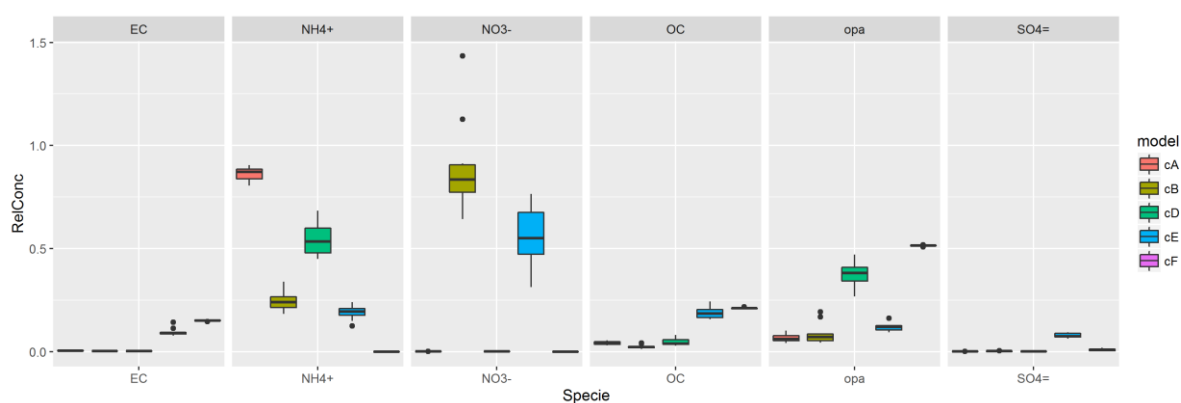
Figure 19. CTM chemical profiles of agriculture source.

5.10 Performance tests

The results of the performance test presented in this section have been computed with the methodology described in section 3. The references for each source have been calculated as the ensemble of all the candidates passing the similarity tests. The result cAs is not presented because only winter simulations were reported and this limited the comparison with the other results. As shown in Figure 20, in the reference site of Lens the majority of the z-scores (>93%) are in the acceptability area indicating a general good ability of models to reproduce average contributions in the reference with a tolerance of 50%.
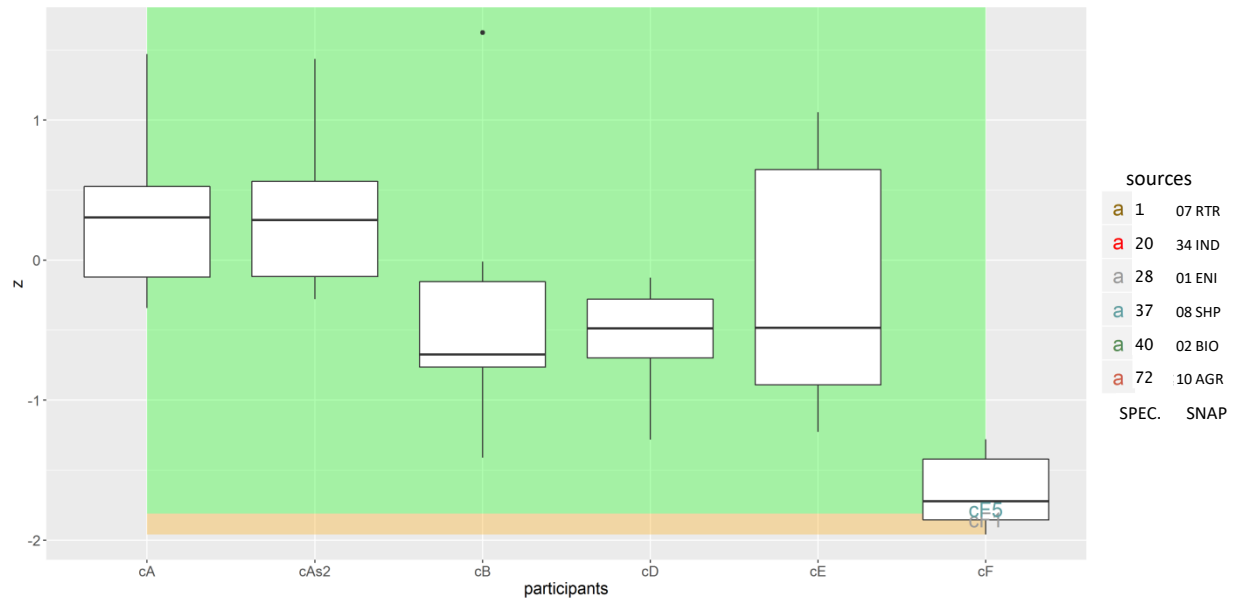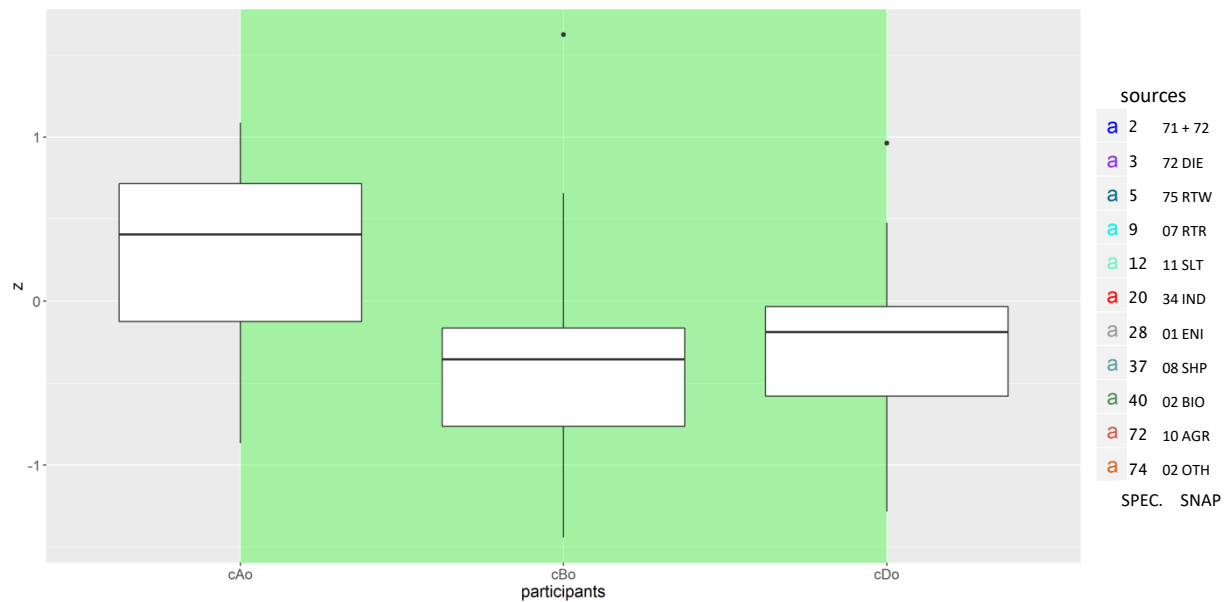
Figure 20. z-scores for the mandatory (a) and optional (b) sets in the site of Lens.

On average, z-scores of cA and cAs2 are generally higher than the reference, while other models rank below it. In the mandatory set, power plant and ship emissions are underestimated in cF, reporting results without ammonium nitrate, likely due to the role of these compounds in the two considered sources (Figure 20a). A relatively high score was recorded in cB for biomass burning but still within the tolerance for this test. This feature may be associated with the relatively high values of organic aerosol estimated by this model in the base case.
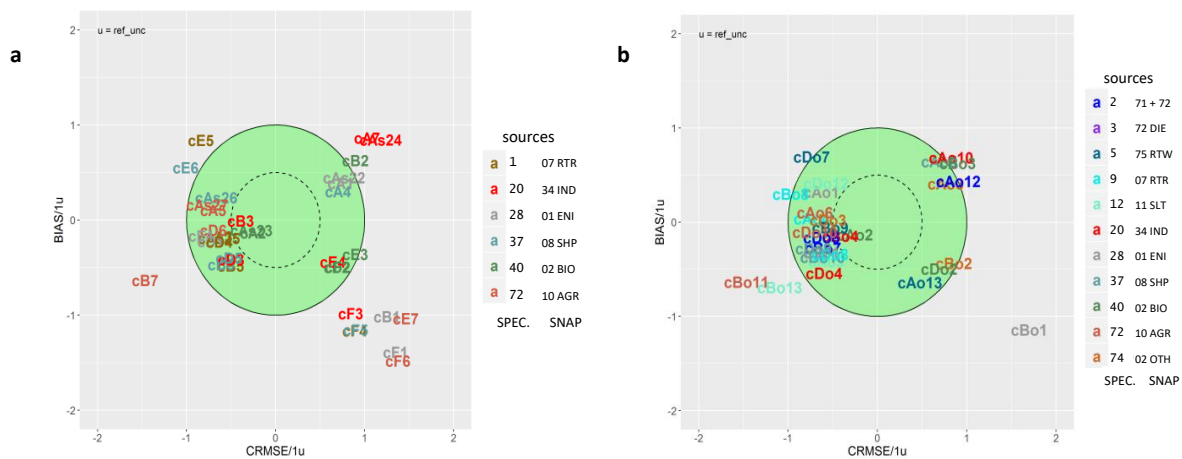
Figure 21. Target plot for the mandatory (a) and optional (b) sets in the site of Lens.

The scores in the Target plot shown in Figure 21 are normalised by (1u) uncertainty of the reference, the same approach adopted for RM evaluation. In Lens, the share of successful candidates ranges between 60% (mandatory) and 73% (optional). In the mandatory set, industry is often out of the acceptability area. The contributions from agriculture are underestimated in three results cB, cE and cF. With the exception of biomass burning, the contributions from cF are always underestimated with a prevailing amplitude problem. In cB energy is underestimated and biomass burning overestimated while cE overestimates ship and traffic contributions.

In the optional set there are only three results and the candidates fall in the acceptability area of the target plot. In the cBo result energy production, sea salt and agriculture are underestimated while biomass burning is overestimated. Moreover, industry is overestimated in cAo.

The general picture described for Lens is observed with some differences in the other sites (see section 7.1). In the mdt set, the ranking of industry in cA and cAs is relatively constant (with the exception of Calais) and the same is applicable to the three abovementioned agriculture candidates. In Paris strong overestimation and low correlation with the reference is observed in cE ship while result cD presents overestimation of traffic. In the harbour area of Calais, most of the results underestimate contributions from ships (cA, cAs, cE and cF) while all industry candidates rank within the acceptability area. However, in Le Havre, the other harbour area of the intercomparison, the results resemble the other sites. In the rural and suburban background sites the performance of the different candidates is quite comparable with the one observed in Lens.

Also in the optional set the overall picture observed in Lens is common to the other sites. In many of them the lack of correlation with the reference of the cB agriculture profile is more problematic than the underestimation. In London cB shows a slight overestimation of the exhaust contribution. In Paris, Gent and one of the rural background sites cD overestimates the road contribution. Unlike the other sites, in Calais and Le Havre cB slightly overestimates sea salt.

The general tendency of some sources in the result cB, which was obtained with a brute force method, to be rejected in this test may be due to the predominance (in the optional

set) of results obtained with tagged species method (2 out of 3). The problems are mostly observed in the following sources: agriculture, energy and sea salt.

Like receptor models, the target test (time series) appears to be more stringent than z scores (overall mean). The CTM source contribution estimates present high comparability with the CTM reference values for SA. The performances in the different receptor sites are quite similar. The comparison between the mandatory and the optional set of sources is not straightforward because of the different number of reported results. In general, there is a higher proportion of accepted candidates in the optional results when compared with the mandatory ones.

5.11 Performance tests with only tagged species results in the reference

In order to test the differences between brute force and tagged species approaches used by CTM to estimate the contributions of sources., in the present section the performance tests are re-calculated using for the construction of the ensemble reference only the results obtained with this kind of models (i.e. CAMx PSAT and LOTOS-EUROS).

The overall picture of z-scores obtained using tagged species models as references is comparable with the values observed using as reference all the models (Figure 22). The medians of the z-scores for every result are in both cases quite similar. This evaluation confirms that candidates of the result cF in the categories power plants and ship are underestimated. The only difference in terms of performance with the previous evaluation is that the candidate cB2 (biomass burning) that was an outlier falling within the acceptability threshold now is in the area of rejection.

Also in the optional set of sources there are minor changes in the z-score tests with tagged species compared to the one resulting from the reference with all models. As expected, the main changes concern the result cBo obtained with a brute force approach where the biomass burning is overestimated. Two of the Biogenic SOA are rejected.

In the site of Lens, the 80% and the 87% of the candidates pass the z-score test in the mandatory and optional sets, respectively.

With this evaluation set up all the candidates obtained with tagged species fall in the area of acceptability (Figure 23). In the result cB only the candidates of industry and ship pass the $RMSE_u$ test, while the only successful candidates of results cE and cF are industry and biomass burning, respectively. Similar situation is observed in the optional set of sources. In this case the result cBo is successful for the following sources: ship, industry, diesel and road dust. In the site of Lens, 51% of the candidates in the mandatory set and 77% of those in the optional set pass the $RMSE_u$ test.
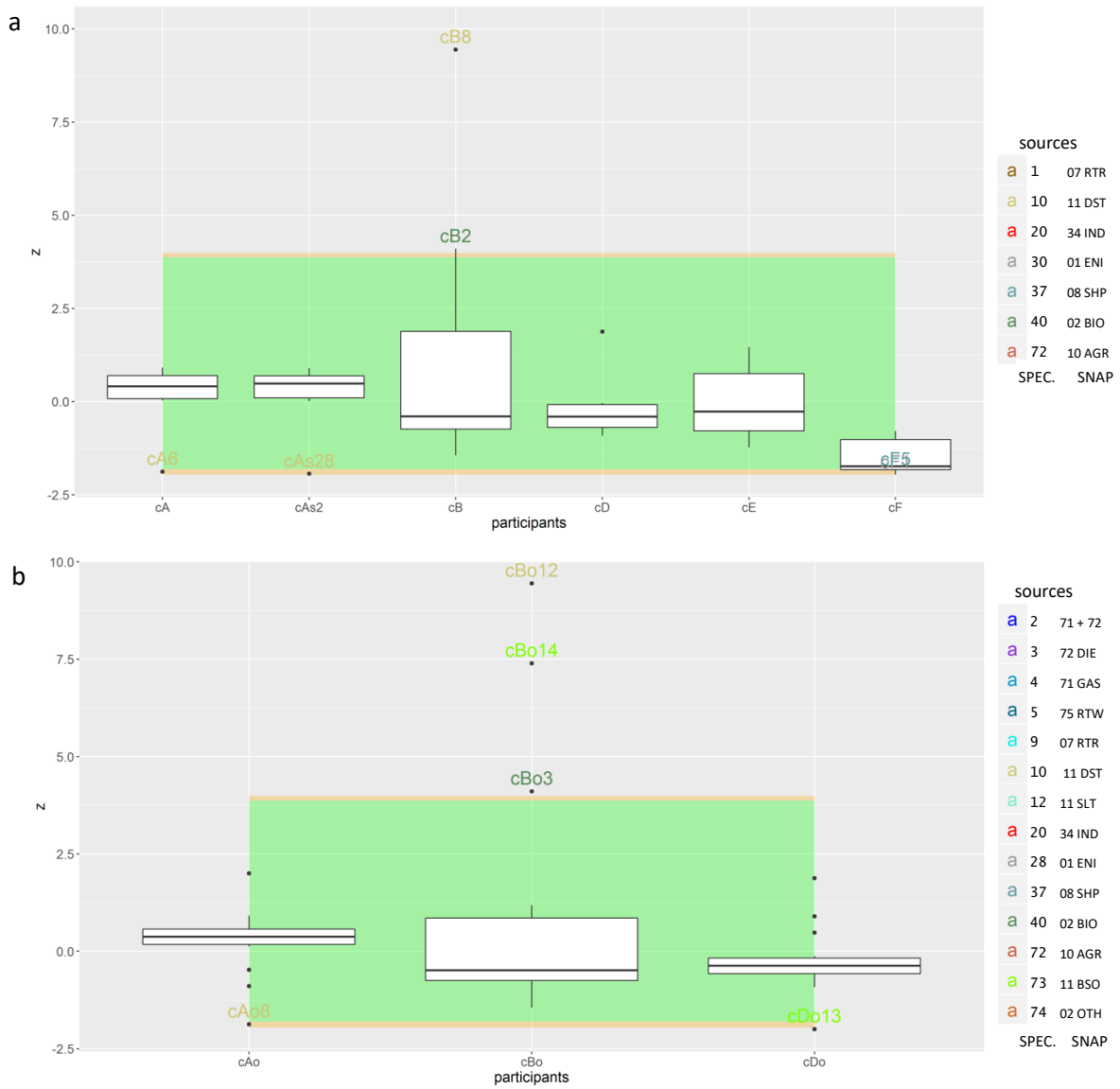
Figure 22. z-scores for the mandatory (a) and optional (b) sets in the site of Lens.
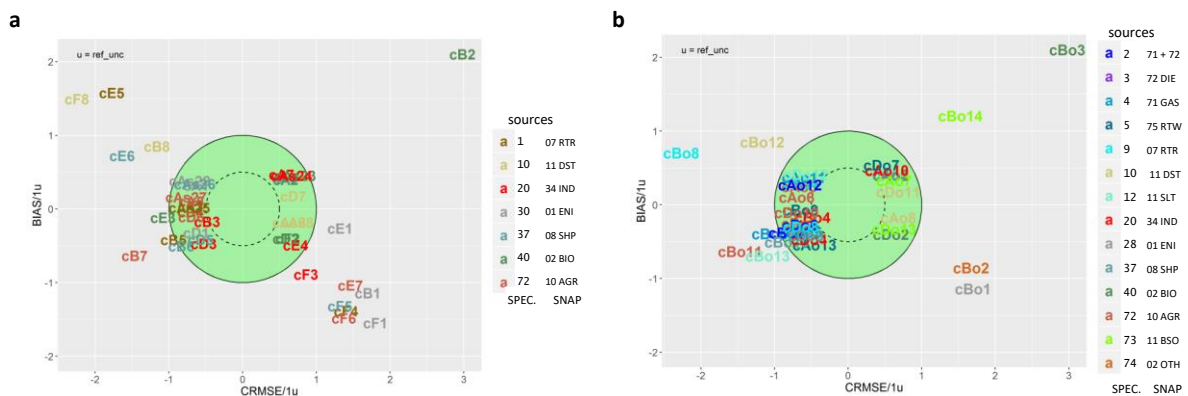
Figure 23. Target plot for the mandatory (a) and optional (b) sets in the site of Lens.

5.12 Sensitivity tests

The goal of the sensitivity test was to evaluate the influence of the reduced horizontal resolution and a different estimation of vertical diffusion coefficients on the CAMx PSAT output. The increased cell dimension in an area close to primary emissions (traffic) was expected to cause a reduction in the concentrations of pollutants associated with that source due to a dilution effect. To that end, CAMx PSAT runs were performed with two different grid steps 7 km (BC) and 20 km (SD).
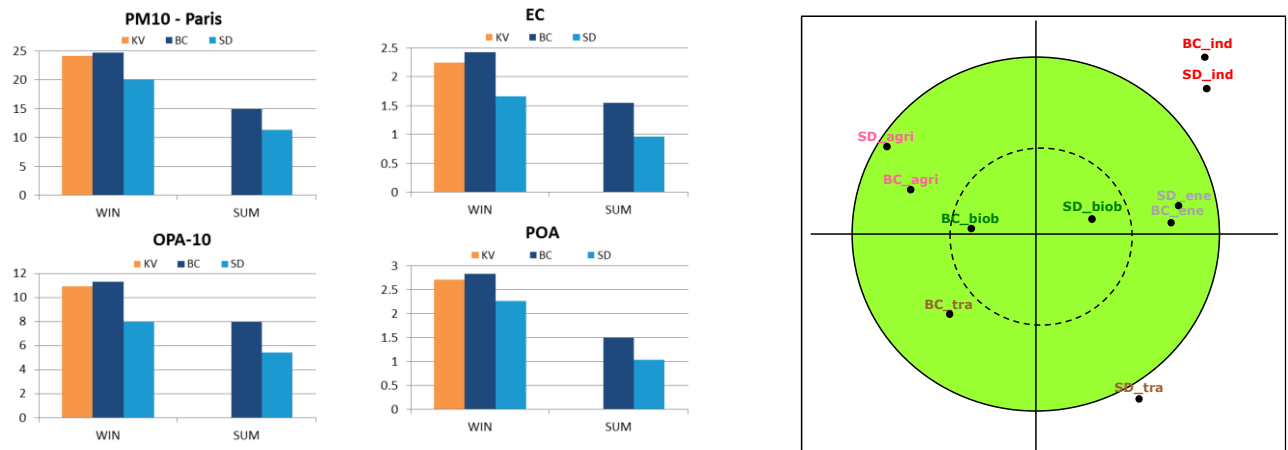
Figure 24 Variations in the concentration of PM_{10} and main chemical components associated with different grid size (left) and the impact on the SA performance (right)

As expected a PM_{10} concentration decrease was observed in the SD run that matched a decrease in Elemental Carbon (EC), Primary Organic Aerosol (POA) and other Primary Anthropogenic Aerosol (OPA-10) compared to the base case (Figure 24, left).

When comparing the performances of CAMX PSAT using two different grid steps it was also observed that the contribution of traffic was underestimated when using low spatial resolution (Figure 24, right). No significant changes in the SA performance were observed in the other tested sources: industry, energy production, biomass burning and agriculture. This behaviour has been attributed to the local impact of the traffic source which is, therefore, most sensitive to the spatial resolution than sources with a more coarse spatial pattern.

Similarly, in London (not shown) the lower spatial resolution leads to worsening in the performance of agriculture (overestimation) but no difference is observed in traffic. Here the explanation is that the traffic network in the surroundings of London is dense enough to minimize the effect of the lower spatial resolution. However, the increased grid step leads to mixing agricultural emissions with those from the city leading to an overestimation of this source in the city.

6 Receptor Models and Chemical Transport Models

In order to compare two families of models that rely on different input data and assumptions it is important to carefully define the source categories in a way that is coherent for both techniques.

In this intercomparison, the source categories for RM were defined on the basis of the repositories of measured source profiles SPECIATE and SPECIEUROPE (Table 1, Section 4.3). On the other hand, the definition of the sources for CTM depends on the emission inventory, which in this case made it possible to define source categories with fuel details for macrosectors 2 and 7 (Table 10, Section 5.6).

In this intercomparison, the source categories for RM were defined on the basis of the repositories of measured source profiles SPECIATE and SPECIEUROPE (Table 1, Section 4.3). On the other hand, the definition of the sources for CTM depends on the emission inventory, which in this case made it possible to define source categories with fuel details for macrosectors 2 and 7 (Table 10, Section 5.6).

In Table 12 the correspondence between the SNAP sectors, defined in the emission inventory used for the present intercomparison, and the source categories, as defined in the SPECIEUROPE database, is indicated.

Table 12. Correspondence between CTM sources (SNAP) and RM sources (SPECIEUROPE) defined for this intercomparison

CTM Mandatory	RM corresp. Mandatory	CTM Optional	RM corresp. Optional
01_ENI	30 fuel oil combustion or 28 power plant	01_ENI	30 fuel oil combustion or 28 power plant
99_OTH		02_OTH	
02_BIO	40 biomass burn.	02_BIO	40 biomass burn.
34_IND	20 industry	34_IND	20 industry
07_RTR	1 traffic	71_RTG	2 exhaust
		72_RTD	
		07_RTR (OTH)	
		07_RTR (OTH)	
		75_RTW	5 road dust
08_SHP	37 ship	08_SHP	37 ship
99_OTH		99_OTH	
10_AGR	NH4 sum	10_AGR	NH4 sum
11_DST	10 dust	11_DST	10 dust
99_OTH		11_SLT	12 marine, 71 aged sea salt
99_OTH		11_BSO	
99_OTH		99_OTH	

The comparison between the two families of models is accomplished only for the reference site of Lens. For the interpretation of the results it is necessary to consider that the estimation for RMs refers to the specific site where the dataset of measured PM_{10} was collected while the CTMs provide an average estimation for a grid cell, roughly corresponding to a 6-7 km grid step, containing the monitoring site.

Another difference to take into account when comparing RMs and CTMs is the different detail about the chemical composition of the PM managed in two families of models. While RMs are based on a relatively detailed information about the chemical composition derived from chemical analyses (in general there are 20 or more chemical species) in the CTMs there is a limited number of chemical families (typically 6 or 7) depending also on the degree of detail in the emission inventories.

One of the most difficult issues when matching the two series of sources is the comparability of the inorganic ions ammonium, sulphate and nitrate, that in the RMs are attributed by definition to one or two secondary sources (ammonium nitrate, ammonium sulphate) while the CTMs allocate these chemical species to the corresponding precursor sources.
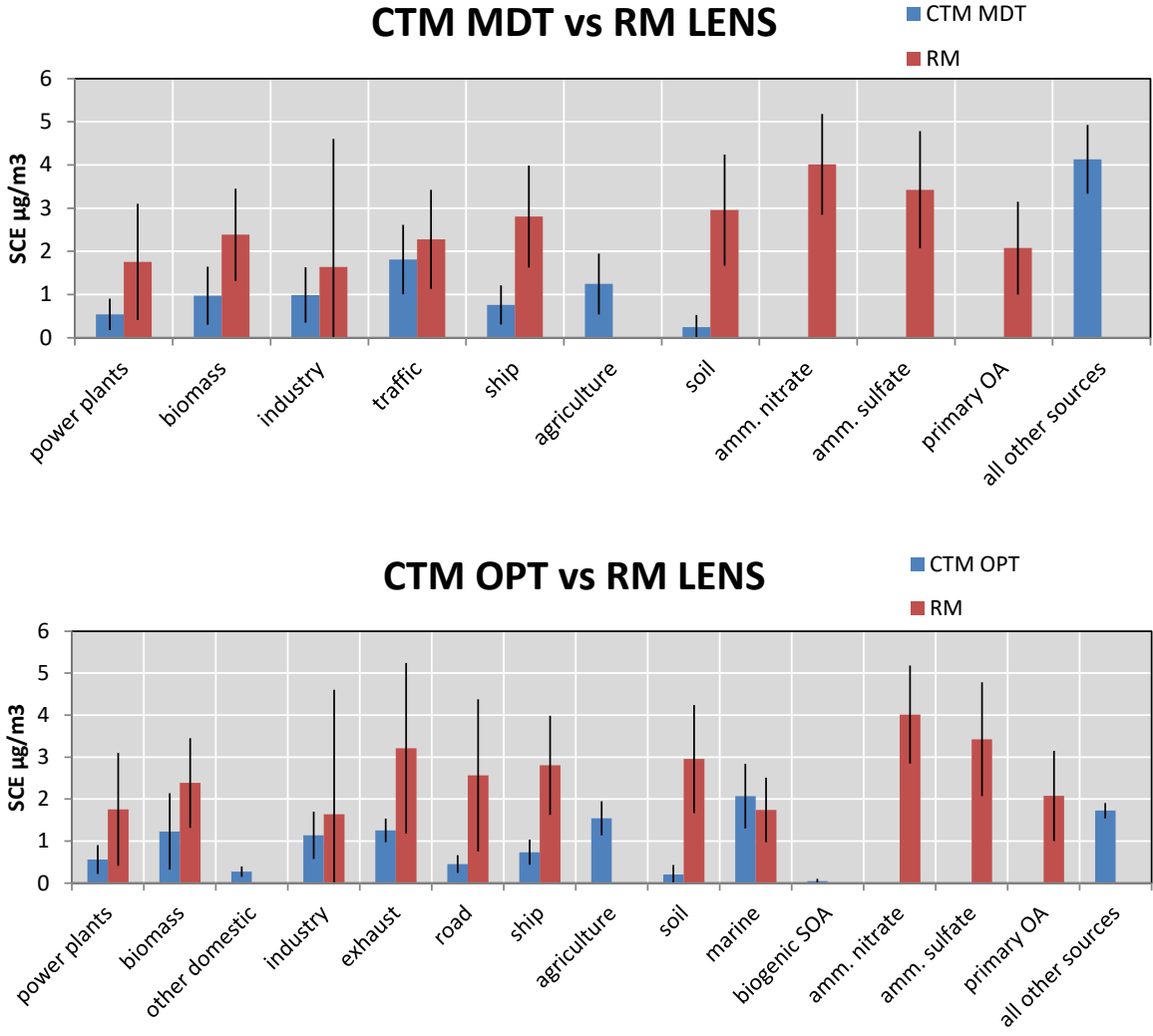
Figure 25. mean and standard deviation of the SCE reported in the present intercomparison obtained wih RM and CTM mandatory (a) and optional (b)sets of sources.

Another important difference between the two families are the trace elements that in RMs commonly range between 10 and 15 while in CTMs are represented by a single chemical entity "other primary aerosols" (OPA). RMs may also have detailed information about the composition of the organic carbon, which is a mix of hundreds of compounds, because they can be used as tracers for specific sources. The CTMs manage a variable number of organic compound groups for the simulation of the gas-to-particle conversion but the output commonly only one species for the particulate fraction is available: organic aerosol (OA) which may be split into primary (POA) and secondary (SOA). Due to the quite different kind of chemical profiles of the sources that derive from the two families of models, in this section there are no tests on the similarity of the chemical profiles.

In Figure 25 are compared the source contributions estimated with RMs and CTMs with both mandatory and optional sets of sources. The CTMs also reported an artificial source including all the mass that is not deriving from the set or sources required under the present intercomparison (all other sources). The SCEs deriving from RM are always higher than the corresponding value reported with CTMs. The only exception is marine aerosol where the two families of models are very close (less than 10% difference). The most important primary sources in RMs are: **exhaust, soil, ship, road dust and biomass burning**. **Traffic, agriculture, industry and biomass burning** are the most important sources in the CTM mandatory set. On the other hand, in the CTM optional set the relevance of sources change due to the apportionment of sources not considered in the previous set and to the split of traffic into its different components. In this set **marine** is the most important source followed by **agriculture, exhaust, biomass burning and industry**.

Traffic and industry are the sources with the closer SCEs in the mandatory set while soil and ship are the most distant. Power plants and biomass burning are rather different but their standard deviations overlap. In the optional set the most comparable SCEs are those of marine and industry and the most divergent are soil, road dust and exhaust. Worth of mention that the sum of the average SCEs attributed to exhaust and road dust in RMs are higher than the average SCE of traffic. This is due to the fact that RMs reported either the traffic as a whole or the exhaust and road separately but none of the results estimated the three of them.

The plots in Figure 25 provide a preliminary understanding about the bias between RMs and CTMs in the different sources, however, to take decision a test with pre-established acceptability criteria that evaluates in detail the average values and the time trends on the basis of different parameters such as the bias, the correlation and the amplitude of the curves is required.

6.1 Complementary tests

The reference for the mass closure test is the gravimetric mass of the PM_{10}. Simple comparison of the target values clearly show that CTM have a considerable fraction of unexplained mass (i.e. gravimetric mass that is not allocated to any specific source). On the other hand, the factor analytical models achieve a quite satisfactory allocation of the PM mass (Figure 26).
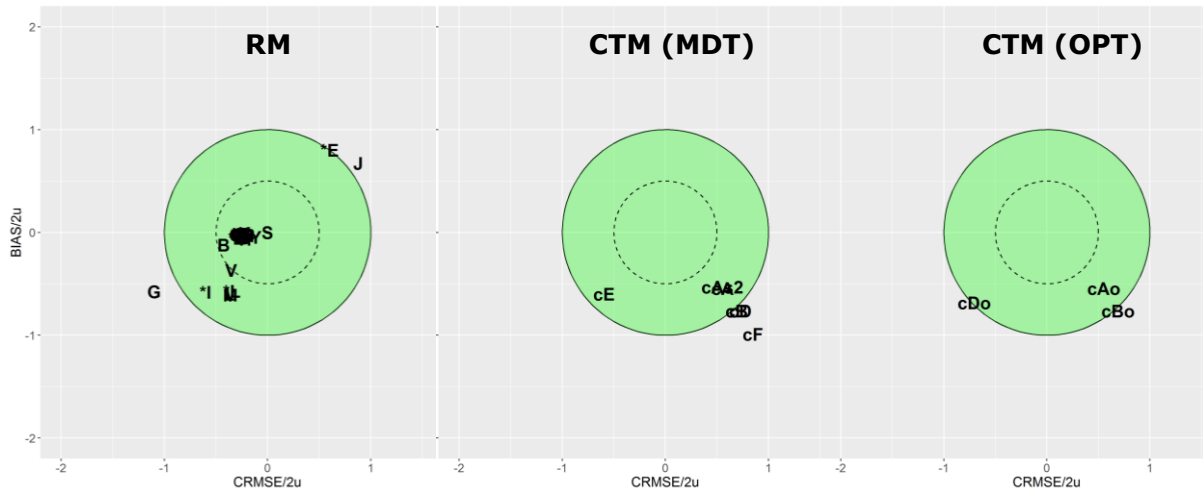
Figure 26. Compared mass closure target plot for RMs and CTMs at the site of Lens.

6.2 Performance tests

In order to compare the RM and the CTM models the approach adopted was to run the performance tests setting the CTM sources as candidates and for the SCE of the corresponding RM source category as reference value.

Even though all the CTMs tend to underestimate the SCE when compared with the RM reference values, the majority of the candidate sources fall in the area of acceptance (Figure 27a). In the mandatory set 83% of the candidates rank in the acceptability area. Soil is the most critical source, only the candidate of cB is successful, followed by power plants and ship. In this test the scores of result cF are on average lower than the others while cB underestimates the contribution from power plants.

Also in the optional set 83% of the candidates score successfully (Figure 27b) and the most critical source is soil with two unsuccessful candidates (cAo and cDo). cAo also fails in road dust while cB underestimates power plants contribution.
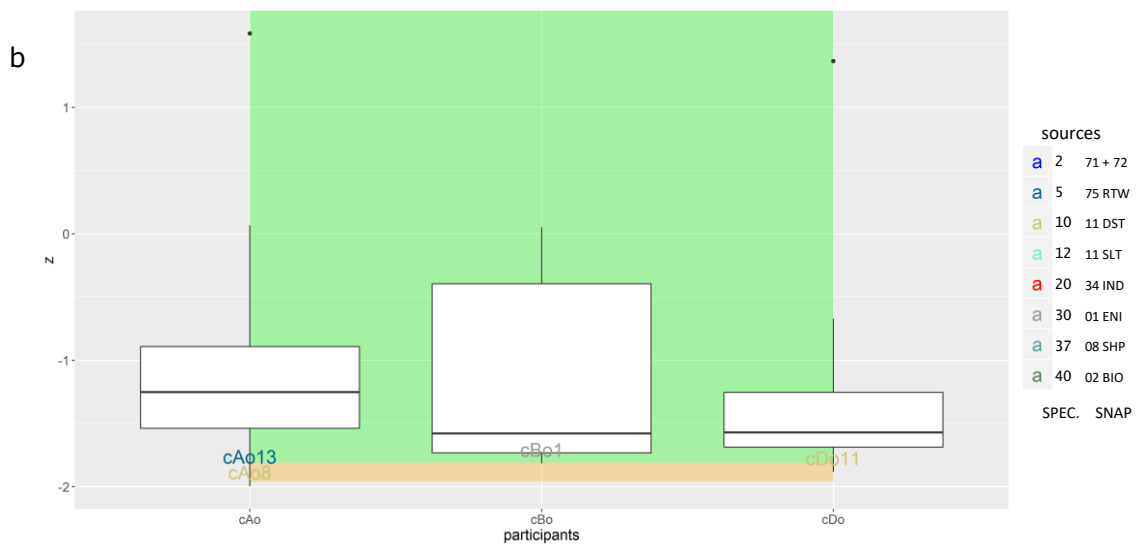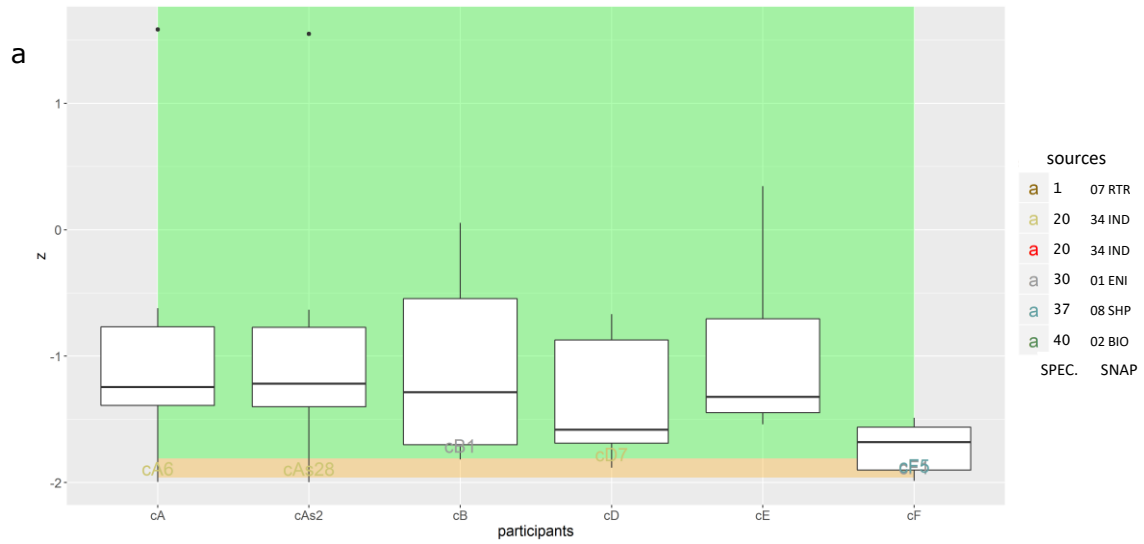
Figure 27. z-scores for the CTM mandatory (a) and optional (b) sets at the site of Lens.
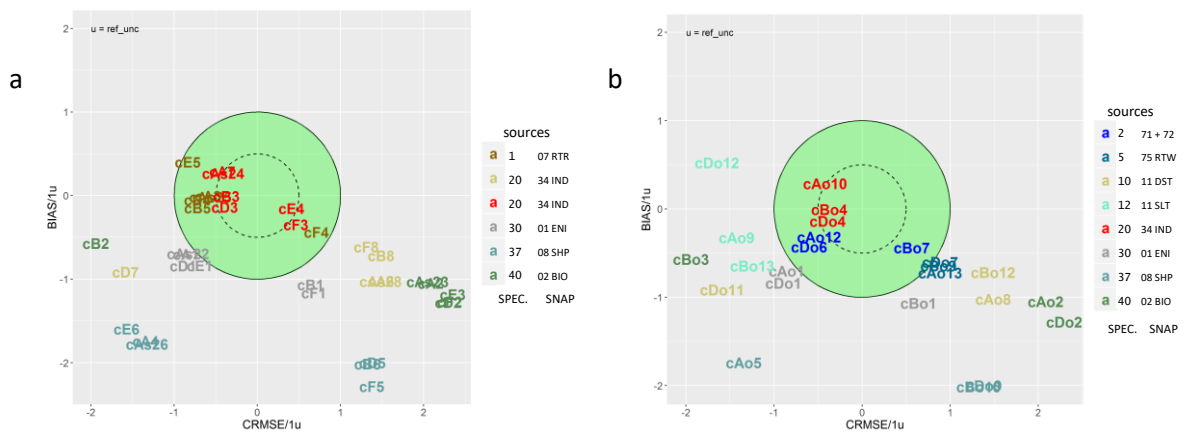
Figure 28. Target plots for the CTM mandatory (a) and optional (b) sets at the site of Lens.

As already observed in the previous sections, the $RMSE_u$ test is more stringent than the z-score test. In the mandatory set only 34% of the candidates are in the acceptability area of the target plot and the rate of success goes down to 25% in the optional set (Figure 28).

In the mandatory set only the candidates of the sources industry and traffic pass the $RMSE_u$ test and the same applies to sources industry and exhaust in the optional set. All the candidates of the other sources fall in the area of rejection in both the mandatory and the optional sets (Figure 28).

The good performance of the candidates in the **industry and exhaust** sources is probably due to the relatively high uncertainty of the RM reference for these categories. The good performance in the **traffic** source category is likely associated to the similar range of SCEs observed in RMs and CTMs (Figure 25). The poor performance of **marine** in this test is indicating that despite the SCE between RMs and CTMs are quite similar, their time trends are poorly correlated.

Soil and road dust sources are known to be little represented in the emission inventories and therefore difficult to be modelled with CTMs. This is likely the cause of the poor comparability between RMs and CTMs which present a sizeable bias component.

The strong bias between the two families of models observed in the source **ship** and to a lesser extent in **power plants seems** to be the cause of the poor **$RMSE_u$** values.

In the **biomass burning** in addition to a moderate bias has been identified a problem of amplitude. It is well known that this source undergoes considerable seasonal excursions and the result of the $RMSE_u$ test points out the two families of models reproduce differently the extent of such variations.

6.3 Contribution of sources to specific PM components

Source apportionment models provide information about the contribution of the pollution sources to the PM_{10} chemical components. In this section are commented the contributions to major species as defined in the CTMs: elemental carbon (EC), organic carbon (OC), nitrate (NH3), sulphate (SO4), ammonium (NH4) and other primary aerosol (OPA) (Figure 29).

EC, also known as black carbon, is a primary pollutant originated in combustion processes which is particularly suitable for comparing RMs and CTMs because chemically stable. A good agreement is observed between RM and CTM on the contribution from the most important sources: traffic, exhaust (i.e. diesel) and biomass burning. In ship, energy production and industry the contributions estimated by RMs is higher. RMs also identify a significant contribution from road dust, and to a lesser extent from soil, which is almost absent in CTMs.

Combustion processes release also OC which is a complex mixture of compounds with different reactivity and partitioning properties. In general, the contributions identified by RMs are much higher than those of CTMs. In addition to combustions processes (biomass burning, traffic, ships, energy, etc.), RMs attribute significant contributions to the secondary processes and also to soil and road dust. Biomass burning and traffic are the main contributors according to CTMs followed by agriculture and industry.

In RMs the inorganic ions (NO3, SO4 and NH4) are mainly associated with the secondary processes leading to the formation of secondary inorganic aerosol (SIA) ammonium nitrate and ammonium sulphate while CTMs allocate them to their precursor sources.
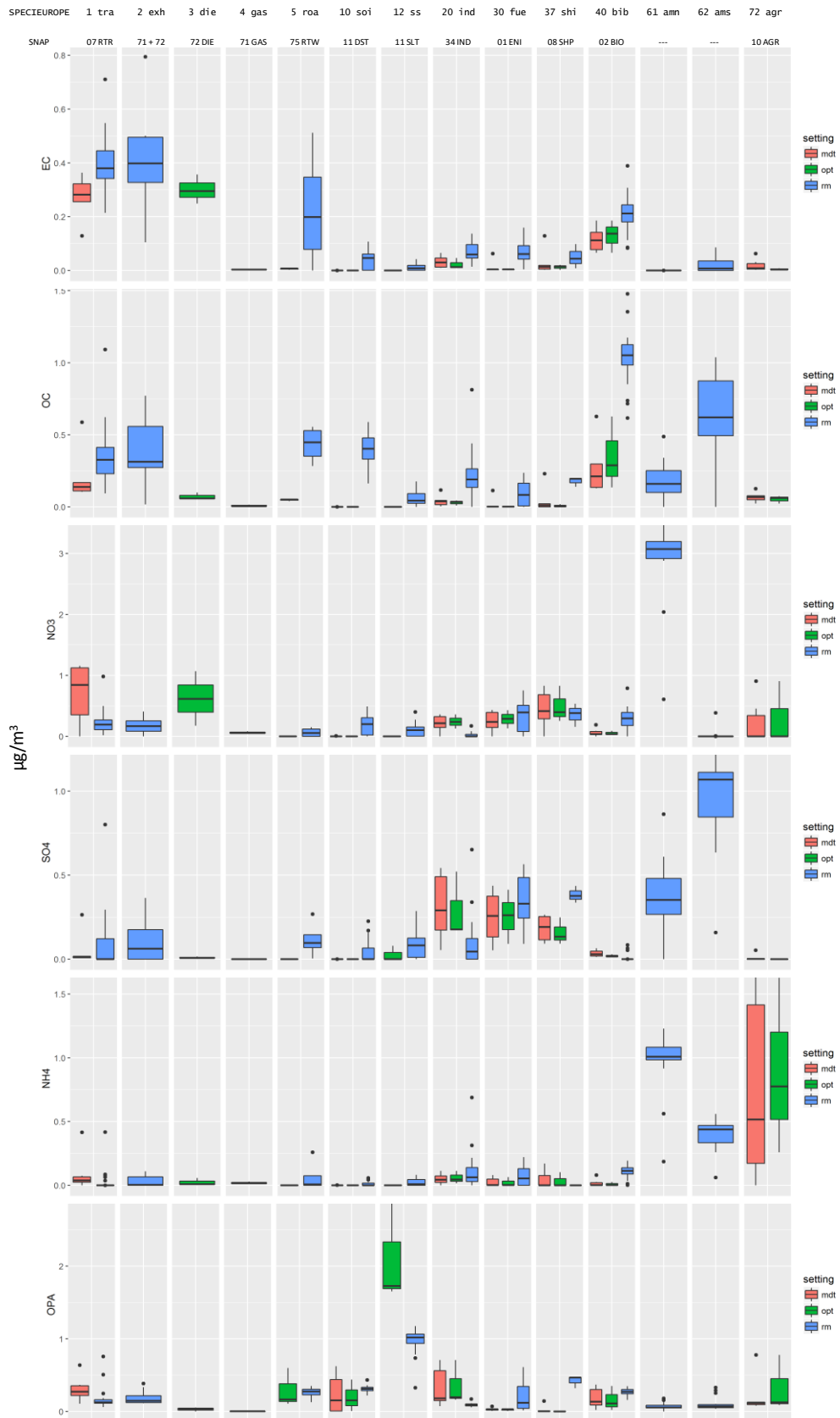
Figure 29. Contributions of sources to the main chemical components of PM_{10} in the site of Lens. EC: elemental carbon, OC: organic carbon, NO3: nitrate, SO4: sulphate, NH4: ammonium, OPA: other primary aerosol, rm: receptor models, mdt: CTM mandatory set and opt: CTM optional set. Source codes are available in Tables 1 and 12

NO3 is attributed to combustion processes such as traffic, exhaust, ship, energy and industry. The presence of OC in the RM ammonium nitrate and ammonium sulphate profiles suggest that both organic and inorganic compounds are produced during aerosol ageing processes. Variable contributions are also allocated to agriculture (see Section 5.9). In CTM results SO4 is mainly attribute to industry, fuel and ships. RMs mainly attribute this species to the secondary formation of SIA and to emissions from ships and energy sector. Finally, NH4 is almost entirely allocated to secondary aerosol processes in RMs while CTMs attribute the emissions of this compound mainly to agriculture.

Also the OPA is treated in a quite different manner in the two families of models. In CTMs it is a pool of minor inorganic components of PM treated as a single species while in RMs this category includes a variety of trace elements, mineral oxides and salts which are used in the identification of the sources. In CTMs the most important contributor to this category is by far sea salt followed by industry, soil, road dust and biomass burning. A variable contribution from agriculture is observed (see section 5.9). Also RMs attribute the highest contribution to sea salt but levels are one half of those reported in CTMs. Other contributing sources to the OPA bulk are: soil, road dust, ships, biomass burning, energy and traffic/exhaust.

7 Concluding remarks

7.1 Synthesis of results

A summary of the performance tests in all the sites is given in Figure 30. In general, the difference in the performance between sites is quite modest, in particular the z-scores, no matter what models were used in the reference. Such homogeneity suggests the geographical pattern for the allocation of sources is rather stable and depends mainly on the emission inventory and meteorological input data.

RM experts reported results obtained mainly with one model (EPA- PMF5). This is probably due to the good performances in previous intercomparisons and the user-friendliness of the tool. RMs show very good performance for overall dataset (91% of z-scores accepted). More difficulties are observed with time series (72% of $RMSE_u$ accepted). Industry appears as the most problematic source for RMs due to high variability among results.

Also the results obtained with CTMs are quite comparable to their ensemble reference using all models for the overall average (>92% of successful z-scores) while the comparability of the time series is more problematic (between 58% and 77% of the candidates' $RMSE_u$ are accepted). In these models a gap is observed between the sum of source contributions and the gravimetric PM_{10} mass likely due to PM underestimation in the base case.

If the tagged species CTM results are used in the reference, the differences between the two CTM approaches appear more evident, particularly in the mandatory set where there are three results obtained with brute force. In this case the percentage of candidates passing the z-score and $RMSE_u$ tests goes down to 50% and 86%, respectively. The percentages are higher in the optional set likely due to the fact that only one brute force result is available in this set.

CTMs show good performance when compared with RMs reference for the overall dataset (83% of the z-scores accepted), more differences are observed when dealing with the time series of the single source contributions. In this case the share of successful $RMSE_u$ ranges between 25% in the optional and 34% in the mandatory set.

Soil and exhaust are the CTM sources with higher negative bias when compared with RM (overall average).

z-score (overall average) (a)

RMSE$_u$ (time series) (b)

z-score (overall average) (c)

RMSE$_u$ (time series) (d)
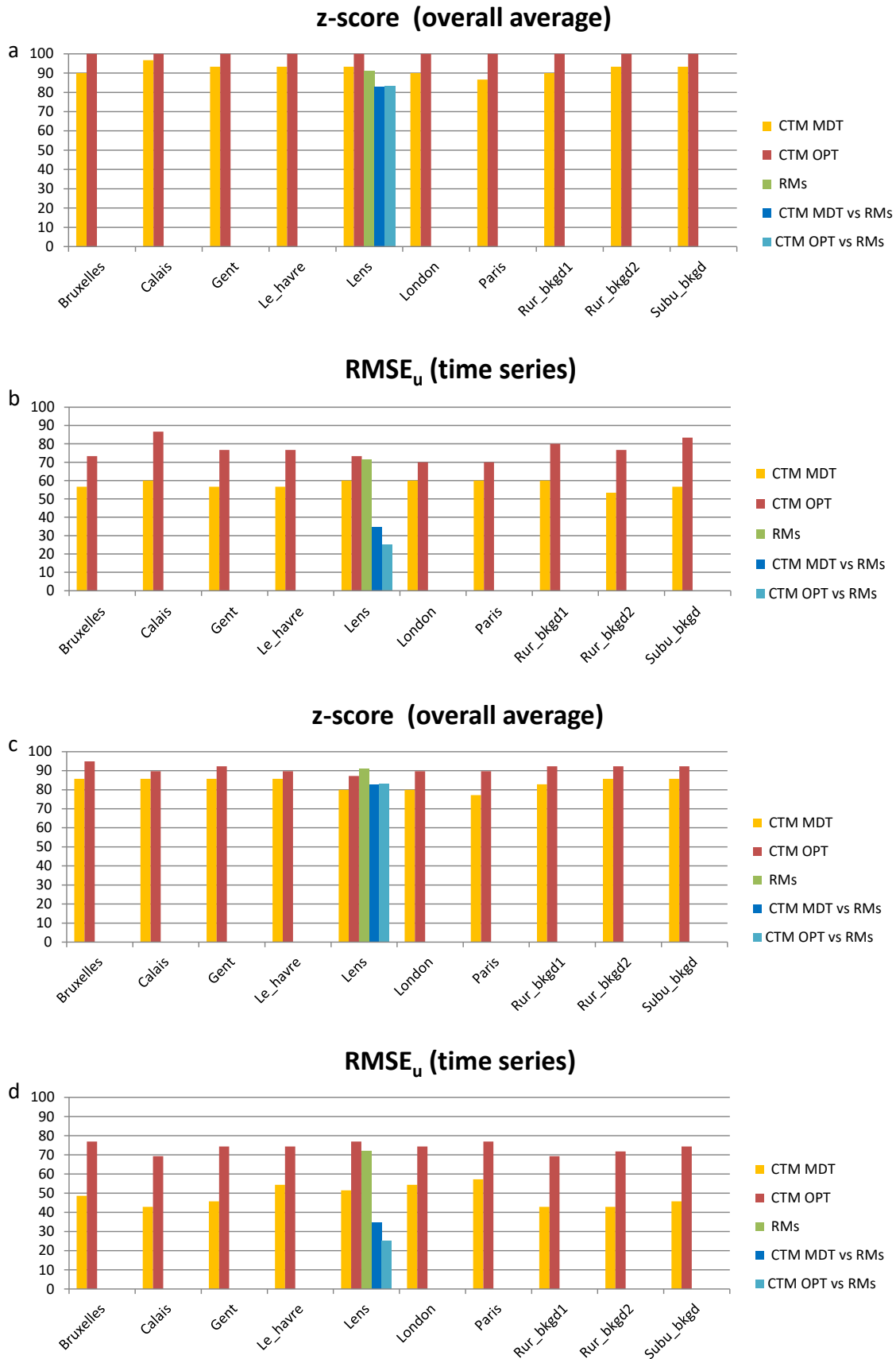
Figure 30. Synthesis of the intercomparison performance tests. a) and b) with all CTMs in the reference, c) and d) with only tagged species CTMs in the reference.

When it comes to the apportionment of specific PM components, both families of models identify traffic and biomass burning as the first two most contributing categories to elemental carbon and the range of SCEs reported by the two families is comparable. Traffic and biomass burning are also important sources of organic carbon even though RMs estimations are higher. In addition, RMs allocate sizeable primary emissions from soil, road dust, industry and energy sector. The comparison for the main inorganic ions (NO3, SO4 and NH4) and OPA is not straightforward because the two families of models treat them in a different way. Comparable contributions of NO3 and SO4 are observed in the energy sector and for NO3 also in ships. On the other hand, NH4 is allocated to agriculture in CTMs while RMs focus on the secondary nature of this species. Both families of models attribute OPA mainly to marine salt even though the estimations by CTMs are the double of those by RMs.

7.2 Conclusions

The high number of participants (40) and high quality of input data for both RM and CTM provided the basis to build up an unprecedented database with key information to support the identification of the factors that influence source apportionment model applications and the behaviour of models in general.

In RMs a convergence towards EPA-PMF5 tool (in part due to the good performances of EPA-PMF tools in previous intercomparisons) has contributed to more comparable results.

In RMs, the industry source category needs better definition because it encompasses a wide range of processes and associated emissions that results in a huge variability between results. More specific allocation within this category (e.g. in subcategories) would lead to a better identification (in terms of chemical profiles and time trends) and consequently to a more accurate quantification of the contributions.

Apparently good performances are observed in the comparison among CTM results when using an ensemble reference including both contributions calculated with brute force approach and those obtained with tagged species approach. Like RMs, models perform better in estimating the overall average contributions that the time trends.

In the performance tests, results obtained by models using brute force approach show higher chances of rejection. The most critical sources for this kind of models are agriculture, biomass burning, traffic and power plants.

When only tagged species results are used as references the differences between this approach and brute force are more evident. The inconsistencies are more dramatic when precursors from different sources are involved in secondary processes (the most emblematic case is the contribution of agriculture to the formation of ammonium nitrate and ammonium sulphate). In the other sources the differences are evident but fall in the area of acceptability. This is probably due to a) the linear behaviour of most sources (all the primary ones and some of the secondary ones) and b) corrections were introduced in the brute force techniques to deal with the lack of match between the sum of the sources and the PM total mass in the base case.

The comparison of CTMs source contributions with the gravimetric mass and the RMs source estimations points out a generalized difficulty to apportion all the PM mass which has been associated with the underestimation of the concentration of this pollutant in the base case model results, with particular reference to the organic fraction.

The comparability between CTM and RM is mostly depending on the source. The $RMSE_u$ test is passed only by candidates of the sources industry and traffic or exhaust. Such a good performance for traffic and exhaust sources in the target plots indicates the two families of models yield comparable SCEs and time trends. On the other hand, the apparently good agreement in the industry is likely due to the high uncertainty of the RM reference.

The most critical sources for CTMs, when compared with RMs, are soil and road dust. Appreciable underestimation, within the tolerance of the test, is observed also in ship, power plants, biomass burning and to a lesser extent exhaust.

The underestimation of SCEs observed in result cF (CHIMERE) is due to the methodological choice of not reporting apportioned nitrate and ammonia. This condition has penalised this results that, however, had obtained a quite satisfactory base case result.

This work demonstrated that the source apportionment assessment methodology is applicable to RMs and CTMs. The results of this study are relevant for the development of the activity of the CEN WG 44.

References

Belis, C.A., Larsen B. Amato F., El Haddad I., Favez O., Harrison R.M., Hopke P., Nava S., Paatero P., Prévot, A., Quass, U., Vecchi, R., and Viana, M., European Guide on Air Pollution Source Apportionment with Receptor Models, p. 88. Publication Office of the European Union, Italy,2014, doi:10.2788/9332.

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*, 2015a, 123, pp. 240-250.

Belis, C.A., Pernigotti, D., Karagulian, F., Pirovano, G., Larsen, B.R., Gerboles, M., Hopke, P.K., 'A new methodology to assess the performance and uncertainty of source apportionment models in intercomparison exercises', *Atmospheric Environment*, 119, 2015b, pp. 35-44.

EEA, 2015. Airbase dataset. https://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-2

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S., 'Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4)', *Geosci. Model Dev.*, 3, 2010, pp. 43-67, doi:10.5194/gmd-3-43-2010.

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. a. M., Helber, R., and Arnone, R. A., Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 76, 2009, pp. 64-82

Kinnison, D., Brasseur, G. P., Walters, S., et al., 'Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model', *J. Geophys. Res.*, 112, 2007, D20302, doi: 10.1029/2006JD007879.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C., 'TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling', *Atmos. Chem. Phys.*, 14, 2014, pp. 10963-10976, doi:10.5194/acp-14-10963-2014.

Kuenen, J. J. P., Visschedijk, A. J. H. and Denier van der Gon, H. A. C., 2015. TNO_SoAp_2011 emission inventory for Source Apportionment studies in support of FAIRMODE.

ISO 13528 ,2005. Statistical methods for use in proficiency testing by interlaboratory comparisons, edited, (ISO) International Organization for Standardization.

Passant, Speciation of UK emissions of NMVOC. 2012. AEAT/ENV/0545 Report.

Pernigotti, D., Belis, C.A., Spanó, L., 'SPECIEUROPE: The European data base for PM source profiles' *Atmospheric Pollution Research*, 7 (2), 2016, pp. 307-314.

Simon, H., Beck, L., Bhave, P.V., Divita, F., Hsu, Y., Luecken, D., Mobley, J.D., Pouliot, G.A., Reff, A., Sarwar, G., Strum, M., 2010. 'The development and uses of EPA's SPECIATE database'. *Atmospheric Pollut. Res., 2010,* pp. 196–206. doi:10.5094/APR.2010.026

Thunis, P., Pederzoli, A., Pernigotti, D., 'Performance criteria to evaluate air quality modeling applications'. *Atmospheric Environment* 59, 2012 pp 476-482.

VMM, CHEMKAR PM_{10} – Stedencampagne, Chemische karakterisering van fijn stof in Antwerpen, Gent, Brugge en Oostende, 2011-2012, 2013, Depotnummer D/2013/6871/021

List of abbreviations and definitions

cp   chemical profile (mass of a species relative to the total PM_{10} mass)

SCE   source contribution estimate (absolute mass apportioned to the different candidates)

sct   source contribution estimate time series (one daily value every 3 days for each candidate)

c2s   candidate contribution to species

PD   pearson distance calculated as 1-pearson correlation coefficient

SID cp   standardised identity distance for chemical profiles

PDcp   pearson distance for chemical profiles

RMSD   root mean square deviation, is a synonym of RMSE

$RMSE_u$ root mean square error normalised by the uncertainty of the reference

RM   receptor model

CTM   chemical transport model

List of figures

List of tables

Annexes

Annex 1. List of participant institutions

AGH-UST, University of Science and Technology (RM)
APPA Trento (RM)
ARIANET (CTM)
Aristotle University of Thessaloniki (RM)
ARPA Emilia-Romagna (RM)
ARPA Lombardia (RM)
ARPA Piemonte (CTM)
ARPA Puglia (RM)
ARPA Veneto (CTM)
CIEMAT (CTM)
Clarkson University, CARES (RM)
CNRS – LGGE* (RM)
Ecole des Mines de Douai – SAGE (RM)
ENEA (RM)(CTM)
Finnish Meteorological Institute (RM)
IDAEA-CSIC (RM)
IIA - CNR (RM)
IMROH (RM)
INERIS* (RM)
INFN (RM)
Institute for Nuclear Research, Atomiki, HAS (RM)
ISAC –CNR (RM)
Institut Scientifique de Service Public - Wallonie (RM)
Istituto Superior Tecnico, Universidade de Lisboa (RM)
LCME* (RM)
LISA-CNRS (CTM)
Miguel Hernandez University (RM)
NCSR "Demokritos" (RM)
Paul Scherrer Institut (RM)
Pontificia Universidad Católica de Chile (RM)
RAMBOLL-ENVIRON (CTM)
RIER, University of Cologne (CTM)
RIVM (RM)
RSE SpA (CTM)
Slovenian Environment Agency, ARSO (RM)
TNO (CTM)
Università degli Studi di Milano (RM)
University College Cork (RM)
University of Aveiro (CTM)
University of Bari (RM)
University of Bologna (RM)
University of Genoa (RM) (CTM)
University of Helsinki (RM)
University of Milano-Bicocca (RM)
Warsaw University of Technology (CTM) (RM)

*LGGE+ (collaboration between INERIS, LCME and LGGE)

Annex 2. Performance evaluation of chemistry transport models

All modelling teams performed their own model performance evaluation (MPE), by comparing CTMs results against observed data of the main chemical species. The air quality data collected and used for MPE were derived from different data set available in the framework of the exercise and concerning the whole Europe as well as the reference site. The observed data set included: 1) a selection of Airbase (EEA, 2015) background stations over the Lens domain; 2) the PM composition data from Lens, also used by RMs; 3) PM composition data from a field campaign held at three Flemish sites (VMM, 2013). In this section, only a subset of 7 sites was used to compare the performance among the different models. The comparison of model performance was limited to PM$_{10}$. Moreover, at Lens site models performance was evaluated also for PM composition. The spatial distribution of the measurement sites is presented in Figure 1.

The measurement sites have been selected in order to cover the different meteorological and emissive features of the Lens domain. Particularly, besides Lens, the following sites were selected: London, Paris (big cities), Gent (middle-size city), Le Havre, Calais (coastal areas) and a rural background station.

Model results were evaluated by means of a few statistical indicators (see Appendix A), namely: the mean bias (MB), the normalised mean bias (NMB), the root mean square error (RMSE) and the Pearson correlation.
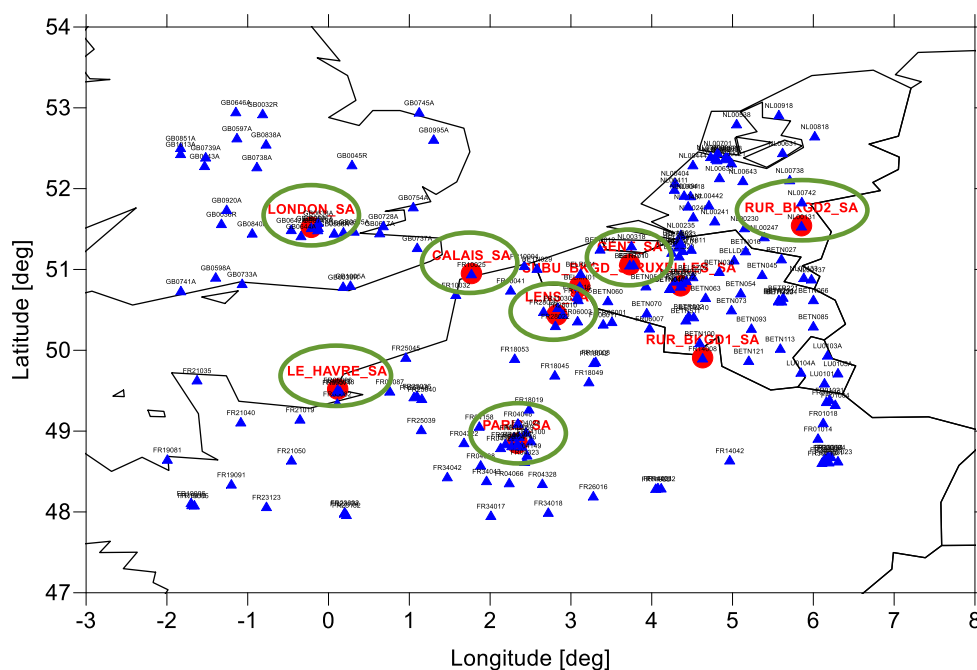

Figure 1 – Position of the available measurement sites (blue). Selected sites are denoted with green rings.

Figure 2 shows the comparison of modelled and observed PM$_{10}$ concentrations at all sites for winter and summer period. During the winter season most of the models tend to underestimate the highest concentrations. The only exception is cB showing a clear overestimation of the observed concentrations ranging between 10 and 40 μg/m^3. PM$_{10}$ underestimation gives rise for all models to a low bias comprised between 30% and 45%. RMSE ranges between 14.4 and 16.7 μg/m^3, roughly corresponding to 58-67% of the observed mean. cB model shows one of the best MB performance but also the worst results with respect to RMSE. This is due to the fact that cB model tends to underestimate the highest observed values, like most of models, but it also overestimates low observed concentrations. This model also shows a poor correlation value (0.32). Best performance in temporal correlation are shown by model cE (0.76), though it shows also the highest bias. Models

underestimate also the summer concentrations, showing a low bias ranging between 31% and 47%. Correlation ranges between 0.44 and 0.67 for most models, except for cB and cD showing correlation values lower than 0.2. Likewise winter period, this is related to the tendency of both models to overestimate the low observed concentrations, while they underestimate high concentration values. For both periods models are closer to each other than to observations. This result points out that most of the discrepancies between models and observations are not related to specific assumptions of each model, but to more common features related either to input data or processes that are poorly described in all models.

Models underestimations are more frequently related to the high observed values, suggesting that also the analysis of source apportionment results should be limited to mean concentrations and not extended to exceedance days that were poorly reproduced by the models.
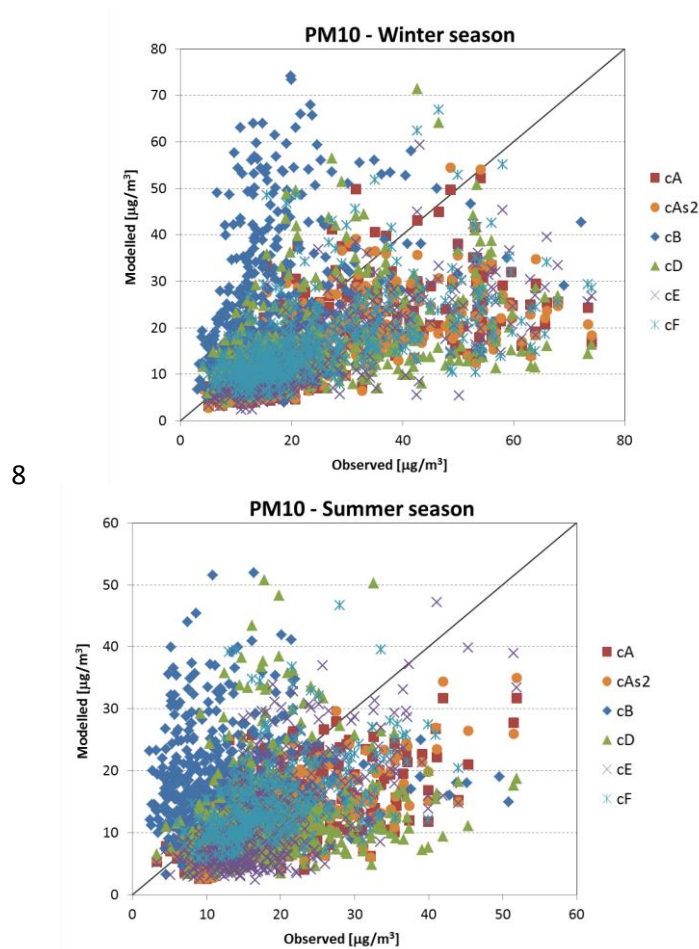

8

Figure 2 – Scatter plot of PM$_{10}$ daily mean concentrations modelled and observed at selected sites for winter (left) and summer (right) episodes.

9

Table 1. Comparison of the model performance evaluated at all selected sites for winter and summer periods.

Winter	cA	cAs	cAs2	cB	cD	cE	cF	Observation
#	651	651	651	651	651	623	651	469
Mean	15.65	15.67	14.82	16.09	16.91	14.35	17.49	24.63
MB	-9.59	-9.60	-10.33	-8.19	-8.24	-11.08	-7.50	
NMB	-39%	-39%	-42%	-33%	-33%	-45%	-30%	
RMSE	14.35	14.41	14.78	16.73	15.85	15.28	14.44	
Correlatio	0.69	0.68	0.70	0.32	0.43	0.76	0.55	

n								
10								
Summer	**cA**	**cAs**	**cAs2**	**cB**	**cD**	**cE**	**cF**	**Observation**
#	644	0	644	644	644	637	427	483
Mean	12.00		11.62	11.65	12.43	12.69	14.88	18.26
MB	-6.42		-6.75	-5.95	-5.59	-5.67	-8.52	
NMB	-35%		-37%	-33%	-31%	-31%	-47%	
RMSE	8.93		8.83	11.90	11.00	9.15	13.15	
Correlation	0.59		0.67	0.02	0.18	0.62	0.44	

Figure 3 shows the winter and summer time series of the PM_{10} daily mean concentrations modelled and observed at each selected site. The observed winter mean concentration ranges between 20 $\mu g/m^3$ in London and 28 $\mu g/m^3$ in Gent and Calais. At Lens site, the observed mean concentration was 25 $\mu g/m^3$. During the winter period, the observed concentration in Lens, as well as at most selected sites, shows the development of three episodes. The first one took place around November 20th, 2011, the second one around January 15th, 2012 and the longest one taking place in the first half of February. During the first and third episode, the observed PM_{10} concentration was higher than 60 $\mu g/m^3$, while during the second one it was around 40 $\mu g/m^3$. Comparable concentrations were observed in Paris and London but also in Le Havre, Calais and at RUR_BKD2 site. This should point out that the temporal evolution of the PM_{10} concentration during the winter period was mainly driven by regional scale processes. Therefore, the selected modelling approach adopting a 7 km horizontal resolution should be adequate to reproduce it as well to perform the source apportionment analysis. Models proved to be able to reproduce the temporal evolution of the PM_{10} concentrations at most sites. Particularly at Lens site, winter episode correlation ranged between 0.7 for model cD and 0.87-0.88 for cA runs. Correlation was satisfactorily also in large urban areas like London (0.69 – 0.90) as well as at Rural Background site (0.60 – 0.75). Model cB and cD showed results less correlated to observations than the other models at coastal sites and to a lesser extent in Paris (correlation values lower than 0.6).

In general, models underestimated the observed mean concentrations at all sites with low biases ranging between 40%-50% at Lens site and between 40%-60% at coastal sites and outside the big cities. Model underestimations were mainly due to the severe episode taking place during the first half of February 2012. In Paris and London NMB was usually lower than 30% and there were models that overestimated the average PM_{10} concentration (cF and cD in Paris and cB in London). In coastal and rural background sites models fail to reproduce high pollution level episodes.

During the summer period the observed concentrations range between 12 $\mu g/m^3$ in London and 24 $\mu g/m^3$ at Rural Background site. The strong difference between winter and summer concentrations in London is probably driven by the reduction of domestic heating emissions as well as by the development of more unstable conditions. Differently, at RUR_BKGD2 site winter and summer concentrations are very similar, confirming that the area is subject to regional scale processes in both seasons. At Lens site the observed concentration was around 18 $\mu g/m^3$, lower than in winter period. At Lens site, all models underestimated the observed concentration, showing a low NMB ranging between 34% (cF) and 58% (cB). Model performance during the summer period is more scattered than winter season, probably due to the greater influence of chemical processes, whose reproduction is handled differently by each model. This should probably influence also the source apportionment results produced by each model.

The stronger influence of chemical processes is clearly pointed out also by correlation values that in Lens range between 0.31 (cD) and 0.81 (cE).

Models were more skilful in reproducing the observed mean concentration in the metropolitan areas of Paris and London than in rural and coastal areas. However, in both cities the model performance for summer period is more scattered than winter season, showing a NMB ranging between -36% (cF) and 65% (cB) in London and -33% (cAs2) and 34% (cD).

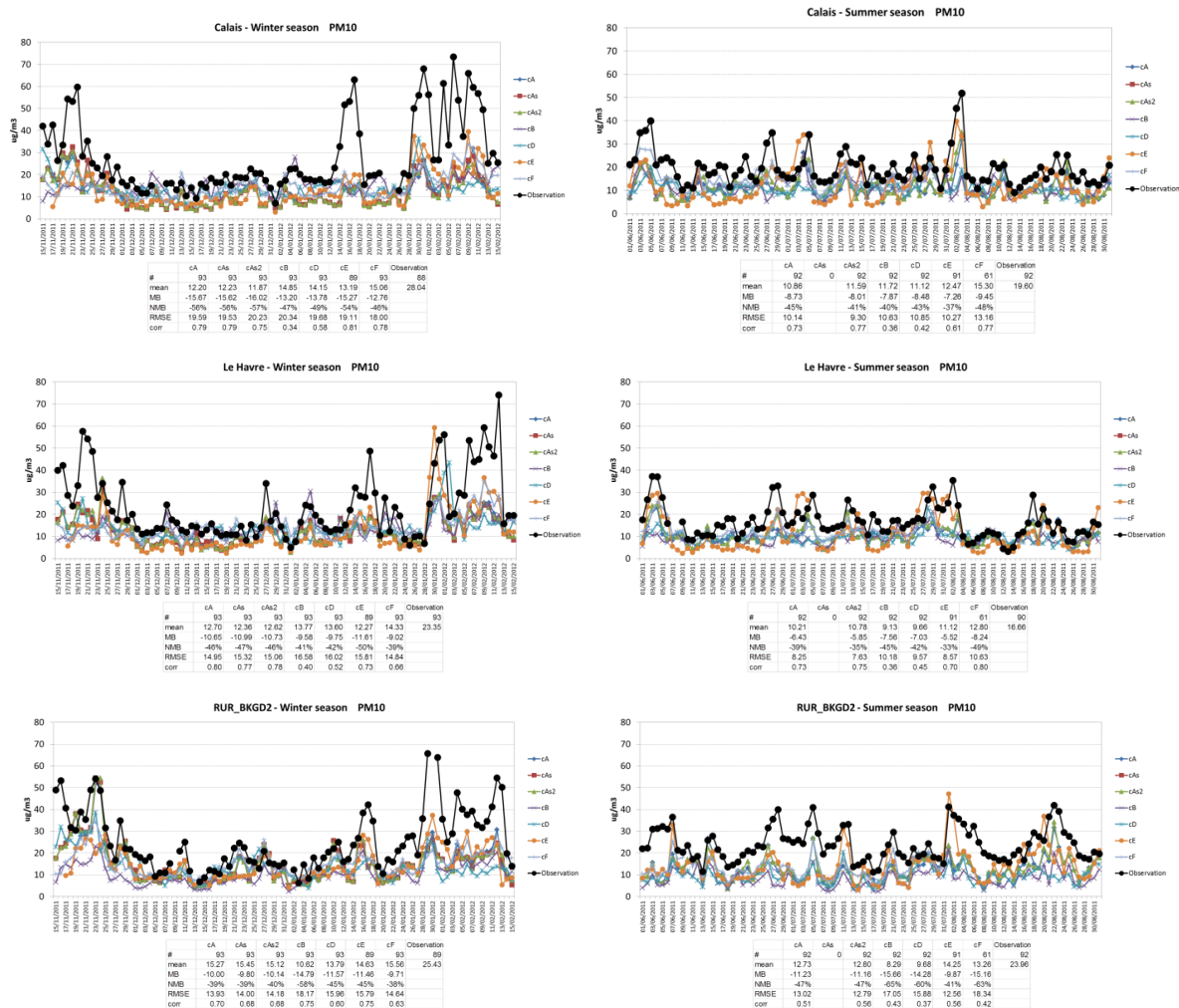Models showed rather scattered results also with respect to correlation, with cE and cF models generally showing the best performance and cB and cD the worst ones.
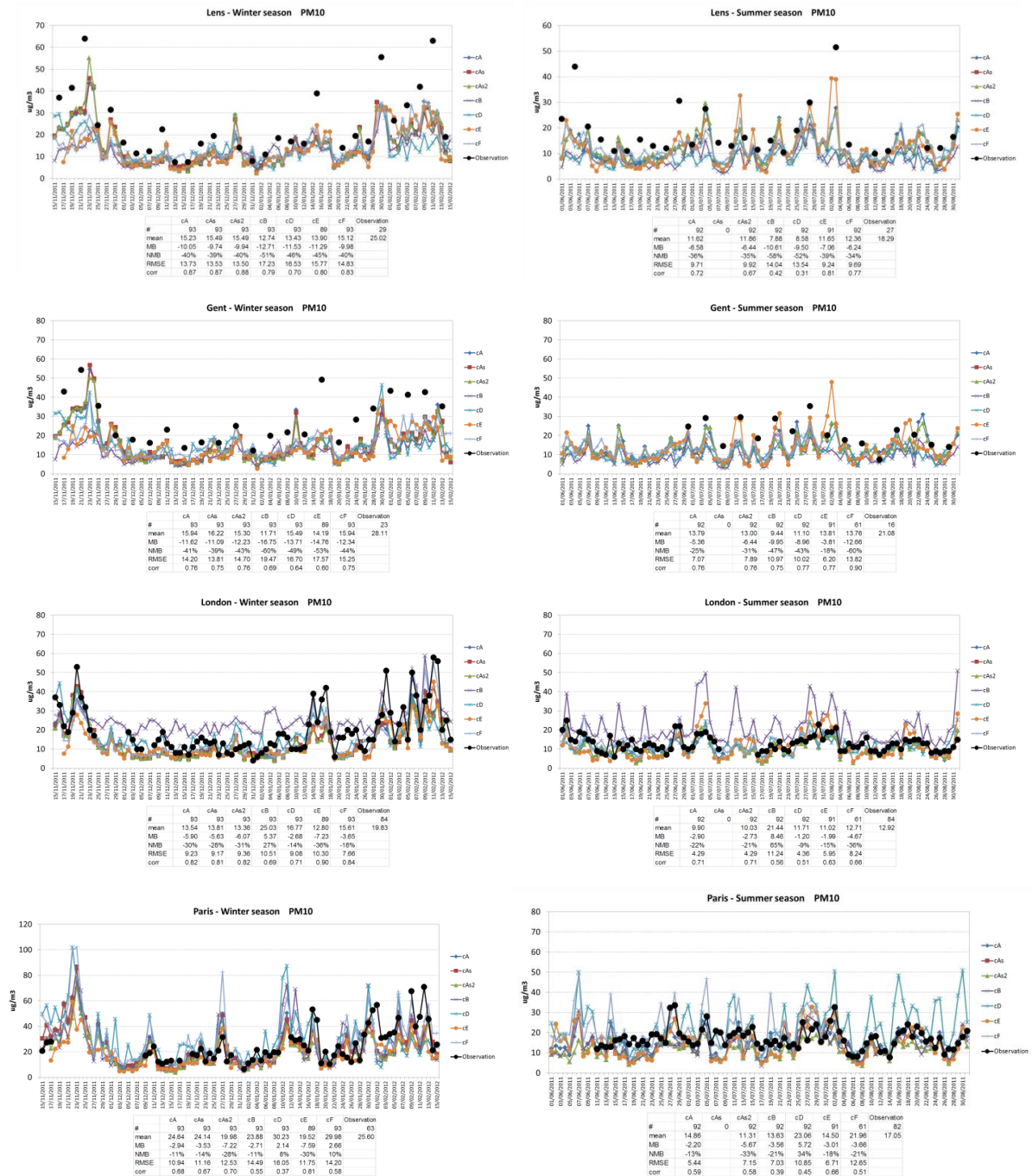
Figure 3 – Comparison of observed and modelled PM₁₀ daily concentrations at selected sites for winter (left) and summer (right) period. The figure also shows the comparison of some performance indicators computed for each period and site. Please refer to the text for additional details.

Models performance was also evaluated by comparison with the PM composition data at the site of Lens (Figure 4). Such analysis is very useful as it allows to better investigate the influence of the different modelling assumptions and input data on the obtained results with respect to the total concentration but also for source contribution estimates.

The first evaluation refers to Elemental Carbon (EC), which is a primary and not reactive compound, therefore it is influenced only by emission and dispersion processes. As expected, all models provided rather similar results with respect to NMB, as they shared the same emissions and meteorological fields. The only exception was model cE that showed a strong overestimation for both episodes.

cB model showed the best performance for NMB (-17% and -13% for winter and summer, respectively), but also the worst results for correlation (around 0.53 for both seasons).

Organic aerosol (OA) is strongly underestimated by most of models that showed a negative NMB greater than 70%, with the exception of cB model that showed slightly better performance (around - 65%) and cE model that showed a good performance (lower underestimation) in winter but in summer period overestimated the observed concentration.

Rather surprisingly correlation values are higher than 0.7 for all models and both periods. This result suggests that models seem able to reproduce the temporal evolution of processed influencing OA concentration, but not their strengths. The strong underestimation of OA can influence also the reliability of source contribution estimates, particularly concerning domestic heating during winter season, biogenic sources during summer season and road transport for both periods.

Sulphate was better reproduced than OA by all models, particularly in cA runs during winter season. Differently, during the summer season the results are more scattered with cA and cF models overestimating the observed concentration, while cB, cD and cE underestimated. cE also shows a very poor correlation in winter (0.13). Such difference among the models could have an influence on the reconstruction of the source contribution from Energy production and Shipping sectors, which represent the main sources of sulphur.

Nitrate was very well reproduced by most models during the winter season, as proved by both NMB, that was frequently close to 0%, and correlation values that were higher than 0.7. This is confirmed also by the analysis of the daily time series showing that models were able to reproduce most of the observed temporal variability. Nitrate concentration during the summer season was correctly reproduced by models cB, cD and cF, while cA and cE clearly overestimated the observed values.

The overestimation of nitrate and sulphate observed in models cA, cE and partially cF gave rise to a corresponding overestimation of the ammonium concentration during the summer season. This holds particularly for cA model that showed a NMB greater than 75% during the warm period. Such overestimation can influence the source contribution estimate of the agriculture sector that represents the main sources of ammonia over the Lens area. Conversely, during the winter season, ammonium is very well reproduced by all models, showing a NMB very close to 0%, with the exception of model cB (-51%), and correlation values higher than 0.70.
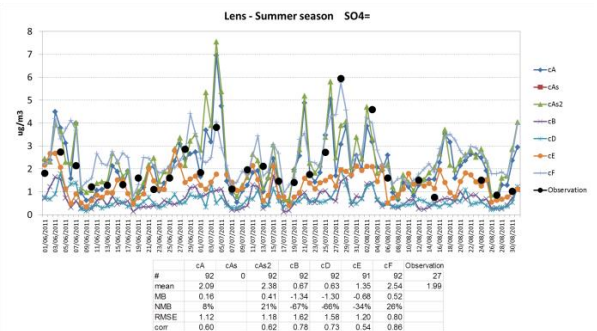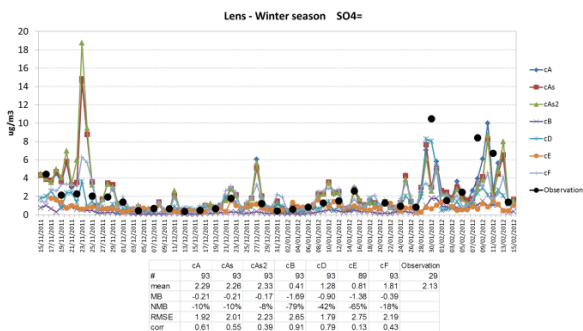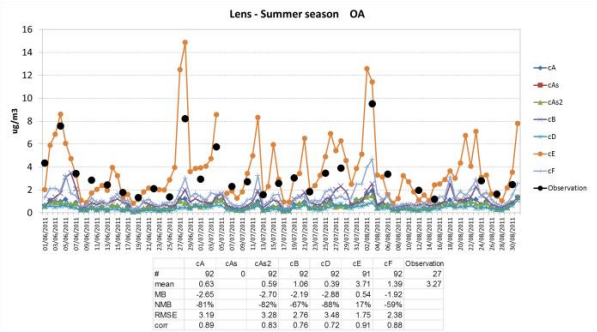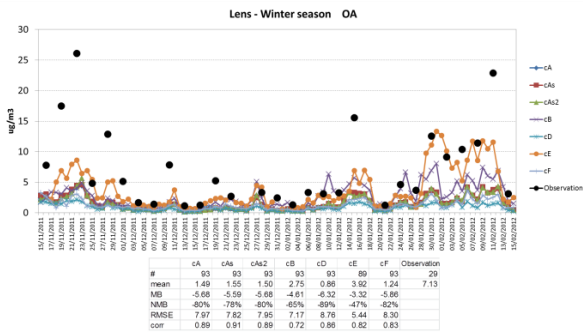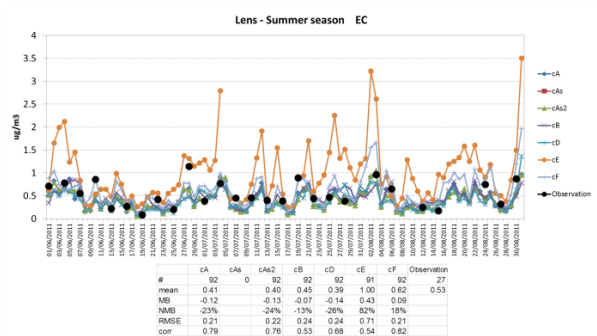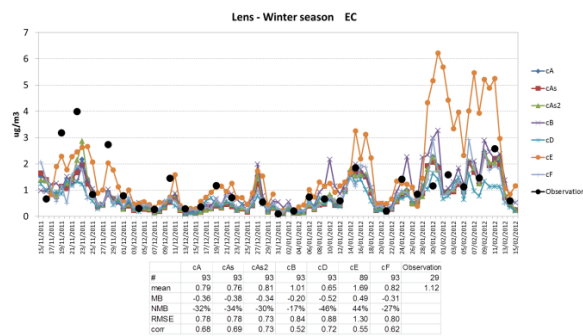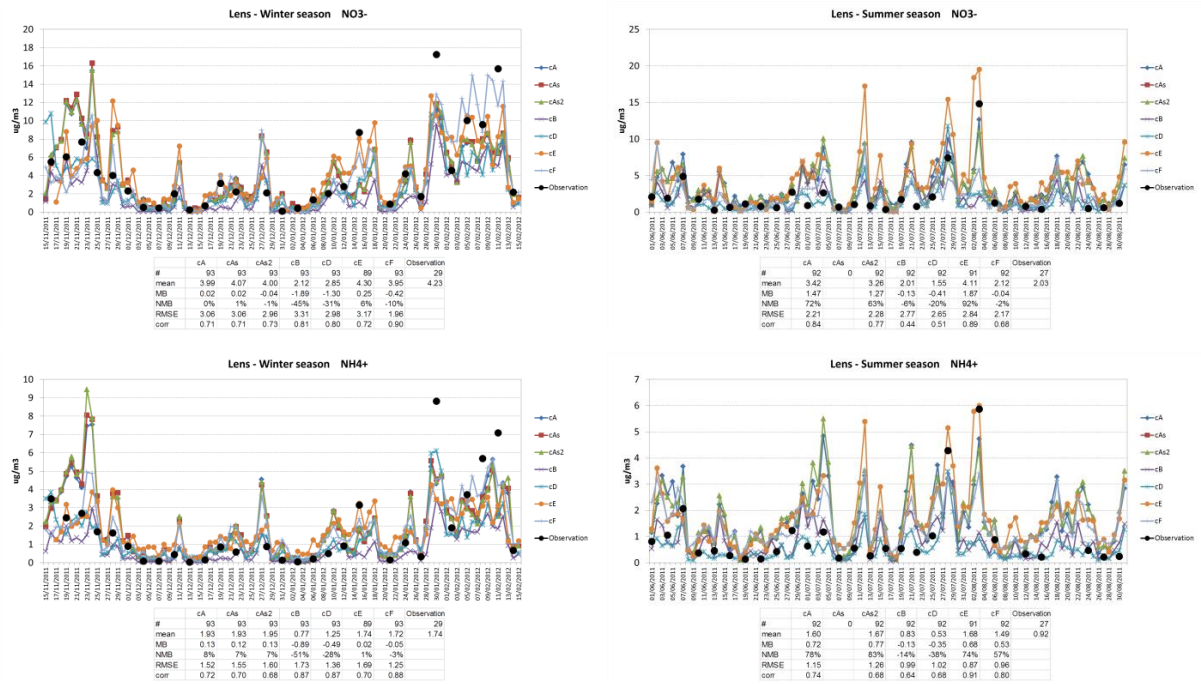
Figure 1 – Comparison of observed and modelled daily concentrations of PM$_{10}$ composition data at Lens site for winter (left) and summer (right) period. The figure also shows the comparison of some performance indicators computed for each period and compound. Please refer to the text for additional details.

Mean Bias

$$MB = \frac{1}{N} \sum_{t=1}^{N} \left(C_{\text{mod}}(x,t) - C_{obs}(x,t) \right) = \overline{C}_{\text{mod}}(x) - \overline{C}_{obs}(x)$$

Normalized Mean Bias (NMB)

$$NMB = \frac{\frac{1}{N} \sum_{t=1}^{N} \left(C_{\text{mod}}(x,t) - C_{obs}(x,t) \right)}{\frac{1}{N} \sum_{t=1}^{N} C_{obs}(x,t)} \qquad (A.1)$$

Root Mean Square Error (RMSE)

$$11 \qquad RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(C_{\text{mod}}(x,t) - C_{obs}(x,t) \right)^2}$$

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub