The European Commission's science and knowledge service

Joint Research Centre

Source apportionment inter-comparison exercise 2015-2016 part 1 RM

C.A. Belis , D. Pernigotti *EC-JRC*

G. Pirovano RSE SpA

Fairmode Technical meeting
Athens, 19-21 June 2017

European Commission

Contributors

O. Favez, J.L. Jaffrezo, J. Kuenen, H. Denier van Der Gon, M. Rezler, M.T. Pay, M. Almeida, F. Amato, A. Aniko, I. Beslic, M. Bove, G. Calori, D. Cesari, C. Colombi, D. Contini, G. De Gennaro, A. Di Gilio, I. El Haddad, H. Elbern, K. Eleftheriadis, G. Foret, M. Garcia Vivanco, S. Gilardoni, M. Grosa, S. Hellebust, R. Hoogerbrugge, P. Hopke, Y. Izadmanesh, H. Jorquera, A. Karppinen, Z. Kertesz, T. Kolesa, P. Lazzeri, F. Lenartz, F. Lucarelli, A. Manders, H. Martins, M. Mircea, D. Mooibroek, S. Nava, D. Oliveira, P. Paatero, M. Paglione, M. Perrone, E. Petralia, A. Pietrodangelo, S. Pilion, P. Pokorna, P. Prati, M. Reizer, V. Riffault, C. Samara, L. Samek, S. Sauvage, F. Scotto, K. Sega, G. Siour, R. Tauler, R. Vecchi, E. Venturini, M. Vestenius, E. Yubero

Information that can be obtained from this IE

- **Overall model performance on the basis of pre-established criteria,**

✓ *for the purposes of air quality management (AQM)*

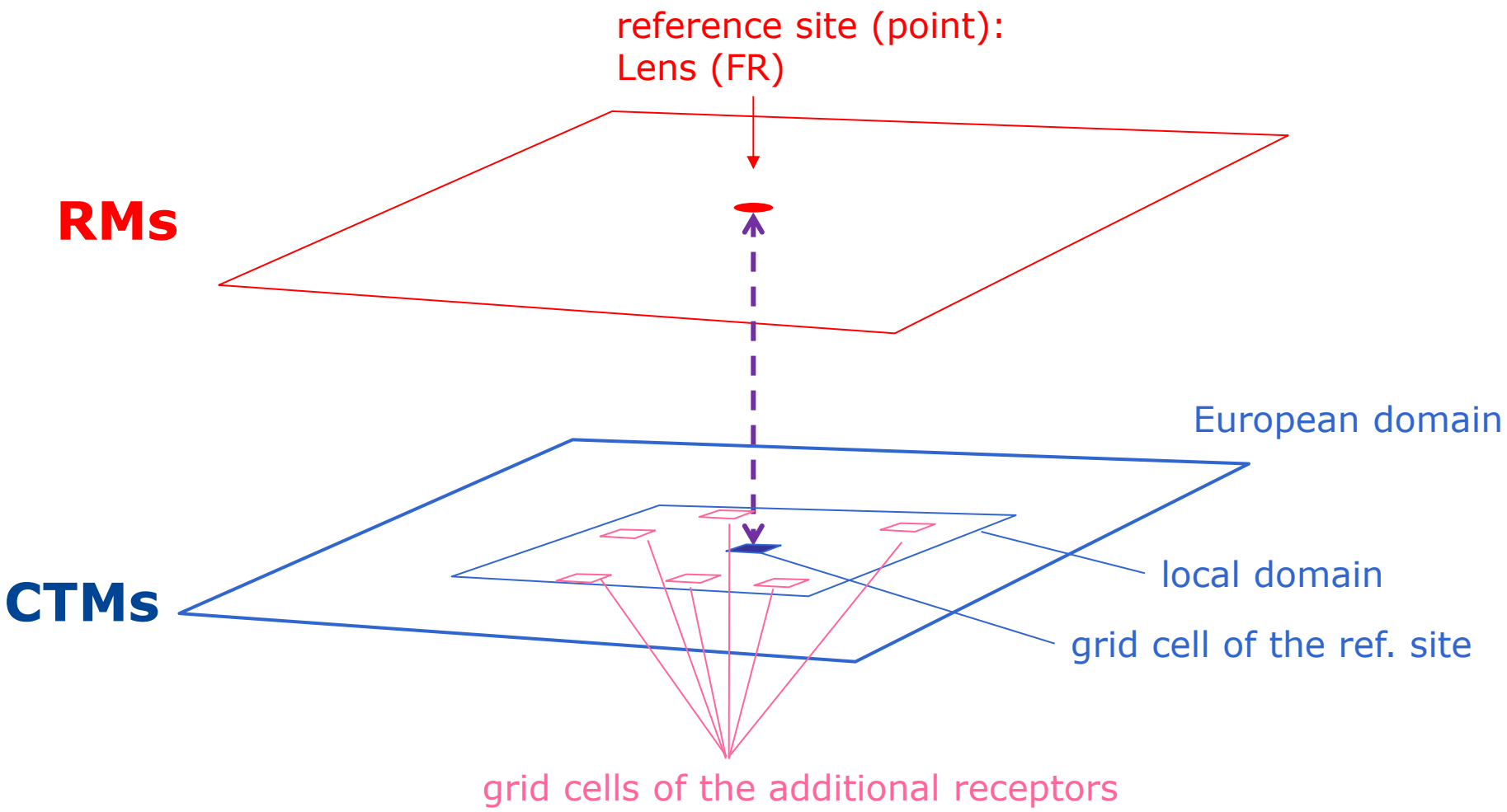
- **Indirect measure of the overall output uncertainty,**
- **More robust SA results (from a single outcome to an ensemble)**
- **Cross-validation of obtained results (to overcome the lack in observed data)**
- **Provide insights to understand the models behavior:**

✓ *influence of specific factors (e.g. input data, type of site, type of pollutant, meteorological conditions, etc…)*

✓ *sensitivity to modelling approaches (e.g. RMs vs SMs) and assumptions*

- **Additional details about SMs performance**
- **Integration of RMs outcomes (e.g. Apportionment of secondary pollutants, source-regions apportionment,…)**

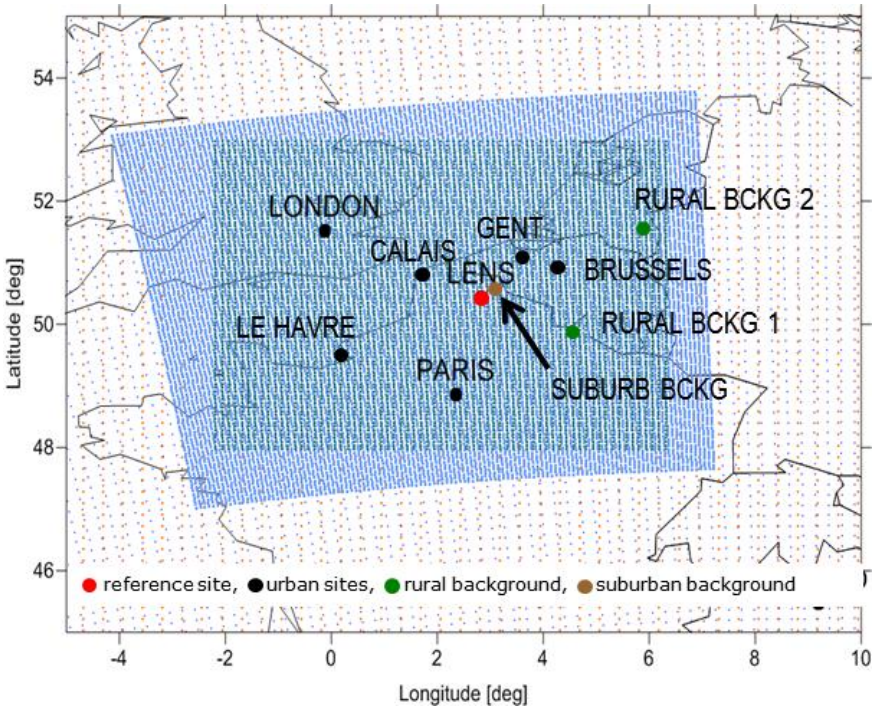
How is the intercomparison organised?



European Commission

Intercomparison outline – Source oriented models (CTM)

- **Common input dataset**
  - ECMWF meteorology
  - TNO emissions
  - MACC chemical fields
- **Centralized MPE (by RSE)**
  - LENS dataset
  - ca. 200 AIRBASE sites
  - Local networks
- **Set of receptors (10)**
  - Lens
  - Urban sites
  - Coastal sites
  - Background sites



8 - 14 source categories

3 + 3 summer/winter months

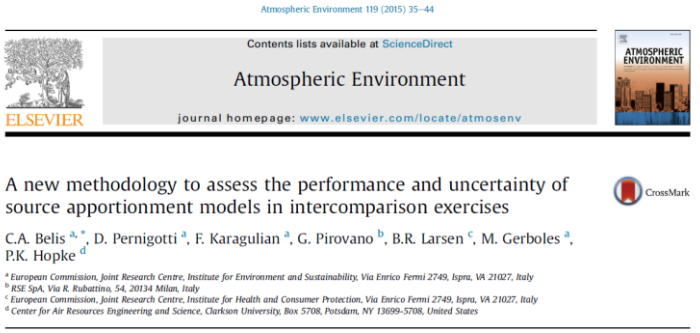
Hourly concentrations

Primary and secondary PM

PM precursors

European Commission

Evaluation Methodology (RM)


Atmospheric Environment 119 (2015) 35–44

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

A new methodology to assess the performance and uncertainty of source apportionment models in intercomparison exercises

C.A. Belis [a,*], D. Pernigotti [a], F. Karagulian [a], G. Pirovano [b], B.R. Larsen [c], M. Gerboles [a], P.K. Hopke [d]

[a] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via Enrico Fermi 2749, Ispra, VA 21027, Italy
[b] RSE SpA, Via R. Rubattino, 54, 20134 Milan, Italy
[c] European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Via Enrico Fermi 2749, Ispra, VA 21027, Italy
[d] Center for Air Resources Engineering and Science, Clarkson University, Box 5708, Potsdam, NY 13699-5708, United States


Atmospheric Environment 123 (2015) 240–250

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

A new methodology to assess the performance and uncertainty of source apportionment models II: The results of two European intercomparison exercises

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

Complementary tests: provide ancillary information about the solutions' performance

- Mass apportionment
- Number of factor/sources

Preliminary tests: test if source/factors belong to a given source category

- Chemical profiles → Pearson distance, SID
- Time-trends → Pearson distance
- Contribution-to-species (%) → Pearson distance

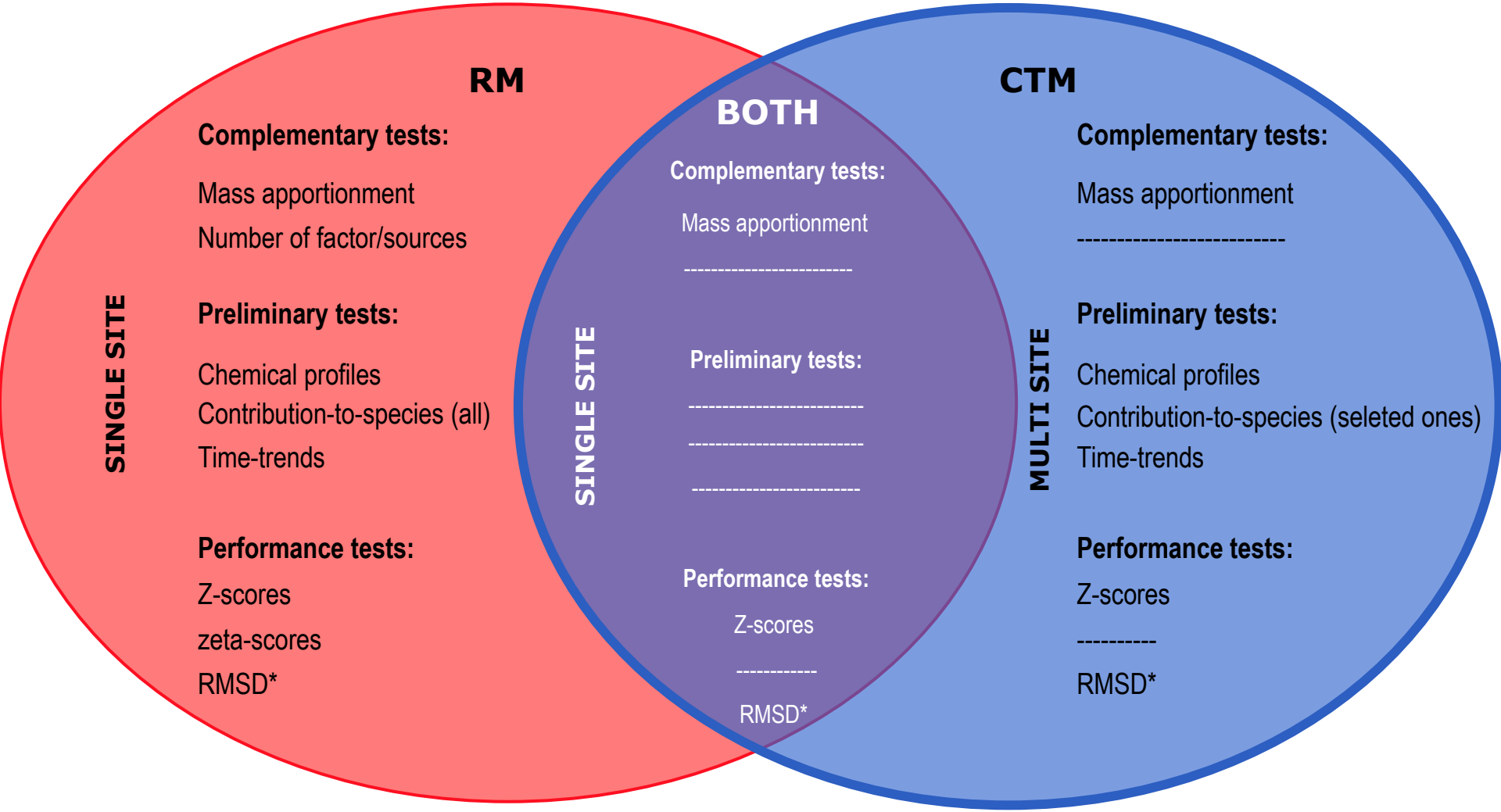
= % of species total matrix (EPA PMF v3) = explained variation (PMF 2) = contribution by species (CMB 8.2) = fingerprints (EPA PMF5)

Performance tests Evaluate if source/factor SCEs fall within an established quality objective

- Z-scores → test solution bias coherence with the quality objective (σ_p)
- RMSEu → test the bias, amplitude and phase of the SCE time trends

European Commission

Evaluation in this IE



European Commission

Evaluation of Receptor Model applications

Intercomparison outline – Receptor oriented models

DATA SET WITH SPECIATED PM (including organic markers)

COUNTRY	France
PERIOD	03.2011 to 03.2012
TIME RES.	every 3 days
DURATION OF SAMPLING	24 hours
TYPE OF SITE	Urban background
	PM10
N SAMPLES	116
IONS	ok (8 species)
EC/OC	ok
TRACE ELEMENTS	ok (25 species)
PAHs	ok (15 species)
LEVO/MANN	ok + galacto
HOPANES	ok (10 species)
N-ALKANES	ok (29 species)
CHOLESTEROL	
SOA MARKERS	ok
OTHER	Pristane, Phytane, Glucose

Intercomparison outline – Receptor oriented models

PARTICIPANTS

AGH-UST	ISAC LE	RIVM
APPATN	FMI	SAGE
ARPA ER	IDAEA_T	UCC
ARPA LO	IDAEA_A	UMH
ARPA PU	IMROH	UNIBO
ARSO	ISSeP	UNIHE
AUTH	IST	UNIMI
CARES	LGGE+	UNMIB
CNR IIA	NCSR	UNIFI
ENEA	PSI	UNIGE
ISAC BO	PUC	WUT

MODELS

EPA PMF5

EPA PMF4

EPA PMF3

PMF2

RCMB

MLPCA-MCR-ALS

ME-2*

*non- USEPA versions

	SOURCE CATEGS, (SPECIEUROPE)
1	traffic
2	exhaust
10	soil
12	marine fresh
20	industry
30	fuel oil
31	coal
37	ship
40	biomass burning
41	wood burning
5	road dust
60	SIA
61	ammonium nitrate
62	ammonium sulphate
66	deicing salt
70	POA
71	aged sea salt
74	combustion

Participants

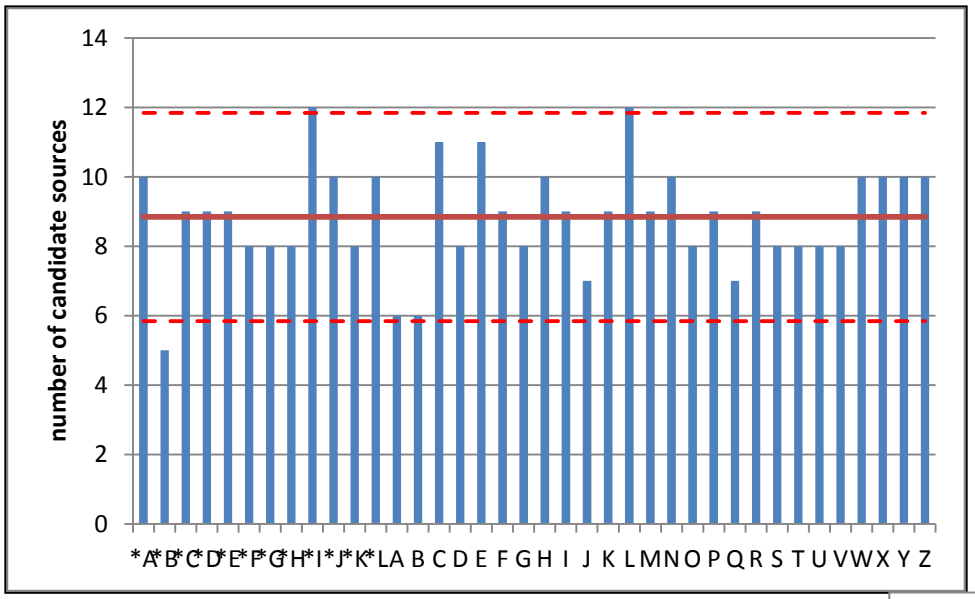
Participant abbrev.	Model	n. of candidates	n. of species	n. of time steps
ATOMKI	PMF5	6	47	108
AGH-UST	PMF5	6	98	116
APPATN	PMF5	11	51	116
ARPA ER	PMF5	8	27	115
ARPA LO	PMF5	11	90	115
ARPA PU	PMF5	9	28	116
ARSO	PMF5	8	89	88
AUTH	RCMB	10	98	116
CARES	PMF5	9	27	116
ISAC_BO	PMF5	7	85	109
CNR IIA	PMF5	9	38	116
ENEA 12	PMF5	12	27	116
ENEA 9	PMF5	9	27	116
INFN	PMF5	10	30	116
FMI	PMF5	8	51	116
UNIGE	PMF5	9	29	116
IDAEA_T	MLPCA	7	92	116
IDAEA_A	PMF5	9	41	116
IMROH	PMF5	8	98	116
ISSeP	PMF5	8	29	116
IST	PMF5	8	32	116
PSI	ME2	8	75	116
LGGE+	PMF5	10	36	116
NCSR_5	PMF5	10	27	116
NCSR_2	PMF2	10	35	116
ISAC_LE3	PMF3	10	34	116
ISAC_LE5	PMF5	10	34	116
PUC	PMF5	5	20	116
RIVM1	ME2	9	48	116
RIVM2	PMF5	9	48	116
RIVM3	PMF5	9	40	116
SAGE	PMF5	8	36	116
UCC	PMF5	8	71	116
UMH	PMF5	8	27	116
UNIBO	PMF4	12	49	116
UNIMI	PMF5	10	27	116
UNMIB	PMF5	8	34	116
WUT	PMF5	10	36	116

Source Categories

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

RM preliminary tests

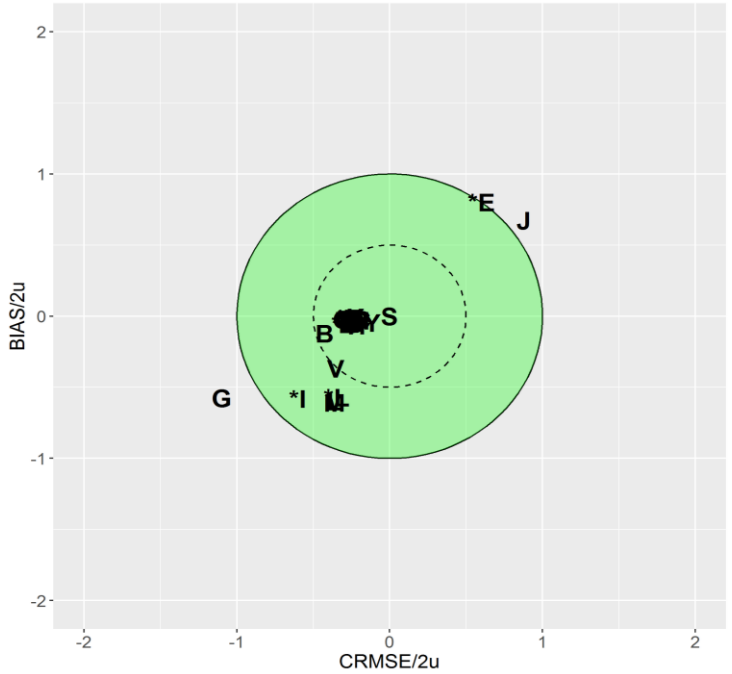
Number of candidates (sources)



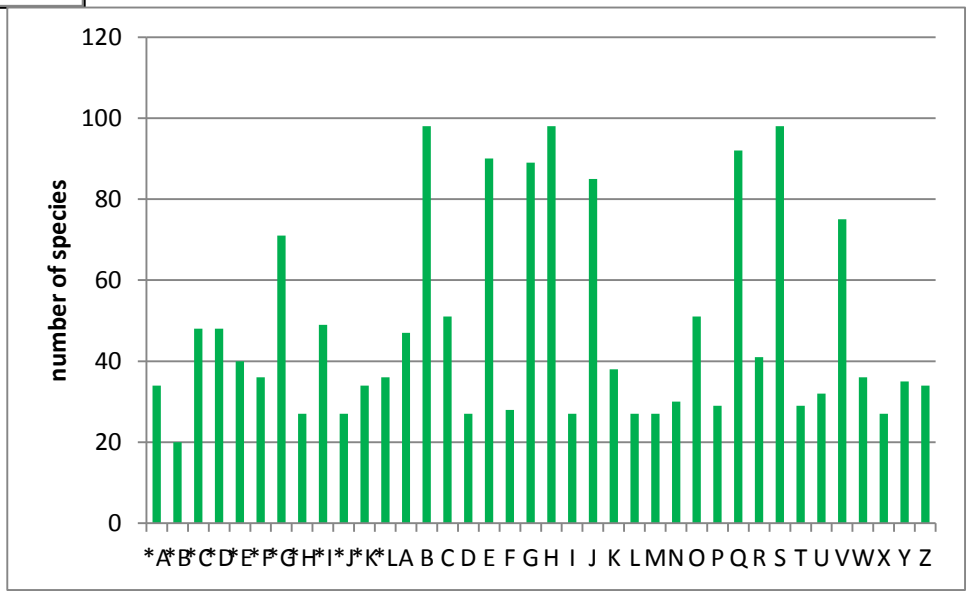
red solid line bar average (9)

Red broken lines represent avg. +-3

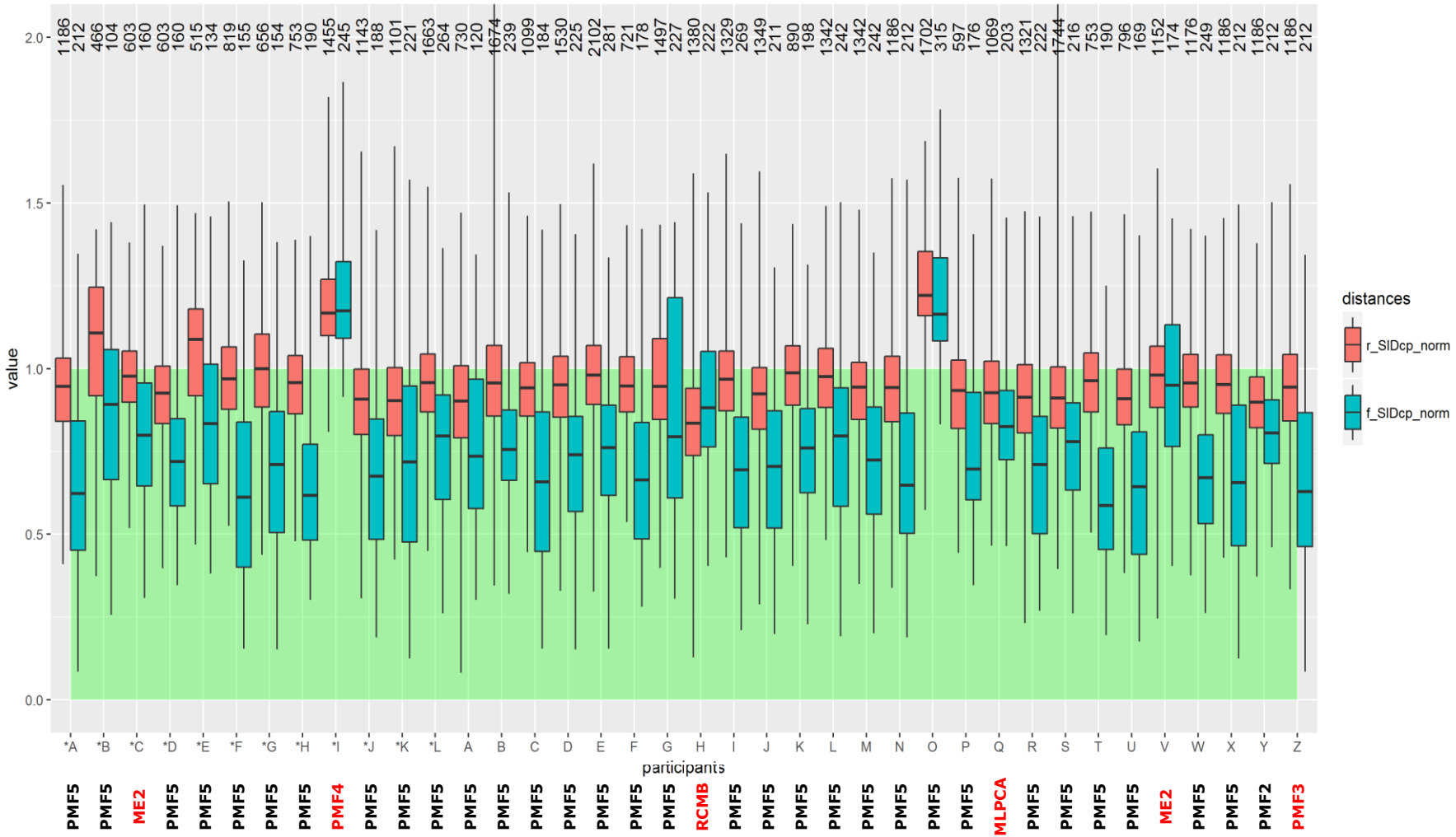
Mass apportionment



Number of species



SID all participants



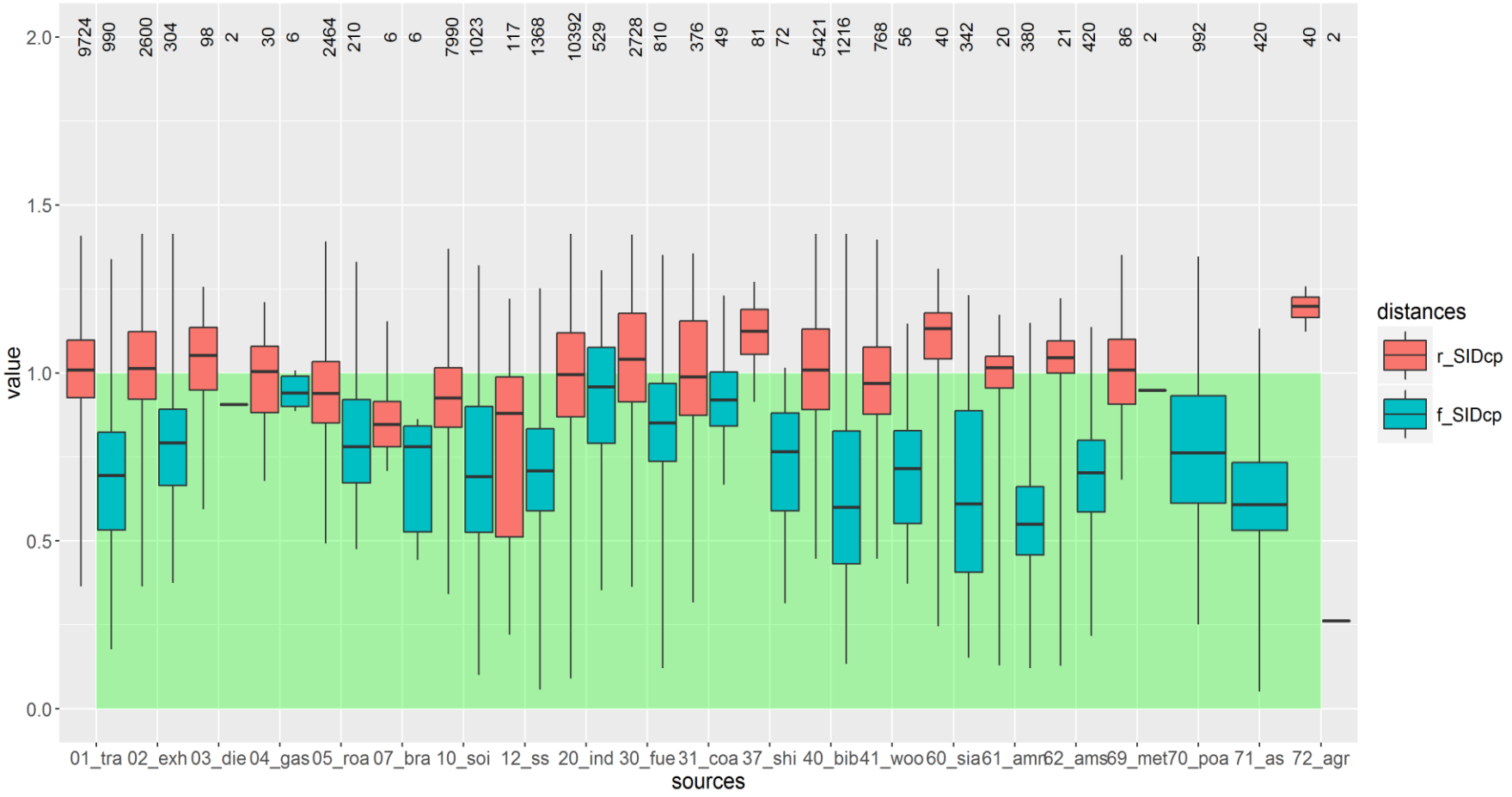
r = distances to the reference chemical profiles (cp) in SPECIATE and SPECIEUROPE

f = distances among the candidate sources

top = number of distances

European Commission

SID by source



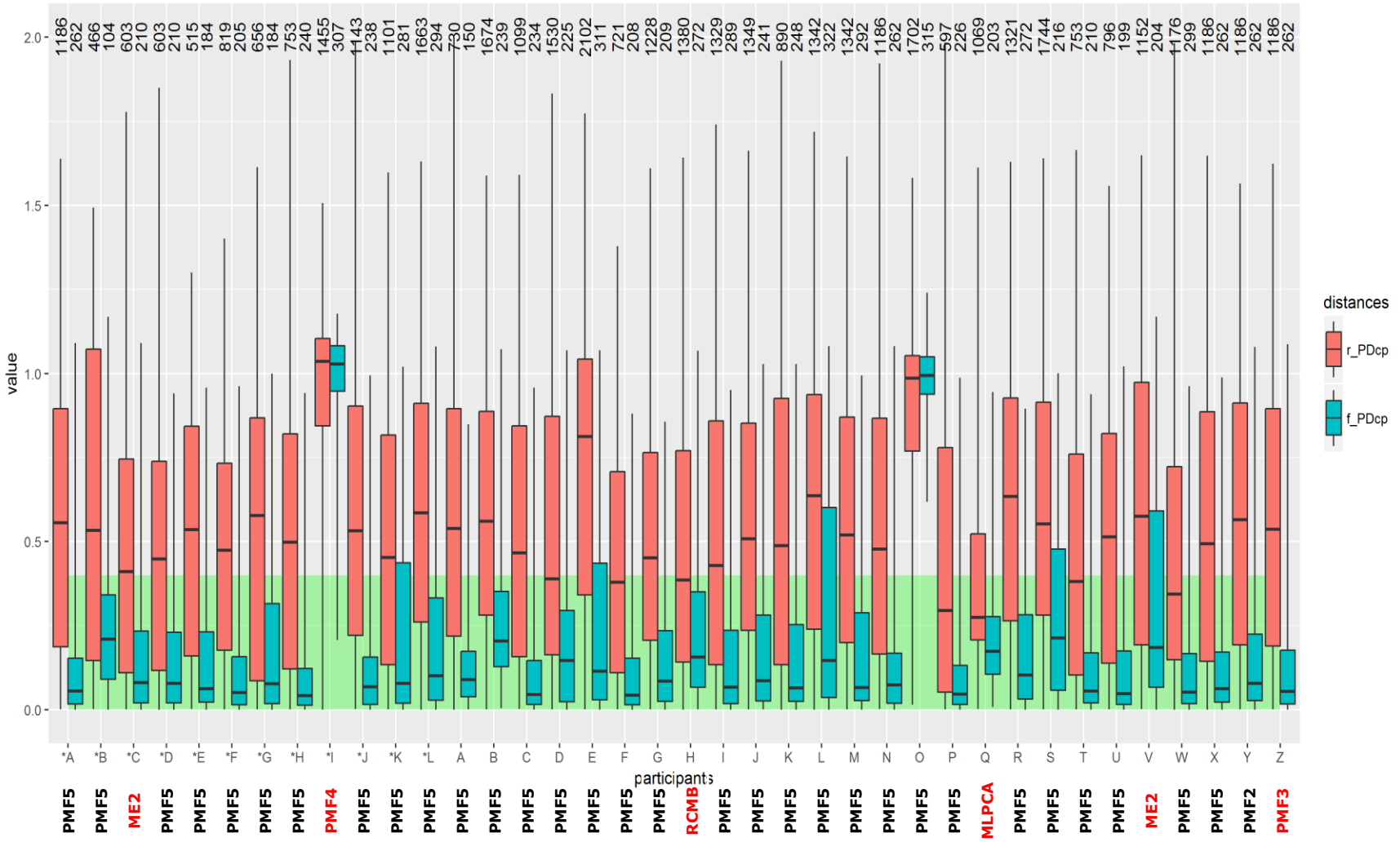
r = distances to the reference chemical profiles (cp) in SPECIATE and SPECIEUROPE

f = distances among the candidate sources

top = number of distances

European Commission

Pearson distance (PD) all participants



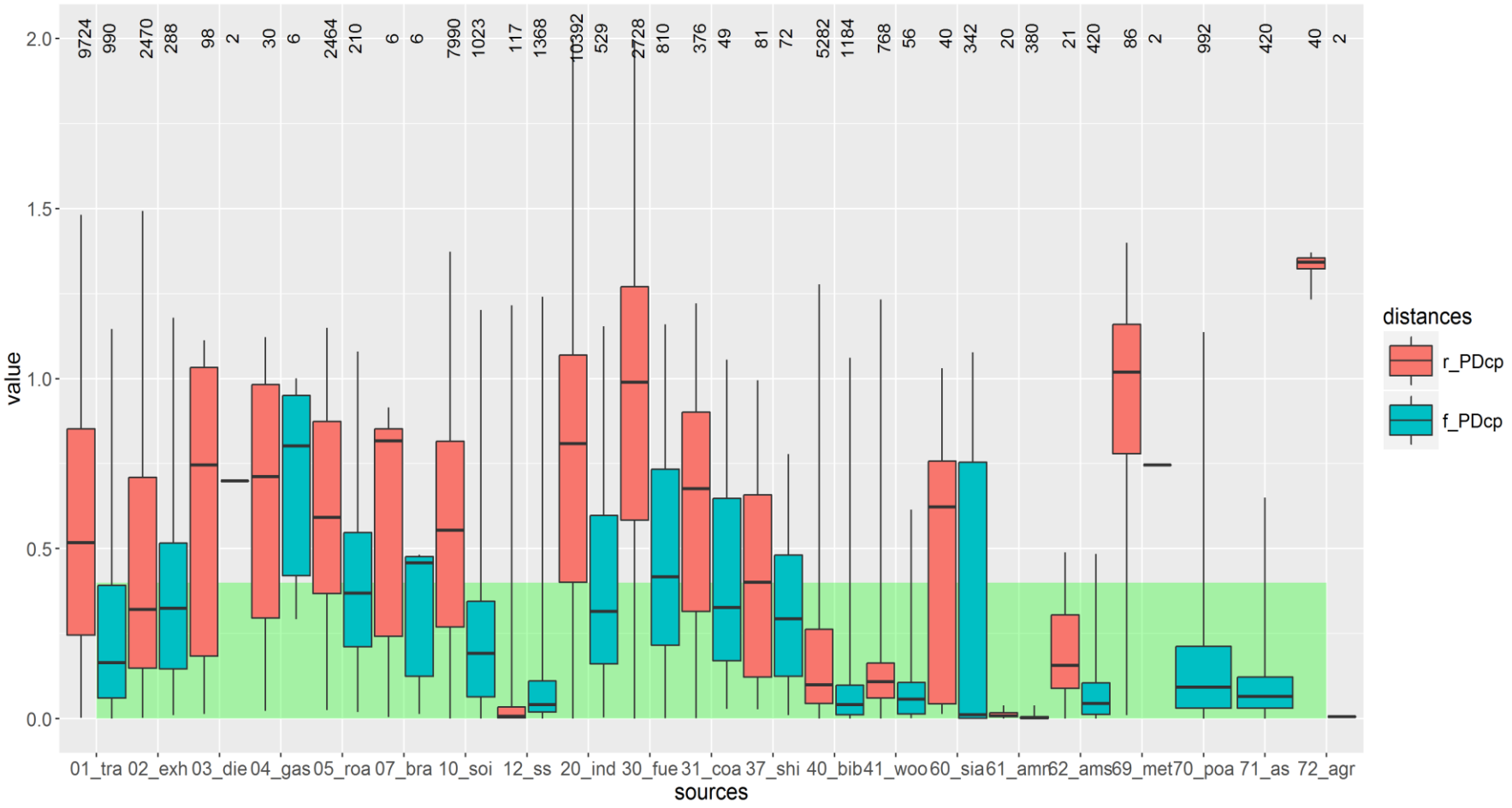
r = distances to the reference chemical profiles (cp) in SPECIATE and SPECIEUROPE

f = distances among the candidate sources

top = number of distances

European Commission

Pearson distance (PD) by source



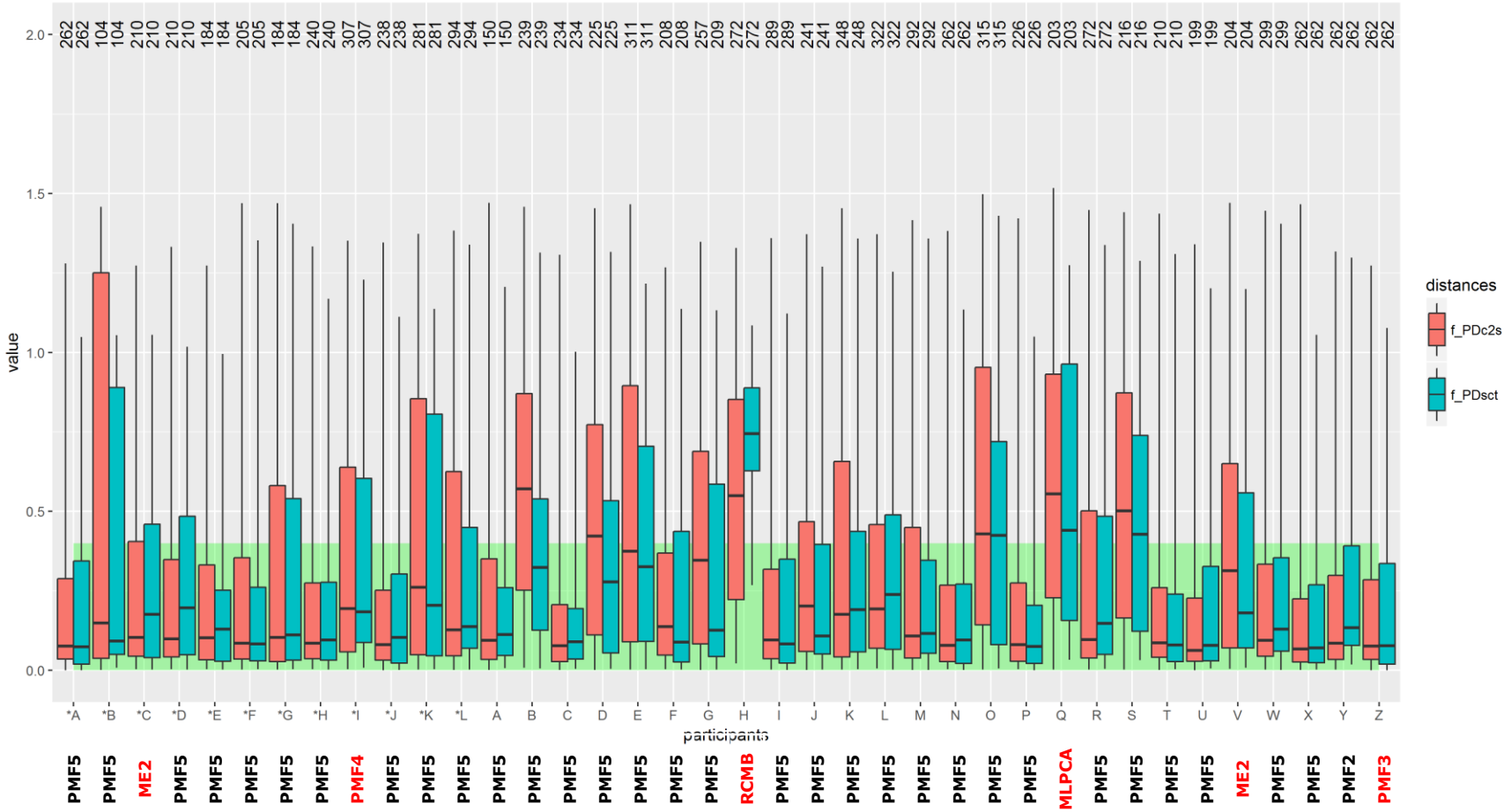
r = distances to the reference chemical profiles (cp) in SPECIATE and SPECIEUROPE

f = distances among the candidate sources

top = number of candidate sources

European Commission

PD time series and contribution-to-species



f = distances among the candidate sources

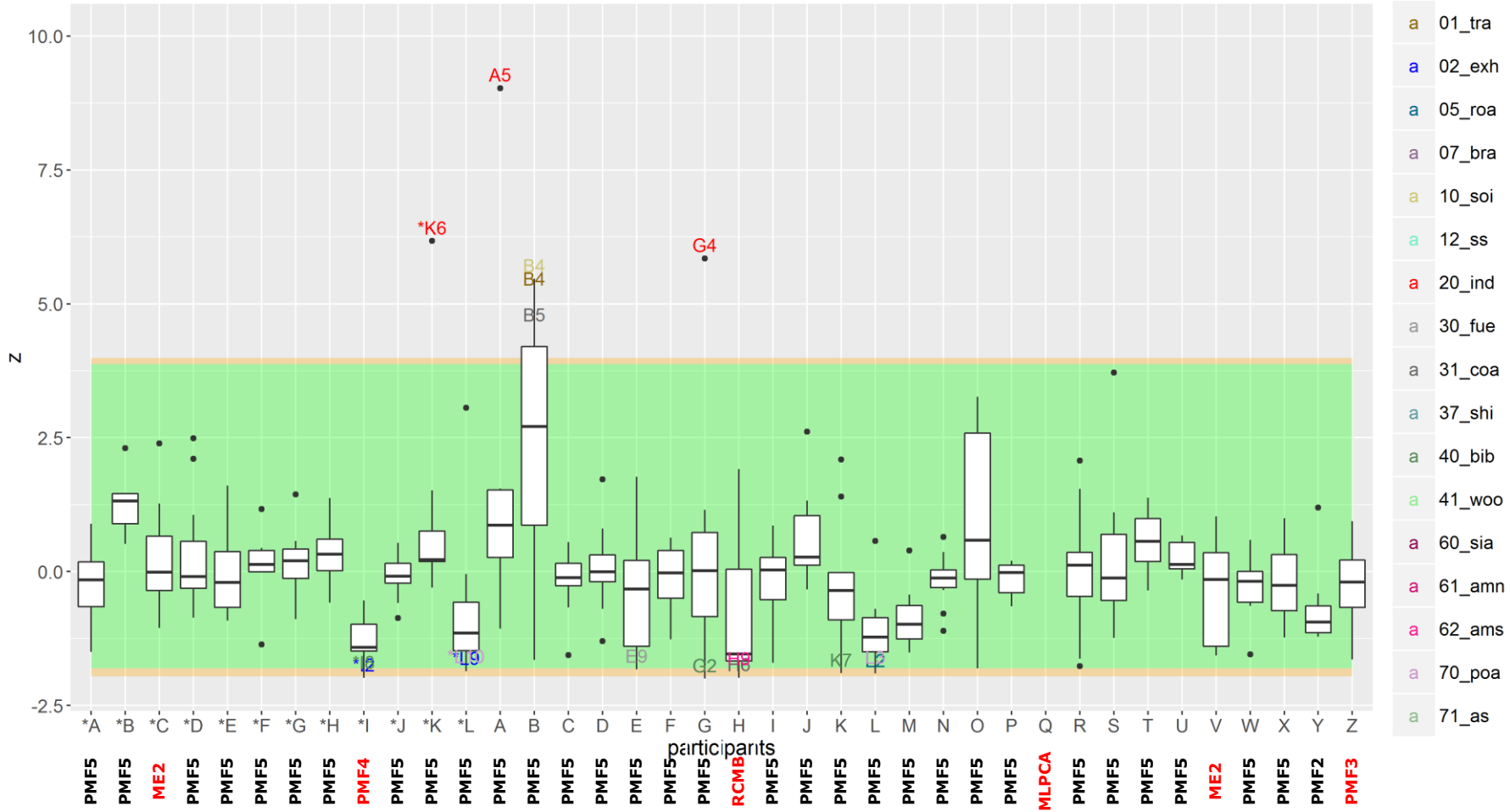
top = number of candidate sources

c2s = contribution to species; sct= source contribution estimate time series

European Commission

RM performance tests

Performance RMs z-score (overall SCE)

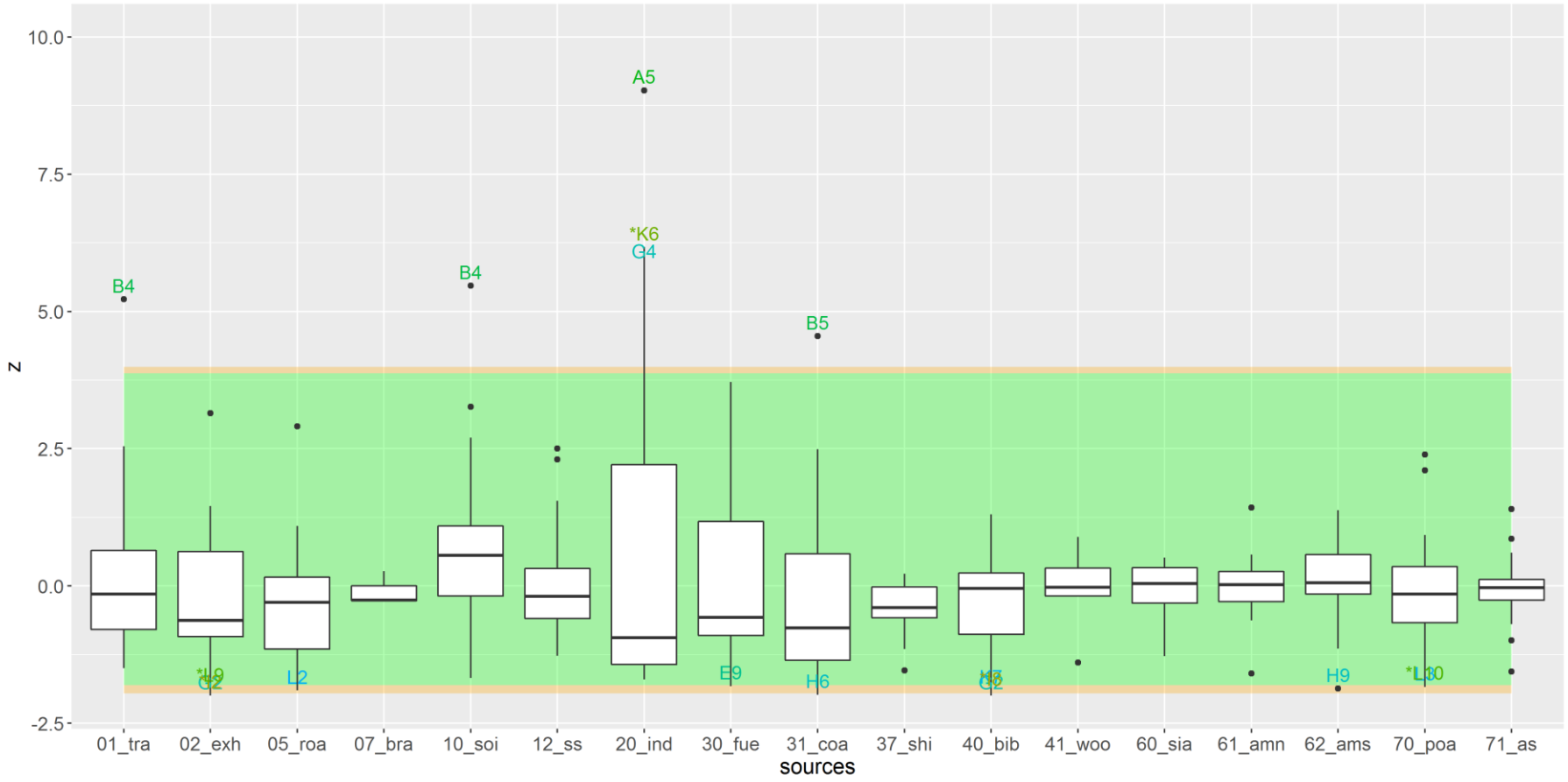


successful candidates: 91%

z score thresholds from kernel curve (Belis et al., 2015); Result Q non plotted (outlier)

European Commission

Performance RMs z-score (overall SCE)

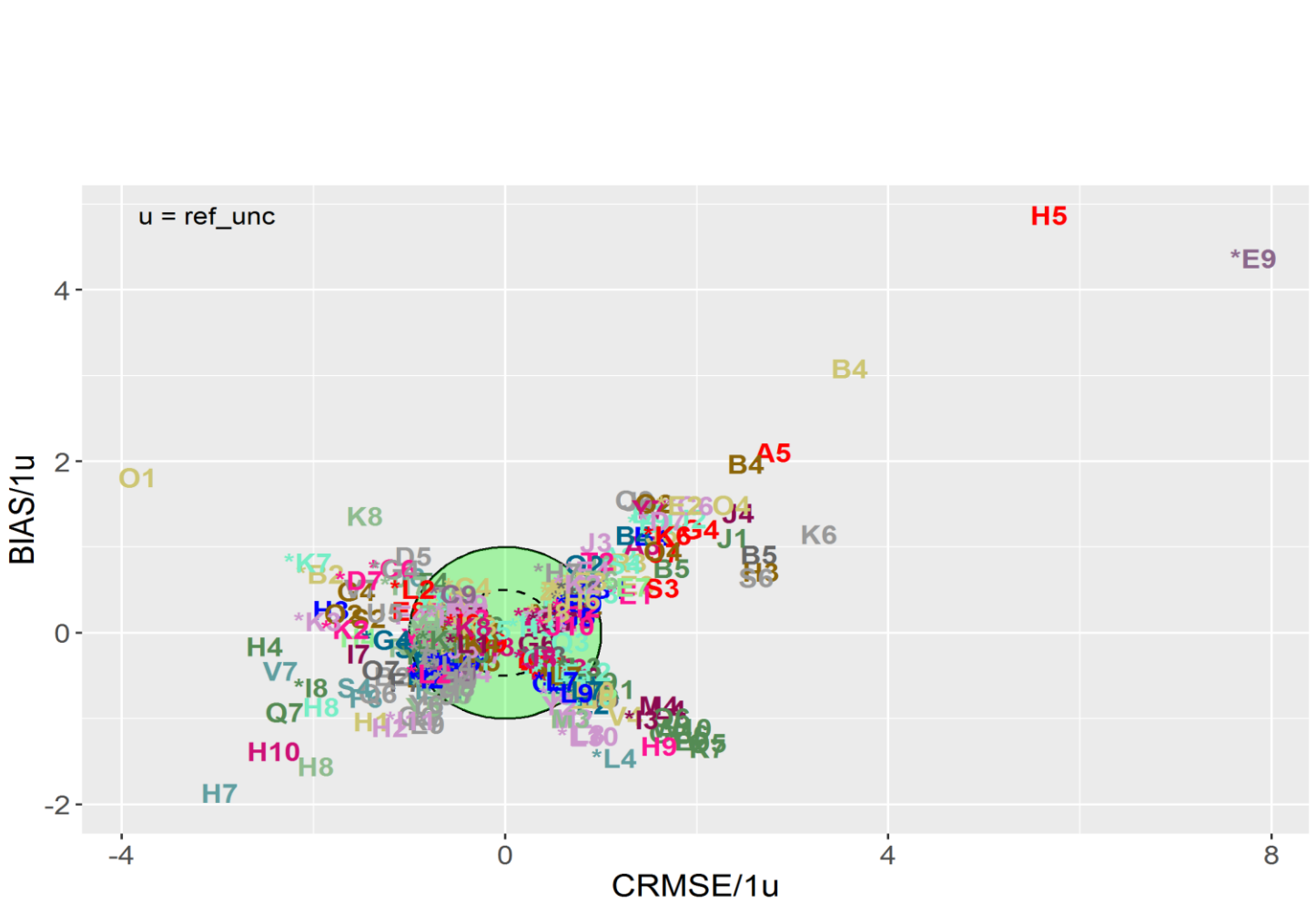


z score thresholds from kernel curve (Belis et al., 2015)

Result Q non plotted (outlier)

European Commission

Performance RMs Target plot (SCE time series)

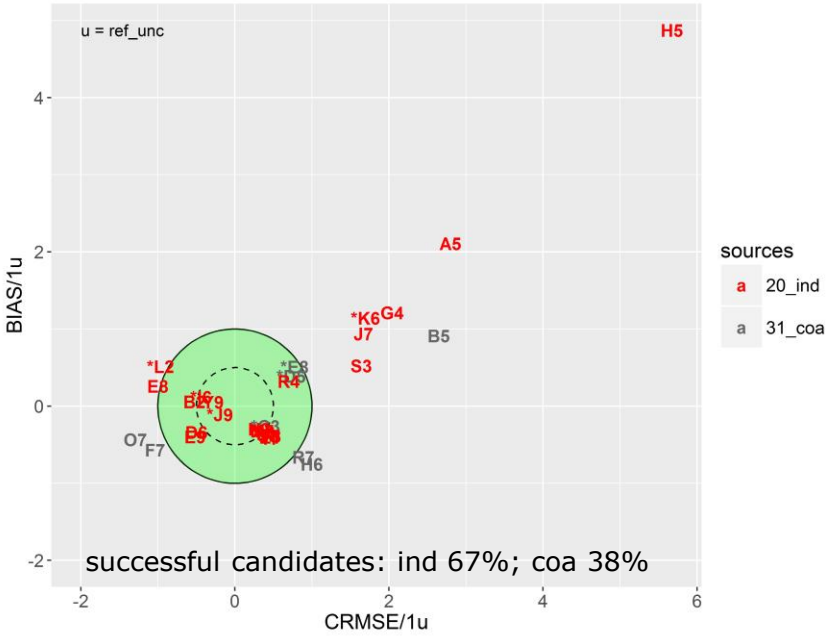


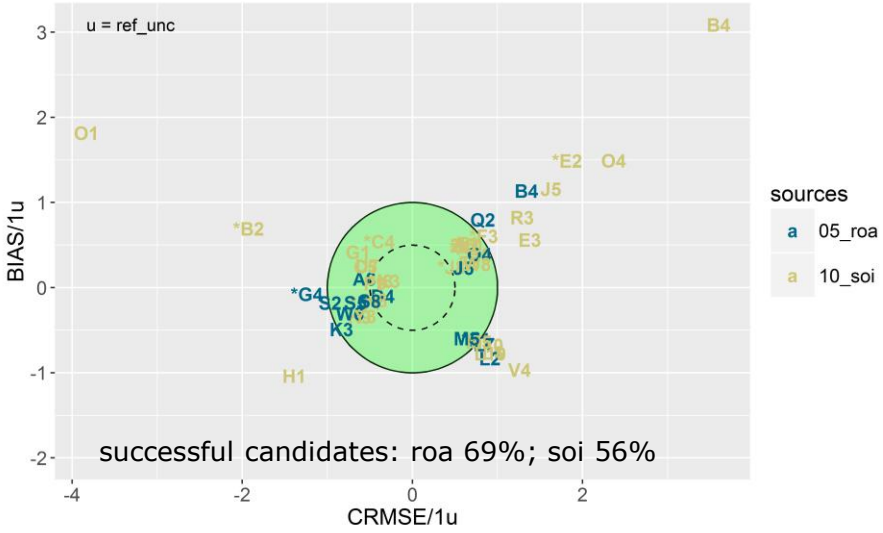
a 01_tra
a 02_exh
a 05_roa
a 07_bra
a 10_soi
a 12_ss
a 20_ind
a 30_fue
a 31_coa
a 37_shi
a 40_bib
a 41_woo
a 60_sia
a 61_amn
a 62_ams
a 70_poa
a 71_as

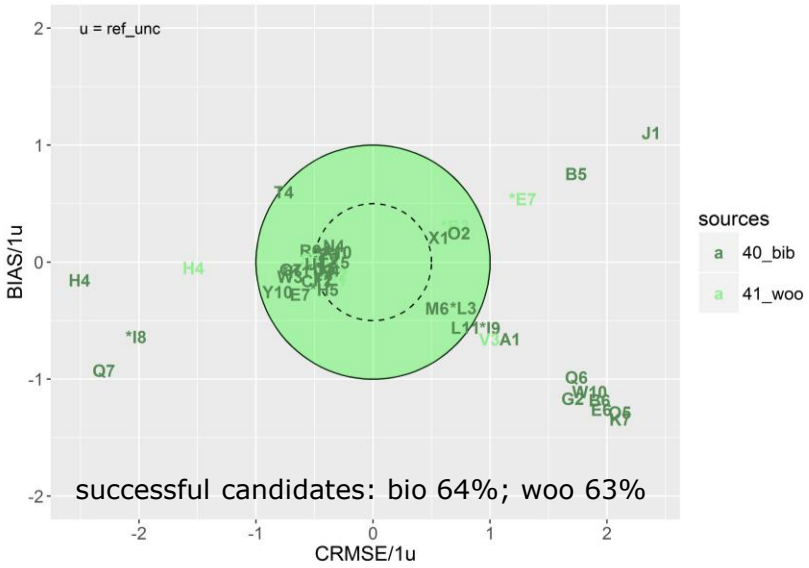
successful candidates: 72%

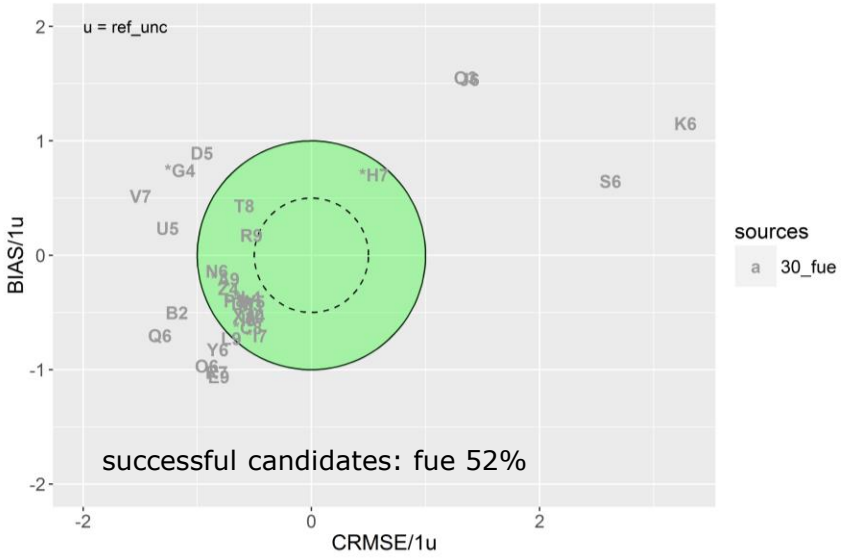
European Commission

Performance RMs Target plot for selected sources





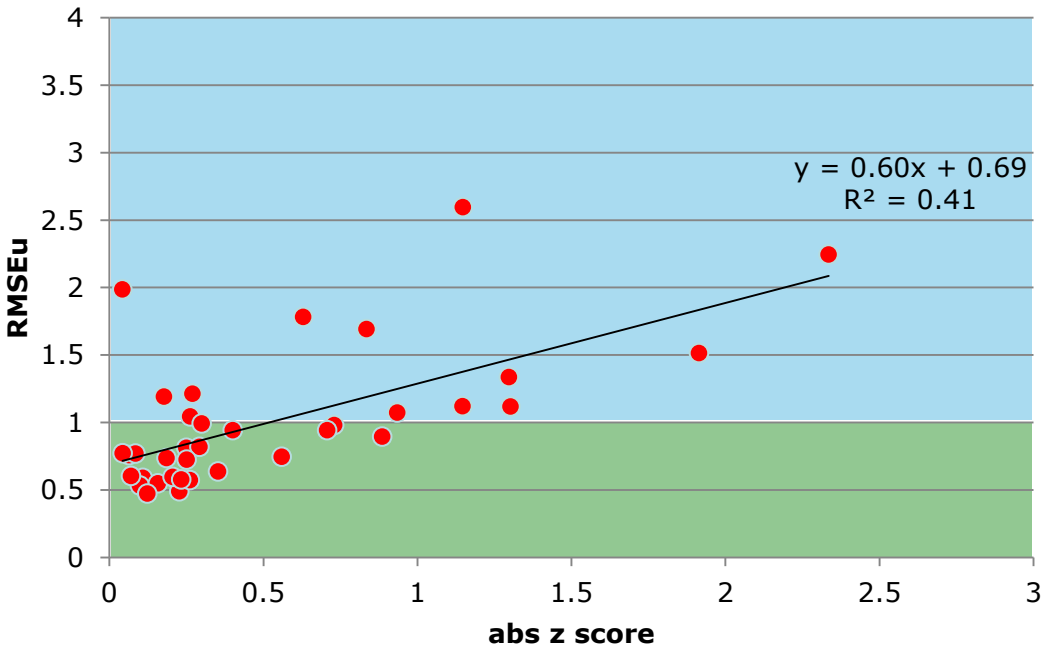




Further analysis of RM results

RM analysis: $RMSE_u$ vs z-score

Average of all scores in every SA result (dots)



area of acceptability for both tests

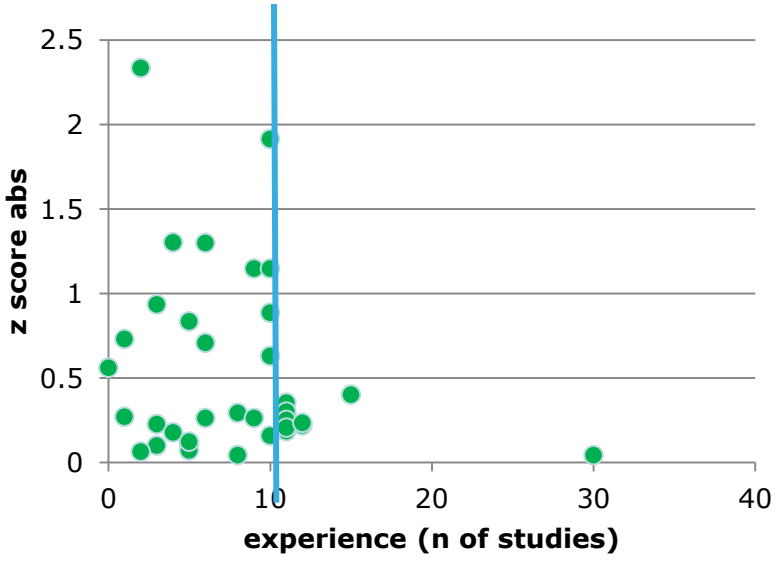
area of rejection for target and OK for z-score

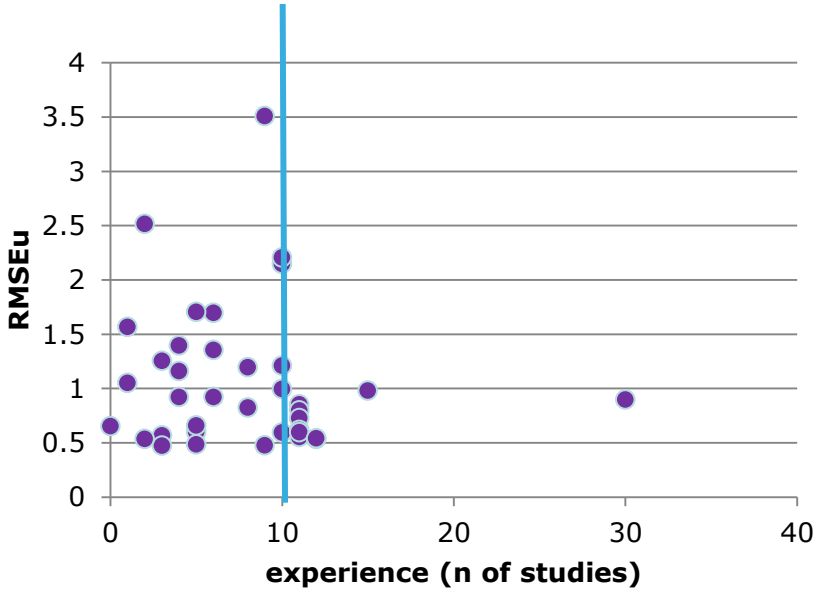
$RMSE_u$ and z-score are correlated.

$RMSE_u$ test is more severe because assess every single time step.

RM analysis: practitioner experience (declared)

Average of all scores in every SA result (dots)



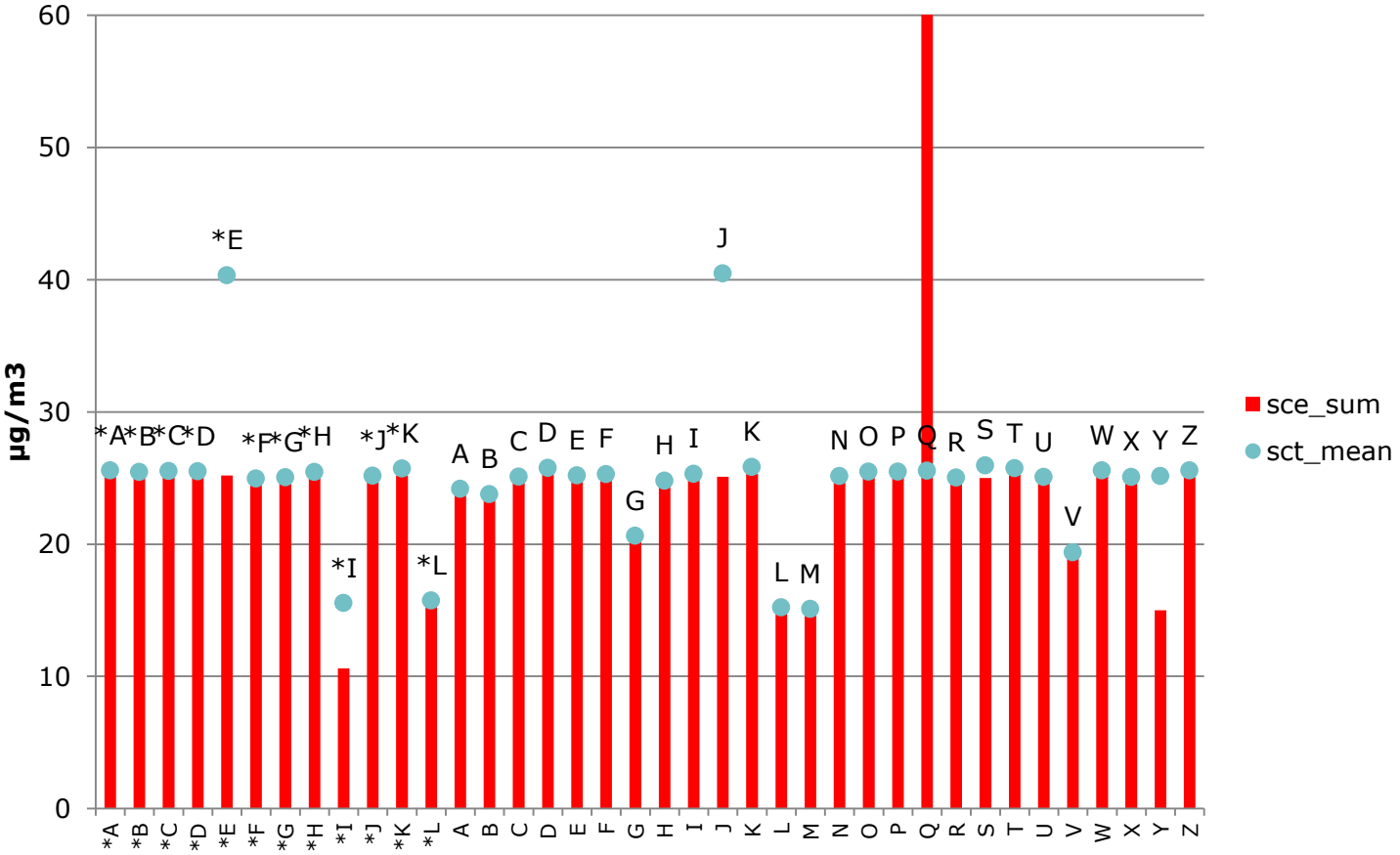


The performance for practitiones that have conducted 10 or less studies is quite variable.

Practitioners declaring to have conducted more than 10 studies always have good performances (low scores).

RM analysis: SCE vs sct

Average of all scores in every SA result (dots)

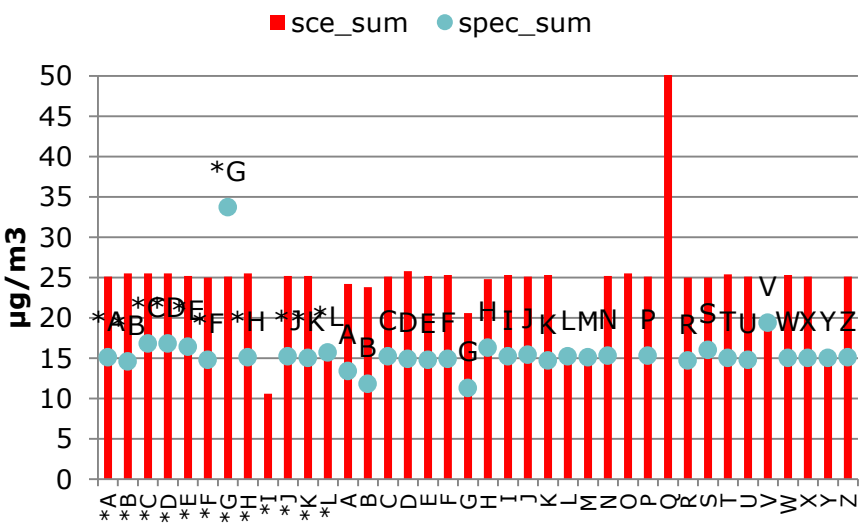


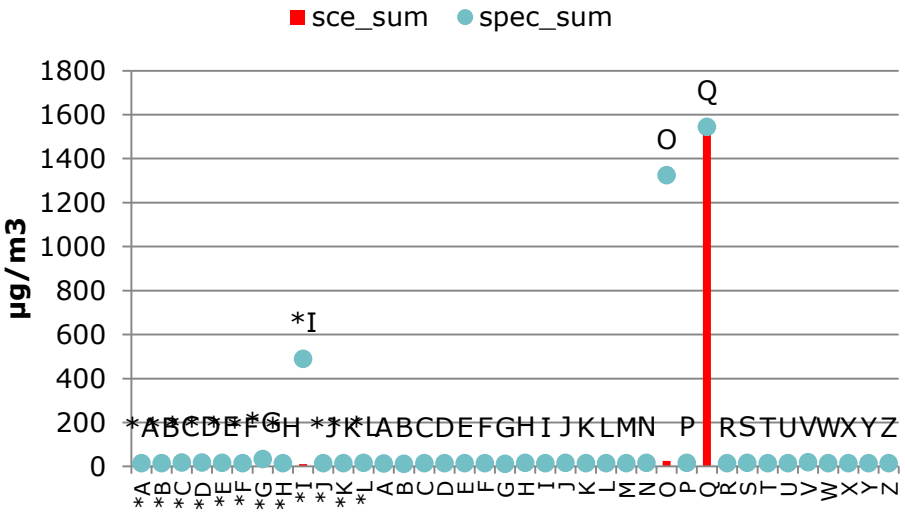
Coherence between the sum of the mass of the sources (sce provided by participants, red bars) and the average mass of the sce time series (blue dots)

European Commission

RM analysis: mass of species and SCEs

Average of all scores in every SA result (dots)



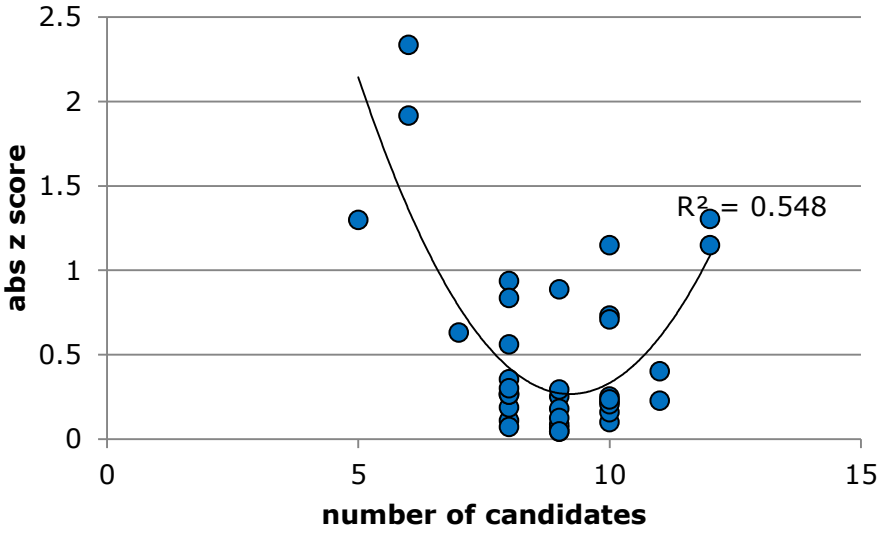


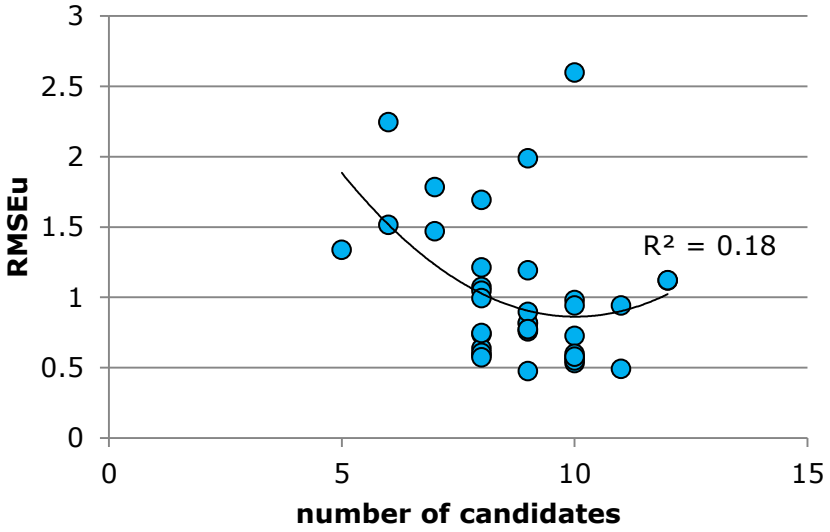
Comparison between the sum of the mass of the sources (sce provided by participants, red bar) and the sum of the mass of the species in the chemical profiles (blue dots)

European Commission

RM analysis: number of factors and performance

Average of all scores in every SA result (dots)



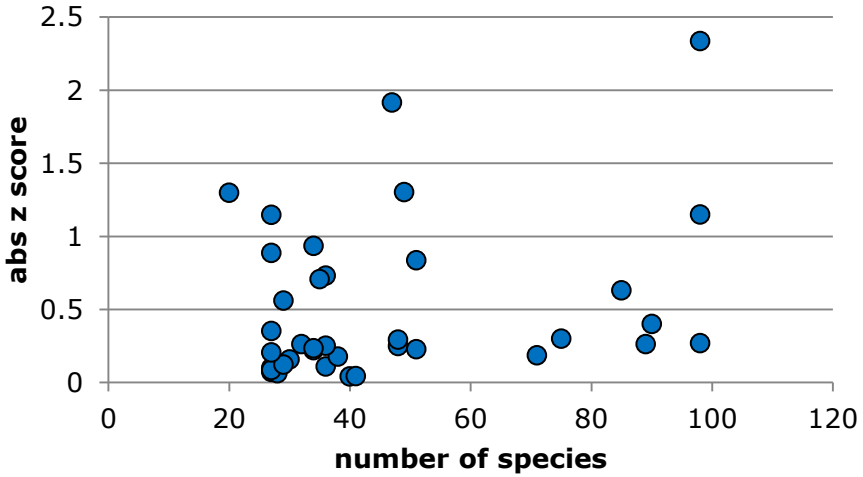


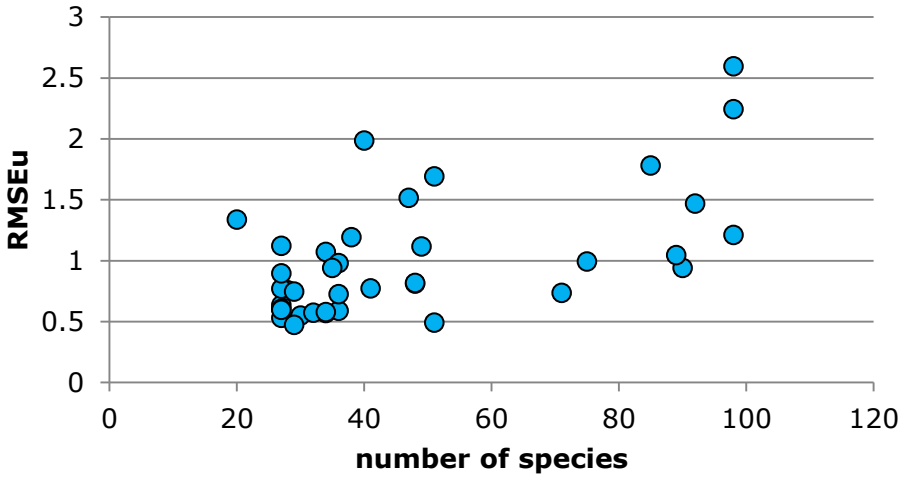
Results (solutions) with a number of candidate sources (factors) near to the average (9) present better z-scores than those with 3 or more sources of difference.

No evident relationship for RMSEu

RM analysis: number of species and performance

Average of all scores in every SA result (dots)





No clear relation ship between number of species and performance

Conclusions of the IE (1)

RMS

- **RMs present results comparable among participants which are also coherent with measured PM.**
- **There is a convergence of users towards one particular model: EPA PMF5.**
- **Industry source category in RM needs better definition because often used to represent a wide variety of different sources.**
- **In industry and soil one third of results show overestimated time series and amplitude problems.**
- **In biomass burning approx. one third of results underestimate time series and shwo prevalent amplitude problem**
- **In fuel oil candidates spread randomly around the reference**
- **Z score and RMSEu are correlated. However, RMSEu has lower success indicating a lower accuracy in single sample than in overall average.**
- **Experienced users >10 studies tend to have better performance.**

Thank you for your attention